



Remington

Essentials of Pharmaceutics

Edited by Linda Felton





Essentials of Pharmaceutics



Free Pharmaceutical Press

e-alerts

Our latest product news *straight to your inbox*
register@ www.pharmpress.com/alerts

Pharmaceutical Press is the publishing division of the Royal Pharmaceutical Society



Essentials of Pharmaceutics

Edited by Linda A. Felton, PhD
Chair, Department of Pharmaceutical Sciences
Associate Professor of Pharmaceutics
College of Pharmacy
University of New Mexico
Albuquerque, NM, USA



University of the Sciences
Philadelphia
College of
Pharmacy



Published by Pharmaceutical Press

1 Lambeth High Street, London SE1 7JN, UK

© Pharmaceutical Press 2012

(PP) is a trade mark of Pharmaceutical Press

Pharmaceutical Press is the publishing division of the Royal Pharmaceutical Society

First edition published 2013

Typeset by Newgen KnowledgeWorks, India

Printed in the United States of America by Edwards Brothers Malloy

ISBN 978 0 85711 105 0

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, without the prior written permission of the copyright holder.

The publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this book and cannot accept any legal responsibility or liability for any errors or omissions that may be made.

A catalogue record for this book is available from the British Library.

This book is adapted from contributions published in *Remington: The Science and Practice of Pharmacy* 22nd edition.

Contents

Preface v		Chapter 16	Tonicity, osmoticity, osmolality, osmolarity 277 <i>Andrew Ingham, MRPharmS, PhD and Cathy Y. Poon, PharmD</i>
About the editor vii		Chapter 17	Chemical kinetics 301 <i>Rodney J. Wigent, PhD</i>
Contributors ix		Chapter 18	Complex formation 317 <i>Thorsteinn Loftsson, MScPharm, MSc, PhD and Marcus E. Brewster, PhD</i>
Chapter 1	Information resources in pharmacy and the pharmaceutical sciences 1 <i>Robin H. Bogner, PhD and Sharon Giovenale, MSLS</i>	Chapter 19	Interfacial phenomena 329 <i>Paul M. Bummer, PhD and Yvonne Perrie, PhD</i>
Chapter 2	Analysis of medicinals 9 <i>Raymond D. Skwirczynski, PhD</i>	Chapter 20	Colloidal dispersions 343 <i>Bill J. Bowman, RPh, PhD, Clyde M. Ofner III, PhD, Hans Schott, PhD and Yvonne Perrie, PhD</i>
Chapter 3	Quality assurance and control 29 <i>John H. Parker, PhD and John E. Enders, PhD, MBA</i>	Chapter 21	Coarse dispersions 371 <i>James Swarbrick, DSc, PhD, Joseph T. Rubino, PhD, RPh and Orapin P. Rubino, PhD</i>
Chapter 4	Stability of pharmaceutical products 37 <i>Allan D. Bokser, PhD and Patrick B. O'Donnell, PhD</i>	Chapter 22	Rheology 393 <i>Lawrence H. Block, PhD</i>
Chapter 5	Bioavailability and bioequivalence 51 <i>Steven B. Johnson, PharmD</i>	Chapter 23	Powders 411 <i>Yi-Bo Wang and Robert O. Williams III, PhD</i>
Chapter 6	Dissolution 63 <i>Vijai Kumar, MS, MBA and Praveen Hiremath, PhD</i>	Chapter 24	Solutions, Emulsions, Suspensions, and Extracts 435 <i>Michael M. Crowley, PhD</i>
Chapter 7	Modern-day drug discovery and development 81 <i>Rick G. Schnatz, PharmD</i>	Chapter 25	Sterilization Processes and Sterility Assurance 469 <i>James Agalloco, BEChE, MSChE, MBA, William G. Lindboem, Jr., PhD and Russell E. Madsen, MS</i>
Chapter 8	The New Drug Approval Process 93 <i>Linda A. Felton, PhD and Dennis W. Raisch, PhD, RPh</i>	Chapter 26	Parenteral Preparations 495 <i>Michael J. Akers, PhD</i>
Chapter 9	Metrology and pharmaceutical calculations 101 <i>Roger L. Schnaare, PhD and Shelly J. Prince, PhD</i>	Chapter 27	Pharmaceutical Compounding – USP <797> Sterile Preparations 533 <i>Catherine Cone, PharmD, BCPS, Linda A. Felton, PhD, Amy Bachyrycz, PharmD</i>
Chapter 10	Statistics 133 <i>Sanford Bolton, PhD and Richard Hirsch, PhD</i>	Chapter 28	Ophthalmic Preparations 541 <i>Masood Chowhan, PhD, John C. Lang, PhD, Paul Missel, PhD</i>
Chapter 11	Molecular structure, properties and states of matter 177 <i>Thomas Rades, PhD, Keith C. Gordon, PhD and Kirsten Graeser, PhD</i>	Chapter 29	Medicated Topicals 565 <i>Lawrence H. Block, PhD</i>
Chapter 12	Thermodynamics 207 <i>Timothy S. Wiedmann, PhD</i>	Chapter 30	Oral Solid Dosage Forms 581 <i>Ahmed Adel Sakr, PhD and Fars K. Alanazi, PhD</i>
Chapter 13	Solutions and phase equilibria 219 <i>Pardeep K. Gupta, PhD</i>		
Chapter 14	Separation methods 241 <i>Lloyd V. Allen, Jr., PhD, RPh</i>		
Chapter 15	Ionic solutions and electrolytic equilibria 257 <i>Barbara R. Conway, PhD</i>		

Chapter 31	Coating of Pharmaceutical Dosage Forms <i>Stuart C. Porter, PhD</i>	611	Chapter 36	Pharmaceutical Excipients <i>William J. Reilly, Jr., MBA</i>	683
Chapter 32	Oral Modified-Release Drug Delivery Systems <i>Ali R. Rajabi-Siahboomi, PhD, Manish S. Rane, PhD and Linda A. Felton, PhD</i>	623	Chapter 37	Basic pharmacokinetics and pharmacodynamics <i>Raymond E. Galinsky, PharmD and Craig K. Svensson, PharmD, PhD</i>	705
Chapter 33	Aerosols <i>John J. Sciarra, PhD and Christopher J. Sciarra, BS, MS Industrial Pharmacy</i>	633	Chapter 38	Drug action and effect <i>Donald N. Franz, PhD</i>	727
Chapter 34	Biotechnology and Drugs <i>Ara H. DerMarderosian, PhD and Zhiyu Li, PhD</i>	651	Chapter 39	Drug absorption, distribution, metabolism and excretion <i>Michael R. Franklin, PhD</i>	739
Chapter 35	Pharmaceutical Packaging <i>C. Jeanne Taborsky, BSCHEM and Kathleen Deiss, RN</i>	667	Chapter 40	Pharmacokinetics/pharmacodynamics in drug development <i>George L. Drusano, MD</i>	755
			Index		763

Preface

Remington's has been in print for more than 100 years and is a well-known reference text for pharmacists, pharmacy students, and pharmaceutical scientists alike. Over the past century, each edition has been updated to reflect the current state of the science and practice of pharmacy. Following the recent publication of the 22nd edition, the publishers asked me, as the editor of the Pharmaceutical Dosage Forms: Manufacturing and Compounding section, to create a version of this reference that focuses exclusively on pharmaceuticals. This new text, entitled 'Remington: Essentials of Pharmaceutics', is truly a historic event in the long and prestigious legacy of the Remington brand and I am honored to be part of this innovative project.

As the name implies, 'Remington: Essentials of Pharmaceutics' focuses on the pharmaceutical aspects of pharmacy practice. Chapters from the original parent text were specifically selected to create this new edition. The text pulls heavily from the Pharmaceutics and Pharmaceutical Dosage Forms sections. Various delivery systems and dosage forms, including powders, tablets and capsules, solutions, topicals and transdermals, aerosols, and ophthalmics, are covered. In addition, parenterals, sterilization processes, and sterile compounding are also presented. One chapter addresses pharmaceutical excipients and another discusses pharmaceutical packaging. Pharmaceutical analysis, product characterization, quality control, stability, bioavailability, and dissolution are also covered. Fundamental scientific concepts including thermodynamics, ionic solutions

and electrolyte equilibria, tonicity, chemical kinetics, rheology, complex formation and interfacial phenomenon are presented. The text also provides an introduction to pharmacokinetics and pharmacodynamics and the principles of absorption, distribution, metabolism and excretion. In addition, some introductory concepts on drug discovery and drug product approval as well as information resources in pharmacy and the pharmaceutical sciences are presented.

While there are many text books on drug delivery systems currently published, many focus on a specific route of delivery or a particular type of delivery system. Some of these books only touch on the underlying scientific principles of such systems while others delve so heavily on the theoretical that the reader may feel lost. In contrast, this new title provides a concise yet detailed resource covering all aspects of pharmaceutics, from the scientific fundamentals to the dosage forms and drug delivery systems to drug product analyses. As with the original parent text, this book will be useful for anyone interested in pharmaceutics and drug delivery. The focus on the essential concepts of pharmaceutics and drug delivery, however, make this text an exceptionally valuable resource for scientists new to the pharmaceutical industry as well as pharmacy and graduate students. With several chapters focused on parenteral products and sterilization, this book is also an excellent reference for compounding pharmacists involved in the preparation of sterile products.

Contributors

James Agalloco, BEChE, MSChE, MBA / President, Agalloco and Associates, Belle Mead, NJ, USA

Michael J. Akers, PhD / Baxter Healthcare Corporation (Retired), Bloomington, IN, USA

Fars K. Alanazi, PhD / Professor of Pharmaceutics, Kayyali Chair for Pharmaceutical Industries, Department of Pharmaceutics, King Saud University, Riyadh, Saudi Arabia

Loyd V. Allen Jr., PhD, RPh / Professor Emeritus, University of Oklahoma School of Pharmacy, Editor-in-Chief *International Journal of Pharmaceutical Compounding*, Edmond, OK, USA

Amy Bachyrycz, PharmD / Assistant Professor, University of New Mexico, Albuquerque, NM, USA

Lawrence H. Block, PhD / Professor Emeritus, Mylan School of Pharmacy, Duquesne University, Pittsburgh, PA, USA

Robin H. Bogner, PhD / Associate Professor, Department of Pharmaceutical Sciences, University of Connecticut School of Pharmacy, Storrs, CT, USA

Allan D. Bokser, PhD / Independent Consultant, San Diego, CA, USA

Sanford Bolton, PhD / College of Pharmacy, University of Arizona, Tucson, AZ, USA

Bill J. Bowman, RPh, PhD / Assistant Professor of Pharmaceutical Sciences, Arizona College of Pharmacy-Glendale, Midwestern University, Glendale, AZ, USA

Marcus E. Brewster, PhD / Distinguished Research Fellow, Johnson & Johnson Pharmaceutical Research and Development, Beerse, BE

Paul M. Bummer, PhD / Associate Professor, Department of Pharmaceutical Sciences, College of Pharmacy, University of Kentucky, Lexington, KY, USA

Masood Chowhan, PhD / Senior Director, Drug Formulation and Delivery, Alcon Laboratories, Inc., Fort Worth, TX, USA

Catherine Cone, PharmD, BCPS / Assistant Professor, Department of Pharmacy Practice and Administrative Sciences, University of New Mexico Health Sciences Center, Albuquerque, NM, USA

Barbara R. Conway, PhD / Professor of Pharmaceutics, Division of Pharmacy and Pharmaceutical Sciences, School of Applied Sciences, University of Huddersfield, Queensgate, Huddersfield, UK

Michael M. Crowley, PhD / President, Theridian Technologies, LLL, Austin, TX, USA

Kathleen Deiss, RN / Quality Assurance Manager, SciRegs International, Inc., Columbia, MD, USA

Ara H. DerMarderosian, PhD / Research Professor of Pharmacognosy, Department of Biological Sciences, University of the Sciences in Philadelphia, Philadelphia, PA, USA

George L. Drusano, MD / Co-Director, Ordway Research Institute, Albany, NY, USA

John E. Enders, PhD, MBA / Chief Operating Officer, Tech Manage Associates, Inc., Dalton, PA, USA

Linda A. Felton, PhD / Chair Department of Pharmaceutical Sciences, Associate Professor of Pharmaceutics College of Pharmacy, University of New Mexico, Albuquerque, NM, USA

Michael R. Franklin, PhD / Professor of Pharmacology and Toxicology, Department of Pharmacology and Toxicology, College of Pharmacy, University of Utah, Salt Lake City, UT, USA

Donald N. Franz, PhD / Professor Emeritus, Department of Pharmacology and Toxicology, University of Utah School of Medicine, Salt Lake City, UT, USA

Raymond E. Galinsky, PharmD / Professor of Pharmaceutics, Department of Industrial and Physical Pharmacy, Purdue University College of Pharmacy, West Lafayette, IN, USA

Sharon Giovenale, MSLS / Pharmacy Librarian, University of Connecticut Pharmacy Library, Storrs, CT, USA

Keith C. Gordon, PhD / Professor of Physical Chemistry, Department of Chemistry, University of Otago, Dunedin, Otago, NZ

Kirsten Graeser, PhD / Formulation Research, F. Hoffmann-La Roche Ltd., Basel, CH

Pardeep K. Gupta, PhD / Associate Professor of Pharmaceutics, Philadelphia College of Pharmacy, University of the Sciences in Philadelphia, Philadelphia, PA, USA

Praveen Hiremath, PhD / Principal Scientist – Pharmaceutical Formulation, Bayer Healthcare, Overland Park, KS, USA

Richard Hirsch, PhD / Faculty for Applied Sciences, Cologne University of Applied Sciences, Cologne, DE

Andrew Ingham, MRPharmS, PhD / Lecturer in Pharmacy/ Pharmaceutical Delivery, School of Life and Health Sciences, Aston University, Birmingham, UK

Steven B. Johnson, PhD / Vice President, Regulatory Affairs, Novo Nordisk, Princeton, NJ, USA

Vijai Kumar, MS, MBA / General Manager and Vice President of Technical Operations and Research and Development, Pharmaceutics International, Hunt Valley, MD, USA

John C. Lang, PhD / Senior Director, AZYP, and Senior Research Scientist/ Adjunct Professor University of Texas at Arlington, Arlington, TX, USA

Thorsteinn Loftsson, MSPharm, PhD / Professor of Physical Pharmacy, Department of Pharmaceutical Sciences, University of Iceland, Hofsvallagata, IS

Zhiyu Li, PhD / Assistant Professor of Pharmaceutical Sciences, Philadelphia College of Pharmacy, University of the Sciences in Philadelphia, Philadelphia, PA, USA

William G. Lindboe, Jr., PhD / Independent Consultant, East Brunswick, NJ, USA

Russell E. Madsen, MS / President, The Williamsburg Group, LLC, Gaithersburg, MD, USA

Paul Missel, PhD / Modeling and Simulation, Alcon Research, Ltd, Fort Worth, TX, USA

Patrick B. O'Donnell, PhD / Senior Director, Pharmaceutical Development, Ambit Biosciences Corporation, San Diego, CA, USA

Clyde M. Ofner III, PhD / Professor of Pharmaceutics, University of the Sciences in Philadelphia, Philadelphia College of Pharmacy, Philadelphia, PA, USA

John H. Parker, PhD / President, Tech Manage Associates, Dalton, PA, USA

Yvonne Perrie, PhD / Head of Pharmacy, Professor in Pharmaceutics/Drug Delivery, Aston University, Birmingham, UK

Cathy Y. Poon, PharmD / Professor and Vice Dean, Department of Pharmacy Practice and Pharmacy Administration, Philadelphia College of Pharmacy, University of the Sciences, Philadelphia, PA, USA

Stuart C. Porter, PhD / Senior Director, Global Pharmaceutical Applications RandD, International Specialty Products, Wayne, NJ, USA

Shelly J. Prince, PhD / Associate Professor of Pharmaceutics, College of Pharmacy, Southwestern Oklahoma State University, Weatherford, OK, USA

Thomas Rades, PhD / Faculty of Health and Medical Sciences, Department of Pharmacy, University of Copenhagen, Copenhagen, DK

Dennis W. Raisch, PhD, MS, RPh / Professor and Chair, Graduate Program in Pharmacoeconomics, Epidemiology, Pharmaceutical Policy, and Outcomes Research, College of Pharmacy, University of New Mexico, Albuquerque, NM, USA

Ali R. Rajabi-Siahboomi, PhD / Vice President and Chief Scientific Officer, Colorcon, Inc., Harleysville, PA, USA

Manish S. Rane, PhD / Formulation Technologies Manager, Colorcon, Inc., West Point, PA, USA

William J. Reilly, Jr., MBA / Director of Technical Services, Lannett Pharmaceuticals, Inc., Philadelphia, PA, USA

Joseph T. Rubino, PhD, RPh / Pharmacist/Consultant, Towaco, NJ, USA

Orapin P. Rubino, PhD / Director, Formulation and Product Development, Glatt Air Techniques, Inc., Ramsey, NJ, USA

Ahmed Adel Sakr, PhD / Distinguished Chaired Professor of Industrial Pharmacy and Pharmaceutics, Board of Trustees, Chair Assistant, Graduate Studies, Research and International Affairs, Future University Egypt, Cairo, Egypt

Roger L. Schnaare, PhD / Professor Emeritus of Pharmacy, Philadelphia College of Pharmacy, University of the Sciences in Philadelphia; Senior Pharmaceutics Fellow, Biosyn, Inc., Philadelphia, PA, USA

Rick G. Schnatz, PharmD / Senior Scientific Liaison, Compounding Expert Committee, Healthcare Quality and Compendia Affairs, United States Pharmacopeia, Rockville, MD, USA

Hans Schott, PhD / Professor Emeritus of Pharmaceutics and Colloidal Chemistry, Temple University, Philadelphia, PA, USA

Christopher J. Sciarra, BS, MS / Industrial Pharmacy Vice President, Scientific, Sciarra Laboratories, Inc., Hicksville, NY, USA

John J. Sciarra, PhD / Professor Emeritus and President, Sciarra Laboratories, Inc., Hicksville, NY, USA

Raymond D. Skwierzynski, PhD / Senior Director, Analytical Development Millennium: The Takeda Oncology Company, Cambridge, MA, USA

Craig K. Svensson, PharmD, PhD / Dean, College of Pharmacy Professor of Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, IN, USA

James Swarbrick, DSc, PhD / President, PharmaceuTech Pinehurst, NC, USA

C. Jeanne Taborsky, BSChem / President and Chief Executive Officer, SciRegs International, Inc., Columbia, MD, USA

Yi-Bo Wang / Graduate Student in Pharmaceutics, The University of Texas at Austin, College of Pharmacy, Austin, TX, USA

Timothy S. Wiedmann, PhD / Professor, Department of Pharmaceutics, University of Minnesota College of Pharmacy, Minneapolis, MN, USA

Rodney J. Wigent, PhD / Professor of Chemistry, Research Professor of Pharmaceutics, Dean, College of Graduate Studies, Director, Research Administration, University of the Sciences in Philadelphia, Philadelphia, PA, USA

Robert O. Williams III, PhD / Johnson and Johnson Centennial Professor of Pharmaceutics, Division Head, Pharmaceutics, The University of Texas at Austin, College of Pharmacy; Editor-in-Chief, Drug Development and Industrial Pharmacy, Austin, TX, USA

Information Resources in Pharmacy and the Pharmaceutical Sciences

Robin H. Bogner, PhD and Sharon Giovenale, MSLS

INTRODUCTION 1

LOCATING APPROPRIATE TYPES OF INFORMATION RESOURCES 1

INTRODUCTION

This chapter describes many significant information resources used by the pharmacy professional student, the pharmaceutical sciences graduate student, the pharmacist, and the pharmaceutical scientist. It is not possible to be comprehensive, since there is a vast array of resources, which is growing exponentially. This chapter arranges the resources in three categories: the original research literature or primary literature, which comprises the largest volume of the literature; the secondary literature, which provides access to the primary literature; and the tertiary literature, which sums up the best practice based on the primary literature. Pharmacy practitioners and PharmD students use the literature to provide evidence-based medicine or the integration of the best research evidence into their practice or studies. Pharmaceutical scientists and their graduate students use the literature to support research in such fields as pharmacology, toxicology, medicinal chemistry, and pharmaceuticals.

LOCATING APPROPRIATE TYPES OF INFORMATION RESOURCES

The nature and depth of the question and the level of experience of the researcher will determine the appropriate type of literature used. Whereas an expert may go directly to a database, a less experienced pharmacy student who is asked to find primary literature to write a paper on a therapeutic or pharmaceutical topic may first need to get general background information from a tertiary resource, such as a textbook, handbook, or subject encyclopedia either online or in print. Once the student is familiar with the basics and the professional vocabulary, he or she will frequently perform a search in either the PubMed/Medline database or the International Pharmaceutical Abstracts database to located relevant review articles (secondary sources) and original research (primary sources). Fortunately, these databases often provide links to the full text of the journal article. A basic search statement for these databases should include synonyms for the search terms, including professional or scientific terms. The terms can be combined using the OR logical operator. The search can be refined by adding a term, using the AND operator. For example, if a student is looking for primary literature on possible future improvements in hpv vaccines, they could try searching for: (future OR advances OR second generation) AND (hvp vaccines OR papillomavirus vaccines). Nesting the parallel terms using parentheses will result in the most relevant articles. Additional limits, such as a date range and language, can easily be applied to the search.

A subject specialist often pursues one or more topics over a longer period and can create a series of search alerts in a database or set up RSS feeds to receive timely, updated search results from a specialized database. More information on search updates can be found in the section on Databases.

PRIMARY LITERATURE

The first formal reports of scientific or clinical findings are found in journals. Journals can be distinguished from trade magazines by the lack of advertisements. Some primary liter-

ature, however, can be found in several trade magazines under sections entitled "Peer-reviewed reports". The peer review process is a rigorous evaluation of manuscripts submitted to the journal. After concluding a study, authors write up their findings in a format consistent with the journal's "instructions for authors." The editor of the journal assigns two or more reviewers who are experts in the particular area of research. The reviewers are volunteers from the scientific community who elect to participate in the peer-review process. In addition to the reviewers' recommendations to the editor on whether to publish the manuscript, they provide feedback to the authors, such as how to make the study design more clear to the reader and any additional experiments to allow the conclusions to be stronger. The peer-review process can take several months before the reviewers, authors, and editor agree on the final text for publication. The process is designed so only those manuscripts with clear scientific merit pass peer-review and are published. Each journal has its own criteria for acceptance of manuscripts, which can include whether the manuscript reports results within the area of research covered by that journal and whether the contribution is important to the scientific community.

The discrimination of the peer-review process is not uniform among all journals. For someone new to a field, it is difficult to discern which journals provide the clearest reports that are reproducible and for which the interpretation of the results is thorough without being too speculative. It is prudent to consult an expert in the field to determine which journals are best. Additionally, a general ranking system has been devised by the Institute for Scientific Information (ISI), which is now part of Thomson Reuters, which measures the "impact" of a journal. The Impact Factor of a journal for a given year is the average number of times articles published in that journal in the preceding two years is cited by others. Although this provides a general guide, the Impact Factor has its critics.

Experts, as well as those new to a field, follow the primary literature in a few publications related to their area of interest. Scanning the titles of the research articles, reviewing the abstracts of key articles, and reading those closely related to a particular area of research is the traditional approach to keep up to date on the general trends, as well as progress in one's own specific area. This can be done by subscribing to print journals or simply accessing the websites of the journals of interest on which the titles and abstracts are generally available without charge. In addition, "open access" journals allow non-subscribers access to the full-length detailed articles. The Directory of Open Access Journals (<http://www.doaj.org/>) provides an up-to-date list. In some cases, journals allow open access to non-subscribers after a period of time, during which only subscribers can view full-length articles on-line.

Of the thousands of journals spanning the areas of clinical and pharmaceutical research, as well as related areas, pharmaceutical scientists often publish in:

Journal of Medicinal Chemistry (Eaton, PA; American Chemical Society)

Journal of Natural Products (Washington, DC; American Chemical Society and American Society of Pharmacognosy)

Journal of Pharmacology and Experimental Therapeutics (Baltimore, MD; Williams & Wilkins)
Pharmaceutical Research (New York, NY; Kluwer Academic Publishers)

Clinical pharmacists refer to major medical and pharmacy journals:

American Journal of Health-System Pharmacy (Bethesda, MD; American Society of Health-System Pharmacists)
JAMA—The Journal of the American Medical Association (Chicago, IL; American Medical Association)
Journal of the American Pharmacists Association (Washington, DC; American Pharmacists Association)
New England Journal of Medicine (Boston, MA; Massachusetts Medical Society)

Clinical pharmacists also refer to journals that focus on a particular disease state or patient population:

Circulation (Dallas, TX; American Heart Association)
American Journal of Veterinary Research (Chicago, IL; American Veterinary Medical Association)
Diabetes (Alexandria, VA; American Diabetes Association)
Pediatrics (Evanston, IL; American Academy of Pediatrics)

A guide to journals and trade magazines recommended for the collections of libraries associated with schools/colleges of pharmacy has been developed and is updated by the Libraries/Educational Resources Section of the American Association of Colleges of Pharmacy. (<http://www.aacp.org/governance/SECTIONS/Libraryinformationscience/Pages/LibraryInformationScienceSpecialProjectsandInformation.aspx>) The 2010 list has 79 entries.

When searching for primary resources, patents and dissertations should also be considered. The purpose of a patent is to protect intellectual property, not disclose research findings. So, information claimed in patents should not be considered equal to that found in peer-reviewed research articles. Similarly, dissertations have not undergone the rigorous peer-review process; however, they can be quite useful to find details of studies that are too lengthy to include in a journal article.

SECONDARY LITERATURE-REVIEWS AND DATABASES

Secondary literature is created by experts in the subject field. These resources are organized to either describe or evaluate the original primary literature and to provide useful access to this literature. Secondary literature consists of review articles, including systematic reviews and meta-analyses and bibliographic databases, such as PubMed.

REVIEWS/SYSTEMATIC REVIEWS

As individual articles, the findings in the primary literature are often more useful to those working in that specialty than to those new to the field. A review article that integrates the information from individual articles from the primary literature may be more useful to someone outside the field or new to the field, as reviews provide the perspective gained from many authors over many years. Reviews provide a good starting place for a literature search into a new area. When beginning a literature search, one can often “refine” his or her search of a database to identify only review articles. Additionally, there is an increasing number of journals consisting of only review articles, such as *Trends in Pharmacological Sciences* (Amsterdam; Elsevier) and *Advanced Drug Delivery Reviews* (Amsterdam; Elsevier).

Systematic reviews are those with a more structured methodology to collect and evaluate the primary literature relevant to a particular research question. The most common type of systematic review in pharmacy literature is the meta-analysis. Using carefully selected sets of data from the primary literature and weighting their findings according to a protocol, the meta-analysis can use the much larger pool of data to produce quantitative findings. The use of statistical measures and the large

number of data can often produce a clearer answer from many sometimes contradictory findings in the primary literature. More information on meta-analysis can be found in *Cochrane Handbook for Systematic Reviews of Interventions*.¹

DATABASES

Bibliographic databases provide access to original research articles or the primary literature and cover a wide range of specialties, depth, and breadth of the literature. It is important for the searcher to consider the goal when selecting a database. If the searcher is unsure of where to begin or how to structure the search, he or she should consult the librarian or information specialist. In the absence of a resource person, there are subject guides and tutorials available online that can assist the searcher. A basic Google search will include non-peer reviewed and non-scholarly resources, so it is more effective to search professional scholarly databases. Occasionally, a Google Scholar search is useful in searching a new interdisciplinary area.

For comprehensive research on clinical or therapeutic topics, the searcher may begin with PubMed, which is available to anyone with an internet connection. This database is produced by the National Center for Biotechnology Information at the US National Library of Medicine (<http://www.pubmed.gov>). If access to subscription databases is unavailable, this may be the only affordable option. PubMed has an excellent thesaurus of professional terms, or a controlled vocabulary named MeSH, which stands for Medical Subject Headings. Common search terms are “mapped” to the MeSH terms, which are also searched by PubMed. For example, if the term “tylenol” is entered in the search box, acetaminophen is also automatically searched. Another useful tool, which requires free registration, is My NCBI. A My NCBI account allows the researcher to run search alerts and set up preferences, such as preferred display formats and colored highlighting of the search terms. While signed into My NCBI, a researcher can save a search in many of the NCBI health databases, including PubMed, and schedule email updates or alerts to run on a monthly, weekly, or even a daily basis. Email will only be received if there are new search results. References can also be saved by topics in online Collections. An author may create a bibliography of his or her own work, which can be linked to the NIH Manuscript Submission System. The data from PubMed is also available through several commercial vendors, such as the Medline database. Although the data is the same, the search platforms are proprietary and can give the searcher varying results while using the same search strategy.

International Pharmaceutical Abstracts (IPA) is a database that is focused on pharmacy and pharmaceutical science and is produced by Thompson Scientific in cooperation with the American Society of Health-System Pharmacists (ASHP). The database indexes and abstracts more than 800 pharmacy and pharmaceutical science related journals and covers drug use and development, as well as pharmacy related topics, including pharmacy education and ethics. IPA is also useful for finding articles on herbs and natural products. In addition to peer reviewed articles, IPA includes meeting abstracts from the meetings of organizations, including the American Pharmacists Association, the American College of Clinical Pharmacy, the American Association of Colleges of Pharmacy, and ASHP.

The Cochrane Library provides reliable and reasonably up-to-date information on the effects of interventions in health care. These interventions may be either therapeutic or diagnostic. There are six separate databases in the Cochrane Library, which may be searched separately or together. The well known Cochrane Systematic Reviews database provides access to the full text of the reviews and the protocols for proposed reviews. There is a Feedback tool that allows users to provide comments and criticisms of Cochrane Reviews and Protocols. The Cochrane Collaboration publishes a list of accepted comments. These comments can be used to further improve and update both Cochrane Reviews and Protocols. Abstracts of non-Cochrane systematic reviews, with a commentary on their

overall quality, are available in the Database of Abstracts of Reviews of Effects. The Cochrane Central Register of Controlled Trials merges the trials listed in Medline, Embase, and other published and unpublished sources. The bibliographic information on publications on the methods used in the conduct of controlled trials is available in the Cochrane Methodology Register. Economic evaluations of healthcare interventions are identified and appraised in the NHS Economic Evaluation Database. Free registration for saved search alerts and RSS feeds from the Cochrane Library is available to the residents of several countries. Cochrane provides a monthly Journal Club publication that introduces a recent Cochrane Review, with a podcast summarizing the review, discussion questions, and key figures and tables in the form of PowerPoint slides. You can also learn more about the Cochrane Library through various social media, including Facebook. The Cochrane Library also includes the *Cochrane Handbook for Systematic Reviews of Interventions*,¹ which provides a thorough methodology for evaluating the clinical research literature and strict guidelines for creating a Cochrane systematic review.

Other heavily used databases in pharmacy and pharmaceutical sciences include such specialized databases as BIOSIS Previews (Biological Abstracts); SciFinder Scholar (Chemical Abstracts and related chemical databases); Toxline and other TOXNET databases, from the US National Library of Medicine; and Embase, a comprehensive, fee-based health science database from Elsevier, which indexes and abstracts many non-English language journals and focuses heavily on pharmaceutical issues. BIOSIS Previews, SciFinder Scholar, and Embase have specialized indexing terms to help the researcher dig deeply into the life science, chemical, and pharmaceutical literature. In addition to indexing journal articles, they also include report, review, patent, and meeting abstracts. The Toxnet databases are a combination of bibliographic and full text databases, which encompass both secondary and tertiary toxicological literature. There is more on these full text databases in the section on Toxicology in the Tertiary literature section.

Important comprehensive citation databases, which allow the researcher to find which articles cited a seminal article, include SciVerse Scopus and Web of Science. Citation searching allows the searcher to see the evolution of an idea and emerging fields of research. SciVerse Scopus from Elsevier includes the data from the Embase database and is quite comprehensive in the sciences and social sciences. SciVerse Scopus also provides links to scientific websites and patents from five patent offices in its search results. Web of Science from Thomson Reuters began several decades ago with the print indexes named Science Citation Index, Social Science Citation Index, and Arts and Humanities Citation Index from the Institute for Scientific Information and has evolved into an extensive, multifunctional set of databases. In both databases, the researcher can save searches and set up search alerts or RSS feeds.

TERTIARY

Tertiary literature provides an introduction to the research literature regarding the practice of pharmacy and the pharmaceutical sciences. It introduces key research findings and accepted concepts. It is usually written by clinical or research specialists within the subject area. These resources are particularly useful to the pharmacy professional and pharmaceutical science graduate students. They comprise textbooks, government monographs, encyclopedias, and other familiar reference works. As we move from print to electronic formats, these resources can provide additional value by allowing students, clinicians, and researchers to make notes, link to related resources, assess learning through online self tests, and share these with interprofessional teams and their peers, colleagues, or instructors. The ability to add more frequent updates to the material online, including videos, will aid the evolution of this literature.

Advantages to tertiary literature include ease of access and well vetted standards of best practice. A major disadvantage

still remaining is that the resources take years to compile and, therefore, can contain dated material.

TEXTBOOKS

As electronic access is making it possible to use the best chapters or sections from multiple resources, the use of textbooks is evolving. Textbooks are often provided as part of electronic packages from vendors, such as AccessPharmacy from McGraw Hill. Electronic book readers also provide greater access and convenience. New editions of classic textbooks are valuable for students. They are also useful to professors or subject specialists, who can read a new edition to see the recent changes in their area. Dipiro's *Pharmacotherapy: A Pathophysiologic Approach*,² *Basic & Clinical Pharmacology*,³ and *Goodman and Gilman's The Pharmacological Basis of Therapeutics*⁴ are all available both in print, as well as electronically through Access Pharmacy. The online resources include additional online updates. The classic title *Foye's Principles of Medicinal Chemistry*⁵ is available through a variety of ebook vendors, as well as in print.

NOMENCLATURE

To locate comprehensive information concerning a drug, the researcher should know the nonproprietary name. Most databases use this name in its controlled vocabulary and may map the brand name to the nonproprietary or generic name. For research into the early development of drugs the chemical names, CAS registry numbers or investigational names are also useful. The two most commonly used resources for naming conventions are the *USP Dictionary of USAN and International Drug Names*⁶ and the World Health Organization's *International Nonproprietary Names* (INN).⁷ The researcher should be aware that the INN name is sometimes different from the US Adopted Name.

The *American Drug Index*⁸ is published annually and includes over the counter drugs, combination products, and drugs currently available in the United States. There is a convenient list of manufacturers and distributors included, along with other useful information at the end of the drug monographs.

The *Merck Index: An Encyclopedia of Chemicals, Drugs, and Biologicals*⁹ provides chemical, nonproprietary, and generic names, as well as graphic chemical structures, CAS Registry Numbers, and chemical formulas. There is a comprehensive cross index to the monographs. The monographs also include useful references to articles, patents, and other tertiary sources on preparation and to comprehensive descriptions of the drug, such as the monograph of the drug in the *Profiles of Drug Substances, Excipients and Related Methodology*.¹⁰ There are also many supplementary tables and organic name reactions. The online version includes structure searching as well.

*Index Nominum: International Drug Directory*¹¹ is a valuable resource for researchers searching for foreign drug substances. Information on over 5,300 drugs includes the therapeutic category and manufacturer or country. There is also information on approximately 10,000 manufactures around the world.

*Japanese Accepted Names for Pharmaceuticals*¹² is a free online database maintained by the Japanese National Institute of Health Sciences (<http://jpub.nihs.go.jp/jan/index.aspx>). It is searchable by the CAS Registry Number, the chemical name, or the Japanese Accepted Name.

PHARMACOPEIAS AND RESOURCES FOR DRUG AND EXCIPIENT STANDARDS

A pharmacopeia is a list of drugs and drug products that describes the purity, strength, method of preparation, and other information. Pharmacopeias are issued or authorized by governments or international agencies. The *United States Pharmacopeia/National Formulary*¹³ (USP/NF) is compiled by a non-governmental organization, the United States Pharmacopeial (USP) Convention (<http://usp.org>), and recognized by the Federal Food, Drug, and Cosmetic (FDC) Act as the official

pharmacopeia of the United States. It contains a list of those drugs, drug products, dietary supplements, excipients, and other relevant compositions for which standards have been agreed to by the USP Convention. The Convention is a representative group of volunteers from the various stakeholders—health-care practitioners, industrial scientists, and educators. The USP/NF is revised yearly and supplemented by updates. The *Pharmacists Pharmacopeia*¹⁴ contains the information from the USP/NF most relevant to practicing pharmacists, along with additional helpful resources, and is updated less often.

The *Japanese Pharmacopeia* is published by the Japanese government through its Pharmaceuticals and Medical Devices Agency (PMDA). It is available in both Japanese and English versions. More information on the Japanese Pharmacopeia can be found at the PMDA website (http://www.pmda.go.jp/english/pharmaco_poeia/about.html).¹⁵ The European Directorate for Quality of Medicines and Healthcare (EDQM) publishes the *European Pharmacopeia*.¹⁶ (More information is available at <http://www.edqm.eu/>). The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) is working with the regulatory authorities of the United States, Japan, and Europe to find and encourage commonalities among these three major pharmacopeias.

Other countries have well-recognized pharmacopeias, such as the *British Pharmacopeia*¹⁷ (BP), published by British Pharmacopoeia Commission Secretariat of the Medicines and Healthcare products Regulatory Agency. *Martindale: The Complete Drug Reference*²² is the new name for Martindale's Extra Pharmacopeia, which lists more than 6,000 drugs found internationally.

Standards for excipients can be found in these pharmacopeias. In addition, the *Handbook of Pharmaceutical Excipients*,¹⁸ although not an official pharmacopeia, is a widely used resource for standards and functionality of excipients. Additionally, since many excipients are also food additives, they can be found in the *Food Chemicals Codex*,¹⁹ published by the United States Pharmacopeial Convention. Finally, the "GRAS list" is a listing of additives that are generally recognized as safe (<http://www.fda.gov/Food/FoodIngredientsPackaging/GenerallyRecognizedasSafeGRAS/GRASSubstancesSCOGSDatabase/default.htm>). The list can be found in searchable form in the Select Committee on GRAS Substances (SCOGS) database. Each of more than 370 entries provides the opinion of the Select Committee and the year in which the opinion was rendered.

DRUG INFORMATION ON PRESCRIPTION PRODUCTS

The most popular resources for finding basic background information on drugs include the following titles: The American Society of Health-System Pharmacists publishes the *AHFS Drug Information: American Hospital Formulary Service*,²⁰ a well recognized reference on prescription drugs. It is organized by therapeutic category, with a general overview of the category, including drug interactions, before the individual listings of the drugs. The succeeding individual drug monographs include the uses, dosage and administration, cautions, pharmacokinetics, chemistry, and stability. Preparations and comparative pricing are included in the online version of this resource. Mobile versions of this resource are also available.

Most pharmacies have traditionally had a loose-leaf copy of *Drug Facts and Comparisons*²¹ on their shelves. Updated monthly, it is a reliable current resource. The drug monographs are arranged by therapeutic category and class hierarchy. The online version, *Facts and Comparisons eAnswers*, includes a variety of tools, including clinical calculators, drug identifiers, and the ability to view comparative data tables and side-by-side monograph summaries of two or more drugs. There is also a mobile version of the databases.

DRUGDEX is part of the *MICROMEDEX* online suite of databases from Thomson Reuters and is often available at hospital pharmacies and health centers. *DRUGDEX Evaluations* mono-

graphs are remarkably thorough and well referenced. Dosing information, pharmacokinetics, cautions, and clinical applications are well covered. Although *DRUGDEX* focuses on prescription drugs, it also includes information on some investigational and nonprescription drugs. There is also a mobile version.

*Martindale: The Complete Drug Reference*²² is a comprehensive listing of worldwide drugs, including herbal medicines. It is also available electronically, through a variety of vendors, and is included in the *MICROMEDEX Healthcare Series*.

Other international drug compendia include *CPS: Compendium of Pharmaceutical Specialties*,²³ from the Canadian Pharmaceutical Association; *Diccionario de Especialidades Farmaceuticas*,²⁴ providing drug and pharmaceutical information from Mexico; *Rote Liste*,²⁵ with complete coverage of drugs and medicines available in Germany; and *Vidal: le dictionnaire*,²⁶ a compendium of French drug and consumer health-care products.

The *Orange Book: Approved Drug Products with Therapeutic Equivalence Evaluations*²⁷ is often simply referred to as the Orange Book. It is available for free at the US Food and Drug Administration website (<http://www.fda.gov/cder/ob>). This government resource is searchable by active ingredient, proprietary name, or by patent number. Therapeutic equivalence, drug approval date, and marketing status (prescription, nonprescription, or discontinued) are listed on a summary page for each drug formulation, as well as patent and exclusivity information.

Trissel's *Handbook on Injectable Drugs*²⁸ summarizes information on parenteral drug stability and compatibility for clinicians. In the most recent edition, there are 372 monographs arranged alphabetically by nonproprietary name. There is also an Appendix with parenteral nutrition formulas.

The *Drug Topics Red Book*²⁹ covers both prescription and nonprescription products. Each drug entry includes product and supplier names, the National Drug Code number, route of administration, strength and quantity, and Average Wholesale Price and Direct Price. This book is used primarily by pharmacists and pharmacy students for drug price information. Additionally, there are many lists, including manufacturers' addresses and phone numbers, state boards of pharmacy, and a product identification guide with color photos.

DRUG INFORMATION ON COMPOUNDED PREPARATIONS

Like manufactured drug products, compounded preparations must be safe and efficacious for the entire period of their use. Existing drugs are prepared or compounded in new dosage forms, sometimes for new routes of administration designed to meet the needs of individual patients. In addition to safety and efficacy, stability, solubility, and permeability are often information required to prepare a new formulation or evaluate an existing formulation. The *United States Pharmacopeia*¹⁴ and the *Pharmacists' Pharmacopeia*³⁰ contain official monographs of compounded preparations that have been evaluated by experts in the field to meet the quality and purity standards for the period prior to the "beyond use date." Additional formulations can be found in standard texts, such as *The Art Science and Technology of Pharmaceutical Compounding*.³¹ Formulations specific for certain patient populations are found in references, such as *Plumbs' Veterinary Drug Handbook*³² and *Pediatric Drug Formulations*.³³ Although many texts provide formulation and preparation information, one must generally search the primary literature to find the limited data on efficacy of these specialized dosage forms. International Pharmaceutical Abstracts (IPA) is the preferred database for finding such information. IPA is also a useful database to find stability information for those drugs and drug preparations not listed in *Trissel's Stability of Compounded Formulations*.³⁴ Journals that often contain new formulations along with stability data are *International Journal of Pharmaceutical Compounding* (Edmond, OK; IJPC) and *American Journal of Health-System Pharmacy* (Bethesda, MD; American Association of Health-System Pharmacists).

DRUG INFORMATION ON NONPRESCRIPTION PRODUCTS

Although there is some information on nonprescription drug products in the prescription drug resources listed previously in this chapter, the most valuable resource for this information is the *Handbook of Nonprescription Drugs: An Interactive Approach to Self-Care*,³⁵ which is available both in print and in PharmacyLibrary, the new electronic package from the American Pharmacists Association. The handbook is arranged into sections beginning with the practitioner's role in self-care, followed by sections on diseases, with chapters arranged in each section by body systems; a section on home medical equipment; and a section on complementary and alternative medicine. To help pharmacy students develop problem-solving skills, case studies are provided in each section. Chapter updates are available for the online version.

The *Physicians' Desk Reference for Nonprescription Drugs, Dietary Supplements and Herbs*³⁶ is a useful supplement to the *Handbook of Nonprescription Drugs: An Interactive Approach to Self-Care*.³⁵ It is available in print and electronically.

HERBAL MEDICINES AND NATURAL PRODUCTS

Herbal medicines and natural products are very popular with consumers, who often perceive the terms "natural" or "herbal" as connoting safety. Hence, there is active research in this area, and students and pharmacists need to be well informed. Luckily, there are a variety of well known textbooks and tertiary databases on this subject. *Trease and Evans' Pharmacognosy*³⁷ is a classic, basic textbook, now in its 16th edition. It focuses on the use of plants in medicine, covering plant taxonomy, commercial production, biological activity, and phytochemical examination and investigation of herbal products. *Tyler's Honest Herbal: A Sensible Guide to the Use of Herbs and Related Remedies*³⁸ provides well-referenced monographs on herbs and their efficacy and safety. It is a good starting point for researchers and useful for consumers, although consumers may view the standards for rating efficacy as too conservative. *Leung's Encyclopedia of Common Natural Ingredients Used in Food, Drugs, and Cosmetics*³⁹ is a more extensive resource covering the source, chemical composition, pharmacology and biological activity, uses, and commercial preparation of each natural ingredient. It is extensively referenced, with appendices providing glossaries of abbreviations and botanical terms and morphological descriptions of plant organs.

Online resources provide search options and special features. Two well known natural product databases are Natural Standard and Natural Medicines Comprehensive Database. Natural Standard provides evidence-based information about alternative and complementary therapies and has a variety of databases. Each therapy is graded according to the evidence for its effectiveness in treating a particular medical condition. Natural Standard databases include Foods, Herbs, and Supplements; Comparative Effectiveness; and Genomics and Proteomics. Natural Medicines Comprehensive Database from Therapeutics Research Center covers natural medicines sold in North America and has evidence-based comprehensive monographs on these medicines, as well as a natural product effectiveness checker. Checkers for interactions and depletions, patient education handouts, and Continuing Education Programs are also available with both databases. Both natural product databases have mobile and print versions and provide RSS feeds.

DRUG INTERACTIONS, SIDE EFFECTS AND ADVERSE REACTIONS

Patients and pharmacy practitioners are very concerned about interactions, side effects, and adverse reactions. These also concern pharmaceutical scientists developing new products.

*Meyler's Side Effects of Drugs: An Encyclopedia of Adverse Reactions and Interactions*⁴⁰ has a long history. According to the Preface of the 15th edition of *Meyler's Side Effects of Drugs:*

An Encyclopedia of Adverse Reactions and Interactions, Leopold Meyler, a Dutch physician, experienced adverse drug effects in the 1940s, while undergoing treatment for tuberculosis. He searched for information on this topic and, finding only a 19th century tome, he went to work to fill this significant gap. The first edition of this work was published in Dutch in 1951, and an English edition was published in 1952. Editions of *Meyler's Side Effects of Drugs: An Encyclopedia of Adverse Reactions and Interactions* are updated by *Side Effects of Drugs Annual*⁴¹ and cover prescription drugs, anesthetics, antiseptics, drugs of abuse, herbal medicines, and devices and methods in alternative medicine. *Meyler's Side Effects of Drugs: An Encyclopedia of Adverse Reactions and Interactions* is available both in print and electronic formats.

*Drug Interactions Analysis and Management*⁴² is written by two well-known authorities in this area, Philip D. Hansten, PharmD and John R. Horn, PharmD. The focus is on the management of drug interactions and gives the reader options to avoid patient harm and decrease risk. Suggestions regarding monitoring and alternative drugs are recommended when appropriate. Specific information on each interaction also includes risk factors, the mechanism of action, a clinical evaluation of cited reports, closely related drugs that may interact in a similar manner, and references. Each interaction is also assigned a Significance Number, which is based on the intervention needed to minimize risk.

An important specialty resource is *Drugs in Pregnancy and Lactation: A Reference Guide to Fetal and Neonatal Risk*.⁴³ This resource provides fetal risk and breast feeding risk summaries, which include animal reproduction data, placental transfer, reports of human pregnancy exposure, and a list of references for more than 1200 drugs. A related database is the Drugs and Lactation Database (LactMed), one of the TOXNET set of databases. It focuses on the levels of the drug in breast milk and infant blood and the possible adverse effects on the nursing infant.

Another special topic is covered by *Food-medication Interactions*,⁴⁴ available either as a spiral bound book or a mobile version. More than 1000 drug monographs are arranged alphabetically and cover drug class or action; side effects; renal, hepatic, cardiac, and pregnancy information; contraindications; monitoring information; and special dietary precautions and nutritional effects. There are many tables, including a useful laboratory value table. Canadian brand names are exclusive to the mobile version.

Medication safety is a major concern. *Drug-induced Diseases: Prevention, Detection, and Management*⁴⁵ provides information for students and health care practitioners to address and improve the standard of care regarding drug-induced diseases. This comprehensive textbook introduces the magnitude of the problem and follows up with sections arranged by anatomical region, such as drug-induced gastrointestinal diseases.

POISONING AND TOXICOLOGY

Two well known textbooks on toxicology are *Casarett and Doull's Toxicology: The Basic Science of Poisons*⁴⁶ and *Goldfrank's Toxicologic Emergencies*.⁴⁷ *Casarett and Doull's Toxicology* covers the various aspects of the subject, including toxicokinetics, carcinogenicity, mutagenicity, developmental toxicology, target organ toxicity, toxic agents, environmental toxicology, food toxicology, analytic toxicology, clinical toxicology, and occupational toxicology. *Goldfrank's Toxicologic Emergencies* presents in-depth information on antidotes, general principles, and techniques to evaluate and manage toxic exposures, biochemical toxicology, organ system principles, and classes of compounds. Case studies and study questions are given, and the answers are provided at the end of the book.

TOXNET is a set of databases made available by the US National Library of medicine (<http://toxnet.nlm.nih.gov/>). They provide indexing and abstracting and full text information concerning toxicology, hazardous chemicals, environmental

health, and toxic releases. Databases may be searched individually or as a set. The Hazardous Substances Databank (HSDB) contains more than 5000 chemical records and is enhanced with additional material on nanomaterials, regulatory requirements, and emergency handling procedures. Other TOXNET databases include the Chemical Carcinogenesis Research Information System (CCRIS), Integrated Risk Information System (IRIS), GENE-TOX, and the previously described LactMed database. The National Cancer Institute provides data on mutagenicity and tumor promotion and inhibition in CCRIS. IRIS contains data in support of human health risk assessment and is compiled by the US Environmental Protection Agency (EPA). GENE-TOX, created by the US EPA, contains genetic toxicology (mutagenicity) test data on more than 3000 chemicals and recommends proper testing protocols and evaluation procedures.

FORMULATION AND MANUFACTURING

Many of the standard textbooks on formulation and manufacturing are now quite dated. There is a need for additional educational material in this area. The National Institute of Pharmaceutical Technology and Education (<http://nipte.org/>), which is dedicated to fundamental research and education in pharmaceutical product development and manufacturing, has developed a draft curriculum in this area; development of educational resources is a planned activity of the group. In the meantime, the *Encyclopedia of Pharmaceutical Technology*⁴⁸ has monographs on many of the aspects related to formulation and the various pharmaceutical manufacturing processes. For detailed information on drug substances necessary for pre-formulation research, Britain's *Profiles of Drug Substances, Excipients, and Related Methodology*¹⁰ provides comprehensive monographs. Additionally, PharmaHub (<http://pharmahub.org/>) has an increasing array of useful databases and process modeling software for formulation and manufacturing.

PRODUCT IDENTIFICATION

Emergency and other medical personnel need to be able to quickly identify capsules or tablets. Many resources include sections on identifying drug or related products. These include the *Physician's Desk Reference*, the *Red Book*, and several online and mobile resources, including Facts and Comparisons eAnswers and Lexi-Comp. The two most well-known titles are *Ident-a-drug Reference*,⁴⁹ which identifies the drug by the color, shape, size, and imprint code, and the *IDENTIDEX System*, a part of *MICROMEDEX*.

CONSUMER DRUG INFORMATION

It is important to know where to find reliable and accurate consumer drug information, especially since consumers frequently use the internet to find information. There are many such resources on the internet. MedlinePlus from the US National Library of Medicine (<http://www.medlineplus.gov>) includes a vast array of features for the consumer, including information on drugs and supplements and a medical dictionary with pronunciation. Drugs.com includes an Interactions Checker. WebMD includes a useful Pill Identifier tool.

An example of expanded access to drug information on the internet is the Drug Information Portal (<http://druginfo.nlm.nih.gov>), which provides a search interface to information on more than 20,000 drugs from US government agencies, including the National Library of Medicine. There is information for researchers, clinicians, students, and consumers.

Two popular print resources include the *Pill Book*⁵⁰ and the *Complete Guide to Prescription and Nonprescription Drugs*.⁵¹ The *Pill Book* includes information on affordable generic alternatives, side effects, adverse effects, drug-drug and drug-food interactions, addictiveness, safe handling of injectables, and when to call the physician. Aside from the usual drug information, the *Complete Guide to Prescription and Nonprescription Drugs* includes the length of time before a drug starts working and warnings regarding premature discontinuation of a drug.

At the beginning of the book, there is a list of commonly used drugs to treat specific diseases or conditions.

PERSONALIZED MEDICINE

As pharmacists become more focused on the individual patient, pharmacogenetics is becoming more important than ever before. Two valuable resources in this area are Ginsberg and Willard's *Essentials of Genomic and Personalized Medicine*⁵² and Catania's *An A-Z Guide to Pharmacogenomics*.⁵³ *Essentials of Genomic and Personalized Medicine* provides an overview of the field, with translational approaches to bring the information into the clinical world. *An A-Z Guide to Pharmacogenomics* introduces the terminology and techniques related to this growing field to students and professionals who have a basic understanding of molecular biology.

PROFESSIONAL ASSOCIATIONS

Associations for almost every specialty of pharmacy in the various regions of the United States and around the world can be found at the Virtual Pharmacy Library (<http://www.pharmacy.org/>) by clicking on the Associations tab. The list of more than 100 pharmacy-related associations is complete, with links to the association's website.

EMERGING TRENDS

The majority of electronic books are still simply copies of the original print versions. Although they can be accessed from anywhere at any time, they are still rather static. According to the *2011 Horizon Report*,⁵⁴ the product of collaboration between the EDUCAUSE Learning Initiative (ELI) and the New Media Consortium, this is rapidly changing, and the newer electronic books with added features will become more in demand than the print. Social features will connect the reader to other researchers, enabling collaboration. Self-directed, interactive experiences and activities are beginning to be provided. Videos and other audiovisual resources will integrate into these electronic resources. Links to supporting materials can enrich the reading experience. Cloud tags will make locating useful chapters or sections vastly simpler.

Mobile resources have been used by pharmacy professional students, pharmacists, and faculty for several years. According to the *2011 Horizon Report*, "Internetcapable mobile devices will outnumber computers within the next year." With the spread of smart phones and other devices, students and scientists will demand more resources and applications compatible with the platforms they have chosen. One example is PubMed on Tap, which allows the user to search PubMed and manage references from the iPod Touch, iPhone, or iPad.

Data planning and management is becoming a major thrust in the sciences. Websites, such as pharmaHUB at pharmahub.org, were created to facilitate the collaborative creation and sharing of pharmaceutical engineering and science information and modeling tools. The PharmaHUB website states that their goal is to "support innovations in product and process development and manufacturing methodology for pharmaceutical products."⁵⁵ This is an area in which the amount of data can be massive, and this data has been isolated in individual research groups or institutes until now. Sharing the data will allow faster progress in the research area.

FOR FURTHER INFORMATION

The American Association of Colleges of Pharmacy Libraries/Educational Resources Section's Basic Resources for Pharmacy Education (<http://www.aacp.org/governance/SECTIONS/libraryinformationscience/Pages/LibraryInformationscienceSpecialProjectsandInformation.aspx>) provides a benchmark to use in selecting many of the resources listed herein and is a good place to explore for more references in a specific pharmacy related field. An excellent book on pharmaceutical resources is Bonnie Snow's *Drug Information: A Guide to Current Resources*.⁵⁶

ACKNOWLEDGMENTS: The previous version of this chapter was written by three colleagues at the University of the Sciences in Philadelphia: Leslie Ann Bowman, Mignon S. Adams and Amy Christopher. Although we have chosen to organize our material in our own way, reviewing their version has provided a valuable point of reference.

REFERENCES

- Higgins J et al. *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester England/Hoboken NJ: Wiley-Blackwell, 2008.
- DiPiro JT. *Pharmacotherapy: A Pathophysiologic Approach*, 8th edn. New York: McGraw-Hill Medical, 2011.
- Katzung BG et al. *Basic Clinical Pharmacology*, 12th edn. New York/London: McGraw-Hill Medical (McGraw-Hill distributor), 2011.
- Goodman LS et al. *Goodman & Gilman's the Pharmacological Basis of Therapeutics*, 12th edn. New York: McGraw-Hill, 2011.
- Foye WO et al. *Foye's Principles of Medicinal Chemistry*, 7th edn. Philadelphia: Lippincott Williams & Wilkins, 2013.
- USP Dictionary of USAN & International Drug Names 2011*. Rockville MD: United States Pharmacopeial Convention, 2011.
- International Nonproprietary Names (INN) for Pharmaceutical Substances: Lists 1-101 of Proposed INN and Lists 1-62 of Recommended INN, Cumulative List no. 13*. Geneva: World Health Organization, 2009.
- Wilson CO. *American Drug Index*. Philadelphia: Lippincott, 2012.
- O'Neil MJ. *The Merck Index: An Encyclopedia of Chemicals, Drugs, and Biologicals*. Whitehouse Station NJ: Merck, 2006.
- Brittain HG, ed. *Profiles of Drug Substances, Excipients, and Related Methodology*. Vol. 37. Amsterdam/Boston: Elsevier Academic Press, 2012.
- Pharmasuisse, ed. *Index Nominum: International Drug Directory*, 20th edn. Stuttgart: MedPharm, 2011.
- Japanese National Institute of Health Sciences. *Japanese accepted names for pharmaceuticals*. www.jpdb.nihs.go.jp/jan/index.aspx (accessed January 2013).
- United States Pharmacopeial Convention. *The United States Pharmacopeia: The National Formulary*. Rockville MD: United States Pharmacopeial Convention, 2012.
- United States Pharmacopeial Convention. *USP Pharmacists' Pharmacopeia*. Rockville MD: United States Pharmacopeial Convention, 2009.
- Pharmaceuticals and Medical Devices Agency, Japan. *About JP*. www.pmda.go.jp/english/pharmacopoeia/about.htm (accessed 2 January 2013).
- Council of Europe. *European Pharmacopoeia*. Strasbourg: Council of Europe, 2010.
- British Pharmacopoeia Commission. *British Pharmacopoeia 2013*. London: Stationery Office, 2013.
- Rowe RC et al. *Handbook of Pharmaceutical Excipients*. London: Pharmaceutical Press, 2012.
- United States Pharmacopeial Convention. *Food Chemicals Codex: By Authority of the United States Pharmacopeial Convention*. Rockville MD: United States Pharmacopeial Convention, 2012.
- American Society of Health-System Pharmacists. *AHFS Drug Information 2012*. Bethesda MD: American Society of Health-System Pharmacists, 2012.
- Facts and Comparisons. *Drug Facts and Comparisons*. St. Louis: Wolters Kluwer Health, 2012.
- Sweetman SC, ed. *Martindale: The Complete Drug Reference*, 37th edn. London/Chicago: Pharmaceutical Press, 2011.
- Repehinsky C, ed. *CPS 2012: Compendium of Pharmaceuticals and Specialties; the Canadian Drug Reference for Health Professionals*. Ottawa ON: Canadian Pharmacists Association, 2012.
- Rodríguez Vásquez A, Domínguez Frías G. *Diccionario De Especialidades Farmacéuticas*, 56a edn. Mexico: Thomson, 2010.
- Bundesverband der Pharmazeutischen Industrie (Germany), Rote Liste Service. *Rote liste: Arzneimittelverzeichnis für Deutschland (einschließlich EU-Zulassungen und bestimmter Medizinprodukte)*. Frankfurt am Main: Verl. Rote Liste, 2012.
- Roquet I, ed. *Vidal De La Famille: Le Dictionnaire Des Médicaments*, 17th edn. Issy-les-Moulineaux: Vidal, 2012.
- Department of Health and Human Services (2012). *Orange Book: Approved Drug Products with Therapeutic Equivalence Evaluations*. www.fda.gov/cder/ob/ (accessed 2 January 2013).
- Trissel LA. *Handbook on Injectable Drugs*, 16th edn. Bethesda MD: American Society of Health-System Pharmacists, 2011.
- Thomson Corporation. *Red Book: Pharmacy's Fundamental Reference*. Montvale NJ: Thomson PDR, 2010.
- United States Pharmacopeial Convention. *Pharmacists' Pharmacopeia*, 2nd edn. Rockville MD: United States Pharmacopeial Convention, 2009.
- Allen LV. *The Art, Science, and Technology of Pharmaceutical Compounding*, 4th edn. Washington DC: American Pharmacists Association, 2012.
- Plumb DC. *Plumb's Veterinary Drug Handbook*, 7th edn. Hoboken NJ: Wiley-Blackwell, 2011.
- Nahata MC, Pai VB. *Pediatric Drug Formulations*, 6th edn. Cincinnati OH: Harvey Whitney Books, 2011.
- Trissel LA. *Trissel's Stability of Compounded Formulations*, 5th edn. Washington DC: American Pharmacists Association, 2012.
- Berardi R. *Handbook of Nonprescription Drugs an Interactive Approach to Self-care*, 17th edn. Washington DC: American Pharmacists Association, 2012.
- Physicians' Desk Reference*, 66th edn. Montvale NJ: Medical Economics Company Inc., 2012.
- Evans WC et al. *Trease and Evans Pharmacognosy*, 16th edn. Edinburgh: Saunders/Elsevier, 2009.
- Foster S, Tyler VE. *Tyler's Honest Herbal: A Sensible Guide to the use of Herbs and Related Remedies*, 4th edn. New York: Routledge, 2009.
- Khan IA, Abourashed EA. *Leung's Encyclopedia of Common Natural Ingredients used in Food, Drugs, and Cosmetics*, 3rd edn. Hoboken NJ: John Wiley & Sons Inc., 2010.
- Aronson JK et al. Side effects of drugs. In: Aronson JK, ed. *Meyler's Side Effects of Drugs: The International Encyclopedia of Adverse Drug Reactions and Interactions*, 15th edn. Oxford: Elsevier, 2006: pages.
- Dukes MNG. *Excerpta Medica Foundation*. In: Aronson JK, ed. *Side Effects of Drugs Annual*. Amsterdam: Elsevier, 2011: pages.
- Hansten PD, Horn JR. *Drug Interactions: Analysis and Management*. St. Louis MO: Wolters Kluwer Health/Facts & Comparisons, 2012.
- Briggs GG et al. *Drugs in Pregnancy and Lactation*, 9th edn. Philadelphia: Wolters Kluwer Health, 2011.
- Pronsky ZM. *Food Medication Interactions*, 16th edn. Birchrunville PA: Food-Medication Interactions, 2010.
- Tisdale JE et al. *Drug-Induced Diseases: Prevention, Detection, and Management*, 2nd edn. Bethesda MD: American Society of Health-System Pharmacists, 2010.
- Casarett LJ et al. *Casarett and Doull's Toxicology: The Basic Science of Poisons*, 8th edn. New York: McGraw-Hill, 2013.
- Goldfrank LR, Flomenbaum N. *Goldfrank's Toxicologic Emergencies*, 9th edn. New York: McGraw-Hill Medical, 2011.
- Swarbrick J, ed. *Encyclopedia of Pharmaceutical Technology*. New York: Informa Healthcare, 2007.
- Therapeutic Research Center. *Ident-a-Drug Reference 2011 for Tablet and Capsule Identification*. Stockton CA: Therapeutic Research Faculty, 2010.
- Silverman HM. *The Pill Book*, 15th edn. New York: Bantam Books, 2012.
- Griffith HW et al. *Complete Guide to Prescription & Nonprescription Drugs*. New York: Penguin Group, 2011.
- Willard HF, Ginsburg GS. *Essentials of Genomic and Personalized Medicine*. Amsterdam: Elsevier/Academic Press, 2010.
- Catania MG. *An A-Z Guide to Pharmacogenomics*. Washington DC: AACCC Press, 2006.

54. New Media Consortium and EDUCAUSE. *2011 Horizon Report*. net.educause.edu/ir/library/pdf/HR2011.pdf (accessed 2 January 2013).
55. PharmaHUB. *About Us*. www.pharmahub.org/about (accessed 2 January 2013).
56. Snow B, Medical Library Association. *Drug Information: Guide to Current Resources*, 3rd edn. New York: Neal-Schuman Publishers, 2008.

Chapter 2

Analysis of Medicinals

Raymond D. Skwierczynski, PhD

SPECIFICATIONS 9

COMPENDIA 9

MONOGRAPHS 10

RELEASE AND STABILITY TESTING 10

UNIVERSAL TESTS 10

METHOD VALIDATION FOR UNIVERSAL TESTS 11

TAXONOMY 15

DOSAGE FORMS—INJECTABLES 16

DOSAGE FORMS—ORAL DRUG PRODUCTS 17

ADDITIONAL DOSAGE FORMS 19

FUTURE DIRECTION 19

APPENDIX A 22

APPENDIX B 24

The aim of this chapter is to provide the pharmaceutical scientist with a general overview of the analytical testing that is performed in order to release a medicinal product for clinical or commercial use and specifications that ensure the quality and performance of the medicinal product being released.

SPECIFICATIONS

The pharmaceutical manufacturer is responsible for ensuring the quality, purity, identity, and strength of each lot of drug product manufactured. One mandatory control strategy is to ensure that the lot “conforms to specification,” which means that the drug product (formulated preparation), when tested to the listed analytical procedures, will meet the listed acceptance criteria.¹ Specifications are critical quality standards.

A specification is a document that is defined as a list of tests, references to analytical procedures, and appropriate acceptance criteria (numerical range or limit).¹ The pharmaceutical manufacturer justifies the information on the specification, which is approved by regulatory authorities. Examples of specifications for hypothetical drug products are provided in Appendix A.

Specifications can be categorized in a couple of manners. One way is by the intended use of the product. For example, is the medicinal an investigational product intended for use in a clinical study, or is it a commercial product that will be marketed? Let's consider the evolution of a hypothetical oral drug product as it enters the clinic for single- and multiple-ascending dose safety studies during Phase 1 clinical studies, then multiple-ascending dose efficacy studies in Phase 2a, and confirmatory efficacy studies in Phase 2b and Phase 3. In order to enter the clinic rapidly, a simple drug product called a drug-in-capsule was chosen for Phase 1. This dosage form consists of neat drug (no excipients) that has been accurately weighed into gelatin capsules. The manufacturing process meets the volume demand for Phase 1, but the dosage form must be changed prior to scaling-up to Phase 2. Hence, the drug-in-capsule is a “lame duck” dosage form. Specifications for investigational products tend to have acceptance criteria that reflect the early life of the product. Here, the acceptance criteria for dissolution of our drug-in-capsule may be “record result.”

Now let's say that our drug-in-capsule drug product will be replaced with a dosage form that is manufactured using a scalable manufacturing process — a tablet for Phase 2. The tablet will first be used in Phase 2 ascending-dose clinical studies, so several dose strengths will need to be manufactured. Let's say that these dose strengths are 10, 25, and 100 mg. The acceptance criterion for the dissolution test may remain as “record result” because the dosage form is new, or the criterion may be a “disaster check” criterion. Although the analytical chemist may

not be able to assign a strict acceptance criterion because of the limited data, a slow-dissolving drug from an immediate-release tablet may jeopardize the interpretation of the clinical results. So a criterion like 70 percent of the drug dissolved after 60 minutes may be appropriate for the 100-mg tablet. The lower dose strengths are usually assigned the same acceptance criterion.

Finally, the tablet has entered late development and will be used for confirmatory studies. The dose ranging Phase 2a studies determined that the efficacious dose was 60 mg, so a new 60-mg tablet was developed for use in the Phase 3 studies. During Phase 3, a robust manufacturing process has been developed at suitable scale, a robust dissolution test also has been developed and validated, and many lots of tablets have been manufactured and analyzed. Now, a suitable acceptance criterion at the time of NDA or MAA submission may be 70 percent of the drug dissolved after 30 minutes.

Another common categorization involves using different acceptance criteria for release testing and stability testing in order to account for changes that may occur during the shelf life. For example, a topical cream may become less viscous upon storage at 25°C to 30°C for two years. The cream may be too fluid to apply when its viscosity is less than 1000 mPa·s. A suitable acceptance criterion for a stability specification may be 2000 mPa·s. Let's say that the development data indicate that in order to ensure a viscosity of 2000 mPa·s at the end of the shelf life, the cream must have a viscosity of 10,000 mPa·s at the time of release. Therefore, the acceptance criteria for the release and stability specifications would be not less than (NLT) 10,000 mPa·s and NLT 2000 mPa·s, respectively, in order to ensure a quality product for its entire shelf life.

COMPENDIA

All drug products, whether commercial or investigational, must meet standards that have been established by Pharmacopeial Conventions or Regulatory Agencies. The United States Pharmacopeial (USP) Convention is a scientific nonprofit organization that sets standards for the quality, purity, identity, and strengths of medicines, food ingredients, and dietary supplements manufactured, distributed, and consumed worldwide.²⁻⁴ USP drug standards are enforceable in the United States by the Food and Drug Administration, and these standards are developed and relied upon in more than 130 countries.^{2,4}

The European Pharmacopeia (Ph. Eur.) Commission establishes official standards that provide a legal and scientific basis for quality control during the development, production, and marketing of medicines in 37 signatory states of the convention.^{5,6} In addition to the signatory states, which comprise 26 countries and the European Union, there are a large number

of observers of Ph. Eur. Twenty-two countries (including Australia, Brazil, Canada, China, the Russian Federation, and the United States) and the World Health Organization are listed as observers. Consequently, the standards developed by Ph. Eur. have an impact on the quality of medicines across the globe.^{5,6}

The USP and Ph. Eur. publish books of pharmacopeial standards. The thirty-fourth revision of the USP and the twenty-ninth edition of the National Formulary, USP 34-NF29, became official on May 1, 2011. The USP-NF is continuously revised. Standard revisions are found in supplements to the USP-NF that are published twice yearly. Accelerated revisions are published in Pharmacopeial Forum (PF) and on the USP website, www.usp.org. A new edition of USP-NF becomes official each year on May 1. Chapters <1> through <999> are enforceable chapters, while Chapters <1000> through <1999> are informational chapters.³

The seventh edition of the *European Pharmacopeia* (Ph. Eur. 7.0) became official on January 1, 2011. This edition will be augmented with eight supplements over a three-year period.⁵ The eighth edition of Ph. Eur. will become official on January 1, 2014. The *European Pharmacopeia* is published in English and French.

The British Pharmacopeia (BP) is the official collection of standards for UK medicinal products and pharmaceutical substances.⁷ The standards are established by the British Pharmacopeia Commission. Canada and Australia also use the BP as their official standards. The BP is recognized in over 100 countries as an internationally acceptable standard and remains an essential reference for all individuals and organizations working within pharmaceutical research, development, manufacture, and testing across the globe.⁸ A new edition of the BP becomes official each year on January 1.

Japan's Ministry of Health, Labor, and Welfare (MHLW) publishes The Japanese Pharmacopeia (JP), which provides an official standard to ensure the quality of medicines in Japan.⁹⁻¹¹ The Japanese Pharmacopeia, Sixteenth Edition (JP 16) became official on April 1, 2011. JP 16 is printed in Japanese only at the time of publication of this chapter. JP 15, however, is available in English. Both editions can be found on The Japanese Pharmacopeia website and are free of charge.

MONOGRAPHS

A monograph is written after a drug product has been registered and established in the marketplace and usually before its patent expires. The USP defines a monograph as a written standard that describes an article (e.g., drug substance, drug product, excipient, compounded preparation).¹² A monograph published in any USP compendium provides the name of a substance; its definition; package, storage, and labeling requirement; and information on tests needed to ensure the substance is of the appropriate identity, strength, quality, and purity.³ The later part of a monograph is similar in scope to a specification.

The monograph gives manufacturers, governments, and scientists a public standard by which to judge an article's quality. Monographs play an important role in meeting the USP's mission by providing standards for substances consumed in a global marketplace—standards that help maintain public health. The USP-NF comprises more than 4000 monographs.¹² The USP contains monographs for drug substances and preparations (drug products); excipient monographs are in the NF.

The *European Pharmacopeia* contains more than 2000 general and specific monographs, including chemical substances, antibiotics, vaccines, dosage forms, herbal drugs, and homeopathic preparations.⁵ One major difference between the Ph. Eur. and USP is illustrated by this example: the USP has monographs for tablets of a specific drug (e.g., amoxine), whereas Ph. Eur. will have only a general monograph for tablets. The BP and JP also contain monographs of drug substances and drug products. The BP, which contains over 3000 monographs, incorporates monographs of the *European Pharmacopeia*.¹³

RELEASE AND STABILITY TESTING

The release and stability testing can be classified in a couple of practical ways. One is by the nature of the testing performed: chemical, physical, and microbial. For example, the assay for the amount of active ingredient in a gel would be a chemical test, the viscosity measurement and the *in vitro* release test would be physical tests, and the microbial limits test would be a microbiological test. This classification system is often used in analytical laboratories that are specialized by the type of test being performed. The chapters in pharmacopeia are also organized by the nature of the test.^{3,5,9,10}

Another classification system that is used is based on the attribute that is being tested. For example, the ICH Guidances and the USP provide two categories for the system: analytical tests and acceptance criteria for assessing (1) general quality attributes and (2) product performance.^{1,14-17} Using the previous gel example, the assay, viscosity, and microbial limits tests would be examples of tests that assess general quality attributes while the *in vitro* release test would assess product performance.

UNIVERSAL TESTS

There are four tests that are generally applicable to all drug products: description, identification, assay, and impurities.

DESCRIPTION

This test is often called “appearance” on a specification and is a qualitative description of the dosage form. For example, the description of a tablet on a specification may read: white, round, biconvex, film-coated tablet, imprinted with “400” on one side.

IDENTIFICATION

The purpose of an identification or identity test is to verify the identity of the active pharmaceutical ingredient(s) (API) in the dosage form. This test should be able to discriminate between compounds of closely related structure that are likely to be present. Infrared and Raman spectroscopy are commonly used techniques. A more practical technique, however, is high-performance liquid chromatography because the identity and assay can be determined using the same analytical method. The ICH Q6A Guidance, however, does not regard identification solely by a single chromatographic retention time as being specific, the definition of which is the ability of the method to assess unequivocally the analyte in the presence of the sample matrix (excipients, impurities, and degradation products). Thus, some laboratories will couple a diode array or mass spectrometric (MS) detector to the HPLC system. The diode array detector provides a UV spectrum of the drug; the MS detector gives the nominal mass of the drug.

ASSAY

This test determines the strength or content of the API in the dosage form and is sometimes called a content test. The method should be stability-indicating, which means that the method is quantitative and specific and can detect chemical changes with respect to time, so that the quantity of the active ingredient(s) can be accurately and precisely measured in the presence of the sample matrix. HPLC is the most common technique used for stability-indicating methods.

IMPURITIES

ICH Guidance Q6A defines an impurity in a drug product as any component that is not the API or an excipient. The most common type of impurities that are measured is related substances, which are process impurities from the new drug substance synthesis, degradation products of the API, or both. The test for related substances is often referred to as a purity test and must be stability indicating. The method can be the same as that used for assay, or it can be a different method that has been developed for measuring low-level impurities. An additional method

to measure chiral purity may be used to quantify enantiomeric impurities in the drug product. In cases in which an organic solvent is used during the manufacture of the drug product, a method for measuring residual solvents is also used.

METHOD VALIDATION FOR UNIVERSAL TESTS

The objective of validation of an analytical method is to demonstrate that it is suitable for its intended purpose.^{3,5,18,19} Therefore, the objective of the analytical method will govern the validation characteristics that need to be evaluated. Typical validation characteristics are specificity, accuracy, precision, linearity, range, robustness, detection limit, quantitation limit, and sample and standard stability. Well-characterized reference materials, with documented purity, should be used throughout the validation study.¹⁹ A validation protocol that describes the validation experiments to be performed and their pre-determined validation performance criteria is required. In early development, a department Standard Operating Procedure (SOP) can be used instead of a protocol, provided that the SOP clearly outlines the validation experiments and validation performance criteria. Let's examine the characteristics evaluated for each of the four universal tests and typical acceptance criteria that accompany each test on the specification.

ASSAY

Acceptance Criteria

The results from an assay are usually expressed as a percentage of the label claim with acceptance criteria that are typically in the range from 90.0 to 110.0 percent label claim. For example, a manufactured lot of 500-mg aspirin tablets with these acceptance criteria would have an average assay value between 450 and 550 mg. The width of these limits is to allow for manufacturing variability and shelf-life stability. In some cases, such as narrow-therapeutic drugs, the acceptance criteria may be from 95.0 to 105.0 percent label claim.

Validation Characteristics

Assay methods are validated for specificity, linearity, range, accuracy, precision, robustness, and sample and standard stability.^{18,19} Validation of detection limit and quantitation limit is not needed unless the same method is used to quantify related substances.

SPECIFICITY

Most assays are performed using HPLC. Specificity can be demonstrated in a number of ways. A common approach is to spike each impurity and excipient into the analytical solution at an appropriate level and compare the results of the spiked samples against those of the unspiked samples to demonstrate that the presence of the impurities or excipients does not affect the assay result. Of course, this approach works only if authentic samples of the impurities are available.

If the impurities are not available, the API, or the dosage form, is subjected to forced degradation conditions, and the degradants and synthetic process impurities are separated from the API, using the method that is being validated, as well as other methods that make use of different stationary phases (chromatographic columns) and chromatographic conditions (mobile phase components and gradients, flow rate). Diode array or MS detection is often used. The goal here is to demonstrate that no impurity co-elutes with the API in the validated method.

The forced degradation conditions that are often used in the analytical laboratory for method development and validation are dissolving the API (with or without excipients) in an aqueous or aqueous-organic solvent and exposing the solution to acidic, basic, and nucleophilic and free radical oxidation conditions. It is important to control the degradation to less than about 10 percent in order to make the forced degradation

experiments representative of the degradation that might be seen during the shelf-life of the drug product and to decrease the probability of the degradants reacting with each other or degradants reacting with the API.

In addition, the API and the dosage form should be exposed to a variety of moderately stressed stability conditions, such as light and prudently selected combinations of heat and humidity in order to degrade the material "purposefully." Some common storage conditions for purposeful degradation of a solid-dosage form include 40°C and 75 percent relative humidity (40/75) in high-density polyethylene (HDPE) bottles, 40/75 in an open dish (containerless), 60/75 in HDPE bottles, and 60/75 in an open dish. Some pharmaceutical companies have formed purposeful-degradation groups that are tasked with identifying "soft spots" on API molecules, understanding the degradation chemistry, identifying impurities, conducting purposeful degradation studies, and proposing stabilization strategies to prevent or inhibit chemical degradation.²⁰

If the API is volatile, gas chromatography (GC) can be used for the assay test. Just as with HPLC, more than one method should be used to demonstrate specificity. Gas chromatography with mass spectrometric detection (GC-MS) is useful to accomplish this goal. In cases in which a non-specific assay, such as titrimetry or UV spectrophotometry, is used, other supporting methods and knowledge of the degradation chemistry can be used to demonstrate overall specificity. The use of non-specific methods to determine assay, however, is not recommended, given the advances in chromatographic and detection technology. The USP is actively modernizing official USP-NF monographs for small-molecule API and their dosage forms that use outdated analytical technology or have non-specific identification and assay tests.²¹

LINEARITY AND RANGE

A linear relationship should be evaluated across the range of the analytical method. For a drug product that has assay acceptance criteria of 90.0 to 110.0 percent label claim, linearity is usually demonstrated minimally from 80.0 to 120.0 percent. Some companies have a policy to demonstrate linearity over an expanded range from 70.0 to 130.0 percent. The linearity can be demonstrated directly on the API or on separate weighings of API and excipient mixtures.

ACCURACY

The accuracy of an analytical method expresses the closeness of agreement between the value found using the method and either an accepted reference value or a conventional true value. Accuracy is also termed *trueness*. A common way that accuracy of a method is evaluated is by performing "spike and recovery" experiments. Here, a known quantity of drug substance is spiked into a placebo matrix and compared to a reference solution that contains a known amount of drug substance. The compatibility of the analyte with the filter used during sample preparation is often formally evaluated at this time. Another common way to validate the accuracy of a method is to compare the results obtained from the method being validated to those obtained from a second well-characterized procedure.

Accuracy should be assessed using at least nine determinations over a minimum of three concentration levels (three different concentrations with three replicates of each concentration) that cover the specified range.

Accuracy can be inferred once precision, linearity, and specificity have been established. This approach is more common in early development activities, such as IND-enabling toxicology studies or Phase 1 clinical studies.

PRECISION

The precision of an analytical method expresses the closeness of agreement between a series of measurements obtained from multiple samplings of the same homogeneous sample under the prescribed conditions.^{18,19} The precision of a method is best

evaluated using authentic homogeneous samples. If authentic samples are not available, artificially prepared samples or a sample solution may be used.

Precision is considered at three levels: repeatability, intermediate precision, and reproducibility, which are intra-assay, within-laboratory, and inter-laboratory measurements, respectively. Repeatability is measured in one of two ways. One is to perform a minimum of six determinations at 100 percent of the test concentration. The other is to perform at least nine determinations over a minimum of three concentration levels (three different concentrations with three replicates of each concentration) that cover the specified range. Depending on company policy, repeatability and intermediate precision may be the only precision evaluation made in early development. Intermediate precision in its simplest sense is a comparison of results obtained by two analytical chemists from the same department (laboratory), preferably using two different chromatographic systems. Intermediate precision, however, can also be a design of experiments (DOE) in which variables such as analytical chemists, chromatographic systems, and days are examined.

Reproducibility is usually assessed prior to or during registration stability studies. A DOE is often used. Variable studies are similar to those used in intermediate precision, except that multiple analytical chemists and chromatographic systems are from two laboratories. These laboratories are often the laboratory that developed the method and the laboratory that will perform the release testing of the drug product.

ROBUSTNESS

The robustness of an analytical method is evaluated prior to or during registration stability studies. The robustness evaluation shows the reliability of the method with respect to deliberate changes made to the method. For example, consider an HPLC assay method for which the pH of the aqueous buffer used in the mobile phase is 4.5. Experiments may be performed using buffers at pH 4.0 and 5.0 in order to evaluate the criticality of mobile phase pH on the assay results. Other variables that may be evaluated in robustness experiments for assay methods are flow rate, injection volume, column temperature, sample tray temperature, column manufacturer, mobile phase composition, extraction time for sample preparation, and solvent (diluent) composition for sample preparation. The numbers of variables can be large, and the variables may be interdependent, so a DOE approach to evaluate robustness is encouraged.

SAMPLE AND STANDARD STABILITY

Regardless of the phase of development, the analytical samples and standards need to be stable for the duration of the analysis. This stability needs to be evaluated formally during method validation.

IMPURITIES: RELATED SUBSTANCES

Acceptance Criteria

The acceptance criteria for related substances on a drug product specification are expressed as less than or equal to a numerical value and are provided for individual impurities and total impurities. The purity results are reported as a percentage, with respect to either the peak area of the API in the drug product or the sum of the integrated peaks on the chromatogram. Purity methods must be stability-indicating. Related substances arising during drug product manufacturing and from the degradation of the drug substance must be monitored. Process impurities originating during the synthesis of the API are normally controlled during drug substance testing and are not listed on the drug product specification unless the synthesis impurity is also a degradation product in the drug product.

The magnitude of the acceptance criterion and how the impurity is listed on the specification are determined by the total daily dose, the dose at which the impurity was qualified in a toxicology study, manufacturing capability, and stability of the

drug product.²² If the daily dose is not more than (NMT) 1.0 g, impurities greater than 0.1 percent must be reported on the Certificate of Analysis (CofA). For daily doses greater than 1.0 g, the impurities greater than 0.05 percent must be reported. The values for all of the reported impurities will be summed, and this sum must meet the acceptance criterion for total impurities.

Some individual impurities are listed on the specification and limited in order to ensure the quality of the drug product. These impurities are called specified impurities. If the level of a specified impurity exceeds the identification threshold listed in ICH Guidance Q3B, then this impurity must be identified using a combination of mass spectrometric and nuclear magnetic resonance (NMR) spectrophotometric techniques. Such impurities are called specified identified impurities. If the level of a specified impurity does not exceed the identification threshold, it does not have to be identified. This specified, unidentified impurity is listed on the specification by its relative retention time, which is calculated by dividing the retention of the impurity by the retention time of the API. Identification thresholds are determined by the maximum daily dose and the total daily intake of the impurity and are provided in Table 2-1. Two examples are provided to help interpret the use of this table.

- Example 1: Consider a tablet for which the maximum daily dose is 750 µg. One percent of the maximum daily dose is 7.5 µg, which is greater than the 5 µg total daily intake (TDI) limit. Therefore, any impurity greater than 5 µg or 0.67 percent must be identified.
- Example 2: Consider a tablet for which the maximum daily dose is 200 µg. One percent of the maximum daily dose is 2 µg, which is lower than 5 µg TDI. Therefore, any impurity greater than 1.0 percent must be identified.

If a specified, identified impurity exceeds the qualification thresholds provided in Table 2-2, the impurity must be qualified in a toxicology study. The qualification threshold calculations from the information in Table 2-2 are analogous to those performed for determining identification thresholds. The acceptance criterion for an individual impurity should not exceed the amount that has been qualified in a toxicology study.

How are meaningful limits set for each impurity? ICH Guidance Q6A recommends that the acceptance criteria be based on safety and efficacy data at hand at the time of filing. The Guidance recognizes that the amount of data batch history data at the time of filing is likely insufficient to assess manufacturing

Table 2-1. Identification Thresholds

Maximum Daily Dose	Threshold
< 1 mg	1.0% or 5 µg TDI, whichever is lower
≥ 1 mg and ≤ 10 mg	0.5% or 20 µg TDI, whichever is lower
> 1 mg and ≤ 2000 mg	0.2% or 2 mg TDI, whichever is lower
> 2000 mg	0.10%

Table 2-2. Qualification Thresholds

Maximum Daily Dose	Threshold
< 10 mg	1.0% or 50 µg TDI, whichever is lower
≥ 10 mg and ≤ 100 mg	0.5% or 200 µg TDI, whichever is lower
> 100 mg and ≤ 2000 mg	0.2% or 3 mg TDI, whichever is lower
> 2000 mg	0.15%

process capability and recommends that limits do not tightly encompass the batch data that are available at the time of filing.

The acceptance criteria for individual impurities that are degradation products are determined by extrapolating the stability data that are available at the time of filing and setting the limit using a statistical analysis that usually involves calculating the upper 95 percent confidence limit.

The acceptance criterion for total impurities is usually established by taking the acceptance criteria for individual impurities into account. If degradation does not occur in the drug product, acceptance criterion for total impurities is set at 2.0 or 3.0 percent. If degradation does occur, this acceptance criterion can be as high as 5.0 or 7.0 percent.

The Guidance, however, recognizes that the development stability data are incomplete at the time of filing and allows for the limit to be raised through a qualification toxicology study. Likewise, the limit can be lowered to a value that is less than the qualified amount if stability data indicate that a lower value is justifiable. As more batch history is obtained, process capability should be re-evaluated. The Guidance allows for the limits to be raised or lowered based on sufficient commercial batch history.

Validation Characteristics

Impurity methods for related substances are validated for specificity, linearity, range, accuracy, precision, robustness, sample and standard stability, and quantitation limit.^{18,19} Usually detection limit is also a validation characteristic that is also determined.

SPECIFICITY

The validation of a related substances impurity method is similar to that for an assay method, regardless if the method is combined with the assay method or a stand-alone impurity method. The USP is actively modernizing official USP-NF monographs for small-molecule API and their dosage forms that do not have a stability-indicating impurity method for related substances.

LINEARITY AND RANGE

A linear relationship should be evaluated across the range of the analytical method. For a drug product, linearity is assessed from the reporting level of an impurity to 120 percent of the specification limit.

ACCURACY

Accuracy of an impurity method is evaluated by performing “spike and recovery” experiments. If reference material for the impurities is not available, it is acceptable in early development to spike drug substance into a placebo matrix and compare the value obtained to a reference solution that contains a known amount of drug substance. The amount of drug substance or impurity reference material spiked should be representative of the expected range for the impurity. The compatibility of the analyte with the filter used during sample preparation is often formally evaluated at this time.

PRECISION

The precision of an impurity method is best evaluated using authentic homogeneous samples. If authentic samples are not available, artificially prepared samples or a sample solution may be used. If the impurity is not available in early development, the drug substance may be substituted. The manners in which repeatability, intermediate precision, and reproducibility are accessed are similar to those used for the assay test. It is important to conduct the evaluation at representative concentrations for the impurities.

ROBUSTNESS

Like the assay method, the robustness of an analytical method is evaluated prior to or during registration stability studies. The variables evaluated in robustness experiments for impurity

methods are often similar to those used for the assay method. A DOE approach to evaluate robustness is encouraged.

SAMPLE AND STANDARD STABILITY

Regardless of the phase of development, the analytical samples and standards need to be stable for the duration of the analysis. Drug substance can be substituted for impurity reference material during early development if no impurity reference material is available. Also, the stability of impurities in working standards can be evaluated as well.

LIMIT OF QUANTITATION (LOQ)

The LOQ for an impurity method can be calculated by measuring the signal-to-noise ratio. LOQ is defined as a signal-to-noise ratio of 10:1. Another way to calculate the LOQ is to measure the standard deviation of the responses and the slope of the calibration in the range examined. The LOQ can be calculated as follows:

$$\text{LOQ} = 10\sigma / s$$

where σ is the standard deviation of the responses and S is the slope of the calibration curve. A practical way to measure LOQ is to prepare a solution at 0.05 or 0.1 percent, which corresponds to the reporting threshold. Make six replicate injections; if the relative standard deviation is less than 20 percent, then the LOQ can be defined as the reporting threshold.

LIMIT OF DETECTION (LOD)

This validation characteristic is generally not required for quantitative impurity methods, but it is easily calculated once the LOQ is known:

$$\text{LOD} = \text{LOQ} / 3$$

IMPURITIES: RESIDUAL SOLVENTS

Acceptance Criteria

Residual solvents are usually controlled by the drug substance specifications and usually do not appear on the drug product specification. Residual solvents in pharmaceuticals are defined as organic volatile chemicals that are used or produced in the manufacture of drug substances or excipients, or in the preparation of drug products. Solvents deliberately used as an excipient in a drug product are not considered residual.²³

There are two cases that may require residual solvents to be controlled on the drug product specification.

Case 1—Cumulative method. Drug products should contain no higher levels of residual solvents than can be supported by safety data. A cumulative method may be used to calculate the residual solvent levels in the drug product from the levels in the ingredients (drug substance and excipients) used to produce the drug product. If the calculation results in a level equal to or below that recommended in ICH Guidance 3C²³, no testing of the drug product for residual solvents is necessary. This calculation is usually performed in development to ensure that the calculated residual solvent levels are significantly lower than the safety limits. If, however, the calculated level is above the safety level, the drug product should be tested to determine whether the formulation process has reduced the relevant solvent level to within the acceptable amount. A method that is specific for each of the solvents of concern will need to be developed and validated. It is rare that the cumulative method results in having to list a residual solvent on the specification.

Case 2—Manufacturing. Drug product should also be tested if a solvent is used during its manufacture. Organic co-solvents, such as tetrahydrofuran (THF), are sometimes used to prepare spray-dried dispersions of low-soluble drugs. Most of the THF will evaporate during the manufacture of the dispersion, but residual THF is expected to be present. The THF can be controlled at the time of manufacture for the dispersion. If it is not,

then THF must be controlled in the drug product specification. THF is regarded as a Class 2 solvent, which means that the exposure to THF should be limited. A Class 2 solvent is considered a non-genotoxic animal carcinogen; a positive causative agent of other irreversible toxicity, such as neurotoxicity or teratogenicity; or an agent suspected of other significant but reversible toxicities. According to the ICH Guidance Q3C, the permitted daily exposure (PDE) for THF is 7.2 mg/day. There are two options to determine the acceptance criterion. Option 1 gives the limit as 720 ppm, which was calculated using the assumption that the product mass consumed each day does not exceed 10 g. Option 1 can be used provided that the total tablet mass consumed per day is less than 10 g and that all other excipients in the tablet also contain less than 720 ppm THF.

If the THF level in the tablet cannot be controlled to 720 ppm because the above conditions are not met or because the residual THF levels in the tablet are greater than 720 ppm, then Option 2 may be used to determine the acceptance criterion. Here the actual total tablet mass consumed each day is used in the calculation. Let's assume this value is 1.5 g. The limit can be calculated as follows:

$$\begin{aligned}\text{Acceptance criterion} &\leq (1000 \mu\text{g THF} / \text{mg THF}) \\ &\quad \times (7.2 \text{ mg THF} / \text{day}) / 1.5 \text{ g tablet} \\ &\leq 4800 \mu\text{g THF} / \text{g tablet} / \text{day} \\ &\leq 4800 \text{ ppm THF} / \text{day}\end{aligned}$$

Therefore, 4800 ppm THF can be used as the acceptance criterion. If manufacturing experience shows that the process capability is lower than 4800 ppm, then a lower acceptance criterion should be used.

Validation Characteristics

Residual solvent analysis is usually determined by gas chromatography (GC) unless the solvent is high boiling. THF is suitable for headspace GC analysis. This method would be validated for specificity, linearity, range, accuracy, precision, robustness, sample and standard stability, and quantitation limit.

SPECIFICITY

Specificity is determined by comparing the chromatogram obtained from a diluted THF solution to that of the drug product matrix. The method is deemed specific if there are no interferences from the matrix observed.

LINEARITY AND RANGE

A linear relationship should be evaluated from 80.0 to 120.0 percent of the acceptance criterion. Some companies have a policy to demonstrate linearity over an expanded range from 70.0 to 130.0 percent.

ACCURACY

Accuracy of a residual solvent method is evaluated by performing "spike and recovery" experiments.

PRECISION

The manners in which repeatability, intermediate precision, and reproducibility are accessed are similar to those used for the impurity test for related substances.

ROBUSTNESS AND SAMPLE AND STANDARD STABILITY

The validation characteristics of a residual solvent method are evaluated in a similar manner to the previously discussed methods.

LIMIT OF QUANTITATION (LOQ)

The LOQ for a residual solvent method can be calculated using either of the two methods previously described. The LOQ

can be verified experimentally if the THF levels are very low by spiking the calculated LOQ amount into the drug product and matrix and measuring the precision of six replicate injections.

LIMIT OF DETECTION (LOD)

The LOD can be determined from the LOQ as previously discussed.

IMPURITIES: ENANTIOMERIC PURITY

Acceptance Criteria

For simplicity, let's assume that the drug substance has only one chiral center and that the drug substance has been synthesized to afford a large enantiomeric excess. The opposite enantiomer is considered an impurity. The opposite enantiomer, however, is excluded from the identification and qualification thresholds to which the related substances are subjected because quantitation of the enantiomer at the threshold levels can be difficult.

The opposite enantiomer is quantified by HPLC using a different method than that used for related substances. The column must have a chiral stationary phase in order to separate the opposite enantiomer from the drug substance. As a result, the chiral purity test and acceptance criterion are separate entries on the drug product specification. In addition, the chiral purity result is usually not included in the calculation for total impurities for related substances.

The acceptance criterion for the amount of enantiomer in a drug product specification is usually the same as that given in the drug substance specification for drug products in clinical development unless there is potential evidence that racemization may occur in the drug product. In this case, the acceptance criterion for the drug product may be greater than that for the drug substance. It is the responsibility of the manufacturer, however, to control the amount of the other enantiomer in the drug product. It is not uncommon to have an acceptance criterion for the enantiomer as high as 1.0 to 2.0 percent, relative to the drug substance.

As development progresses or after registration, drug product release and stability data are used to establish the acceptance criterion. If racemization has been shown to be insignificant, the amount of the enantiomer does not need to be controlled in the drug product. Then the test is removed from the specification or sunsetted.

Validation Characteristics

Drug product enantiomeric purity methods are validated for specificity, linearity, range, accuracy, precision, robustness, sample and standard stability, and quantitation limit. Sometimes detection limit is also a validation characteristic that is determined.

SPECIFICITY

The method should be capable of separating the two enantiomers with baseline resolution and ensuring that there are no co-eluting peaks from other molecules present in the sample.

It should be noted that a stereospecific identity test for the drug substance is usually not needed in the drug product release specification. If drug product release and stability data indicate that racemization is insignificant, the stereospecific identity is more appropriately addressed as part of the drug substance specification. On the other hand, the enantiomeric impurity testing will serve as an identity test if racemization in the drug product is a concern.

LINEARITY AND RANGE

A linear relationship should be evaluated across the range of the analytical method. For a drug product, linearity is assessed from the LOQ to 120 percent of the specification limit. Some companies will assess linearity in the range from the LOQ to 2.0 percent (relative to the drug substance).

ACCURACY

The accuracy of an impurity method is evaluated by performing “spike and recovery” experiments. If reference material for the impurities is not available, the accuracy is inferred from the linearity and the precision data.

PRECISION

The precision of an enantiomeric purity method is usually evaluated using authentic homogeneous samples. If authentic samples are not available, artificially prepared samples may be used. The precision should be evaluated in the range of interest. The manners in which repeatability, intermediate precision, and reproducibility are accessed are similar to those used for the related substances impurity test.

ROBUSTNESS AND SAMPLE AND STANDARD STABILITY

The validation characteristics of an enantiomeric purity method are evaluated in a similar manner to the previously discussed methods.

LIMIT OF QUANTITATION (LOQ)

The LOQ for an enantiomeric purity method can be calculated using either of the two methods previously described. The LOQ can be verified experimentally by spiking the enantiomer into the drug product matrix at the calculated LOQ or at 0.05 percent, whichever is greater, and measuring the precision of six replicate injections.

LIMIT OF DETECTION (LOD)

The LOD can be determined from the LOQ as previously discussed.

IMPURITIES: ELEMENTAL IMPURITIES

The European Medicines Agency issued a draft guideline on the specification limits for residual metal in 2008.²⁴ This guideline applies to drug substances and new and existing drug products. Pharmaceutical companies, however, have been given up to five years for the implementation of the guideline to existing drug products. The USP has formed an Expert Panel to revise general chapters concerning elemental impurities.^{25,26} Revised chapters appear in *Pharmacopeial Forum*.^{27,28}

IDENTIFICATION

Acceptance Criteria

The acceptance criteria for an identification test depend on the technique used. For chromatographic tests in which retention time of the sample is compared to that of a reference material, the acceptance criterion is usually given as a relative retention time (RRT), which is defined as the retention time of the standard divided by the retention time of the reference material. This criterion is commonly $RRT = 1.00 \pm 0.05$.

The acceptance criterion for a spectroscopic test is usually given as “consistent with reference standard.”

Validation Characteristics

Identification methods are not validated in the typical sense. Nevertheless, sound scientific judgment must be used. The chromatographic method, which is the same as that used for assay determination, has been shown to be free from interferences. A spectroscopic identification test must be consistent with the structure of the molecule, and the reference standard must have been fully characterized by MS and NMR spectroscopy.

DESCRIPTION

Description is usually a visual test that is dependent on the physical appearance of the dosage form. Validation is not required. Examples of acceptance criteria for several dosage forms are given below.

A capsule to be used in a clinical study: Opaque, red, size 3 capsule

A commercially available 10-mg capsule: Blue and white, opaque capsules, imprinted with “ABC” on one half and “10mg” on the other half in black ink

A tablet to be used in a clinical study: White to off-white, round, biconvex tablets

A commercially available 500-mg tablet: Off-white, film-coated tablets, ovaloid, biconvex with beveled edges, debossed with “500” on one side, with score on the other side, and “LL” on each side of the score

A small-volume injectable solution: Clear, colorless solution; free from particles or foreign matter

Lyophilized drug for injection: White to off-white cake or powder; free from particles or foreign matter

Topical cream: White to off-white cream with a uniform appearance

Topical gel: Clear, colorless to pale yellow gel

TAXONOMY

The Dosage Forms Expert Committee of the USP developed a compendial taxonomy that was introduced in *Pharmacopeial Forum* 29(5) in 2002 in order to provide a uniform way of categorizing pharmaceutical dosage forms; a tool for finding information and linking specific dosage form monographs and general chapters; a rational linkage between dosage forms and their compendia specifications; and a linkage between dosage form drug substance(s) and their compendial use.²⁹ The Committee’s overall motivation was to give the manufacturer and the regulator clearer knowledge of the specifications generally needed to demonstrate compliance. The taxonomy is based on a tier concept.

Tier 1 delineates the dosage form according to the route of administration by which the drug substance is delivered. Five routes of administration have been identified:^{29,30}

- Oral
- Parenteral
- Inhalation
- Transdermal/Topical
- Mucosal

The USP General Chapters—Dosage Forms Expert Committee has been tasked to update or write General Chapters for each route.³⁰

Tier 2 is based on the general type of dosage form and its physical properties. For example, tablets and capsules are examples of oral drug products. USP Chapter <1151>, “Pharmaceutical Dosage Forms”, was rewritten to provide descriptions of dosage forms commonly used to administer the drug substance. A glossary is provided as a resource on nomenclature. USP Chapter <1151> gives general principles involved in the manufacture or compounding of these dosage forms and recommendations for proper use and storage. This informational chapter becomes official in USP 35-NF 30.^{14,15}

Tier 3 gives the type of release pattern: immediate; extended; or delayed. Tablets, for example, can be subdivided into each of the three release patterns. Product performance tests may be different for each release pattern.

Product quality and performance tests for representative dosage forms are provided below. These tests are presented for pedagogical purposes. The reader is asked to consult *Pharmacopeial Forum* for in-process revisions of these chapters and the United States Pharmacopeia and European Pharmacopeia for official versions of the quality and performance tests for the various routes of delivery.

DOSAGE FORMS—INJECTABLES

There are five general types of injectable presentations.³¹

1. [DRUG] Injection—Liquid preparations that are drug substances or solutions thereof.

2. [DRUG] for Injection—Dry solids that, upon the addition of suitable vehicles, yield solutions conforming in all respects to the requirements for *Injections*.
3. [DRUG] Injectable Emulsion—Liquid preparations of drug substances dissolved or dispersed in a suitable emulsion medium.
4. [DRUG] Injectable Suspension—Liquid preparations of solids suspended in a suitable liquid medium.
5. [DRUG] for Injectable Suspension—Dry solids that, upon the addition of suitable vehicles, yield preparations conforming in all respects to the requirements for *Injectable Suspensions*.

Product quality and performance tests for [Drug] Injection and [Drug] for Injection will be described below. Typical specifications for injectable dosage forms that are used in clinical development are provided in Tables A-1 and A-2, and representative monographs for these dosage forms are given in Appendix B.

[DRUG] INJECTION

Product Quality: Universal Tests

Appearance, identification, assay, and impurity (related substances) are universal product quality tests for this dosage form. If the drug substance is chiral, enantiomeric purity may be a universal product quality test if racemization has not been shown to be insignificant.

Product Quality: Dosage Form Specific Tests

Tests that are specific for Injection solutions include pH, particulate matter, sterility, and endotoxin. All of these tests are compendial, so validation is not needed. The test should be verified using the drug product so that there is a degree of confidence in the results obtained from the compendial test and that the test has performed suitably as intended. USP-NF Chapter <1226>, "Verification of Compendial Procedures," should be consulted for further information.³²

pH

For compendial purposes, pH is defined as the value given by a suitable, properly standardized, potentiometric instrument (pH meter) capable of reproducing pH values to 0.02 pH unit using an indicator electrode sensitive to hydrogen activity, the glass electrode, and a suitable reference electrode. This definition indicates that pH is an experimental value.³³ (The theoretical definition of pH is negative 1 times the base ten logarithm of the hydrogen ion activity.)

The pH of the solution may affect the solubility and stability of the drug substance and may be a critical quality attribute that needs to be controlled. Specification will usually specify a pH range of 0.4 to 0.6 pH units, with the target pH in the middle of the range. See USP <791> and Pharm Eur 2.2.3 for further information on this test.^{33,34}

PARTICULATE MATTER

Particulate matter in injections and parenteral infusions consists of extraneous mobile undissolved particles, other than gas bubbles, unintentionally present in the solutions.³⁵ The presence of visible particles will be determined from the Appearance Test. Two methods are used to test for subvisible particles and are provided in USP <788>, "Particulate Matter in Injections." These tests are harmonized with those provided in the European Pharmacopeia and the Japanese Pharmacopeia.

Method 1, Light Obscuration Particle Count Test, is the preferred method for observing sub-visible particles. However, this method may not work for drug products for which clarity is reduced or viscosity is increased. Method 2, Microscopic Particle Count Test, provides an alternative procedure for these drug products. It may be necessary to run both methods on injection and infusion solutions in order to reach a conclusion on the conformance to the requirements. The acceptance criteria for these methods depend on the volume of the drug product.

STERILITY

The [DRUG] Injection must meet the requirements under sterility testing. A typical acceptance criterion is "no growth." The ICH Steering Committee, based on the evaluation by the Q4B Expert Working Group (EWG), recommends that the official pharmacopoeial texts, Ph. Eur. 2.6.1. Sterility, JP 4.06 Sterility Test, and USP <71> Sterility Tests, can be used interchangeably in the ICH regions subject to the conditions detailed below.³⁶

BACTERIAL ENDOTOXINS TEST

The Bacterial Endotoxins Test is a test to detect or quantify bacterial endotoxins from gram-negative bacteria using Limulus Amebocyte Lysate (LAL), obtained from the horseshoe crab. There are three types of techniques for this test, and any of them can be used to test for the presence of endotoxins:

1. The gel-clot technique, which is based on gel formation;
2. The turbidimetric technique, based on the development of turbidity after cleavage of an endogenous substrate; and
3. The chromogenic technique, which is based on the development of color after cleavage of a synthetic peptide-chromogen complex.

Proceed by any one of these techniques, unless otherwise indicated in the monograph. In case of dispute, the final decision is based on the gel-clot technique, unless otherwise indicated in the monograph.³⁷

The endotoxin limit for parenteral drugs, defined on the basis of dose, equals K/M, where K is a threshold pyrogenic dose of 5 endotoxin units (EU) per kg of body weight, and M is equal to the maximum recommended bolus dose of product per kg of body weight. When the product is to be injected at frequent intervals or infused continuously, M is the maximum total dose administered in a single hour period. The endotoxin limit (acceptance criterion) for parenteral drugs is specified in drug product specification in units such as EU/mg of drug substance.³⁷

UNIFORMITY OF DOSAGE FORMS

If the dose for the [DRUG] Injection is determined by the volume of solution delivered, a test for uniformity of the solution dosage form is not required. If the entire contents of the drug product are to be delivered, however, a uniformity test is needed. This test is a weight uniformity in which the weight of liquid in individual containers is weighed and an acceptance value is calculated.^{38–41} The USP chapter for Uniformity of Dosage Form is not harmonized with the Uniformity of Dosage Forms chapter in the European Pharmacopeia.^{38–41}

Product Performance: Dosage Form Specific Tests

There are no performance tests for [DRUG] Injection dosage forms.

[DRUG] FOR INJECTION

Product Quality: Universal Tests

The universal tests for [DRUG] for Injection are identical to those for [DRUG] Injection. The acceptance criterion, however, is descriptive of a solid instead of a solution. For example, a typical acceptance criterion is "white to off-white cake or powder; free from particles or foreign matter." If the drug substance is chiral, enantiomeric purity may be a universal product quality test if racemization has not been shown to be insignificant.

If a drug product exists in [DRUG] Injection and [DRUG] for Injection formats, different methods are often used for assay, impurity, and enantiometric purity. The chromatographic conditions and parameters will likely be identical, but the sample preparations are different because the physical forms of the two dosage forms are different.

Product Quality: Dosage Form Specific Tests

UNIFORMITY OF DOSAGE FORM

Uniformity of Dosage Forms is a product quality test for [DRUG] for Injection dosage forms. Weight uniformity may be used for single-dose containers; otherwise, a suitable analytical method must be used to determine uniformity of content.^{38–41} Typically, an isocratic, non-stability-indicating method is used for content uniformity. The method is validated for specificity, accuracy, precision, linearity, and sample and standard stability. The performance criteria for accuracy and precision are less stringent for this method than performance criteria for assay and purity methods. The European and United States Pharmacopeias, procedures for the Uniformity of Dosage Forms Test are harmonized for this dosage form.

WATER DETERMINATION

The product quality tests for [DRUG] for Injection may include water determination if water content is a critical quality attribute. Karl Fischer titrations and loss on drying are common techniques to determine moisture content.⁴²

Product Performance: Dosage Form Specific Tests

RECONSTITUTION TIME

A common performance test for [DRUG] for Injection dosage forms is reconstitution time. Here, the time to dissolve the powder in a specified volume of diluent is measured by a stopwatch. The vial is vigorously shaken after the addition of the diluent and observed every 5 to 10 seconds. The time for which the solid contents take to dissolve is recorded and must be less than the acceptance criterion.

Product Quality: Dosage Form Specific Tests—Reconstituted Drug Product

Appearance, pH, and particulate matter are typical quality tests for reconstituted solutions. The acceptance criteria for these tests are similar for a [DRUG] for Injection and [DRUG] Injection dosage forms.

DOSAGE FORMS—ORAL DRUG PRODUCTS

There are many types of oral dosage forms: tablets, capsules, powders, granules, medicated gums, solutions, and suspensions are common dosage forms that are available by prescription or over-the-counter.^{14,15,43} An example of a specification for an immediate-release tablet that is used in clinical development is given in Table A-3. Representative monographs for immediate-, delayed-, and extended-release tablets are provided in Appendix B. The product quality and performance tests for these types of tablets are discussed below.

IMMEDIATE-RELEASE TABLETS

Product Quality: Universal Tests

The universal tests for an immediate-release tablet are identical to those for injectable dosage forms. The acceptance criterion, however, is descriptive of a solid unit-dose. For example, a typical acceptance criterion is “white to off-white round convex tablets.” If the drug substance is chiral, enantiomeric purity may be a universal product quality test if racemization has not been shown to be insignificant.

Product Quality: Dosage Form Specific Tests

Tests that are specific for immediate-release tablets include uniformity of dosage forms, tablet friability, tablet breaking force, water determination, X-ray powder diffraction, microbial enumeration tests, and tests for specified microorganisms.

UNIFORMITY OF DOSAGE FORM

Each tablet should have a drug substance content within a narrow range around the label claim (LC). The magnitude of this

range is governed by an acceptance limit, which is designed to provide within a stated probability that a single tablet taken from an entire manufacturing lot would pass the content uniformity test. Details on how to calculate the acceptance value (AV) are provided in the United States Pharmacopeia³⁸ and the European Pharmacopeia.³⁹

In most cases,^{38,39,44} the content of the tablet is measured using an appropriate analytical method, such as an isocratic non-stability-indicating HPLC method. First, 30 tablets from a manufacturing lot are randomly selected. Ten of these tablets are individually tested. The average ($\bar{M}$) and standard deviation (s) are calculated from the results.

The acceptance value and $\bar{M}$ are typically calculated as follows: $AV = \bar{M} + 2.4s$

$$\begin{aligned} M &= \bar{X} && \text{if } \bar{X} < 98.5\% \text{ LC;} \\ M &= 98.5 - \bar{X} && \text{if } 98.5 \leq \bar{X} \leq 101.5\% \text{ LC; and} \\ M &= \bar{X} - 101.5 && \text{if } \bar{X} > 101.5\% \text{ LC.} \end{aligned}$$

If AV is less than or equal to 15.0, which is called L1, the lot of tablets passes the uniformity of dosage form test. If AV is greater than L1, then 20 more tablets must be individually tested. The average and standard deviation for all 30 tablets are calculated. The AV for the 30 tablets is calculated: $AV = \bar{M} + 2.0s$. The equations to calculate $\bar{M}$ are given above. The lot passes the uniformity of dosage form test if the AV for the 30 tablets is less than or equal to 25.0 (L2), if no individual tablet content is less than 0.75 $\bar{M}$, and if no individual tablet content is greater than 1.25 $\bar{M}$.

The USP allows for a weight variation test to be substituted for a content uniformity assay for uncoated or film-coated tablets if the tablet label claim is not less than 25 mg of active ingredient and if the drug load is not less than 25 percent.³⁸ Drug load is defined as the weight percent of the tablet composition that is an active substance.

UNIFORMITY OF CONTENT AND UNIFORMITY OF MASS

The general monograph for tablets in the European Pharmacopeia (01/2008:0478) requires compliance to two additional quality tests: uniformity of content and uniformity of mass.⁴⁵ The uniformity of content test⁴¹ applies only to tablets with less than 2 mg of active substance or less than 2 percent drug load. The lot passes this test if the individual content of each of 10 tablets measured in the uniformity of dosage forms test is between 85 and 115 percent of the average content. If this criterion is not met and if the individual content of each of the 10 tablets is between 75 and 125 percent of the average, the content of 20 additional tablets is measured. If the individual contents of all 30 tablets are between 75 and 125 percent of the average, and if the content of 29 of the 30 tablets is between 85 and 115 percent of the average, then the lot passes the uniformity of content test.

The uniformity of mass test⁴⁰ is performed by weighing 20 tablets individually. An average mass, $\bar{M}$, is calculated. The acceptance limits for this test are determined from the value of $\bar{M}$.

If $\bar{M} \leq 80 \text{ mg}$	All tablet masses must be between $0.8\bar{M}$ and $1.2\bar{M}$ and at least 18 of the individual masses must be between $0.9\bar{M}$ and $1.1\bar{M}$.
If $\bar{M} \leq 80\text{--}250 \text{ mg}$	All tablet masses must be between $0.85\bar{M}$ and $1.15\bar{M}$ and at least 18 of the individual masses must be between $0.925\bar{M}$ and $1.075\bar{M}$.
If $\bar{M} \geq 250 \text{ mg}$	All tablet masses must be between $0.9\bar{M}$ and $1.1\bar{M}$ and at least 18 of the individual masses must be between $0.95\bar{M}$ and $1.05\bar{M}$.

TABLET FRIABILITY

The tablet friability⁴⁶ test usually does not appear in a specification or monograph. It is, however, an important test that is performed during development of the tablet to ensure that the tablet will not break or chip during coating, packaging, or routine handling during distribution to the patient. The test is harmonized across the European, Japanese, and United States Pharmacopeia and is described in detail in their chapters. Ten tablets or 650 g of tablets, whichever is greater, are placed in a tablet friability apparatus and rotated 100 times at 25 rpm. The lot passes the test if no tablet breaks and if the total weight loss after tumbling is less than the acceptance limit for the product. This limit is usually not more than 1.0 percent weight loss, but it varies depending on the nature of the product. Tablet friability is monitored during the tablet compression operation in manufacturing.

TABLET BREAKING FORCE

The tablet breaking force test,⁴⁷ or tablet hardness test as it is sometimes called, is often found on a specification during development to ensure that proper control during manufacturing exists. Variation in tablet hardness can affect *in vitro* and *in vivo* performance because tablet disintegration and dissolution may vary. The test is often performed during developmental stability studies and as an in-process control during the tablet compression operation in manufacturing.

WATER DETERMINATION

The product quality tests for a tablet may include water determination if water content is a critical quality attribute. Karl Fischer titrations and loss on drying are common techniques to determine moisture content.⁴²

X-RAY POWDER DIFFRACTION

The crystal form or polymorph of the drug may affect the solubility, dissolution rate, and bioavailability of the drug. X-ray diffraction is often performed to ensure that the crystal form has not changed during manufacturing or storage of the tablet.⁴⁸ If a tablet is made using an amorphous dispersion of the drug substance, X-ray diffraction is performed to ensure that the drug has not crystallized. An example of the impact of crystal form on product performance was observed in 1998 for Norvir (ritonavir) soft gelatin capsules.⁴⁹

MICROBIAL ENUMERATION TESTS AND TESTS FOR SPECIFIED MICROORGANISMS

Typical microbiological tests that appear in an oral solid-dosage form specification include total aerobic microbial count (TAMC), total yeast and molds count (TYMC), and the absence of *E. coli*. The tests are described in detail in the United States Pharmacopeia^{50,51} and the European Pharmacopeia.^{52,53} Although these tests appear in a specification, they are generally not included in a monograph.

Product Performance: Dosage Form Specific Tests

There are two performance tests for immediate-release tablets: dissolution and disintegration.

DISSOLUTION

A dissolution test is performed by monitoring the amount of drug dissolved as a function of time. Usually the test is performed using Apparatus 1 or Apparatus 2.^{54,55} When Apparatus 1 is used, the dosage form is placed in a basket, which is lowered into a specified volume (usually 500 or 900 mL) of medium equilibrated at $37.0 \pm 0.5^\circ\text{C}$. The medium is chosen so the solubility of the active ingredient is about 3 to 4 times greater than the concentration of the drug in medium after complete dissolution of the active substance has occurred.⁵⁶ The dosage form is rotated at a pre-determined speed, usually between 50 and 75 rpm. Six dosage forms are tested simultaneously. Apparatus 1 is

usually used for capsules. Apparatus 2, usually used for tablets, makes use of a paddle to create hydrodynamic flow over the tablet, which rests at the bottom of the dissolution flask.

During tablet development when a dearth of release and stability data exist and when the formulation or manufacturing process is subject to change, it is common to record the entire dissolution profile or the amount of drug dissolved at specified intervals, such as 10, 20, 30, 40, 50, and 60 minutes or 15, 30, 45, and 60 minutes. An isocratic, non-stability-indicating HPLC method or a UV-Vis spectrophotometric method is often used to measure the drug dissolved. Prior to registration, dissolution for an immediate-release tablet becomes a single-point test. The acceptance criterion for this test is a function of Q, which is expressed as percent label claim of drug dissolved at the specified time.^{54,55}

Like the Uniformity of Dosage Forms Test, the dissolution test is a staged test. There are three levels. In the first level of testing called S1, six tablets are tested. Each tablet must be $Q + 5$ percent (absolute percentage points) dissolved at a specified time.^{54,55} In the USP monograph for capecetabine tablets, the Q value is specified as 80 percent and the time point is 30 minutes. Here, 85 percent of the capecetabine in each tablet must be dissolved at 30 minutes.

If this criterion is not met, then six additional tablets are tested in level 2 (S2). In order to pass the S2 acceptance criteria, the average of all 12 tablets must be equal to or greater than Q (80 percent LC for capecetabine tablets), and no tablet has less than $Q - 15$ percent (65 percent LC for capecetabine tablets).^{54,55}

If these criteria are not met, then Level 3 or S3 testing must be performed by testing 12 additional tablets. To pass S3, the average of all 24 tablets must be equal to or greater than Q (80 percent LC for capecetabine tablets). Two additional criteria must be met as well: no more than 2 tablets are less than $Q - 15$ percent (65 percent LC for capecetabine tablets), and no tablet is less than $Q - 25$ percent dissolved (55 percent LC for capecetabine tablets).^{54,55}

DISINTEGRATION

One tablet is placed in each of six tubes of a basket rack assembly for a disintegration apparatus and immersed in a specified medium equilibrated at $37 \pm 2^\circ\text{C}$. The basket is raised and lowered a distance of 55 ± 2 mm at a rate of 29 to 32 cycles per minute for a defined period of time, usually 15 minutes for uncoated tablets and 30 minutes for coated tablets. The lot passes the test if all of the tablets disintegrate completely at the end of the specified time. If one or two of the tablets fail to disintegrate completely, the test is repeated on 12 additional tablets. The disintegration test is passed if at least 16 of the 18 tablets have disintegrated completely.^{57,58} Automated disintegration apparatuses exist to record the exact time that each tablet disintegrated and are often used during tablet development and optimization and during release and stability testing.

Disintegration may not be included in a specification or monograph if dissolution is performed. Disintegration, however, is frequently performed as an in-process control during tablet manufacturing. A disintegration test may be used in place of a dissolution test when a drug is rapidly dissolving.¹

DELAYED-RELEASE TABLETS

Product Quality: Universal Tests and Dosage Form Specific Tests

A delayed-release tablet is designed to release the drug after the tablet leaves the acidic environment of the stomach. The quality tests for a delayed-release tablet are identical to those for an immediate-release tablet. Tablet friability and tablet breaking force are usually controlled on the tablet core prior to the coating operation. Hardness may also be tested on the enteric-coated tablet.

Product Performance: Dosage Form Specific Tests

DISSOLUTION

A dissolution test is a two-part test. First, the tablets are exposed to an acid medium that is followed by exposure to a buffer medium. The medium used for an acid stage is 0.1N HCl, and the duration of this stage is 2 hours. The tablets are transferred to a buffer medium, usually 0.2M phosphate buffer at pH 6.8, but other buffers and pH targets may be used if justified. The duration of the buffer stage is usually 45 minutes, but this duration may vary depending on the product.^{54,55}

An aliquot of the acid medium from each vessel is analyzed. The acceptance criteria for the acid stage are as follows and there are three levels:

Level 1 (A1) testing is passed if no individual value exceeds 10 percent dissolved. If the A1 criteria are not met, then the dissolution test is performed on 6 additional tablets for Level 2 (A2) testing. Level A2 criteria are passed if the average of all 12 tablets in the acid stage is not more than 10 percent dissolved and if no individual tablet is greater than 25 percent dissolved.^{54,55}

Level 3 testing is performed if the A2 criteria are not met. The A3 criteria are passed if the average of all 24 tablets in the acid stage is not more than 10 percent dissolved and if no individual tablet is greater than 25 percent dissolved.^{54,55}

The acceptance criteria for the buffer stage are based on a Q value at a specified time. The B1, B2, and B3 criteria are identical to those for the immediate release S1, S2, and S3 criteria.^{54,55}

DISINTEGRATION

For enteric-coated tablets, disintegration is performed on six tablets for one hour using simulated gastric fluid or 0.1 N HCl as the immersion medium. The tablets are observed for no evidence of disintegration, cracking, or softening. The test is repeated on the same tablets using simulated intestinal fluid or a phosphate buffer as the immersion medium. The duration in the medium varies but is usually between 15 and 60 minutes. The same acceptance criteria apply to immediate- and delayed-release tablets.^{57,58} Disintegration is usually performed as an in-process test on the cores during manufacturing.

EXTENDED-RELEASE TABLETS

Product Quality: Universal Tests and Dosage Form Specific Tests

An extended-release or prolonged-release tablet is designed to release the drug over a defined time period, usually 4 to 12 hours. The quality tests for an extended-release tablet are identical to those for an immediate- or delayed-release tablet.

Product Quality: Dosage Form Specific Tests

Dissolution is the only performance test for extended-release tablets. Disintegration is usually not performed.

DISSOLUTION

A dissolution test is generally the same as that used for an immediate-release tablet, except that three time points are specified. The time points used for acetaminophen extended-release tablets that are 30 minutes, 90 minutes, and 4 hours. The acceptance ranges are as follows:

Between 40 and 60 percent acetaminophen dissolved at 30 minutes;

Between 55 and 85 percent acetaminophen dissolved at 90 minutes; and

Not less than 80 percent acetaminophen dissolved at 4 hours.

Six tablets are analyzed in Level 1 (L1); acceptance criteria are met if no individual value lies outside each of the stated ranges and no individual value is less than the final time period.^{54,55} If the L1 criteria are not met, then six additional tablets are analyzed in Level 2 (L2). The L2 criteria are met if these three conditions are met:

1. The average values of the 12 tablets lies within each of the stated ranges and is not less than the stated range of the final time point;
2. None is more than 10 percent of labeled content outside each of the stated ranges; and
3. None is more than 10 percent of the labeled content below the stated amount at the final test time.^{54,55}

If the L2 criteria are not met, then 12 additional tablets are tested in Level 3 (L3). The L3 criteria are met if these five conditions are met:

1. The average values of the 24 tablets lies within each of the stated ranges and is not less than the stated range of the final time point;
2. Not more than 2 of the 24 tablets are more than 10 percent of labeled content outside each of the stated ranges;
3. Not more than 2 of the 24 tablets are more than 10 percent of the labeled content below the stated amount at the final test time;
4. None of the 24 tablets is more than 20 percent of labeled content outside each of the stated ranges; and
5. None of the 24 tablets is more than 20 percent of the labeled content below the stated amount at the final test time.^{54,55}

ADDITIONAL DOSAGE FORMS

The product quality and performance tests for many dosage forms are summarized in “Chart 4: Non Complex Active Drug Products” in the Guide to General Chapters to USP 34,⁵⁹ USP <1151>,^{14,15} and the Dosage Forms chapter of the European Pharmacopeia.⁴³

The USP General Chapters-Dosage Forms Expert Committee is charged to revise product quality chapters for five routes of administration.³⁰ These General Chapters have been recently or are in the process of being revised:

- <2> Oral Drug Products—Product Quality Tests³⁰
- <3> Topical and Transdermal Drug Products—Product Quality Tests^{16,17,30,62}
- <725> Topical and Transdermal Drug Products—Product Performance Tests^{30,60,62}
- <1724> Topical and Transdermal Drug Products—Product Performance Tests^{30,61,62}
- <5> Inhalation and Nasal Drug Products—General Information and Product Quality Tests^{30,63,65}
- <601> Aerosols, Nasal Sprays, Metered-Dose Inhalers, and Dry Powder Inhalers^{30,64,65}

FUTURE DIRECTION

The biggest change in the analysis of medical products in the coming years will likely be in the implementation of the Quality by Design (QbD) paradigm and corresponding ICH Guidelines⁶⁶⁻⁶⁹ into everyday pharmaceutical development. The information and knowledge gained from pharmaceutical development studies and manufacturing experience provide scientific understanding to support the establishment of the design space, specifications, and manufacturing controls. The use of meaningful analytical methods and specifications will be an integral part of QbD because quality is expected to be built into the design of the product. Analytical tests, especially those that test product performance, will be expected to be discriminating. For example, dissolution tests will be expected to tell when a manufacturing or formulation change will affect the product performance.

REFERENCES

1. International Conference on Harmonization, Topic Q6A. Specifications: Test Procedures and Acceptance Criteria for New Drug Substances and New Drug Products: Chemical Substances, 2000.

2. Williams RL. Medicine quality faces challenges. *Chemical and Engineering News* 2011; 89(26): 58–61.
3. United States Pharmacopeial Convention. *United States Pharmacopeia*, 34th edn, and *National Formulary*, 29th edn. Baltimore: United Book Press, 2011.
4. About USP. www.usp.org/about-usp (Accessed 5 February, 2012).
5. European Directorate for the Quality for Medicines and Healthcare of the Council of Europe. *European Pharmacopeia*, 7th edn. Strasbourg, France, 2010.
6. Membership and observership. www.edqm.eu/site/Membership-Observership-608.html (Accessed 5 February, 2012).
7. British Pharmacopoeia Commission. *British Pharmacopoeia* 2012. London, England: Stationery Office, 2011.
8. Welcome to the website of the British Pharmacopoeia. www.pharmacopoeia.gov.uk (Accessed 5 February, 2012).
9. Japan Ministry of Health, Labour, and Welfare. *The Japanese Pharmacopeia*, 16th edn. Tokyo, Japan, 2011.
10. Japan Ministry of Health, Labour, and Welfare. *The Japanese Pharmacopeia*, 15th edn (in English). Tokyo, Japan, 2006.
11. JP online. www.pmda.go.jp/english/pharmacopoeia/online.html (Accessed 5 February, 2012).
12. United States Pharmacopeial Convention. USP Monograph Backgrounder. Rockville, MD, 2008.
13. Monographs. www.pharmacopoeia.gov.uk/monographs.php (Accessed 5 February, 2012).
14. In-process revision: <1151>, Pharmaceutical dosage forms. *Pharmacopeial Forum* 2010; 36(6).
15. United States Pharmacopeial Convention. Chapter <1151>, Pharmaceutical dosage forms. *United States Pharmacopeia*, 35th edn, and *National Formulary*, 30th edn. Baltimore: United Book Press, 2012.
16. In-process Revision: <3>, Topical and transdermal drug products—product quality tests. *Pharmacopeial Forum* 2010; 36(6).
17. United States Pharmacopeial Convention. Chapter <3>, Topical and transdermal drug products—product quality. *United States Pharmacopeia*, 35th edn, and *National Formulary*, 30th edn. Baltimore: United Book Press, Baltimore, 2012.
18. International Conference on Harmonization, Topic Q2A. Guideline for Industry Text on Validation of Analytical Procedures, 1995.
19. United States Food and Drug Administration. Guidance for Industry: Q2B Validation of Analytical Procedures: Methodology, 1995.
20. Skwierczynski RD, Smith SW. Preformulation and Stabilization of Pharmaceuticals, University of Wisconsin—Extension Services in Pharmacy Short Course, 2001–2011.
21. USP seeks submission of proposals for monograph modernization. www.usp.org/usp-nf/development-process/monograph-modernization (Accessed 5 February, 2012).
22. International Conference on Harmonization, Topic Q3B. Guidance for Industry Q3B(R2) Impurities in New Drug Products, Revision 2, 2006.
23. International Conference on Harmonization, Topic Q3C. Guidance for Industry Q3C(R4) Impurities: Guideline for Residual Solvents, Revision 4, 2009.
24. European Medicines Agency Committee for Medicinal Products for Human Use (CHMP). Guideline on the Specification Limits for Residues of Metal Catalysts or Metal Reagents, 2008.
25. Hot topics: elemental impurities. www.usp.org/usp-nf/hot-topics/hot-topics-elemental-impurities (Accessed 5 February, 2012).
26. Chapter <231>, Heavy metals. United States Pharmacopeial Convention. *United States Pharmacopeia*, 34th edn, and *National Formulary*, 29th edn. Baltimore: United Book Press, 2011.
27. In-process revision: <232> elemental impurities—limits. *Pharmacopeial Forum* 2011; 37(3).
28. In-process revision: <233> elemental impurities—procedures. *Pharmacopeial Forum* 2011; 37(3).
29. Marshall K *et al.* Development of a compendial taxonomy and glossary for pharmaceutical dosage forms. *Pharmacopeial Forum* 2002; 29(5).
30. 2010–2015 general chapters—Dosage Forms Expert Committee. www.usp.org/council-experts-expert-committees-overview/expert-committees/general-chapters-dosage-forms. (Accessed 12 February, 2012). Also see the link to the Work Plan for this Expert Committee. http://www.usp.org/sites/default/files/usp_pdf/EN/members/workplansgcdf.xls.xls.
31. Chapter <1>, Injections. United States Pharmacopeial Convention. *United States Pharmacopeia*, 34th edn, and *National Formulary*, 29th edn. Baltimore: United Book Press, 2011.
32. Chapter <1226>, Verification of compendial procedures. United States Pharmacopeial Convention. *United States Pharmacopeia*, 34th edn, and *National Formulary*, 29th edn. Baltimore, United Book Press, 2011.
33. Chapter <791>, pH. United States Pharmacopeial Convention. *United States Pharmacopeia*, 34th edn, and *National Formulary*, 29th edn. Baltimore: United Book Press, 2011.
34. Chapter 2.2.3, Potentiometric determination of pH. European Directorate for the Quality for Medicines and Healthcare of the Council of Europe. *European Pharmacopeia*, 7th edn. Strasbourg, France, 2010.
35. Chapter <788>, Particulate matter in injections. United States Pharmacopeial Convention. *United States Pharmacopeia*, 34th edn, and *National Formulary*, 29th edn. Baltimore: United Book Press, 2011.
36. Chapter <71>, Sterility tests. United States Pharmacopeial Convention. *United States Pharmacopeia*, 34th edn, and *National Formulary*, 29th edn. Baltimore: United Book Press, 2011.
37. Chapter <85>, Bacterial endotoxins test. United States Pharmacopeial Convention. *United States Pharmacopeia*, 34th edn, and *National Formulary*, 29th edn. Baltimore: United Book Press, 2011.
38. Chapter <905>, Uniformity of dosage forms. United States Pharmacopeial Convention. *United States Pharmacopeia*, 34th edn, and *National Formulary*, 29th edn. Baltimore: United Book Press, 2011.
39. Chapter 2.9.40, Uniformity dosage units. European Directorate for the Quality for Medicines and Healthcare of the Council of Europe. *European Pharmacopeia*, 7th edn. Strasbourg, France, 2010.
40. Chapter 2.9.5, Uniformity of mass of single-dose preparations. *European Pharmacopeia*, 7th edn. Strasbourg, France: European Directorate for the Quality for Medicines and Healthcare of the Council of Europe, 2010.
41. Chapter 2.9.6, Uniformity of content of single-dose preparations. European Directorate for the Quality for Medicines and Healthcare of the Council of Europe. *European Pharmacopeia*, 7th edn. Strasbourg, France, 2010.
42. Chapter <921>, Water determination. United States Pharmacopeial Convention. *United States Pharmacopeia*, 34th edn, and *National Formulary*, 29th edn. Baltimore: United Book Press, 2011.
43. Dosage forms. European Directorate for the Quality for Medicines and Healthcare of the Council of Europe. *European Pharmacopeia*, 7th edn. Strasbourg, France, 2010: 705–742.
44. For the special case in which an overage of the drug substance has been added for stability reasons, the USP <901> and Ph. Eur. 2.9.40 should be consulted.
45. Tablets. European Directorate for the Quality for Medicines and Healthcare of the Council of Europe. *European Pharmacopeia*, 7th edn. Strasbourg, France, 2010: 736–738.
46. Chapter <1216>, Tablet friability. United States Pharmacopeial Convention. *United States Pharmacopeia*, 34th edn, and *National Formulary*, 29th edn. Baltimore: United Book Press, 2011.
47. Chapter <1217>, Tablet breaking force. United States Pharmacopeial Convention. *United States Pharmacopeia*, 34th edn, and *National Formulary*, 29th edn. Baltimore: United Book Press, 2011.

48. Chapter <941>, Characterization of crystalline and partially crystalline solids by x-ray powder diffraction. United States Pharmacopeial Convention. *United States Pharmacopeia*, 34th edn, and *National Formulary*, 29th edn. Baltimore: United Book Press, 2011.
49. Bauer J *et al.* Ritonovir: an extraordinary example of conformational polymorphism. *Pharm Res* 2001;18: 859–866.
50. Chapter <61>, Microbiological examination of nonsterile products: microbial enumeration tests. United States Pharmacopeial Convention. *United States Pharmacopeia*, 34th edn, and *National Formulary*, 29th edn. Baltimore: United Book Press, 2011.
51. Chapter <62>, Microbiological examination of nonsterile products: tests for specified microorganisms. United States Pharmacopeial Convention. *United States Pharmacopeia*, 34th edn, and *National Formulary*, 29th edn. Baltimore: United Book Press, 2011.
52. Chapter 2.6.12, Microbiological examination of non-sterile products: microbial enumeration tests. European Directorate for the Quality for Medicines and Healthcare of the Council of Europe. *European Pharmacopeia*, 7th edn. Strasbourg, France, 2010.
53. Chapter 2.6.13, Microbiological examination of non-sterile products: test for specified micro-organisms. European Directorate for the Quality for Medicines and Healthcare of the Council of Europe. *European Pharmacopeia*, 7th edn. Strasbourg, France, 2010.
54. Chapter <711>, Dissolution. United States Pharmacopeial Convention. *United States Pharmacopeia*, 34th edn, and *National Formulary*, 29th edn. Baltimore: United Book Press, 2011.
55. Chapter 2.9.3, Dissolution test for solid dosage forms. European Directorate for the Quality for Medicines and Healthcare of the Council of Europe. *European Pharmacopeia*, 7th edn. Strasbourg, France, 2010.
56. Chapter <1092>, The dissolution procedure: development and validation. United States Pharmacopeial Convention. *United States Pharmacopeia*, 34th edn, and *National Formulary*, 29th edn. Baltimore: United Book Press, 2011.
57. Chapter <701>, Disintegration. United States Pharmacopeial Convention. *United States Pharmacopeia*, 34th edn, and *National Formulary*, 29th edn. Baltimore: United Book Press, 2011.
58. Chapter 2.9.1, Disintegration of tablets and capsules. European Directorate for the Quality for Medicines and Healthcare of the Council of Europe. *European Pharmacopeia*, 7th edn. Strasbourg, France, 2010.
59. Chart 4: Non-complex active drug products. In: The Guide to General Chapters United States Pharmacopeial Convention. *United States Pharmacopeia*, 34th edn, and *National Formulary*, 29th edn. Baltimore: United Book Press, 2011: 19.
60. In-process revision: <725> Topical and transdermal drug products—product performance tests. *Pharmacopeial Forum* 2009; 35(3).
61. In-process revision: <1724> Topical and transdermal drug products—product performance tests. *Pharmacopeial Forum* 2011; 37(5).
62. USP forwards proposed general chapters on topical and transdermal drug products. General announcement posted on 29 May 2009. www.usp.org/usp-nt/notices/retired-compendial-notices/usp-forwhhttpdrstagingusporgnode1117editgroupbulards-proposed-general-chapters-topical (Accessed 12 February, 2012).
63. In-process revision: <5> Inhalation and nasal drug products—general information and product quality tests. *Pharmacopeial Forum* 2011; 37(4).
64. In-process revision: <601> Nasal sprays, metered-dose inhalers, and dry powder inhalers. *Pharmacopeial Forum* 2011; 37(4).
65. Aerosols—Inhalation and Nasal Drug Products Workshop. www.usp.org/meetings-courses/workshops/past-usp-workshops/aerosols-inhalation-and-nasal-drug-products-workshop (Accessed 12 February, 2012).
66. ICH Harmonized Tripartite Guideline Q8(R2). Pharmaceutical Development, Revision 2, 2009.
67. ICH Harmonized Tripartite Guideline Q9. Quality Risk Management, 2005.
68. ICH Harmonized Tripartite Guideline Q10, Pharmaceutical Quality Risk Management, 2008.
69. ICH Harmonized Tripartite Draft Guideline Q11. Development and Manufacture of Drug Substances, 2011.
70. USP monograph: Nafcillin injection. United States Pharmacopeial Convention. *United States Pharmacopeia*, 34th edn, and *National Formulary*, 29th edn., Baltimore: United Book Press, 2011: 3588–3589.
71. USP monograph: Nafcillin for injection. United States Pharmacopeial Convention. *United States Pharmacopeia*, 34th edn, and *National Formulary*, 29th edn. Baltimore: United Book Press, 2011: 3589.
72. USP monograph: Capecitabine tablets. United States Pharmacopeial Convention. *United States Pharmacopeia* 34th edn, and *National Formulary*, 29th edn. Baltimore: United Book Press, 2011: 2148–2149.
73. USP monograph: Naproxen delayed-release tablets. United States Pharmacopeial Convention. *United States Pharmacopeia*, 34th edn, and *National Formulary*, 29th edn., Baltimore: United Book Press, 2011: 3604.
74. USP monograph: Acetaminophen extended-release tablets. United States Pharmacopeial Convention. *United States Pharmacopeia*, 34th edn, and *National Formulary*, 29th edn. Baltimore: United Book Press, 2011: 1723–1724.

APPENDIX A EXAMPLE SPECIFICATIONS

Table A-1. ABC1234 Injection Specifications

Attribute	Method	Acceptance Criteria
Appearance	Visual	Clear, colorless solution and essentially free of particles or foreign matter
Identification	Reversed-phase HPLC Method ABC-2012-0001	Conforms to Reference Standard (RRT 1.00 ± 0.05)
Assay	Reversed-phase HPLC Method ABC-2012-0001	90.0–110.0% of label claim
Impurities	Reversed-phase HPLC Method ABC-2012-0001	Specified Identified Impurities ABC0001 (RRT ~ 0.37): $\leq 0.30\%$ area ABC0002 (RRT ~ 1.15): $\leq 1.00\%$ area ABC0003 (RRT ~ 1.58): $\leq 3.00\%$ area Specified Unidentified Impurities RRT ~ 1.08: $\leq 0.20\%$ area Unspecified Unidentified Impurities Any other individual Impurity $\leq 0.10\%$ area Total Impurities $\leq 4.00\%$ area $\leq 1.00\%$ area
Enantiomeric purity	Chiral HPLC Method ABC-2012-0002	
pH	USP <791>	5.5 ± 0.3
Particulate matter	USP <788>	Light Obscuration Particle Count particles per container $\geq 10 \mu\text{m}$: ≤ 6000 particles per container $\geq 25 \mu\text{m}$: ≤ 600 Microscopic Particle Count^a particles per container $\geq 10 \mu\text{m}$: ≤ 3000 particles per container $\geq 25 \mu\text{m}$: ≤ 300
Sterility	USP <71>	No Growth
Endotoxin	USP <85>	$< 3.5 \text{ EU/mg of ABC1234}$

^aMicroscopic particle count will be performed only if the light obscuration particle count exceeds the acceptance criteria.

Table A-2. ABC1234 Injection Specifications

Attribute	Method	Acceptance Criteria
Appearance	Lyophilized Cake Visual	White to off-white cake or powder, essentially free of foreign matter
Identification	Reversed-phase HPLC Method ABC-2012-0003	Conforms to Reference Standard (RRT 1.00 ± 0.05)
Assay	Reversed-phase HPLC Method ABC-2012-0003	90.0–110.0% of label claim (5.0 mg ABC1234/vial)
Impurities	Reversed-phase HPLC Method ABC-2012-0003	Specified Identified Impurities ABC0001 (RRT ~ 0.37): $\leq 0.30\%$ area ABC0002 (RRT ~ 1.15): $\leq 1.00\%$ area ABC0003 (RRT ~ 1.58): $\leq 2.00\%$ area Specified Unidentified Impurities RRT ~ 1.08: $\leq 0.20\%$ area Unspecified Unidentified Impurities Any other individual impurity $\leq 0.10\%$ area Total Impurities $\leq 3.00\%$ area $\leq 1.00\%$ area
Enantiomeric purity	Chiral HPLC Method ABC-2012-0004	
Moisture content	USP <921>	Record result
Content uniformity	USP <905>; Reversed-phase HPLC Method ABC-2012-0005	Conforms to USP and Ph. Eur. requirements

Attribute	Method	Acceptance Criteria
Sterility	USP <71>	No Growth
Endotoxin	USP <85>	< 3.5 EU/mg of ABC1234
Reconstitution time	Visual	≤ 120 seconds
Reconstituted Solution		
Appearance	Visual	Clear, colorless solution, essentially free of particles and foreign matter
pH	USP <791>	5.0 ± 0.3
Particulate matter	USP <788>	Light Obscuration Particle Count particles per container ≥10 µm: ≤ 6000 particles per container ≥25 µm: ≤ 600 Microscopic Particle Count^a particles per container ≥10 µm: ≤ 3000 particles per container ≥25 µm: ≤ 300

^aMicroscopic particle count will be performed only if the light obscuration particle count exceeds the acceptance criteria.

Table A-3. ABC5678 Tablet (10- and 100-mg) Specifications

Attribute	Method	Acceptance Criteria
Appearance	Visual	White to off-white round convex tablets
Identification	Reversed-phase HPLC Method ABC-2012-0006	Conforms to Reference Standard (RRT 1.00 ± 0.05)
Assay	Reversed-phase HPLC Method ABC-2012-0006	90.0–110.0% of label claim
Impurities	Reversed-phase HPLC Method ABC-2012-0006	Specified Identified Impurities ABC0001 (RRT ~ 0.37): ≤ 0.30% area ABC0002 (RRT ~ 1.15): ≤ 1.00% area ABC0003 (RRT ~ 1.58): ≤ 1.50% area Specified Unidentified Impurities RRT ~ 1.08: ≤ 0.20% area Unspecified Unidentified Impurities Any other individual impurity ≤ 0.10% area Total Impurities ≤ 3.00% area ≤ 1.00% area
Enantiomeric purity	Chiral HPLC Method ABC-2012-0007	≤ 1.00% area
Content uniformity	USP <905>; Reversed-phase HPLC Method ABC-2012-0008	Conforms to USP and Ph. Eur. requirements
Dissolution	USP <711>; Reversed-phase HPLC Method ABC-2012-0009	Record result
Moisture content	USP <901>	Record result
Tablet breaking force	USP <1217>	Record result
X-ray powder diffraction	USP <941>	Consistent with a reference pattern
Microbial Enumeration Tests	USP <61> or Ph. Eur. 2.6.12	Total Aerobic Microbial Count ≤ 10 ³ CFU/g Total Yeast and Molds Count ≤ 10 ² CFU/g
Microbial Specified Microorganisms Test	USP <62> or Ph. Eur. 2.6.13	Absence of <i>E. coli</i>

Appendix B Sample Monographs

NAFCILLIN INJECTION⁷⁰

Nafcillin Injection is a sterile isoosmotic solution of Nafcillin Sodium and one or more buffer substances in Water for Injection. It contains dextrose as a tonicity-adjusting agent. It contains an amount of nafcillin sodium equivalent to not less than 90.0 percent and not more than 120.0 percent of the labeled amount of nafcillin ($C_{21}H_{22}N_2O_5S$). It contains no antimicrobial preservatives.

PACKAGING AND STORAGE

Preserve in *Containers for Injections* as described under *Injections* <1>. Maintain in the frozen state.

LABELING

It meets the requirements for *Labeling* under *Injections* <1>. The label states that it is to be thawed just prior to use, describes conditions for proper storage of the resultant solution, and directs that the solution is not to be refrozen.

USP REFERENCE STANDARDS <11>

USP Endotoxin RS
USP Nafcillin Sodium RS

IDENTIFICATION

The retention time of the major peak for nafcillin in the chromatogram of the *Assay preparation* corresponds to that in the chromatogram of the *Standard preparation*, as obtained in the *Assay*.

BACTERIAL ENDOTOXINS <85>

It contains not more than 0.13 USP Endotoxin Unit per mg of nafcillin.

STERILITY <71>

It meets the requirements when tested as directed for *Membrane Filtration* under *Test for Sterility of the Product to be Examined*.

pH <791>: between 6.0 and 8.5.

Particulate matter <788>: meets the requirements for small-volume injections.

ASSAY

Acetic acid solution, 0.05 M Sodium acetate, Diluent, Mobile phase, Standard preparation, Resolution solution, and Chromatographic system—Proceed as directed in the *Assay* under *Nafcillin Sodium*.

Assay preparation

Allow one container of Injection to thaw, and mix. Transfer an accurately measured volume of Injection, equivalent to about 40 mg of nafcillin, to a 100-mL volumetric flask, dilute with *Diluent* to volume, and mix.

Procedure

Proceed as directed for *Procedure* in the *Assay* under *Nafcillin Sodium*. Calculate the quantity, in mg, of nafcillin ($C_{21}H_{22}N_2O_5S$) in each mL of the Injection taken by the formula:

$$0.1(C/V)(r_u/r_s)$$

in which *C* is the concentration, in µg per mL, of nafcillin in the *Standard preparation*; *V* is the volume, in mL, of Injection taken to prepare the *Assay preparation*; and *r_u* and *r_s* are the nafcillin peak responses obtained from the *Assay preparation* and the *Standard preparation*, respectively.

NAFCILLIN FOR INJECTION⁷¹

Nafcillin for Injection contains an amount of Nafcillin Sodium equivalent to not less than 90.0 percent and not more than 120.0 percent of the labeled amount of nafcillin ($C_{21}H_{22}N_2O_5S$).

PACKAGING AND STORAGE

Preserve in *Containers for Sterile Solids* as described under *Injections* <1>.

USP REFERENCE STANDARDS <11>

USP Endotoxin RS
USP Nafcillin Sodium RS

CONSTITUTED SOLUTION

At the time of use, it meets the requirements for *Constituted Solutions* under *Injections* <1>.

IDENTIFICATION

The retention time of the major peak for nafcillin in the chromatogram of the *Assay preparation* corresponds to that in the chromatogram of the *Standard preparation*, as obtained in the *Assay*.

BACTERIAL ENDOTOXINS <85>

It contains not more than 0.13 USP Endotoxin Unit per mg of nafcillin.

STERILITY <71>

It meets the requirements when tested as directed for *Membrane Filtration* under *Test for Sterility of the Product to be Examined*.

pH <791>: between 6.0 and 8.5, in the solution constituted as directed in the labeling.

Water, Method I <921>: between 3.5 percent and 5.3 percent.

Particulate matter <788>: meets the requirements for small-volume injections.

OTHER REQUIREMENTS

It meets the requirements for *Uniformity of Dosage Units* <905> and *Labeling* under *Injections* <1>.

ASSAY

Acetic acid solution, 0.05 M Sodium acetate, Diluent, Mobile phase, Standard preparation, Resolution solution, and Chromatographic system—Proceed as directed in the *Assay* under *Nafcillin Sodium*.

Assay preparation 1

(where it is represented as being in a single-dose container)—Constitute Nafcillin for Injection in a volume of water, accurately measured, corresponding to the volume

of solvent specified in the labeling. Withdraw all of the withdrawable contents, using a suitable hypodermic needle and syringe, and dilute quantitatively with *Diluent* to obtain a solution having a concentration of about 0.4 mg of nafcillin ($C_{21}H_{22}N_2O_5S$) per mL.

Assay preparation 2

(where the label states the quantity of nafcillin in a given volume of constituted solution)—Constitute Nafcillin for Injection in a volume of water, accurately measured, corresponding to the volume of solvent specified in the labeling. Dilute an accurately measured volume of the constituted solution quantitatively with *Diluent* to obtain a solution having a concentration of about 0.4 mg of nafcillin ($C_{21}H_{22}N_2O_5S$) per mL.

Procedure

Proceed as directed for *Procedure* in the Assay under *Nafcillin Sodium*. Calculate the quantity, in mg, of nafcillin

($C_{21}H_{22}N_2O_5S$) in the portion of constituted Nafcillin for Injection taken by the formula:

$$(C / 1000)(L / D)(r_u / r_s)$$

In which *L* is the labeled quantity, in mg, of nafcillin in the portion of Nafcillin for Injection taken; *D* is the concentration, in mg per mL, of nafcillin in *Assay preparation 1* or *Assay preparation 2*, as appropriate, based on the volume of constituted Nafcillin for Injection taken and the extent of dilution; and the other terms are as defined therein.

Perform the above procedure on ten containers where it is represented as being in a single-dose container and, if necessary, on ten containers where the label states the quantity of nafcillin in a given volume of constituted solution. Use the individual results to determine the *Uniformity of dosage units* and the average thereof as the Assay value.

CAPECITABINE TABLETS⁷²

DEFINITION

Capecitabine Tablets contain NLT 93.0 percent and NMT 105.0 percent of the labeled amount of capecitabine ($C_{15}H_{22}FN_3O_6$).

IDENTIFICATION

- **A. INFRARED ABSORPTION <197K>**
Analytical wave number: 1500–1760 cm^{-1}
Sample: Grind 1 tablet to a fine powder with a mortar and pestle. Mix 1 mg of this sample with 300 mg of potassium bromide.
- **B.** The retention time of the major peak of the *Sample solution* corresponds to that of the *Standard solution*, as obtained in the Assay.

ASSAY

PROCEDURE

- **Diluent:** Methanol, acetonitrile, and water (7:1:12)
- **Solution A:** 0.1 percent mixture of glacial acetic acid in water
- **Solution B:** Methanol, acetonitrile, and *Solution A* (7:1:12)
- **Solution C:** Methanol, acetonitrile, and *Solution A* (16:1:3)
- **Mobile phase:** See the gradient table below.

Time (min)	Solution B b> (%)	Solution C (%)
0	100	0
5	100	0
20	49	51
30	49	51
31	100	0
40	100	0

[NOTE—The following solutions may be sonicated as necessary.]

System suitability solution: Includes 0.6 $\mu g/mL$ of USP Capecitabine RS, 0.6 $\mu g/mL$ of USP Capecitabine Related Compound A RS, 0.6 $\mu g/mL$ of USP Capecitabine Related Compound B RS, and 0.6 $\mu g/mL$ of USP Capecitabine Related Compound C RS in *Diluent*

Standard solution: 0.6 mg/mL of USP Capecitabine RS in *Diluent*

Sample solution: Equivalent to 0.6 mg/mL of capecitabine, from powdered tablets (NLT 20), in *Diluent*. [NOTE—Pass through a PVDF membrane filter of 0.45- μm pore size, and use the filtrate.]

Chromatographic system

(See *Chromatography* <621>, *System Suitability*.)

Mode: LC

Detector: UV 250 nm

Column: 4.6-mm x 25-cm; 5- μm packing L1

Column temperature: 40°

Autosampler temperature: 5°

Flow rate: 1 mL/min

Injection size: 10 μL

System suitability

Samples: *System suitability solution* and *Standard solution* [NOTE—For the purpose of peak identification, the approximate relative retention times are given in *Impurity Table 1*. The relative retention times are measured with respect to capecitabine.]

Suitability requirements

Resolution: NLT 1.0 between Capecitabine Related Compound A and Capecitabine Related Compound B, *System suitability solution*

Tailing factor: NMT 1.5, *Standard solution*

Relative standard deviation: NMT 2.0 percent, *Standard solution*

Analysis

Samples: *Standard solution* and *Sample solution*. Calculate the percentage of $C_{15}H_{22}FN_3O_6$ in the portion of tablets taken:

$$\text{Result} = (r_u / r_s) (C_s / C_u) 100$$

- r_u = peak response from the *Sample solution*
- r_s = peak response from the *Standard solution*
- C_s = concentration of USP Capecitabine R_s in the *Standard solution* (mg/mL)
- C_u = nominal concentration of capecitabine in the *Sample solution* (mg/mL)
- Acceptance criteria: 93.0 percent–105.0 percent

PERFORMANCE TESTS

• DISSOLUTION <711>

Medium: Water; 900 mL, degassed

Apparatus 2: 50 rpm

Time: 30 min

STANDARD SOLUTIONS

For tablets labeled to contain 150 mg: 17 mg of USP Capecitabine RS in 100 mL of *Medium*

For tablets labeled to contain 500 mg: 28 mg of USP Capecitabine RS in 50 mL of *Medium*

Sample solution: Pass a portion of the solution under test through a fiberglass filter of 0.45- μ m pore size.

Analysis: Determine the amount of $C_{15}H_{22}FN_3O_6$ dissolved by selecting a wavelength with appropriate sensitivity between 300 and 330 nm on portions of the *Sample solution*, suitably diluted with *Medium*, if necessary, in comparison with the appropriate *Standard solution*, using a 1-mm quartz cell. Calculate the percentage of $C_{15}H_{22}FN_3O_6$ dissolved in each tablet:

$$\text{Result} = (A_u / A_s) C_s (V/L) 100$$

- A_u = absorbance of the *Sample solution*
- A_s = absorbance of the *Standard solution*
- C_s = concentration of capecitabine in the *Standard solution* (mg/mL)
- V = volume of medium, 900 mL
- L = Tablet label claim (mg)

Tolerances: NLT 80 percent (Q) of the labeled amount of $C_{15}H_{22}FN_3O_6$ is dissolved.

Uniformity of Dosage Units <905>: Meet the requirements.

IMPURITIES

ORGANIC IMPURITIES

• PROCEDURE

Diluent, Solution A, Solution B, Solution C, Mobile phase, System suitability solution, Standard solution, Sample solution, and Chromatographic system:

Proceed as directed in the *Assay*.

ANALYSIS

Samples: *Standard solution* and *Sample solution*. Calculate the percentage of each impurity in the portion of tablets taken:

$$\text{Result} = (r_u / r_s) (C_s / C_u) 100 / F$$

- r_u = peak response for each impurity from the *Sample solution*
- r_s = peak response for capecitabine from the *Standard solution*
- C_s = concentration of USP Capecitabine RS in the *Standard solution* (mg/mL)
- C_u = nominal concentration of capecitabine in the *Sample solution* (mg/mL)
- F = relative response factor for each impurity, from *Impurity Table 1*

ACCEPTANCE CRITERIA

Individual impurities: See *Impurity Table 1*

Total degradation products: NMT 2.0 percent

ADDITIONAL REQUIREMENTS

- **PACKAGING AND STORAGE:** Preserve in tight containers. Store at controlled room temperature.
- **USP REFERENCE STANDARDS <11>**
USP Capecitabine RS
USP Capecitabine Related Compound A RS
5'-Deoxy-5-fluorocytidine.
 $C_9H_{12}FN_3O_4$ 245.21 USP Capecitabine Related Compound B RS
5'-Deoxy-5-fluorouridine.
 $C_9H_{11}FN_2O_5$ 246.19 USP Capecitabine Related Compound C RS
2',3'-O-Carbonyl-5'-deoxy-5-fluoro-N⁴-(pentyloxy-carbonyl)cytidine.
 $C_{16}H_{20}FN_3O_7$ 385.34.

Table B-1. Impurity Table 1

Name	Relative Retention Time	Relative Response Factor	Acceptance Criteria, NMT (%)
Capecitabine Related Compound A	0.18	1.05	1.0
Capecitabine Related Compound B	0.19	0.81	1.0
2',3'-Di-O-acetyl-5'-deoxy-5-fluorocytidine*	0.36	0.89	—
5'-Deoxy-5-fluoro-N ⁴ -(2-methyl-1-butyloxy-carbonyl)cytidine + 5'-Deoxy-5-fluoro-N ⁴ -(3-methyl-1-butyloxy-carbonyl)cytidine*	0.95	1.01	—
Capecitabine	1.00	1.00	—
[1-[5-Deoxy-3-O-(5-deoxy- β -D-ribofuranosyl)- β -D-ribofuranosyl]-5-fluoro-2-oxo-1,2-dihydropyrimidin-4-yl]-carbamic acid pentyl ester*	1.06	1.00	—
[1-[5-Deoxy-2-O-(5-deoxy- β -D-ribofuranosyl)- β -D-ribofuranosyl]-5-fluoro-2-oxo-1,2-dihydropyrimidin-4-yl]-carbamic acid pentyl ester*	1.09	1.00	—
Capecitabine Related Compound C	1.11	0.91	0.5
[1-[5-Deoxy-3-O-(5-deoxy- α -D-ribofuranosyl)- β -D-ribofuranosyl]-5-fluoro-2-oxo-1,2-dihydropyrimidin-4-yl]-carbamic acid pentyl ester*	1.20	1.00	—
2',3'-Di-O-acetyl-5'-deoxy-5-fluoro-N ⁴ -(pentyloxy-carbonyl)cytidine*	1.37	0.85	—
Individual unspecified degradation product	—	1.00	0.1

*These impurities are process impurities and are not included in the total degradation products.

NAPROXEN DELAYED-RELEASE TABLETS⁷³

Naproxen Delayed-Release Tablets contain not less than 90.0 percent and not more than 110.0 percent of the labeled amount of naproxen ($C_{14}H_{14}O_3$).

PACKAGING AND STORAGE

Preserve in well-closed containers, and store at controlled room temperature.

USP REFERENCE STANDARDS <11>

USP Naproxen RS

IDENTIFICATION

A: *Ultraviolet Absorption* <197U>

Test solution

Use the solution under test as obtained in the *Buffer stage* of the *Dissolution* test.

Standard solution

Use the Standard solution prepared as directed in the *Buffer stage* of the *Dissolution* test.

B: The retention time of the major peak in the chromatogram of the *Assay preparation* corresponds to that in the chromatogram of the *Standard preparation*, as obtained in the *Assay*.

Dissolution, Delayed-Release Dosage Forms, Method B <711>

ACID STAGE

Acid stage medium: 0.1 N hydrochloric acid; 1000 mL.

Apparatus 2: 50 rpm.

Time: 2 hours.

Procedure

Determine the amount of $C_{14}H_{14}O_3$ dissolved by employing UV absorption at the wavelength of maximum absorbance at about 332 nm on filtered portions of the solution under test, suitably diluted with *Acid stage medium*, if necessary, in comparison with a Standard solution having a known concentration of USP Naproxen RS in the same *Acid stage medium*.

Tolerances

Not more than 10 percent (*Q*) of the labeled amount of $C_{14}H_{14}O_3$ is dissolved in 2 hours.

BUFFER STAGE

Buffer stage medium: 0.2 M phosphate buffer, pH 6.8; 1000 mL.

Apparatus 2: 50 rpm.

Time: 45 minutes.

Procedure

Determine the amount of $C_{14}H_{14}O_3$ dissolved by employing UV absorption at the wavelength of maximum absorbance at about 332 nm on filtered portions of the solution under test, suitably diluted with *Buffer stage medium*, if necessary, in comparison with a Standard solution having a known concentration of USP Naproxen RS in the same *Buffer stage medium*.

Tolerances

Not less than 80 percent (*Q*) of the labeled amount of $C_{14}H_{14}O_3$ is dissolved in 45 minutes.

Uniformity of dosage units <905>: meet the requirements.

PROCEDURE FOR CONTENT UNIFORMITY

Mobile phase, Diluent A, Diluent B, and Chromatographic system

Proceed as directed in the *Assay*.

Standard solution

Transfer about 12.5 mg of USP Naproxen RS, accurately weighed, to a 50-mL volumetric flask, dilute with *Diluent A* to volume, and mix well. Transfer 10 mL of this solution to a 25-mL volumetric flask, dilute with *Diluent B* to volume, and mix.

Test solution

Transfer 1 tablet to a 200-mL volumetric flask, and add about 140 mL of *Diluent B*. Shake by mechanical means for 15 minutes, sonicate for 15 minutes, dilute with *Diluent B* to volume, and mix. Pass a portion of this solution through a filter having a porosity of 0.45 μ m, pipet 2.0 mL of the filtrate for a 500-mg tablet and 2.5 mL for a 375-mg tablet into a 50-mL volumetric flask, dilute with *Mobile phase* to volume, and mix.

ASSAY

Mobile phase

Prepare a filtered and degassed mixture of 1 percent acetic acid solution and acetonitrile (11 : 9).

Diluent A

Use acetonitrile and water (9 : 1).

Diluent B

Use acetonitrile and water (1 : 1).

Standard stock preparation

Transfer about 12.5 mg of USP Naproxen RS, accurately weighed, to a 25-mL volumetric flask. Dissolve in and dilute with *Diluent A* to volume, and mix.

Standard preparation

Accurately transfer 10.0 mL of the *Standard stock preparation* into a 50-mL volumetric flask, and dilute with *Mobile phase* to volume, and mix.

Assay preparation

Weigh and powder 20 tablets. Accurately weigh an amount of the powder, equivalent to about 250 mg of naproxen, into a 100-mL volumetric flask, and add about 70 mL of *Diluent B*. Shake by mechanical means for 15 minutes, sonicate for 15 minutes, dilute with *Diluent B* to volume, and mix. Pass this solution through a filter having a porosity of 0.45 μ m, transfer 2.0 mL of the filtrate into a 50-mL volumetric flask, dilute with *Mobile phase* to volume, and mix.

Chromatographic system (see *Chromatography* <621>—The liquid chromatograph is equipped with a 254-nm detector and a 4.6-mm $\times$ 25-cm column that contains 5- μ m packing L1. The flow rate is about 1.0 mL per minute. Chromatograph the *Standard preparation*, and record the peak responses as directed for *Procedure*: the tailing factor of the naproxen peak is not more than 1.5, and the relative standard deviation for replicate injections of the *Standard preparation* is not more than 2.0 percent.

Procedure

Separately inject equal volumes (about 50 μ L) of the *Standard preparation* and the *Assay preparation* into the chromatograph, record the chromatograms, and measure the responses for the naproxen peak. Calculate the quantity, in mg, of naproxen ($C_{14}H_{14}O_3$) in the portion of tablets taken by the formula:

$$2500(r_u/r_s)$$

in which *C* is the concentration, in mg per mL, of USP Naproxen RS in the *Standard preparation*; and r_u and r_s are

the peak responses obtained from the *Assay preparation* and the *Standard preparation*, respectively.

ACETAMINOPHEN EXTENDED-RELEASE TABLETS⁷⁴

Acetaminophen Extended-Release Tablets contain not less than 90.0 percent and not more than 110.0 percent of the labeled amount of acetaminophen ($\text{C}_8\text{H}_9\text{NO}_2$).

PACKAGING AND STORAGE

Preserve in tight containers.

LABELING

Where the tablets are gelatin-coated, the label so states. When more than one *Dissolution* test is given, the labeling states the *Dissolution* test used only if *Test 1* is not used.

USP REFERENCE STANDARDS <11>

USP Acetaminophen RS Identification

A: *Infrared Absorption* <197K>—Use a portion of powdered tablets.

B: The retention time of the major peak in the chromatogram of the *Assay preparation* corresponds to that in the chromatogram of the *Standard preparation*, as obtained in the *Assay*.

DISSOLUTION <711>

TEST 1

Medium: simulated gastric fluid TS (without enzyme); 900 mL.

Apparatus 2: 50 rpm.

Times: 15 minutes, 1 hour, and 3 hours.

Procedure

Determine the amount of $\text{C}_8\text{H}_9\text{NO}_2$ dissolved from UV absorbances at 280 nm, using a filtered portion of the solution under test in comparison with a Standard solution having a known concentration of USP Acetaminophen RS in the same *Medium*.

Tolerances

The percentages of the labeled amount of $\text{C}_8\text{H}_9\text{NO}_2$ dissolved at the times specified conform to *Acceptance Table*.

Time	Amount dissolved
15 minutes	between 45% and 65%
1 hour	between 60% and 85%
3 hours	not less than 85%

FOR GELATIN-COATED TABLETS

Medium, Apparatus, and Procedure

Proceed as directed above.

Times: 30 minutes, 90 minutes, and 4 hours.

Tolerances

The percentage of the labeled amount of $\text{C}_8\text{H}_9\text{NO}_2$ dissolved at the times specified conform to *Acceptance Table*.

Time	Amount dissolved
30 minutes	between 40% and 60%
90 minutes	between 55% and 85%
4 hours	not less than 80%

TEST 2

If the product complies with this test, the labeling indicates that it meets USP *Dissolution Test 2*.

Medium, Apparatus, and Procedure

Proceed as directed for *Test 1*.

Tolerances

The percentages of the labeled amount of $\text{C}_8\text{H}_9\text{NO}_2$ dissolved at the times specified conform to *Acceptance Table*.

Time	Amount dissolved
15 minutes	between 40% and 60%
1 hour	between 55% and 75%
3 hours	not less than 80%

Uniformity of dosage units <905>: meet the requirements.

ASSAY

Mobile phase

Prepare a mixture of water and phosphoric acid (9 : 1). Combine 1 mL of this solution with a mixture of water and methanol (700 : 300). Filter and degas. Make adjustments if necessary. See *System Suitability* under *Chromatography* <621>.

Standard preparation

Dissolve an accurately weighed quantity of USP Acetaminophen RS in methanol, and dilute quantitatively, and stepwise if necessary, with *Mobile phase* to obtain a solution having a known concentration of about 0.65 mg per mL.

Assay preparation

Transfer 10 tablets into a 250-mL volumetric flask containing 50 mL of water and a magnetic stir bar. Stir at least 30 minutes or until the coating has dissolved. Add 150 mL of methanol, and stir for 45 minutes. Tablet cores should be disintegrated at least 15 minutes prior to ending the stirring. Remove the magnetic stir bar and rinse into the flask with methanol. Dilute with methanol to volume, mix well, and centrifuge. Transfer 5 mL of the clear supernatant to a 200-mL volumetric flask, dilute with *Mobile phase* to volume, and mix well.

Chromatographic system (see *Chromatography* <621>—The liquid chromatograph is equipped with a 295-nm detector and a 3.9-mm x 15-cm column that contains packing L1. The flow rate is about 2.0 mL per minute. Chromatograph the *Standard preparation*, and record the peak responses as directed for *Procedure*: the tailing factor is not more than 3.0, and the relative standard deviation for replicate injections is not more than 2.0 percent.

Procedure

Separately inject equal volumes (about 20 μL) of the *Standard preparation* and the *Assay preparation* into the chromatograph, record the chromatograms, and measure the responses for the major peaks. Calculate the quantity, in mg, of acetaminophen ($\text{C}_8\text{H}_9\text{NO}_2$) in each tablet taken by the formula:

$$1000C(r_u/r_s)$$

In which C is the concentration, in mg per mL, of USP Acetaminophen RS in the *Standard preparation*; and r_u and r_s are the acetaminophen peak responses obtained from the *Assay preparation* and the *Standard preparation*, respectively.

Chapter 3

Quality Assurance and Control

John H. Parker, PhD and John E. Enders, PhD, MBA

QA AND QC: ORGANIZATION/RESPONSIBILITIES 29

QUALITY BY DESIGN 30

DOCUMENTATION 31

QUALITY IN PHARMACEUTICAL BIOTECHNOLOGY 31

CONTEMPORARY ISSUES 32

FDA MODERNIZATION 32

APPENDIX A 34

The pharmaceutical industry, as a vital segment of the healthcare system, conducts research and manufactures and markets pharmaceutical and biological products and medical devices used for the acute/chronic treatment and diagnosis of disease. Recent advances in drug discovery, primarily in the field of biotechnology and in the required controls over manufacturing processes, are presenting new challenges to the control of quality and to the systems that operate internally in the industry. The external regulations established by the federal Food and Drug Administration (FDA) and other regulatory bodies also add to these challenges. The evolving role of the industrial quality professional requires more extensive education including food and drug law, business, as well as the traditional science/technology coursework.

The pursuit of quality is being approached through the application of quality systems including risk-based assessment and continuous improvement, whereby management and labor join forces to build quality into products while helping to ensure the company's financial success. This changed emphasis is directed toward defect prevention (proactive) rather than defect detection (after the fact).

Quality assurance (QA) and quality control (QC) departments develop and follow standard internal operating procedures directed toward assuring the quality, safety, purity, and effectiveness of drug products. The FDA has issued a primary regulation to the industry entitled *Current Good Manufacturing Practice for Finished Pharmaceuticals* (commonly referred to as the cGMPs or GMPs) (see Appendix). Numerous guidelines have been issued relative to specific dosage forms and operations such as aseptic manufacturing, validation and stability testing, etc., which impose significant compliance requirements. These guidelines also serve as the basis for compliance investigations conducted by the FDA and are used in regulatory agency inspections of facilities and operations. Emphasis is being placed on the inspection of quality systems as part of the regulatory pre-approval program when reviewing New Drug Applications (NDAs) and Biological License Applications (BLAs).

QA AND QC: ORGANIZATION/RESPONSIBILITIES

Industry, to ensure compliance with these government regulations and with their own internal policies and procedures, has developed very sophisticated quality organizations with well-defined responsibilities. It has been accepted that QA and QC have different functions within an organization; although both are considered part of the Quality Unit as identified in 21 CFR (see Appendix). QC most commonly functions to test and measure material and product. QA establishes systems for ensuring the quality of the product. Firms must decide upon the exact roles they wish QC and QA to perform in operations and put these definitions in writing.

QA FUNCTIONS AND RESPONSIBILITIES

The QA department within any organization, because of its responsibilities, normally will report to a relatively high-level

administrator within a company, depending on its size. In smaller companies they may report to the chief executive officer or the president. In larger corporations, they will sometimes report to the president or executive vice-president or chief of operations. In any case, however, responsibility for quality, as currently dictated by FDA, ultimately resides with top management which has responsibility for ensuring that appropriate resources are provided to meet all quality and compliance requirements. In all cases, the QA department will be independent of the economic issues associated with manufacturing and distribution of the product.

The QA department is responsible for ensuring that the quality policies adopted by a company are followed. In some organizations, QA serves as the primary contact with regulatory agencies and is the final authority for product acceptance (release) or rejection. It is customary for QA to play a major role in the identification and preparation of the necessary policies and standard operating procedures (SOPs) relative to the control of quality. Where QA has responsibility for final product release, it must determine that the product meets all the applicable specifications and that it was manufactured according to internal standards and cGMPs. QA departments now tend to work as a team member with the other functional groups within the firm rather than simply to serve a police function, a largely outdated role of QA.

A second major responsibility of the QA department is the quality monitoring or audit function. Through this activity, it is able to determine if operations have adequate systems, facilities, and written procedures to control the quality of products produced. Thus, the QA function not only determines that the procedures are current and correct, but that properly trained operators are following them. Combining this review of SOPs with an audit of facilities and operations, including those of contract manufacturing and testing subcontractors, will give company management an inside report on its level of compliance and will allow the necessary changes and/or corrections to be made prior to a product failure or being reported as a deficiency during an inspection by an FDA investigator. This is consistent with the top-level management review component of the quality systems approach currently emphasized by FDA during inspections. Senior management of a company looks to the QA function to assess operations continually and to advise and guide them toward full compliance with all applicable internal and external regulations. Organizationally, the Quality department(s) should report, as directed by the GMPs, to someone other than the person responsible for production.

QC FUNCTIONS AND RESPONSIBILITIES

Quality Control is responsible for the day-to-day control of quality within a company. This department is staffed with scientists and technicians responsible for the sampling and analytical testing of incoming raw materials and inspection of packaging components, including labeling. QC conducts in-process testing when required, performs environmental monitoring, inspects operations for compliance, and conducts the required release

tests on finished dosage form. Finally, QC is responsible for monitoring product quality through distribution, including testing of product complaint samples, evaluating product stability, and so on.

Many companies have the heads of QC and production report to a common higher level of management, but with QC being independent of production. This higher-level management may be the same or different individuals, but it allows independent operation of both functions without direct conflict arising when reaching a final decision on the acceptability of final products.

The analytical control laboratory must be staffed with persons who are trained academically and are, through experience, capable of performing the often complex analyses used to evaluate the acceptability of the materials tested. The equipment and instrumentation in the laboratory must be suitable for performing the testing in an accurate and efficient manner. Detailed specifications must be available, as well as validated test methods against which products and raw materials will be evaluated. The specifications detail the limits for acceptance of the article, based on identified critical parameters.

The testing and acceptance of only high-quality raw materials is essential for the production of uniformly acceptable products. Quality Control plays a major role in the selection and qualification of vendors from whom these materials are purchased. Testing of representative samples is required, and in many cases, an audit of the vendor's operation is necessary to determine their suitability and degree of compliance with GMPs and other relevant standards prior to their being approved. The vendor audit frequently is organized by QA, with technical support from research, QC, and manufacturing.

At various critical in-process steps in production, it may be necessary to sample and test product against criteria previously established. Often, in-process alert or action levels will be identified for the critical in-process parameters as a means of process control. These alert or action levels are normally set such that they are more restrictive than the final acceptance limits, but serve as an in-process control by providing early warnings of conditions that could lead to an out-of-control situation and thus will allow timely corrective action before such conditions occur. The acceptance criteria for such testing are established using Quality by Design (QBD) approaches which identify the Design Space within which the process will perform satisfactorily. Trending of analytical data is also useful in providing early warning signals that the process is moving out of control. It should be noted, however, that materials, which have reached the alert or action level criteria, are still acceptable for use in manufacturing, since they have not exceeded an out-of-limit rejection level.

Quality Control is responsible, as part of its testing and inspection functions, for monitoring the environmental conditions under which products are manufactured and/or held. Different levels of control are established depending on the intended use of the dosage form. Parenteral and ophthalmic products must be produced in a controlled environment that is designed to ensure their sterility. Monitoring of air and water systems is critical in confirming that they are being controlled and that the levels of particulates, microbial matter, and other contaminants are within pre-established limits. The United States Pharmacopeia (USP) contains monographs and specifications on Water Used for Pharmaceutical Purposes. Formerly, the Federal Government Standard 209E, *Airborne Particulate Cleanliness Classes in Cleanrooms and Clean Zones*, established acceptable limits for particulates in a controlled environment, but it is no longer considered applicable to the pharmaceutical industry, having been replaced by international standards. Federal standards are currently not enforced for environmental quality, but guidance is available in the FDA Concept Paper, *Sterile Drug Products Produced by Aseptic Processing*, published in September 2004.¹ In addition, reference is made to the Baseline Pharmaceutical Engineering Guide, Vol. 3, *Sterile Manufacturing Facilities*, published by the International Society

of Pharmaceutical Engineering (ISPE) in partnership with the FDA, in June 2000.² Generally, conditions listed as Class ISO 5 (formerly Class 100) (or equivalent) are maintained in areas where parenteral products are filled into clean, sterile containers. The ISO 5 classification is defined as an area that can be controlled to contain fewer than 100 particles, 0.5 microns and larger, per cubic foot of air. In addition, manufacturers must establish limits for the presence of viable microorganisms in the environment and appropriately monitor the air quality in the filling area.

Another major element of control is the environmental monitoring of the areas in which nonsterile products are manufactured, such as liquids, tablets, and capsules. The objective here is first to determine an acceptable level of particulates and microbial contaminants and then to control them to this level. If particulate levels are found to be excessive, steps must be taken to bring them within acceptable limits so as not to compromise the quality of the product. This monitoring and control of the environment will further ensure the quality and stability of the product by preventing the products from being exposed to a hostile environment.

Control of packaging components, especially those that come into direct contact with a product, is required. These materials must be inspected and tested against rigid specifications to ensure that they meet predetermined functional standards. This includes evaluation of compatibility of the product with the packaging materials. Labeling is understandably a critical component, not just in its original design and acceptance, but also with regard to secure storage and issuance to ensure accountability. Furthermore, final product labeling must be 100% inspected to ensure that it is correct. Often automated imaging methods are employed to conduct these inspections.

QUALITY BY DESIGN

The production of quality pharmaceutical products requires embracing the principles of Total Quality Management (TQM). Although the term TQM has fallen out of favor in recent years and has been replaced by other, though similar designations, such as Total Product Quality (TPQ), the principles of TQM will serve to improve productivity and customer satisfaction. The quality function is part of a team composed of research, production, marketing/sales, and customer service. In the competitive markets of today, it is critical to improve quality and service continually while minimizing costs and maximizing resource utilization to help contain overall healthcare costs. The concept of TQM requires the total commitment of senior-level management and supervision of all departments, operators, suppliers, and customers. Its basic principle is one of continually striving for process improvement that begins with product development and only concludes when feedback and follow-up have been completed on customer complaints and suggestions. In many firms the QA department has the responsibility to organize and implement programs with these objectives in mind.

Quality must be designed into products, beginning with research and development phases. Quality by Design as described in the International Committee on Harmonization (ICH) Q8 (R2) Guidance for Industry³ is becoming more common in the development of products and processes. Product quality criteria are established, and detailed specifications are written. Meticulous, written procedures must be prepared for production and control, and processes must be rigorously validated. Raw materials must be characterized and then purchased from reputable, approved suppliers to ensure that, when the materials are incorporated into the finished dosage form, they will provide products of uniformly high quality. Facilities must be designed, constructed, and controlled to provide the proper stable environment for protecting the integrity of products. Equipment must be selected that is efficient and can be cleaned readily and sanitized, to aid in preventing cross-contamination of one product with another. Single use equipment is commonly used in

the industry, particularly in manufacture of biological products. Product contact components are discarded after use. Personnel must be trained properly so that their personal habits, clothing, and job performance do not compromise product quality. The directions that they use must be in writing, approved by responsible individuals, and strictly followed. Training programs must be thoroughly documented and include an evaluation of mastery of the procedures employed.

Distribution departments are responsible for controlling the shipping and handling of products, using inventory-control systems most often based on the *first in–first out* principle. They select modes of distribution that will protect products from adverse handling or environmental conditions while in transit to distribution points and to customers. Products that require shipment under refrigerated or frozen conditions also require that validation of shipping conditions be rigorously completed. Furthermore, they must maintain accurate records of distribution to ensure that any product recall, if required, will be effectively and thoroughly conducted.

The marketing department must be sensitive to the customer's needs and be responsive to complaints. The quality department should be kept informed of real or potential problems as reported from the marketplace so that they may conduct investigations of product complaints, as appropriate, to determine the cause of the condition described in the complaint.

Involved with each of the operations described above, QA is ever present and gives approval only after assessing and being assured that the entire production process has been completed satisfactorily and that all the aspects of the GMPs have been satisfied.

In the pharmaceutical industry, QBD or the equivalent system, therefore, can be looked upon as a combined team effort to develop, produce, market, distribute, and control products that are safe and will be effective for the time they remain in the marketplace. Such a program ultimately will assure the professional dispenser and the final consumer that each lot of every product conforms to certain specifications and that each unit has met all the requirements, both internal and external, of the industry and will fulfill the declarations made in its labeling.

DOCUMENTATION

The saying, “If it wasn’t documented, it wasn’t done,” describes the linkage between written records of action taken and the quality operation. These written documents include those found in the product development phase as well as those associated with the actual manufacture and testing of individual batches. The former will consist of research and development reports, technology transfer reports, nonclinical and clinical study reports, and the validation records required when the FDA conducts its pre-approval investigations. Elements of these documents will include raw material and final product specifications along with appropriate validated test methods, technology-transfer documents, and production scale-up support data as well as stability study reports. Specific critical pieces of equipment must be identified along with the process and product qualification/validation records. The *Master Production Batch Record* (MPBR) is often the document that facilitates the orderly transition from product and process development to commercial-scale production, since it is the document that captures the process as described by the product development documentation.

The *Production Batch Record* (PBR), an exact copy of the approved MPBR, is used along with written SOPs to produce individual batches of product that are assigned specific code or lot numbers. The PBR provides a historical record of every step, beginning with the receipt of raw materials and package components and continuing through each phase of production. Recording charts or computer printouts of significant operations such as autoclaving, drying, air-particulate monitoring, lyophilizing, and so on, become part of the batch history. More and more companies are using electronic PBRs to capture the

data collected during the production process. After a batch has been completed, including final analytical and physical testing, there is one final additional step that must be completed prior to approving the batch for distribution. All documentation relating to the production of the specific batch is given a final review. This is often a two-step process in which each of the required documents is checked for accuracy and completeness by manufacturing. Any discrepancy must then be investigated and a written explanation made. This is followed by a final review by QA to ensure that all documents are complete and that all issues have been appropriately resolved. Only after this final review by QA has been completed satisfactorily may the batch be approved for release. Once the batch has been approved, accurate distribution records are required to trace the batch in the marketplace, which would facilitate, if the need arose, recalling the batch.

QUALITY IN PHARMACEUTICAL BIOTECHNOLOGY

Because of the physical/chemical nature of the proteinaceous products derived from pharmaceutical biotechnology, unique quality considerations prevail that are associated with early research and development synthesis, clinical product scale-up, and commercial manufacture. This is particularly challenging for QC departments due to some gross differences between biological therapeutic agents and chemical drug products. In contrast to small-molecule pharmaceuticals, biotechnology-derived drugs are obtained from living organisms and often consist of complex mixtures of protein and other substances, frequently are heat labile, and, finally, are highly susceptible to microbiological contamination. In the context of this discussion, pharmaceutical biotechnology products are defined to include proteins and peptides produced by recombinant DNA (rDNA) techniques and monoclonal anti-body/hybridoma (Mab) technology. The former refers to the fact that these products are often produced by microorganisms or mammalian cells containing hybrid DNA, most often produced by joining pieces of DNA from different sources to gain the appropriate expressed product; the latter involves production of a single clone from hybrid cells, using hybridoma technology that fuses different cells to make the desired antibody. Currently, many Mab products are produced by mammalian cell culture in which the product of interest is expressed in large quantities. These Mab products are often human protein molecules in highly purified and well-characterized form. See also Chapter 34, *Biotechnology and Drugs*. In general, the objective of the pharmaceutical biotechnology manufacturing process is to produce a product essentially free of contamination. The attributes that the product should possess are sterility and the absence of pyrogens, unwanted organisms, by-products of the manufacturing process, and degradation by-products. The same requirements for adherence to cGMPs exist throughout the production process for recombinant proteins and monoclonal antibodies as apply to other pharmaceutical products.

Consideration must also be given to the design of the delivery system for the biotechnology-derived drug, especially because of the lability of many of these products. In accomplishing this task, often the more conventional manufacturing processes may be employed such as sterile filtration, aseptic handling, and, in some instances, lyophilization. However, for many of these products, new delivery systems are being evaluated as described below.

Characterization of the products produced by pharmaceutical biotechnology is a rapidly changing technical challenge, but tremendous advances have been made in the past decade. The FDA considers most of these products (other than some of the vaccines) to be well characterized and has changed the approach to license approval with this improved understanding of product specifications in mind. Guidance documents in the ICH Q5 series and ICH Q10 outline QC or testing considerations that address the structure, potency, and purity of proteins and

the analysis of contaminants resulting from the manufacturing process or degradation. Full characterization includes physical and chemical stability. Satisfactory stability of the product is a requirement for controlled manufacture and an acceptable shelf life following distribution. Because many of these products are temperature labile, the maintenance of cold chain shipping conditions is a challenge that must be addressed during the distribution and handling processes.

CONTEMPORARY ISSUES

As a demonstration of the pace of change in the pharmaceutical industry, technologies that were on the cutting edge as recently as 10 years ago have become rather commonplace, as organizations routinely employ them to increase productivity and reduce costs while maintaining product quality. Statistical process control based on an understanding of process design space permits improved real-time control, thereby reducing end-product failures. Qualification and potential certification of suppliers of goods and services adds to the thrust of building in quality and allows the reduction in inventory costs by following just-in-time purchasing and receipt principles. Finally, in practically all facets of research, development, and operations, automation and computerization, including robotics, are work methods that have an impact on our daily lives by increasing productivity and raising the standards of quality by enhancing reproducibility. This application of computerized manufacturing and control has led to increasingly rigorous requirements for the appropriate design and validation of computer-controlled processes. This is evident in the publishing of 21 CFR Part 11 and Good Automated Manufacturing Practice (GAMP) 4, which describe the requirements for electronic records and signatures and validation respectively (see Appendix).

Today, the pace of change in the pharmaceutical industry continues unabated. There is increased emphasis on analytical chemistry as it relates to the entire drug discovery, development, and manufacturing sequence. This, coupled with new clues as to potential drug targets uncovered through the Human Genome Project, new histochemical methods to identify targets at a cellular level, and new approaches to computer modeling of potential drug compounds has led to many new candidates for development. This is evident in the new drug submissions made by industry to the FDA and in the depth and complexity of the review process associated with these new compounds. More technically sophisticated instrumental methods of analysis, assisted by computer interfacing, provide greater sensitivity along with the ability to analyze the results more efficiently and effectively. From these advances flows the requirement for more stability-indicating assay methods as well as increased emphasis on the impurities in drug substances and products, such as organic volatile impurities and even ordinary impurities derived from raw materials used in production. Taken as a whole, accurate mass balance of the parent compound, degradants, and impurities is an expectation. Compendia such as the USP and ICH guidance are including these concepts in general chapters and individual drug monographs. The evolution of high performance liquid chromatography (HPLC) methods, often coupled with mass spectrometry, and other even more sensitive technologies, and the wide acceptance of these techniques has increased the trend toward painstakingly thorough characterization of products, so that it is becoming the exception rather than the rule. Additionally, HPLC facilitates the recent focus on optical purity through chiral separation to support improvements in asymmetric synthesis, with the intent of producing the single, therapeutically active compound.

The move toward the elimination of tests that require animals is exemplified first by the replacement of the rabbit pyrogen test some 15 years ago with the bacterial endotoxin (LAL) method and the move, particularly in the EU, to using only a single species in nonclinical toxicology testing. It may be expected that similar chemical and biochemical approaches will

be developed to eliminate the use of animal testing in pre-clinical drug evaluation and toxicological testing as well.

The concept of parametric release of end-product is being applied on the basis of complete knowledge and control of all phases of the process including such things as information on suppliers, process and product validation, operator training, and thorough process knowledge coupled with statistical process controls. The sum of these prospectively managed and controlled quality activities results in greater real-time control and, hence, a diminished need for end-product confirmatory testing.

In the operations area, newer types of dosage forms, such as liposomes and transdermal devices, are demanding innovative manufacturing and control procedures and practices. Nanotechnology also shows promise in the development of novel delivery modalities. Novel routes of administration such as pulmonary inhalation for the administration of insulin are also being explored. Automation and computerization continue to increase manufacturing yields and concomitant tighter tolerances. As a result, there is a renewed interest in functional testing of raw materials, to keep up with these manufacturing advances.

All of these advances will, no doubt, be refined with the passage of time and continuing diligent efforts on the part of industrial professionals and regulators. This will lead to ever-new issues affecting the industrial quality professional and, hence, the challenge and reward of such an exciting endeavor.

FDA MODERNIZATION

The FDA Modernization Act (FDAMA) of 1997 was introduced with the intent of improving the review process and thereby making new drug products available to the public faster than in the past. This effort continues both in form and detail and, in general, has shortened the review and approval timeline. A recent decision by the FDA to merge many of the product categories regulated by the Center for Biologics (CBER) into the Center for Drugs (CDER) is intended to further expedite the review process. In addition, there have been a number of new initiatives by the FDA over the last 5 to 10 years to work more collaboratively with the pharmaceutical industry to encourage better definition and understanding of processes and products.

GMP REGULATIONS

In March 1979, the FDA issued revised GMP regulations. These regulations, still in effect today, present the minimum requirements to be met by industry when manufacturing, packaging, and holding human and veterinary drugs. The Federal Food, Drug, and Cosmetic Act states that a drug is deemed to be adulterated unless the methods used in its manufacture, processing, packing, and holding, as well as the facilities in which it was produced and the controls used during its production, conform to the GMPs so that the drug will meet the safety requirements of the Act and that it has the correct identity and strength to meet the quality and purity characteristics that it is represented to possess. Through the intervening years, additional regulations and guidelines have been issued to supplement the original drug GMPs (see References/Bibliography for examples). The ICH, of which FDA is a part, has continued to work to provide global standards for manufacturing and testing of products, and the FDA has already adopted many of these. A number of other guidelines or concept papers have been prepared by various organizations within the industry itself, such as the Pharmaceutical Manufacturers Association (PhRMA), the Parenteral Drug Association (PDA), the International Society of Pharmaceutical Engineering (ISPE), and others. A partial listing is provided in the Bibliography. In addition, as of January 2011, the FDA has become a member of The Pharmaceutical Inspection Co-operation Scheme (PIC/S) whose mission is "to lead the international development, implementation and maintenance of harmonized Good Manufacturing Practice (GMP) standards

and quality systems of inspectorates in the field of medicinal products.”⁴

REFERENCES

1. *Sterile Drug Products Produced by Aseptic Processing*. Rockville, MD: FDA, Sep 2004.
2. *Sterile Manufacturing Facilities*. Baseline Pharmaceutical Engineering Guide, Vol.3. ISPE in partnership with the FDA, June 2000.
3. *Q8 (R2) Pharmaceutical Development*, International Conference on Harmonisation – Quality (ICH). Final guidance. 2009.
4. PIC/S (Pharmaceutical Inspection Co-operation Scheme). Retrieved September 15, 2011, from <http://picscheme.org>.

BIBLIOGRAPHY

- Process Validation: General Principles and Practices*. Rockville, MD: FDA, Jan 2011.
- Submission of Documentation in Applications for Parametric Release of Human and Veterinary Drug Products Terminally Sterilized by Moist Heat Processes*. Rockville, MD: FDA, Feb 2010.
- Container and Closure System Integrity Testing in Lieu of Sterility Testing as a Component of the Stability Protocol for Sterile Products*. Rockville, MD: FDA, Feb 2008.
- Quality Systems Approach to Pharmaceutical CGMP Regulations*. Rockville, MD: FDA, Sep 2006.
- PAT – A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance*. Rockville, MD: FDA, Sep 2004.
- Points to Consider for Aseptic Processing—2003* (Supplement Volume 57, #2). Philadelphia: PDA, 2003.
- FDA Guidance for Industry Part 11, Electronic Records: Electronic Signatures – Scope and Application*. Rockville, MD: FDA, Dec 2003.
- Risk-Based Approach to 21 CFR Part 11—ISPE White Paper*. Tampa, FL: ISPE, 2003.
- General Principles of Software Validation; Final Guidance for Industry and FDA Staff*. Rockville, MD: FDA, Jan 2002.
- Points to Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use*. Rockville, MD: FDA, Feb 1997.

Points to Consider on Plasmid DNA Vaccines for Preventive Infectious Disease Indications. Rockville, MD: FDA, Dec 1996.

Human and Veterinary Drugs—Current Good Manufacturing Practice for Finished Pharmaceuticals. 21 CFR 211:2011. Rockville, MD: USP/NF, USPC.

Good Manufacturing Practice for the Manufacture, Packing, Storage and Installation of Medical Devices. 21 CFR 211:2011. Rockville, MD: USP/NF, USPC.

Quality System Regulation in the Manufacturing of Medical Devices. 21 CFR 820:2011. Rockville, MD: USP/NF, USPC.

Good Laboratory Practice for Non-clinical Laboratory Studies. 21 CFR 58:2011. Rockville, MD: USP/NF, USPC.

WEB SITES

FDA Website at <http://www.fda.gov> has many current guidance documents available for review and/or downloading.

GMP Training Website at <http://gmpttraining.com> has training programs and materials for FDA compliance training.

PIC/S Website at <http://www.picscheme.org> has lists of several cGMP guidance documents.

International Conference on Harmonization Website at <http://www.ich.org> has lists of several quality guidance documents such as Q7A - Good Manufacturing Practice Guidance for Active Pharmaceutical Ingredients and Q8 (R2) – Pharmaceutical Development.

Other Websites providing useful references: ISPE (International Society for Pharmaceutical Engineering) at <http://www.ispe.org>; DIA (Drug Information Association) at <http://www.diahome.org>; RAPS (Regulatory Affairs Professionals Society) at <http://www.raps.org>; and PDA (Parenteral Drug Association) at <http://www.pda.org>.

The current GMP regulations and the additional guides and guidelines listed in the References/Bibliography should be read and understood thoroughly by those involved in or interested in pursuing QC and QA responsibilities. The scope of the present regulation is given in the following outline (Appendix A), along with a brief interpretation of each subpart.

APPENDIX A 21 CFR FOOD AND DRUGS

PART 211 CURRENT GOOD MANUFACTURING PRACTICE FOR FINISHED PHARMACEUTICALS

SUBPART A GENERAL PROVISIONS

- 211.1 The regulations in this part contain the minimum current good manufacturing practice for preparation of drug products for administration to humans or animals.
- 211.3 (Definitions) The scope of the regulations are explained for human prescription and OTC drug products including drugs used to produce medicated animal feed. Reference is made to Part 210.3 of the chapter that gives definitions for all significant terms used in the regulations.

SUBPART B ORGANIZATION AND PERSONNEL

- 211.22 (Responsibilities of Quality (QC) unit) The QC unit has total responsibility for ensuring that adequate systems and procedures exist and are followed to ensure product quality.
- 211.25 (Personnel qualifications) Personnel, either supervisory or operational, must be qualified by training and experience to perform their assigned tasks.
- 211.28 (Personnel responsibilities) The obligations of personnel engaged in the manufacture of drug products concerning their personal hygiene, clothing, and medical status are defined.
- 211.34 (Consultants) The qualifications (education, training, and experience, or any combination thereof) of consultants must be sufficient for the project to which they are assigned.

SUBPART C BUILDINGS AND FACILITIES

Buildings and facilities are considered acceptable only if they are suitable for their intended purpose and can be cleaned and otherwise maintained. Design and construction features for lighting, pest control, and maintenance are described below.

- 211.42 (Design and construction features)
- 211.44 (Lighting)
- 211.46 (Ventilation, air filtration, air heating and cooling)
- 211.48 (Plumbing)
- 211.50 (Sewage and refuse)
- 211.52 (Washing and toilet facilities)
- 211.56 (Sanitation)
- 211.58 (Maintenance)

SUBPART D EQUIPMENT

Equipment must be designed, constructed, of adequate size, suitably located, and able to be maintained and cleaned to be considered suitable for its intended use. Regarding the use of automatic equipment, data processors, and computers, there is a need for input/output verification and for proper calibration of recorders, counters, and other electrical or mechanical devices.

- 211.63 (Equipment design, size, and location)
- 211.65 (Equipment construction)
- 211.67 (Equipment cleaning and maintenance)
- 211.68 (Automatic, mechanical, and electronic equipment)
- 211.72 (Filters) Filters to be used liquid filtration shall not release fibers into products.

SUBPART E CONTROL OF COMPONENTS AND DRUG PRODUCT CONTAINERS AND CLOSURES

211.80 (General requirements) Written procedures must be available that describe the receipt, identification, storage, handling, sampling, testing, and approval or rejection of components (raw materials) and drug products.

- 211.82 (Receipt and storage of untested components, drug product containers, and closures)

- 211.84 (Testing and approval or rejection of components, drug product containers, and closures)
- 211.86 (Use of approved components, drug product containers, and closures) Materials shall be rotated so that the oldest approved stock is used first.
- 211.87 (Retesting of approved components, drug product containers, and closures) Materials shall be retested or reexamined as necessary after storage for long periods or after exposure to air, heat or other conditions that might adversely affect the component, drug product container, closure.
- 211.89 (Rejected components, drug product containers, and closures) These items shall be identified and controlled to prevent their use in manufacturing.
- 211.94 (Drug product containers and closures) Containers and closures (product contact materials) must protect the product and must be nonreactive with or additive to the product, suitable for their intended use, and controlled using written procedures.

SUBPART F PRODUCTION AND PROCESS CONTROLS

- 211.100 (Written procedures and deviations) Written standard operating procedures (SOPs) for each production process and control procedure are necessary. Any deviation from an SOP must be investigated, recorded, and approved prior to final product acceptance.
- 211.101 (Charge-in of components) The procedures used to formulate a batch shall be written and followed.
- 211.103 (Calculation of yield) Actual yields and theoretical yields shall be determined. All products are to be formulated to provide not less than 100% of the required amount of active ingredient. The identity and quantity of each component incorporated into a batch must be recorded.
- 211.105 (Equipment identification) Equipment shall be properly identified.
- 211.110 (Sampling and testing of in-process materials and drug products) Significant in-process steps are to be identified and appropriate sampling, testing, and approvals obtained before proceeding further in the production cycle. Rejected material must be controlled.
- 211.111 (Time limitations on production) If required, time limitations will be placed on in-process steps.
- 211.113 (Control of microbiological contamination) Appropriate procedures are to be prepared for the control and prevention of microbiological contamination. The sterilization process must be validated.
- 211.115 (Reprocessing) Reprocessing of product is allowed providing there are written procedures covering the methods and QC unit review to be used.

SUBPART G PACKAGING AND LABELING CONTROL

- 211.122 (Materials examination and usage criteria) Labeling and packaging materials are to be received, identified, stored, sampled, and tested following detailed written procedures.
- 211.125 (Labeling issuance) Strict control shall be exercised over labeling for use in drug product labeling operations.
- 211.130 (Packaging and labeling operations) There shall be written procedures designed to ensure that correct labels, labeling, and packaging materials are used for drug products. Special controls must be exercised over labeling to ensure that only the correct labels are issued to packaging for a specific product and that the quantities used are reconciled with the quantity issued.

- 211.132 (Tamper-resistant packaging requirements for over-the-counter (OTC) human drug products) Provides details of tamper-resistant packaging.
- 211.134 (Drug product inspection) Packaged and labeled products shall be inspected for correct labels.
- 211.137 (Expiration dating) As supported by appropriate stability studies, products on the market shall bear an expiration date.

SUBPART H HOLDING AND DISTRIBUTION

- 211.142 (Warehousing procedures) Describes the requirements for warehousing holding product under appropriate conditions of light, temperature, and humidity.
- 211.150 (Distribution procedures) Written procedures describing product distribution shall be prepared.

SUBPART I LABORATORY CONTROLS

- 211.160 (General requirements) The general requirements for laboratory control mechanisms are described.
- 211.165 (Testing and release for distribution) There shall be written procedures in the form of specifications, standards, sampling plans, and test procedures that are used in a laboratory for controlling components and finished drug products. Acceptance criteria for sampling and approval shall be adequate to support release of product for distribution.
- 211.166 (Stability testing) There shall be a written testing program designed to assess the stability characteristics of drug products. The results of this testing shall be used in assigning appropriate storage conditions and expiration dates.
- 211.167 (Special testing requirements) There are special testing requirements for sterile and/or pyrogen-free ophthalmic ointment and controlled-release dosage form products.
- 211.170 (Reserve samples) Reserve sample quantity and retention times are described.
- 211.173 (Laboratory animals) Animals used in any testing shall be maintained and controlled in a manner suitable for use.
- 211.176 (Penicillin contamination) Drug products cannot be marketed if, when tested by a prescribed procedure, found to contain any detectable levels of penicillin.

SUBPART J RECORDS AND REPORTS

- 211.180 (General requirements) Describes record retention time and availability for inspection.

- 211.182 (Equipment cleaning and use log) A written record of major equipment cleaning, maintenance, and use shall be included in major equipment logs.
- 211.184 (Component, drug product container, closure, and labeling records) Deals with the issues of the receipt, testing, and storage of components, drug product containers, and closures. Details the various records and documents that should be generated during the manufacture of drug products and that are to be available for review.
- 211.186 (Master production and control records) A master production record must be prepared for each drug product, describing all aspects of its manufacture, packaging, and control. Individual batch records are derived from this approved master.
- 211.188 (Batch production and control records) Batch production and control records with information about the production and control of each batch are prepared.
- 211.192 (Production record review) All drug product batch records shall be reviewed and approved by the QC unit (QA/QC) before the batch is released.
- 211.194 (Laboratory records) Complete records of any laboratory testing shall be maintained to include raw data, test procedures and results, initials or signatures of personnel performing the test or reviewing the results of tests, periodic equipment calibration, and stability test results.
- 211.196 (Distribution records) Distribution records include warehouse shipping logs, invoices, bills of lading, and all documents associated with distribution. These records should provide all the information necessary to trace lot distribution to facilitate product retrieval if necessary.
- 211.198 (Complaint files) Written records of complaints received from consumers and professionals are to be maintained along with the report of investigations and responses.

SUBPART K RETURNED AND SALVAGED DRUG PRODUCTS

- 211.204 (Returned drug products) Records are to be maintained of drug products returned from distribution channels and the reason for their return. These data can be used as part of the total lot accountability, should the need arise, to trace its distribution and/or for its recall.
- 211.208 (Drug product salvaging) Drug products that have been stored improperly are not to be salvaged.

Chapter 4

Stability of Pharmaceutical Products

Allan D. Bokser, PhD and Patrick B. O'Donnell, PhD

INTRODUCTION 37

REGULATORY REQUIREMENTS 38

PRODUCT STABILITY 39

PREDICTING SHELF LIFE 45

APPROXIMATIONS IN ASSESSING
PRODUCT STABILITY 45

PHARMACEUTICAL CONTAINERS 47

CLOSURES 48

INTRODUCTION

Stability of a pharmaceutical product may be defined as the capability of a particular formulation, in a specific container/closure system, to remain within its physical, chemical, microbiological, therapeutic, and toxicological specifications at a defined storage condition. Pharmaceutical products are expected to meet their specifications for identity, purity, quality, and strength throughout their defined storage period at specific storage conditions. Assurances that the packaged product will be stable for its anticipated shelf life must come from an accumulation of valid data on the drug in its commercial package. These stability data include selected parameters that, taken together, form the stability profile.

The stability of a pharmaceutical product is investigated throughout the various stages of the development process. The stability of a drug substance is first assessed in the preformulation stage. At this stage, pharmaceutical scientists determine the drug substance and its related salts stability/compatibility with various solvents, buffered solutions and excipients considered for formulation development. Suitable analytical methods must be employed in order to ensure the likelihood that this assessment will be successful. Optimization of a stable formulation of a pharmaceutical product is built (using statistical design) upon the information obtained from the preformulation stage and continues during the formulation development stages.

Typically, the first formulation development stage may be for preclinical studies or as late as the preparation of a “first in human” formulation which is often a non-elegant formulation optimized for short-term dose-ranging clinical studies. The second major formulation development stage occurs to support Phase II clinical studies (proof of concept phase). The pharmaceutical product developed at this stage is usually the prototype for the commercial product. Therefore, the pharmaceutical product will be formulated based in part on the stability information obtained from the previous formulations and must meet stability requirements for longer-term clinical studies. In the final formulation development state for Phase III clinical studies, the formulation must be truly representative of what the commercial pharmaceutical product will be in order to avoid delays in approval. In addition to building on the clinical requirements of the drug, the commercial pharmaceutical product must also incorporate the commercial or the final market image of the product, which includes the container closure system. The stability of this product must be demonstrated to the appropriate regulatory agencies in order to assign an expiration period and date for the product. This expiration period allows for the assignment of an expiration date based on the manufacture date of each lot of drug product.

Once a pharmaceutical product has gained regulatory approval and is marketed, the pharmacist must understand the proper storage and handling of the drug. In some cases, a pharmacist may need to prepare stable compounded preparations from this product.

Most drug products are not shipped directly from the manufacturer to a pharmacy. Typically, a drug product is shipped from a manufacturer to a distribution center. From the distribution center the drug product is then shipped to a wholesaler. From the wholesaler, the drug product may be shipped to the distribution center for a pharmacy chain or directly to the pharmacy. Finally, the drug product is dispensed by the pharmacy to the patient. Dispensing of the drug product may be at a hospital, a clinic, and a traditional “brick and mortar” pharmacy or from a mail-order pharmacy. Therefore, the stability typically must also assess the robustness of the drug product through its supply chain. It is not unusual for temperature excursions to occur during these transfers of control.

Inventory control, or holding, of each drug is important for a wholesaler or pharmacy. A drug must be within its expiration dating throughout its use by the patient. Solid oral dosages may be dispensed in the commercial packaging or in a pharmacy supplied container closure system. Most prescriptions are supplied to patients for up to 30 or 90 days by traditional and mail-order pharmacies, respectively. Inventory control of product by wholesalers and pharmacies must assess how much dating must remain on a product for it to be useful for its customer. This causes the actual holding of a product to be shorter than the expiration date. Under normal circumstances it is unusual for a pharmacy to accept any product with less than 6 month dating remaining on a product.

Much has been written about the development of a stable pharmaceutical product. Comprehensive treatments of all aspects of pharmaceutical product stability have been published by Connors *et al.*¹, Carstensen² and more recently by Allen.³ This chapter will outline the appropriate steps from preformulation to drug approval to assure that the pharmaceutical product developed is stable. Requirements for compounded products will also be discussed.

The United States Pharmacopeia (USP) General Chapter <1191>⁴ defines the stability of a pharmaceutical product as “extent to which a product retains, within specified limits, and throughout its period of storage and use (i.e., its shelf life), the same properties and characteristics that it possessed at the time of its manufacture.” There are five types of stability that must be considered for each drug (Table 4-1).

The use of kinetic and predictive studies for establishing credible expiration dating for pharmaceutical products is now accepted worldwide. Scientifically designed studies using reliable, meaningful, and specific stability-indicating assays, appropriate statistical concepts, and a computer to analyze the resulting data are used to determine an accurate and realistic shelf life. In this way the maximum amount of valid information is obtained to establish a reliable, defensible expiration date for each formulation. The assigned expiration date is a direct application and interpretation of the knowledge gained from the stability study.

Although there are exceptions, 90% of labeled potency generally is recognized as the minimum acceptable potency

Table 4-1. Types of Stability

Type of Stability	Conditions Maintained Throughout the Shelf Life of the Drug Product
Chemical	Each active ingredient retains its chemical integrity and labeled potency, within the specified limits.
Physical	The original physical properties, including appearance, palatability, uniformity, dissolution, and suspendability are retained.
Microbiological	Sterility or resistance to microbial growth is retained according to the specified requirements. Antimicrobial agents that are present retain effectiveness within the specified limits.
Therapeutic	The therapeutic effect remains unchanged.
Toxicological	No significant increase in toxicity occurs.

level over the shelf life of a drug product. Exceptions to this minimum potency include drugs with active pharmaceutical ingredients that have a narrow therapeutic thresholds and biologics.

The stability of a commercial pharmaceutical product is expressed as an expiration date (expiry). Expiration dating is defined, therefore, as the time in which a drug product in a specific packaging configuration will remain stable when stored under recommended conditions. This date is usually calculated by adding the established expiration period to the date of manufacture. The date of manufacture is many times defined as the date in which the active pharmaceutical ingredient is first combined with a drug product excipient.

An expiration date, which is expressed traditionally in terms of month and year, denotes the last day of the month. In the United States, the expiration date shall appear on the immediate container and the outer retail package. However, when single-dose containers are packaged in individual cartons, the expiration date may be placed on the individual carton instead of the immediate product container. If a dry product is to be reconstituted at the time of dispensing, expiration dates are assigned to both the dry mixture and the reconstituted product. Tamper-resistant packaging is to be used where applicable.

REGULATORY REQUIREMENTS

Stability study requirements and expiration dating are covered in the Current Good Manufacturing Practices (cGMPs),⁵ the USP,⁶ and the FDA and ICH guidelines.

GOOD MANUFACTURING PRACTICES

The GMPs⁵ as stated in US 21 CFR part 211.166 requires that there shall be a written testing program designed to assess the stability characteristics of drug products. The results of such stability testing shall be used to determine appropriate storage conditions and expiration dating. The latter is to ensure that the pharmaceutical product meets applicable standards of identity, strength, quality, and purity at time of use. Regulatory agencies of other regions have similar requirements. For example Canadian regulations require that every distributor and importer shall monitor, by means of a continuing program, the stability of the drug in the package in which it is sold.⁷ The European Union GMPs states “after marketing, the stability of the medicinal product should be monitored according to a continuous appropriate program that will permit the detection of any stability issue (e.g., changes in levels of impurities or dissolution profile) associated with the formulation in the marketed package.”⁸ These regulations, which apply to both human and

veterinary drugs, are updated periodically in light of current knowledge and technology.

COMPENDIA

The USP also contains extensive stability and expiration dating information. Included are a discussion of stability considerations in dispensing practices and the responsibilities of both the pharmaceutical manufacturer and the dispensing pharmacist. It now is required that product labeling of official articles provide recommended storage conditions and an expiration date assigned to the specific formulation and package. Official storage conditions as defined by the USP 34⁹ are as follows:

- *Freezer* – a place where temperature is maintained thermostatically between -25°C and -10°C .
- *Cold* – any temperature not exceeding 8°C and *refrigerator* is a cold place where the temperature is maintained thermostatically between 2 and 8°C .
- *Controlled Cold Temperature* – temperature maintained thermostatically between 2° and 8°C .
- *Cool* is defined as any temperature between 8 and 15°C .
- *Room Temperature* – the temperature prevailing in a working area.
- *Controlled Room Temperature* – temperature maintained thermostatically between 20 and 25°C .
- *Warm* – any temperature between 30 and 40°C .
- *Excessive Heat* – any heat above 40°C .
- *Protect from Freezing* – where, in addition to the risk of breakage of the container, freezing subjects an article to loss of a product to a loss of strength, potency or to destructive alteration its characteristics, the container label should bear appropriate instructions to protect the product from freezing.
- *Dry Place* – denotes a place that does not exceed 40% average relative humidity (RH) at Controlled Room Temperature or the equivalent water vapor pressure at other temperatures.

Most drug products are stored at controlled room temperature and are labeled as such. As is noted above in USP 34, the definition of controlled room temperature was a “temperature maintained thermostatically between 20 and 25°C (68 and 77°F).” This definition was established to harmonize with international drug standards efforts. The usual or customary temperature range is identified as 20 to 25°C , with the possibility of encountering excursions in the 15 to 30°C range and with the introduction the mean kinetic temperature (MKT).

The mean kinetic temperature is calculated using the following equation:

$$T_k = \left[\frac{\Delta H / R}{- \ln \left(\frac{e^{-\Delta H / RT_1} + e^{-\Delta H / RT_2} + \dots + e^{-\Delta H / RT_{n-1}} + e^{-\Delta H / RT_n}}{n} \right)} \right] \quad (1)$$

in which T_k is the mean kinetic temperature; ΔH is the heat of activation, $83.144 \text{ kJ} \cdot \text{mole}^{-1}$; R is the universal gas constant, $8.3144 \times 10^{-3} \text{ kJ} \cdot \text{mole}^{-1} \cdot \text{degree}^{-1}$; T_1 is the value for the temperature (in degrees Kelvin ($^{\circ}\text{K}$)) recorded during the first time period, T_2 is the value for the temperature recorded during the second time period, e.g., second week; T_{n-1} is the value of the second to last time period, and T_n is the value for the temperature recorded during the n th time period. Typically, the time period is in days or weeks.

Mean kinetic temperature determine the thermal expose of a material. This allows an acceptable estimation to assess if a temperature excursion (or series of excursions) adversely affected a material.

FDA and ICH Guidelines provide recommendations for

1. the design of stability studies to establish appropriate expiry and product storage requirements

- the submission of stability information for investigational new drugs, biologics, new drug applications, and biological product license applications.

Thus, the guidelines represent a framework for the experimental design and data analysis as well as the type of documentation needed to meet regulatory requirements in the drug-development process.

The International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) is an ongoing project that brought together regulatory authorities and experts from the pharmaceutical industry from three regions of the world; Europe, Japan, and the US since its inception in 1990. The first conference (ICH1) took place in November 1991 in Brussels, and the second conference (ICH2) in Orlando, FL, in October 1993. These conferences continue to provide an open forum for discussion and resulted in the creation of an extensive set of guidelines dealing with the many aspects of safety, quality, and efficacy of medicinal products. The following ICH Harmonized Tripartite Guidelines related to stability are summarized below.

- Q1A (R2) provides the general requirements for *Stability Testing of New Drug Substances and Products*. The main thrust of the stability guideline centers on criteria for setting up stability protocols, shown in Table 4-2. ICH Q1A is considered the parent guideline for stability.¹⁰
- Q1B provides the general requirements for assessing the intrinsic photostability characteristics of new drug substances and drug products.¹¹
- Q1C provides the reduced general requirements for stability testing of new drug products dosage forms that use active ingredients already approved in another dosage form.¹²
- Q1D provides approaches that allows for bracketing and matrixing designs for stability testing. This is especially important when a new drug product is to be commercialized with multiple strengths in multiple packaging configurations.¹³
- Q1E provides acceptable approaches for evaluating stability data to assess the shelf life of the product in a registration application.¹⁴
- Q1F was withdrawn. It provided definitions for testing conditions to register drugs marketed in Climatic Zones III and IV.¹⁵ However, countries in Climatic Zone IV believed that a larger safety margin was required than presented in this guidance.

It should be noted that all ICH guidelines provide approaches that are now common in the industry for assessing stability of a drug substance or drug product. However, these guidelines also make it clear that alternative approaches can be used when there are scientifically justifiable reasons.

Table 4-2a. Stability Protocols

Conditions	Minimum Time Period at Submission
Long-term testing 25°C ± 2°C/60% RH ± 5%	12 month
Accelerated testing 40°C ± 2°C/75% RH ± 6%	6 month
Alternate testinga 30°C ± 2°C/65% RH ± 5%	12 month

Required if *significant change* occurs during 6-mo storage under conditions of accelerated testing.

Table 4-2b. Example Stability Pull Schedule for a Solid Oral Dose for Zone I and II

Storage Conditions	Durations (Months)								
	0	1	3	6	9	12	18	24	36
25 °C/60 % RH	R*		X	X	X	X, Y	X	X	X
30 °C/65 % RH			O	O	O	O			
40 °C/75% RH		X	X	X, Y					

*From Release testing if testing is within 30 days of stability set down
 R= Release Tests; Appearance (visual); Identity; Assay (HPLC); Impurities (HPLC); Dissolution (USP <711>); Moisture Content (Karl Fischer); Uniformity of Dosage Unit
 X = Tests at Every Stability Pull; Appearance (visual); Assay (HPLC); Impurities (HPLC); Dissolution USP <711>
 O = Pull and test only after 40 °C/75% is out of specification; Appearance (visual); Assay (HPLC); Impurities (HPLC); Dissolution Y = Additional tests periodically performed; Moisture Content (Karl Fischer)

The FDA and ICH guidelines are readily available on the internet. The pharmaceutical scientist can now access a complete listing of both FDA and ICH publications. To view the publications, go to: <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm> for the FDA guidance webpage, and <http://www.ich.org/products/guidelines/quality/article/quality-guidelines.html> for the ICH guideline webpage.

PRODUCT STABILITY

Many factors affect the stability of a pharmaceutical product including the intrinsic stability of the active ingredient(s), the potential interaction between active and inactive ingredients, the manufacturing process, the dosage form, the container-liner-closure system and the environmental conditions encountered during shipment, storage and handling and length of time between manufacture and usage.

Classically, pharmaceutical product stability evaluations have been separated into studies of chemical (including biochemical) and physical stability of formulations. Realistically, there is no absolute division between these two arbitrary divisions. Physical factors—such as heat, light, and moisture—may initiate or accelerate chemical reactions, whereas every time a measurement is made on a chemical compound, physical dimensions are included in the study.

One type of time-related chemical stability failure is a decrease in therapeutic activity of the preparation to below some arbitrary labeled content. A second type of chemical stability failure is the appearance of a toxic substance, formed as a degradation product upon storage of the formulation. The numbers of published cases reflecting this second type are few. However, it is possible, though remote, for both types of stability failures to occur simultaneously within the same pharmaceutical product. Thus, the use of stability studies with the resulting application of expiration dating to pharmaceuticals is an attempt to predict the approximate time at which the probability of occurrence of a stability failure may reach an intolerable level. This estimate is subject to the usual Type 1 or alpha error (setting the expiration too early so that the product will be destroyed or removed from the market appreciably earlier than actually is necessary) and the Type 2 or beta error (setting the date too late so that the failure occurs in an unacceptably large proportion of cases). Thus, it is obligatory that the manufacturer clearly and succinctly defines the method for determining the

degree of change in a formulation and the statistical approach to be used in making the shelf life prediction. An intrinsic part of the statistical methodology must be the statements of value for the two types of error. For the safety of the patient a Type 1 error can be accepted, but not a Type 2 error.

One type of time related physical stability failures may affect the availability or rate of drug release of a product. This type of physical stability failure may cause the active ingredient not to be released or a higher rate of drug release (dose dumping). Another type of time related physical stability failures are appearance related. These may just cause the drug product not to appear pharmaceutically elegant or may be an artifact of another physical or chemical stability failure.

In this treatment, physical and chemical stability are discussed along with those dosage form properties that can be measured and are useful in predicting shelf life. The effect of various physical and chemical phenomena of pharmaceuticals also is treated.

Knowledge of the physical stability of a formulation is very important for three primary reasons. First, a pharmaceutical product must appear fresh, elegant, and professional, for as long as it remains on the shelf. Any changes in physical appearance such as color fading or haziness can cause the patient or consumer to lose confidence in the product. Second, since some products are dispensed in multiple-dose containers, potency of the active ingredient over time must be ensured for each individual dose. A cloudy solution or a broken emulsion can lead to a non-uniform dosage pattern. Third, the active ingredient must be bioavailable to the patient throughout the expected shelf life of the preparation. A breakdown in the physical system can lead to non-availability or "dose dumping" of the medication to the patient. In the case of metered-dose inhaler pulmonary aerosols, particle aggregation may result in inadequate lung deposition of the medication.

The chemical causes of drug deterioration have been classified as incompatibility, oxidation, reduction, hydrolysis, racemization, and other mechanisms. In the latter category, decarboxylation, deterioration of hydrogen peroxide and hypochlorites and the formation of precipitates have been included.

PHARMACEUTICAL DOSAGE FORMS

As the various pharmaceutical dosage forms present unique stability problems, they are discussed separately in the following section.

Tablets

Stable tablets retain their original size, shape, weight, and color under normal handling and storage conditions throughout their shelf life. In addition, the *in vitro* availability of the active ingredients should not change appreciably with time.

Excessive powder or solid particles at the bottom of the container, cracks or chips on the face of a tablet, or appearance of crystals on the surface of tablets or on container walls are indications of physical instability of uncoated tablets. Hence, the effect of mild, uniform, and reproducible shaking and tumbling of tablets should be studied. The recommended test for such studies is the determination of tablet friability as described in the USP. Tablet Friability USP <1216> describes the recommended apparatus and the test procedure. After visual observation of the tablets for chips, cracks, and splits, the intact tablets are sorted and weighed to determine the amount of material worn away by abrasion. In general, a maximum weight loss of not more than 1% of the weight of the tablets being tested is considered acceptable for most products. The results of these tests are comparative rather than absolute and should be correlated with actual stress experience. Packaged tablets also should be subjected to cross-country shipping tests as well as to various *drop tests*.

Tablet hardness (or resistance to crushing or fracturing) can be assessed by commercially available hardness testers. As

results will vary with the specific make of the test apparatus used, direct comparison of results obtained on different instruments may not necessarily be made. Thus, the same instrument should be used consistently throughout a particular study.

Color stability of tablets can be followed by an appropriate colorimeter or reflectometer with heat, sunlight, and intense artificial light employed to accelerate the color deterioration. It is still not unusual for color assessment to be performed visually. Caution must be used in interpreting the elevated temperature data, as the mechanism for degradation at that temperature may differ from that at a lower temperature. It is not always proper to assume that the same changes will occur at elevated temperatures as will be evidenced later at room temperature. Cracks, mottling, or tackiness of the coating indicates evidence of instability of coated tablets.

Typically, dissolution is the *in vitro* test performed to estimate bioavailability for a tablet regardless of the solubility of the active ingredients. Disintegration has been relegated to an in-process test or used to help dissolution. Dissolution tests should be run in an appropriate medium at 37°C. Actual dissolution conditions, including medium, are developed during the clinical development phase of a product. The dissolution method developed has to demonstrate a correlation that is relevant to the bioavailability of the dosage form. Dissolution profiles are examined during development to provide sufficient information to define a single sample time point with a minimum concentration for immediate release product. Controlled release drug products require a dissolution profile with concentration ranges at set sampling points for product assessment. When no significant change (such as a change in the polymorphic form of the crystal) has occurred, an unaltered dissolution-rate profile of a tablet formulation usually indicates constant *in vivo* bioavailability.

Uniformity of weight, odor, texture, drug and moisture contents, and humidity effect may also be studied during a tablet stability test.

Gelatin Capsules

Hard gelatin capsules are the type used by pharmaceutical manufacturers in the production of the majority of their capsule products. The pharmacist in the extemporaneous compounding of prescriptions may also use hard gelatin capsules. Soft gelatin capsules are prepared from shells of gelatin to which glycerin or a polyhydric alcohol such as sorbitol has been added to render the gelatin elastic or plastic-like. Gelatin is stable in air when dry but is subject to microbial decomposition when it becomes moist or when it is maintained in aqueous solution. Normally hard gelatin capsules contain between 13 and 16% moisture. If stored in a high humidity environment capsule shells may soften, stick together or become distorted and lose their shape. On the other hand, in an environment of extreme dryness, gelatin capsules may harden and crack under slight pressure. Gelatin capsules should be protected from sources of microbial contamination. Encapsulated products, like all other dosage forms, must be packaged properly.

Because moisture may be absorbed or released by gelatin capsules depending on the environmental conditions, capsules offer little physical protection to hygroscopic or deliquescent materials enclosed within a capsule when stored in an area of high humidity. It is not uncommon to find capsules packaged in containers along with a packet of desiccant material as a precautionary measure.

Dissolution development and requirements for capsules are similar to tablets. The capsule shell can affect dissolution test results but not be relevant to bioavailability. Both hard and soft gelatin capsules exposed to excessive heat and moisture may exhibit delayed or incomplete dissolution due to cross-linking of the gelatin in the capsule shell. The cross-linking of gelatin capsules is an irreversible chemical reaction. Cross-linking may also occur in capsules that are exposed to aldehydes and peroxides. Although cross-linked capsules may fail dissolution due to

pellicle formation, digestive enzymes will dissolve the capsules. For hard or soft gelatin capsules that do not conform to the dissolution specification, the dissolution test may be repeated with the addition of enzymes. Where water or a medium with a pH less than 6.8 is specified as the medium in the individual monograph, the same medium specified may be used with the addition of purified pepsin that results in an activity of 750,000 units or less per 1000 mL. For media with a pH of 6.8 or greater, pancreatin can be added to produce not more than 1750 USP units of protease activity per 1000 mL.

Suspensions

A stable suspension can be redispersed homogeneously with moderate shaking and can be poured easily throughout its shelf life, with neither the particle size distribution, the crystal form, nor the physiological availability of the suspended active ingredient changing appreciably with time.

Most stable pharmaceutical suspensions are flocculated; that is, the suspended particles are bonded together physically to form a loose, semi rigid structure. The particles are said to uphold each other while exerting no significant force on the liquid. Sedimented particles of a flocculated suspension can be redispersed easily at any time with only moderate shaking.

In non-flocculated suspensions, the particles remain as individuals unaffected by neighboring particles and are affected only by the suspension vehicle. These particles, which are smaller and lighter, settle slowly. Once they have settled, they often form a hard, difficult-to-disperse sediment. Non-flocculated suspensions can be made acceptable by decreasing the particle size of the suspended material or by increasing the density and viscosity of the vehicle, thus reducing the possibility of settling.

When studying the stability of a suspension, a differential manometer is used to determine if the suspension is flocculated. If the suspension is flocculated, the liquid will travel the same distance in the two side arms. With non-flocculated suspensions, the hydrostatic pressures in the two arms are unequal; hence, the liquids will be at different levels.

The history of settling of the particles of a suspension may be followed by a Brookfield viscometer fitted with a Helipath attachment. This instrument consists of a rotating T-bar spindle that descends slowly into the suspension as it rotates. The dial reading on the viscometer is a measure of the resistance that the spindle encounters at various levels of the sedimented suspension. This test must be run only on fresh, undisturbed samples.

An electronic particle counter and sizer, such as a Coulter counter, or a microscope may be used to determine changes in particle size distribution. Crystal form alterations may be detected by microscopic, near-IR or Raman examination and, when suspected, must be confirmed by X-ray powder diffraction.

All suspensions should be subjected to cycling temperature conditions to determine the tendency for crystal growth to occur within the suspension. Shipping tests, namely transporting bottles across the country by rail or truck are also used to study the stability of suspensions.

Solutions

A stable solution retains its original clarity, color, and odor throughout its shelf life. Retention of clarity of a solution is a main concern of a physical stability program. As visual observation alone under ordinary light is a poor test of clarity, a microscope light should be projected through a diaphragm into the solution. Undissolved particles will scatter the light, and the solution will appear hazy. Although the Coulter counter also can be used, light-scattering instruments are the most sensitive means of following solution clarity.

Solutions should remain clear over a relatively wide temperature range such as 4 to 47°C. At the lower range an ingredient may precipitate due to its lower solubility at that temperature, whereas at the higher temperature the flaking of particles from

the glass containers or rubber closures may destroy homogeneity. Thus, solutions should be subjected to cycling temperature conditions.

The stability program for solutions also should include a study of pH changes, especially when the active ingredients are soluble salts of insoluble acids or bases. Among other tests are observations for changes in odor, appearance, color, taste, light-stability, pourability, viscosity, isotonicity, gas evolution, microbial stability, specific gravity, surface tension, and pyrogen content, in the case of parenteral products.

When solutions are filtered, the filter medium may absorb some of the ingredients from the solution. Thus, the same type of filter should be used for preparing the stability samples as will be used to prepare the production-size batches.

For dry-packaged formulations reconstituted prior to use, the visual appearance should be observed on both the original dry material and on the reconstituted preparation. The color and odor of the cake, the color and odor of the solution, the moisture content of the cake, and the rate of reconstitution should be followed as a part of its stability profile.

Emulsions

A stable emulsion can be redispersed homogeneously to its original state with moderate shaking and can be poured at any stage of its shelf life. Although most of the important pharmaceutical emulsions are of the oil in water (O/W) type, many stability test methods can be applied to either an O/W or water in oil (W/O) emulsion.

Two simple tests are used to screen emulsion formulations. First, heating to 50 to 70°C and observing its gross physical stability either visually or by turbidimetric measurements can determine the stability of an emulsion. Usually the emulsion that is the most stable to heat is the one most stable at room temperature. However, this may not be true always, because an emulsion at 60°C may not be the same as it is at room temperature. Second, the stability of the emulsion can be estimated by the *coalescence time* test. Although this is only a rough quantitative test, it is useful for detecting gross differences in emulsion stability at room temperature.

Emulsions also should be subjected to refrigeration temperatures. An emulsion stable at room temperature has been found to be unstable at 4°C. It was reasoned that an oil-soluble emulsifier precipitated at the lower temperature and disrupted the system. An emulsion chilled to the extent that the aqueous base crystallizes is damaged irreversibly.

The ultracentrifuge also is used to determine emulsion stability. When the amount of separated oil is plotted against the time of centrifugation, a plateau curve is obtained. A linear graph results when the oil flotation (creaming) rate is plotted versus the square of the number of centrifuge revolutions per minute. The flotation rate is represented by the slope of the line resulting when the log distance of emulsion-water boundary from the rotor center is plotted against time for each revolution per minute.

For stability studies, two batches of an emulsion should be made at one time on two different sizes of equipment. One should be a bench-size lot and the other a larger, preferably production-size, batch. Different types of homogenizers produce different results, and different sizes of the same kind of homogenizer can yield emulsions with different characteristics.

Ointments

Ointments have been defined as high-viscosity suspensions of active ingredients in a non-reacting vehicle. A stable ointment is one that retains its homogeneity throughout its shelf life period. The main stability problems observed in ointments are *bleeding* and changes in consistency due to aging or changes in temperature. When fluid components such as mineral oil separate at the top of an ointment, the phenomenon is known as *bleeding* and can be observed visually. Unfortunately, as there is no known way to accelerate this event, the tendency to *bleed* cannot be predicted.

An ointment that is too soft is messy to use, whereas one that is very stiff is difficult to extrude and apply. Hence, it is important to be able to define quantitatively the consistency of an ointment. This may be done with a penetrometer, an apparatus that allows a pointed weight to penetrate into the sample under a measurable force. The depth of the penetration is a measure of the consistency of an ointment. Consistency also can be measured by the Helipath attachment to a high-viscosity viscometer or by a Burrell Severs rheometer. In the latter instrument the ointment is loaded into a cylinder and extruded with a measured force. The amount extruded is a measure of the consistency of the ointment.

Ointments have a considerable degree of structure that requires a minimum of 48 hours to develop after preparation. As rheological data on a freshly made ointment may be erroneous, such tests should be performed only after the ointment has achieved equilibrium. Slight changes in temperature (1 or 2°C) can affect the consistency of an ointment greatly; hence, rheological studies on ointments must be performed only at constant and controlled temperatures.

Among the other tests performed during the stability study of an ointment are a check of visual appearance, color, odor, viscosity, softening range, consistency, homogeneity, particle size distribution, and sterility. Undissolved components of an ointment may change in crystal form or in size with time. Microscopic examination or an X-ray diffraction measurement may be used to monitor these parameters.

In some instances it is necessary to use an ointment base that is less than ideal, to achieve the required stability. For example, drugs that hydrolyze rapidly are more stable in a hydrocarbon base than in a base containing water, even though they may be more effective in the latter.

Transdermal Patches

A typical transdermal patch consists of a protective backing, a matrix containing active drug, an adhesive that allows the patch to adhere to the skin and a release liner to protect the skin adhering adhesive. Therefore, the transdermal patch must deliver drug as labeled, adhere properly to both the backing and to the patient's skin. In addition, the transdermal patch must be pharmaceutically elegant through the shelf life of the product. For a transdermal patch, this means that the release line peels easily with minimal transfer of adhesive onto the release liner and that the adhesive does not ooze from the sides of the patch. Therefore, the typical stability related tests for transdermal patches are appearance, assay, impurities, drug release per USP<724>, and backing peel force.

Metered-Dose Aerosols Drug Products

A metered dose inhalation product comprises an aerosol can containing a propellant and drug, and a mouthpiece used to present an aerosolized drug to the patient. There are many drug contact components in a metered-dose inhalation product. Therefore, the drug may be in contact with materials that could allow plasticizer leach into the propellant. The typical stability related tests for metered-dose aerosols include appearance, assay, impurities, plume geometry, emitted dose, particle size distribution of the emitted dose, and number of doses per unit. In addition, stability studies on leachables may be required. Shelf life of metered-dose aerosols drug products may also be dependent on the orientation that the drug product is stored. Typically most canister type products are tested at least in the upright orientation.

Dry-Powdered Inhalation Products

A dry-powdered inhalation product consists of drug with excipients delivered in a dry powdered form. The delivery system for a dry-powdered inhalation product may be a separate device or integrated with the active. A dry-powdered dosage must reproducibly deliver a specific amount of drug at a particle size that can be deposited into the lungs. Particles too large will get

trapped in the throats and particles too small will just be carried out of the lungs on the next expiration. The typical stability related tests for dry powder inhalation products include appearance, assay, impurities, emitted dose, particle size distribution of the emitted dose, and water content.

Nasal Inhalation Products

A nasal inhalation product consists of drug with excipients delivered from a delivery system. The delivery system for a nasal inhalation product may be a separate device or integrated with the active. A nasal inhalation product must reproducibly deliver a specific amount of drug at a particle size and plume that can be deposited into the nasal membrane. Particles too large will not be absorbed into nasal membrane or run out of the nose; and poor spray pattern will deposit the drug ineffective in the nasal cavity. The typical stability related tests for nasal inhalation products include appearance, assay, impurities, spray content uniformity, particle (droplet) size distribution of the emitted dose, spray pattern or /and plume geometry, leachables, weight loss and preservative content. Sterility and microbial testing may be required periodically for stability testing.

INCOMPATIBILITY

Typically, physicochemical stability is assessed at the preformulation stage of development. A drug substance candidate is treated with acid, base, heat, light, and oxidative conditions to assess its inherent chemical stability. Binary mixtures of the drug substance with individual excipients are also investigated at the preformulation stage. These tests are performed to determine the drug substance sensitivity to degrade or reactivity with common pharmaceutical excipients. The most common reactions observed for drug substances from these tests include: hydrolysis, epimerization (racemization), decarboxylation, dehydration, oxidation, polymerization, photochemical decomposition and addition. All drug substances have the potential to degrade by at least one of the reactions mentioned above. With an understanding of the stability/reactivity of a drug substance in the preformulation stage, it is possible to formulate the drug product to minimize drug decomposition. Numerous examples are described in other sections of this book, and the literature is replete with illustrations.

Although undesirable reactions between two or more drugs are said to result in a *physical, chemical, or therapeutic* incompatibility, physical incompatibility is somewhat of a misnomer. It has been defined as a physical or chemical interaction between two or more ingredients that leads to a visibly recognizable change. The latter may be in the form of a gross precipitate, haze, or color change.

On the other hand, a chemical incompatibility is classified as a reaction in which a visible change is not necessarily observed. Since there is no visible evidence of deterioration, this type of incompatibility requires trained, knowledgeable personnel to recognize it.

A therapeutic incompatibility has been defined as an undesirable pharmacological interaction between two or more ingredients that leads to:

1. Potentiation of the therapeutic effects of the ingredients.
2. Destruction of the effectiveness of one or more of the ingredients.
3. Occurrence of a toxic manifestation within the patient.

CHEMICAL REACTIONS

The most frequently encountered chemical reactions, which may occur within a pharmaceutical product, are described below.

Oxidation-Reduction

Oxidation is a prime cause of product instability, and often, but not always, the addition of oxygen or the removal of hydrogen is involved. When molecular oxygen is involved, the reaction

is known as auto-oxidation because it occurs spontaneously, though slowly, at room temperature.

Oxidation, or the loss of electrons from an atom, frequently involves free radicals and subsequent chain reactions. Only a very small amount of oxygen is required to initiate a chain reaction. In practice, it is easy to remove most of the oxygen from a container, but very difficult to remove it all. Hence, nitrogen and carbon dioxide frequently are used to displace the head-space air in pharmaceutical containers to help minimize deterioration by oxidation.

As an oxidation reaction is complicated, it is difficult to perform a kinetic study on oxidative processes within a general stability program. The redox potential, which is constant and relatively easy to determine, can, however, provide valuable predictive information. In many oxidative reactions, the rate is proportional to the concentration of the oxidizing species but may be independent of the concentration of the oxygen present. The rate is influenced by temperature, radiation, and the presence of a catalyst. An increase in temperature leads to an acceleration in the rate of oxidation. If the storage temperature of a preparation can be reduced to between 0–5° C, usually it can be assumed that the rate of oxidation will be at least halved.

The molecular structures most likely to oxidize are those with a hydroxyl group directly bonded to an aromatic ring (e.g., phenol derivatives such as catecholamines and morphine), conjugated dienes (e.g., vitamin A and unsaturated free fatty acids), heterocyclic aromatic rings, nitroso and nitrite derivatives, and aldehydes (e.g., flavorings). Products of oxidation usually lack therapeutic activity. Visual identification of oxidation, for example, the change from colorless epinephrine to its amber colored products, may not be visible in some dilutions or to some eyes.

Oxidation is catalyzed by pH values that are higher than optimum, polyvalent heavy metal ions (e.g., copper and iron), and exposure to oxygen and UV illumination. The latter two causes of oxidation justify the use of antioxidant chemicals, nitrogen atmospheres during ampoule and vial filling, opaque external packaging, and transparent amber glass or plastic containers.

Trace amounts of heavy metals such as cupric, chromic, ferrous, or ferric ions may catalyze oxidation reactions. As little as 0.2 mg of copper ion per liter considerably reduces the stability of penicillin. Similar examples include the deterioration of epinephrine, phenylephrine, lincomycin, isoprenaline, and procaine hydrochloride. Adding chelating agents to water to sequester heavy metals and working in special manufacturing equipment (e.g., glass) are some means used to reduce the influence of heavy metals on a formulation. Parenteral formulations should not come in contact with heavy metal ions during their manufacture, packaging, or storage.

Hydronium and hydroxyl ions catalyze oxidative reactions. The rate of decomposition for epinephrine, for example, is more rapid in a neutral or alkaline solution with maximum stability (minimum oxidative decomposition) at pH 3.4. There is a pH range for maximum stability for any antibiotic and vitamin preparation, which usually can be achieved by adding an acid, alkali, or buffer.

Oxidation may be inhibited by the use of antioxidants, called negative catalysts. They are very effective in stabilizing pharmaceutical products undergoing a free-radical-mediated chain reaction. These substances, which are easily oxidizable, act by possessing lower oxidation potentials than the active ingredient. Thus, they undergo preferential degradation or act as chain inhibitors of free radicals by providing an electron and receiving the excess energy possessed by the activated molecule.

The ideal antioxidant should be stable and effective over a wide pH range, soluble in its oxidized form, colorless, nontoxic, nonvolatile, nonirritating, effective in low concentrations, thermostable, and compatible with the container-closure system and formulation ingredients.

The commonly used antioxidants for aqueous systems include sodium sulfite, sodium metabisulfite, sodium bisulfite,

sodium thiosulfate, and ascorbic acid. For oil systems, ascorbyl palmitate, hydroquinone, propyl gallate, nordihydroguaiaretic acid, butylated hydroxytoluene, butylated hydroxyanisole, and alpha-tocopherol are employed.

Synergists, which increase the activity of antioxidants, are generally organic compounds that complex small amounts of heavy metal ions. These include the ethylenediamine tetraacetic acid (EDTA) derivatives, dihydroethylglycine, and citric, tartaric, gluconic, and saccharic acids. EDTA has been used to stabilize ascorbic acid, oxytetracycline, penicillin, epinephrine, and prednisolone.

Reduction reactions are much less common than oxidative processes in pharmaceutical practice. Examples include the reduction of gold, silver, or mercury salts by light to form the corresponding free metal.

Hydrolysis

Drugs containing esters (e.g., cocaine, physostigmine, aspirin, tetracaine, procaine, and methyl dopa), amides (e.g., dibucaine), imides (e.g., amobarbital), imines (e.g., diazepam), and lactam (e.g., penicillins, cephalosporins) functional groups are among those prone to hydrolysis.

Hydrolysis reactions are often pH dependent and are catalyzed by either hydronium ion or hydroxide ions (specific-acid or specific-base catalysis, respectively). Hydrolysis reactions can also be catalyzed by either a Brønsted acid or a Brønsted base (general-acid or general-base catalysis, respectively). Sources of Brønsted acid or base include buffers and some excipients. Sometimes, it is necessary to compromise between the optimum pH for stability and that for pharmacological activity. For example, several local anesthetics are most stable at a distinctly acid pH, whereas for maximum activity they should be neutral or slightly alkaline. Small amounts of acids, alkalines, or buffers are used to adjust the pH of a formulation. Buffers are used when small changes in pH are likely to cause major degradation of the active ingredient.

Obviously, the amount of water present can have a profound effect on the rate of a hydrolysis reaction. When the reaction takes place fairly rapidly in water, other solvents sometimes can be substituted. For example, barbiturates are much more stable at room temperature in propylene glycol–water than in water alone.

Modification of chemical structure may be used to retard hydrolysis. In general, as it is only the fraction of the drug in solution that hydrolyzes, a compound may be stabilized by reducing its solubility. This can be done by adding various substituents to the alkyl or acyl chain of aliphatic or aromatic esters or to the ring of an aromatic ester. In some cases less-soluble salts or esters of the parent compound have been found to aid product stability. Steric and polar complexation has also been employed to alter the rate of hydrolysis. Caffeine reduces the rate of hydrolysis and thus promotes stability by complexation with local anesthetics such as benzocaine, procaine, or tetracaine.

Esters and β -lactams are the chemical bonds that are most likely to hydrolyze in the presence of water. For example, the acetyl ester in aspirin is hydrolyzed to acetic acid and salicylic acid in the presence of moisture, but in a dry environment the hydrolysis of aspirin is negligible. The aspirin hydrolysis rate increases in direct proportion to the water vapor pressure in an environment.

The amide bond also hydrolyzes, though generally at a slower rate than comparable esters. For example, procaine (an ester) will hydrolyze upon autoclaving, but procainamide will not. The amide or peptide bond in peptides and proteins varies in the lability to hydrolysis. The lactam and azomethine (or imine) bonds in benzodiazepines are also labile to hydrolysis. The major chemical accelerators or catalysts of hydrolysis are adverse pH and specific chemicals (e.g., dextrose and copper in the case of ampicillin hydrolysis).

The rate of hydrolysis depends on the temperature and the pH of the solution. A much-quoted estimation is that for each

10°C rise in storage temperature, the rate of reaction doubles or triples. As this is an empiricism, it is not always applicable.

When hydrolysis occurs, the concentration of the active ingredient decreases while the concentration of the decomposition products increases. The effect of this change on the rate of the reaction depends on the order of the reaction. With zero-order reactions the rate of decomposition is independent of concentration of the ingredient. Although weak solutions decompose at the same absolute rate as stronger ones, the weaker the solution, the greater the proportion of active ingredient destroyed in a given time; i.e., the percentage of decomposition is greater in weaker solutions. Increasing the concentration of an active ingredient that is hydrolyzing by zero-order kinetics will slow the percentage decomposition.

With first-order reactions, which occur frequently in the hydrolysis of drugs, the rate of change is directly proportional to the concentration of the reactive substance. Thus, changes in the concentration of the active ingredient have no influence on the percentage decomposition.

The degradation of many drugs in solution accelerates or decelerates exponentially as the pH is decreased or increased over a specific range of pH values. Improper pH ranks with exposure to elevated temperature as a factor most likely to cause a clinically significant loss of drug, resulting from hydrolysis and oxidation reactions. A drug solution or suspension, for example, may be stable for days, weeks, or even years in its original formulation, but when mixed with another liquid that changes the pH, it may degrade in minutes or days. It is possible that a pH change of only one unit (e.g., from 4 to 3 or 8 to 9) could decrease drug stability by a factor of ten or greater.

A pH-buffer system, which is usually a weak acid or base and its salt, are common excipients used in liquid preparations to maintain the pH in a range that minimizes the drug degradation rate. The pH of drug solutions may also be either buffered or adjusted to achieve drug solubility. For example, pH in relation to pKa controls the fractions of the usually more soluble ionized and less soluble nonionized species of weak organic electrolytes.

Interionic (Ion N+ – Ion N –) Compatibility

The compatibility or solubility of oppositely charged ions depends mainly on the number of charges per ion and the molecular size of the ions. In general, polyvalent ions of opposite charge are more likely to be incompatible. Thus, an incompatibility is likely to occur upon the addition of a large ion with a charge opposite to that of the drug.

As many hydrolytic reactions are catalyzed by both hydronium and hydroxyl ions, pH is an important factor in determining the rate of a reaction. The pH range of minimum decomposition (or maximum stability) depends on the ion having the greatest effect on the reaction. If the minimum occurs at about pH 7, the two ions are of equal effect. A shift of the minimum toward the acid side indicates that the hydroxyl ion has the stronger catalytic effect and *vice versa* in the case of a shift toward the alkaline side. In general, hydroxyl ions have the stronger effect. Thus, the minimum is often found between pH 3 and 4.

The influence of pH on the physical stability of two-phase systems, especially emulsions, is also important. For example, intravenous fat emulsion is destabilized by acidic pH.

Decarboxylation

Pyrolytic solid-state degradation through decarboxylation usually is not encountered in pharmacy, as relatively high heats of activation (25 to 30 kcal) are required for the reaction. However, solid *p*-aminosalicylic acid undergoes pyrolytic degradation to *m*-aminophenol and carbon dioxide. The reaction, which follows first-order kinetics, is highly pH-dependent and is catalyzed by hydronium ions. The decarboxylation of *p*-aminobenzoic acid occurs only at extremely low pH values and at high temperatures. Some dissolved carboxylic acids, such as *p*-aminosalicylic acid, lose carbon dioxide from the carboxyl

group when heated. The resulting product has reduced pharmacological potency. β -Keto decarboxylation can occur in some solid antibiotics that have a carbonyl group on the β -carbon of a carboxylic acid or a carboxylate anion. Such decarboxylations will occur in the following antibiotics: carbenicillin sodium, carbenicillin free acid, ticarcillin sodium, and ticarcillin free acid.

Racemization

Racemization, or the action or process of changing from an optically active compound into a racemic compound or an optically inactive mixture of corresponding *R* (*rectus*) and *S* (*sinister*) forms, is a major consideration in pharmaceutical stability. Optical activity of a compound may be monitored by polarimetry and reported in terms of specific rotation. Chiral high performance liquid chromatography (HPLC) has been used in addition to polarimetry to confirm the enantiomeric purity of a sample.

Epimerization

Members of the tetracycline family are most likely to incur epimerization. This reaction occurs rapidly when the dissolved drug is exposed to a pH of an intermediate range (higher than 3), and it results in the steric rearrangement of the dimethylamino group. The epimer of tetracycline, epitetracycline, has little or no antibacterial activity.

In general, racemization follows first-order kinetics and depends on temperature, solvent, catalyst, and the presence or absence of light. Racemization appears to depend on the functional group bound to the asymmetric carbon atom, with aromatic groups tending to accelerate the process.

PHOTOCHEMICAL REACTIONS

Photolytic degradation can be an important limiting factor in the stability of pharmaceuticals. A drug can be affected chemically by radiation of a particular wavelength only if it absorbs radiation at that wavelength and the energy exceeds a threshold. Ultraviolet radiation, which has a high energy level, is the cause of many degradation reactions. Exposure to, primarily, Ultraviolet (UV) illumination may cause oxidation (photo-oxidation) and scission (photolysis) of covalent bonds. Nifedipine, nitroprusside, riboflavin, and phenothiazines are very labile to photo-oxidation. In susceptible compounds, photochemical energy creates free radical intermediates, which can perpetuate chain reactions.

If the absorbing molecule reacts, the reaction is said to be photochemical in nature. When the absorbing molecules do not participate directly in the reaction, but pass their energy to other reacting molecules, the absorbing substance is said to be a photosensitizer.

As many variables may be involved in a photochemical reaction, the kinetics can be quite complex. The intensity and wavelength of the light and the size, shape, composition, and color of the container may affect the velocity of the reaction.

The photodegradation of chlorpromazine through a semiquinone free-radical intermediate follows zero-order kinetics. On the other hand, alcoholic solutions of hydrocortisone, prednisolone, and methylprednisolone degrade by reactions following first-order kinetics.

Colored-glass containers most commonly are used to protect light-sensitive formulations. Yellow-green glass gives the best protection in the ultraviolet region, whereas amber confers considerable protection from ultraviolet radiation but little from infrared radiation. Riboflavin is best protected by a stabilizer that has a hydroxyl group attached to or near the aromatic ring. The photodegradation of sulfacetamide solutions may be inhibited by an antioxidant such as sodium thiosulfate or metabisulfite.

A systematic approach to photostability testing is recommended covering, as appropriate, studies such as tests on the drug substance, tests on the exposed drug product outside of the immediate pack; and if necessary, tests on the drug product in the immediate pack. The ICH Q1B guidance discusses

the minimum requirements for assessing photostability. Drug substance is first assessed by exposing sample powder having a depth of not more than 3 mm to an overall illumination of not less than 1.2 million lux hours and an integrated near ultraviolet energy of not less than 200 watt hours/square meter. If the drug substance shows sensitivity to photodegradation, then the drug product will need to be tested as well. The testing of drug product uses the same light exposure that was used to test drug substance. The drug product should be tested directly exposed to light and in its container closure system.

ULTRASONIC ENERGY

Ultrasonic energy, which consists of vibrations and waves with frequencies greater than 20,000 Hz, promotes the formation of free radicals and alters drug molecules. Changes in prednisolone, prednisone acetate, or deoxycorticosterone acetate suspensions in an ultrasonic field have been observed spectrometrically in the side chain at C-17 and in the oxo group of the A ring. With sodium alginate, in an ultrasonic field, it has been reported that above a minimum power output, degradation increased linearly with increased power.

IONIZING RADIATION

Ionizing radiation, particularly gamma rays, has been used for the sterilization of certain pharmaceutical products. At the usual sterilizing dose, 2.5 mRad, it seldom causes appreciable chemical degradation. In general, formulations that are in the solid or frozen state are more resistant to degradation from ionizing radiation than those in liquid form. For example, many of the vitamins are little affected by irradiation in the solid state but are decomposed appreciably in solution. On the other hand, both the liquid- and solid-state forms of atropine sulfate are affected seriously by radiation.

PREDICTING SHELF LIFE

ICH RECOMMENDED EVALUATION

The shelf life of a commercial drug product must be determined in the commercial container closure at the defined storage conditions. The FDA and ICH Q1A (R2) guidances require at least 12 months stability data at the time of submission. Most products require at least 24 months shelf life to be commercially viable. The ICH Q1E recommends how the 12 months data may be used to predict long-term stability. Figures 4-1 and 4-2 show trending graphs with double-sided and single-sided 95%

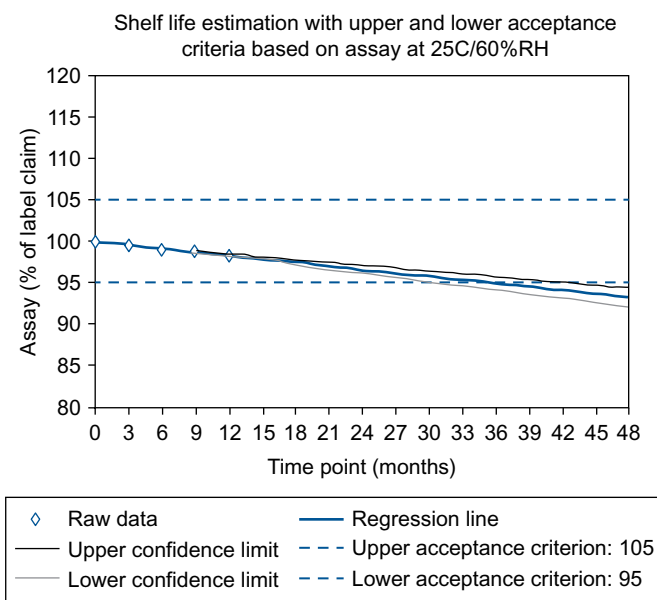


Figure 4-1. Typical Two-sided Shelf Life Estimation plot.

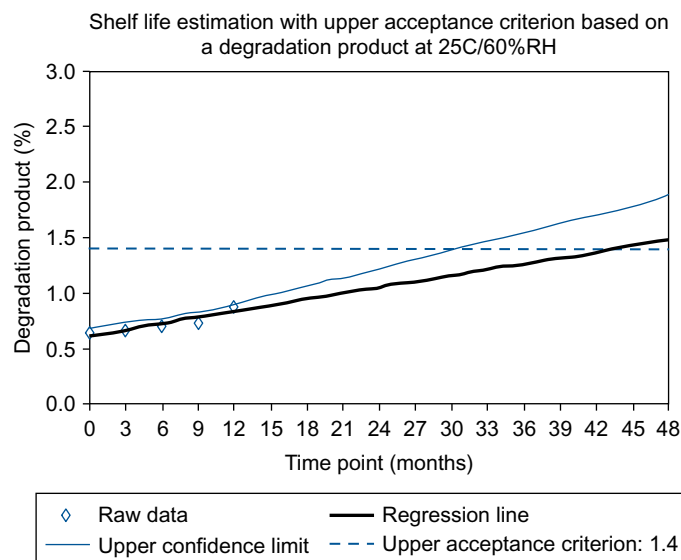


Figure 4-2. Typical One-sided Shelf Life Estimation plot.

confidence limits plots, respectively. The expiration of a product is the time where the confidence line intersects with the acceptance limit. Trend analysis of data need only be performed on test data that show a change related to time.

APPROXIMATIONS IN ASSESSING PRODUCT STABILITY

ESTIMATION OF TEMPERATURE EFFECT

In early development, a shelf life prediction of a clinical material, especially a Phase I material, may be based on a very limited amount of sample and a very limited amount of time to make the evaluation. One way to estimate long-term storage for a material is by extrapolating data from studies performed at elevated conditions. An understanding of potential activation energy is needed to make the long-term stability prediction. Many may have heard of the estimate that for every 10°C decrease in storage temperature the shelf life doubles. This is only true, however, if the activation energy of the reaction(s) that cause degradation is 15 kcal/moles. The activation energy, E_a , for many chemical processes related to the degradation of a drug substance/product is typically within the range of 10 to 25 kcal/moles. The equation below shows a way of calculating the Q_{AT} value that may be used to estimate the affect of temperature on shelf life (Table 4-3).

$$Q_{AT} = \exp \left[\frac{E_a}{R} \left(\frac{\Delta T}{(T+\Delta T)(T)} \right) \right] \quad (2)$$

The technique of estimating the shelf life of a formulation from its accumulated stability data has evolved from examining the data and making an educated guess through plotting the time-temperature points on appropriate graph paper and crudely extrapolating a regression line to the application of rigorous

Table 4-3. An Approximation for Estimating the Temperature Effects

E_a (kcal/mole)	Q5 (30 to 25 °C)	Q10 (35 to 25 °C)	Q15 (40 to 25 °C)
10	1.32	1.73	2.24
15	1.52	2.27	3.36
20	1.75	2.99	5.04
25	2.01	3.93	7.55

physical-chemical laws, statistical concepts, and computers to obtain meaningful, reliable estimates.

A simple means of estimating shelf life from a set of computer-prepared tables has been described by Lintner *et al.*⁶ This system was developed to select the best prototype formulation on the basis of short-term stability data and predict both estimated and minimum shelf life values for the formulation. It is a middle-ground approach between the empirical methods and the modern, rigorous statistical concepts. All calculations can be made readily by hand, and the estimated values can be obtained easily from appropriate tables. The system assumes that:

1. Shelf life predictions can be made satisfactorily for lower temperatures using the classical Arrhenius model from data obtained at higher temperatures.
2. The energy of activation of the degradation reaction is between 10 and 20 kcal/mol (this is a safe assumption, as Kennon¹⁶ has noted that rarely are drugs with energies of activation below 10 kcal/mol used in pharmacy, and for values as high as 20 kcal/mol, the error in the shelf life prediction will be on the conservative side).
3. The rate of decomposition will not increase beyond that already observed.
4. The standard deviation of the replicated assays is known or can be estimated from the analytical data.

This concept further assumes that the degradation reaction follows zero- or pseudo-zero-order kinetics. For data corresponding to a zero-, first-, or second-order degradation pattern, it is impossible to distinguish one order from another with usual analytical procedures, when the total degraded material is not large. In addition, shelf life calculations assuming zero-order kinetics are more conservative than those for higher orders.

This middle-ground system is useful in creating the experimental design for the stability study. The formulator has the opportunity to study various combinations of parameters to try to optimize the physical-statistical model. The effect of improving the assay standard deviation, running additional replicates, using different time points, and assuming various degradation rates and energies of activation on the stability of the test formulation can be checked.

McMinn and Lintner later developed and reported on an information-processing system for handling product stability data.¹⁷ This system saves the time of formulators in analyzing and interpreting their product stability data, in addition to minimizing the amount of clerical help needed to handle an ever-increasing assay load. For products such as those of vitamins, for example, where large overages are required, the statistical portions of this advanced technique aid the manufacturer to tailor the formula composition to obtain the desired and most economical expiration dating.

This system stores both physical and chemical data and retrieves the information in three different formats (one of which was designed specifically for submitting to regulatory agencies). It analyzes single-temperature data statistically by analysis of covariance and regression or multiple-temperature data by weighted or unweighted analysis using the Arrhenius relationship; provides estimates of the shelf life of the preparation with the appropriate confidence intervals; preprints the assay request cards that are used to record the results of the respective assay procedures and to enter the data into the system; and produces a 5-year master-stability schedule as well as periodic 14-day schedules of upcoming assays.

As mentioned above, a portion of the advanced system analyzes the stability data obtained at a single temperature by analysis of covariance and regression. This analysis is based on the linear (zero-order) model

$$Y_{ij} = \beta_i X_{ij} + \alpha_i + \varepsilon_{ij} \quad (3)$$

where Y_{ij} is the percentage of label of the j th stability assay of the i th lot, X_{ij} is the time in months at which Y_{ij} was observed,

β_i and α_i are the slope and intercept, respectively, of the regression line of the i th lot, and ε_{ij} is a random error associated with Y_{ij} . The random errors are assumed to be identically and independently distributed normal variables with a zero mean and a common variance, σ^2 .

A summary of the regression analysis for each individual lot and for the combination of these lots, plus a summary of the analyses of covariance and deviation from regression is prepared by the computer.

Because the computer combines, or pools, the stability data from the individual lots, irrespective of the statistical integrity of this step, the pooled data are examined for validity by the F test. The mean square of the regression coefficient (slope) is divided by the mean square of the deviation within lots, and similarly, the adjusted mean (y intercept) is divided by the common mean square to give the respective F ratios. The latter values then are compared with the critical 5% F values. When the calculated F values are smaller than the critical F values, the data may be combined, and the pooled data analyzed.

A printout for the combined lots as well as for each individual lot provides the estimated rate of degradation and its standard error in percentage per month for each ingredient. The *Student's t* test value is calculated from these estimates and tested for significance from zero. When the t value is significant, the printout contains an estimate of the shelf life with the appropriate confidence interval. When the t value is not significantly different from zero, estimates of the minimum and projected shelf life values are made. In addition, coordinates of the calculated least-squares regression line with appropriate confidence limits for the mean and individual predicted assays are printed.

Plots of the resulting least-squares line containing the individual data points also are printed by the computer. For the calculation of X_0 , $\hat{Y}$ equals $\langle \text{obar} \rangle Y + \langle \text{caret} \rangle \beta (X_0 - \langle \text{obar} \rangle X_{..})$, where $\langle \text{caret} \rangle \beta$ is the least-squares estimate of the slope, and $\langle \text{obar} \rangle X_{..}$ is the mean time of assay.

The sample variance for this estimate, $S^2(\hat{Y})$, is equal to

$$S_{y..x}^2 \left[\frac{1}{N} + \frac{\left(X_0 - \overline{X_{..}} \right)^2}{\sum \left(X_{ij} - \overline{X_{..}} \right)^2} \right] \quad (4)$$

where N is the number of assays. The 95% confidence interval is equal to $Y \pm t_{0.05S(y)}$.

For cases in which the slope of the best fitting line is positive and significantly different from zero (resulting, e.g., from solvent evaporation), the statement "no degradation has been detected and hence no shelf life estimate is made" is printed. When the computed line has a positive slope but not significantly different from zero, only the minimum shelf life value is calculated.

Traditionally, extensive stability data are collected at the recommended storage temperatures (usually refrigerator and/or room temperature) to be placed on the label of the package. However, elevated-temperature data are very valuable in determining the shelf life of a product. In practice, multiple levels of thermal stress are applied to the formulation so that appropriate shelf life estimates can be made for normally expected marketing conditions. In cases in which data from accelerated studies are used to project a tentative expiration date that is beyond the date supported by actual shelf life studies, testing must continue until the tentative expiration date is verified.

The effect of temperature variation on the rate of a reaction could be expressed by an integrated form of the Arrhenius equation

$$k = se^{-E_A / RT} \quad (5)$$

$$\log \frac{k_2}{k_1} = \frac{E_a}{2.303R} \left(\frac{T_2 - T_1}{T_2 * T_1} \right) \quad (6)$$

A weighted modification of this model has been incorporated into the previously described computerized system. Each

printout contains a statement concerning the acceptability of the Arrhenius assumption with its appropriate probability level, the slope and intercept for the Arrhenius line, the estimated apparent energy of activation with its 95% confidence limits, plus the estimated shelf life values at selected temperatures.

The analysis of first-order stability data is based on the linear model

$$Y_{ij} = \alpha_i + \beta_i X_{ij} + \varepsilon_{ij} \quad (7)$$

where Y_{ij} is the natural logarithm of the assay value for the j th observation of the i th temperature, X_{ij} is the elapsed time in months for the assay sample for the i th temperature, β_i and α_i are the slope and intercept, respectively, and ε_{ij} is a random error associated with Y_{ij} . The errors are assumed to be distributed identically and independently, normally with a zero mean and variance σ^2 .

For orders other than first, Y_{ij} represents the concentration raised to the power of 1 minus the order.

The estimated rate constant (i.e., the negative slope) is

$$-b_i = -\sum_j (Y_{ij} - Y_i)(X_{ij} - X_i) / \sum_j (X_{ij} - X_i)^2 \quad (8)$$

The standard error of the estimated rate constant is

$$S_{-b_i} = \frac{S(X/Y)}{\left[\sum (x_{ij} - X_i)^2 \right]^{1/2}} \quad (9)$$

where $S(X/Y)$, the residual standard error, is equal to

$$S(X/Y) = \left\{ \frac{1}{N-2} \left[\sum_{j=1}^{12} (Y_{ij} - Y_i)^2 - \frac{\left[\sum (X_{ij} - X_i)(Y_{ij} - Y_i)^2 \right]^2}{\sum (X_{ij} - X_i)^2} \right] \right\}^{1/2} \quad (10)$$

According to the Arrhenius relationship, faster degradation occurs at the higher temperatures; hence, assays for the high-temperature data usually are run more often but for a shorter period of time. The effect of simple least-squares analysis of this type of data is to force the Arrhenius equation through the low temperature data and essentially ignore the high-temperature information. Thus, much more credence is placed in the point estimates of the low temperature than is warranted. In addition, the usual confidence limits on extrapolated degradation rates at refrigerator or room temperature cannot be made validly. For these reasons, Bentley¹⁸ presented a method based on a weighted least-squares analysis to replace the un-weighted approximation. He also developed a statistical test for the validity of the Arrhenius assumption, which is computed easily from the results of the unweighted method.

To make shelf life estimates from elevated temperature data, two storage temperatures are obviously the minimum. As the accuracy of the extrapolation is enhanced by using additional temperatures, a minimum of four different temperatures is recommended for most product stability studies. With the current use of computers to do the bulk of stability calculations, including weighted least-squares analysis, the temperatures, and storage conditions need not be selected for arithmetic convenience.

It is not necessary to determine the mechanism of the degradation reaction. In most cases, it is necessary only to follow some property of degradation and to linearize this function. Either the amount of intact drug or the amount of a formed degradation product may be followed. It usually is impractical to determine the exact order of the reaction. With assay errors in the range of 2 to 5%, at least 50% decomposition must occur before the reaction order can be determined. As the loss with pharmaceuticals generally is less, zero-order kinetics should be assumed, unless the reaction order is known from previous work. In any case, replication of stability assays is advisable.

The batches of drugs used for a stability study should be representative of production run material or at least material of a known degree of purity. The quality of the excipients also should be known, as their impurities or even their moisture

content can affect product stability deleteriously. Likewise, the samples of the formulation taken for the stability study must be representative of the lot.

Specific, stability-indicating assay methods must be used, to make meaningful shelf life estimates. The reliability and specificity of the test method on the intact molecule and on the degradation products must be demonstrated.

PHARMACEUTICAL CONTAINERS

The official standards for containers apply to articles packaged by either the pharmaceutical manufacturer or the dispensing pharmacist unless otherwise indicated in a compendial monograph. In general, repackaging of pharmaceuticals is inadvisable. However, if repackaging is necessary, the manufacturer of the product should be consulted for potential stability problems.

A pharmaceutical container has been defined as a device that holds the drug and is, or may be, in direct contact with the preparation. The immediate container is described as that which is in direct contact with the drug at all times. The liner and closure traditionally have been considered to be part of the container system. The container should not interact physically or chemically with the formulation so as to alter the strength, quality, or purity of its contents beyond permissible limits.

The choice of containers and closures can have a profound effect on the stability of many pharmaceuticals. Now that a large variety of glass, plastics, rubber closures, tubes, tube liners, etc are available, the possibilities for interaction between the packaging components and the formulation ingredients are immense. Some of the packaging elements themselves are subject to physical and chemical changes that may be time-temperature dependent.

Frequently, it is necessary to use a well-closed or a tight container to protect a pharmaceutical product. A *well-closed container* is used to protect the contents from extraneous solids or a loss in potency of the active ingredient under normal commercial conditions. A *tight container* protects the contents from contamination by extraneous materials, loss of contents, efflorescence, deliquescence, or evaporation and is capable of tight re-closure. When the packaging and storage of an official article in a well-closed or tight container is specified, water-permeation tests should be performed on the selected container.

In a stability program, the appearance of the container, with special emphasis on the inner walls, the migration of ingredients onto/into the plastic or into the rubber closure, the migration of plasticizer or components from the rubber closure into the formulation, the possibility of two-way moisture penetration through the container walls, the integrity of the tac-seal, and the back-off torque of the cap, must be studied.

GLASS

Traditionally, glass has been the most widely used container for pharmaceutical products to ensure inertness, visibility, strength, rigidity, moisture protection, ease of re-closure, and economy of packaging. Although glass has some disadvantages, such as the leaching of alkali and insoluble flakes into the formulation, these can be offset by the choice of an appropriate glass. As the composition of glass may be varied by the amounts and types of sand and silica added and the heat treatment conditions used, the proper container for any formulation can be selected.

According to USP 34, glass containers suitable for packaging pharmacopeial preparations may be classified as either Type I, Type II, Type III, or type NP. Containers of Type I borosilicate glass are generally used for preparations that are intended for parenteral administration, although Type II treated soda-lime glass may be used where stability data demonstrates its suitability. Containers of Type III and Type NP are intended for packaging articles intended for oral or topical use.

New, unused glass containers are tested for resistance to attack by high-purity water by use of a sulfuric acid titration to

determine the amount of released alkali. Both glass and plastic containers are used to protect light-sensitive formulations from degradation. The amount of transmitted light is measured using a spectrometer of suitable sensitivity and accuracy.

Glass is generally available in flint, amber, blue, emerald green, and certain light-resistant green and opal colors. The blue-, green-, and flint-colored glasses, which transmit ultraviolet and violet light rays, do not meet the official specifications for light-resistant containers.

Colored glass usually is not used for injectable preparations, since it is difficult to detect the presence of discoloration and particulate matter in the formulations. Light-sensitive drugs for parenteral use usually are sealed in flint ampoules and placed in a box. Multiple-dose vials should be stored in a dark place.

Manufacturers of prescription drug products should include sufficient information on their product labels to inform the pharmacist of the type of dispensing container needed to maintain the identity, strength, quality, and purity of the product. This brief description of the proper container, such as light-resistant, well-closed, or tight, may be omitted for those products dispensed in the manufacturer's original container.

PLASTICS

Plastic containers have become very popular for storing pharmaceutical products. Polyethylene, polystyrene, polyvinyl chloride, and polypropylene are used to prepare plastic containers of various densities to fit specific formulation needs.

Factors such as plastic composition, processing and cleaning procedures, contacting media, inks, adhesives, absorption, adsorption, and permeability of preservatives also affect the suitability of a plastic for pharmaceutical use. Hence, biological test procedures are used to determine the suitability of a plastic for packaging products intended for parenteral use and for polymers intended for use in implants and medical devices. Systemic injection and intracutaneous and implantation tests are employed. In addition, tests for nonvolatile residue, residue on ignition, heavy metals, and buffering capacity were designed to determine the physical and chemical properties of plastics and their extracts.

The high-density polyethylene (HDPE) containers, which are used for packaging capsules and tablets, possess characteristic thermal properties, a distinctive infrared absorption spectrum, and a density between 0.941 and 0.965 g/cm³. In addition, these containers are tested for light transmission, water-vapor permeation, extractable substances, nonvolatile residue, and heavy metals. When a stability study has been performed to establish the expiration date for a dosage form in an acceptable high-density polyethylene container, any other high-density polyethylene container may be substituted provided that it, too, meets compendial standards and that the stability program is expanded to include the alternative container.

Materials from the plastic itself can leach into the formulation, and materials from the latter can be absorbed onto, into, or through the container wall. The barrels of some plastic syringes bind various pharmaceutical preservatives. However, changing the composition of the syringe barrel from nylon to polyethylene or polystyrene has eliminated the binding in some cases.

A major disadvantage of plastic containers is the two-way permeation or *breathing* through the container walls. Volatile oils and flavoring and perfume agents are permeable through plastics to varying degrees. Components of emulsions and creams have been reported to migrate through the walls of some plastics, causing either a deleterious change in the formulation or collapse of the container. Loss of moisture from a formulation is common. Gases, such as oxygen or carbon dioxide in the air, have been known to migrate through container walls and affect a preparation.

Solid dosage forms, such as penicillin tablets, when stored in some plastics, are affected deleteriously by moisture penetration from the atmosphere into the container.

Single unit dosage packaging in the form of blister packages are often used to package capsule and tablet dosage forms. A typical blister package is comprised of a polymeric film that is molded to have a cavity into which the dosage form is placed. The polymer film is then heat bonded to a paper backed foil liner.

As with plastic bottles, the blister package will allow a certain amount of moisture vapor permeation to occur, and this must be a consideration when selecting the type of film used for the package. The choice of packaging materials used depends on the degree to which the product needs to be protected from light, heat, and moisture. Each material has different resistance to each of these elements and will affect the shelf life and storage conditions of the packaged pharmaceutical.

Polyvinylchloride (PVC) offers the least resistance to moisture vapor permeation. Polyvinylidenechloride (PVdC) has characteristics similar to PVC but offers superior resistance to moisture vapor permeation. Aclar, which is a polychlorotrifluoroethylene (PVC-CTFE) film has the lowest water vapor permeability and thus offers the best protection from moisture.

METALS

The pharmaceutical industry was, and to a degree still is, a tin stronghold. However, as the price of tin constantly varies, more-collapsible aluminum tubes are being used.

A variety of internal linings and closure fold seals are available for both tin and aluminum tubes. Tin tubes can be coated with wax or with vinyl linings. Aluminum tubes are available with lining of epoxy or phenolic resin, wax, vinyl, or a combination of epoxy or phenolic resin with wax. As aluminum is able to withstand the high temperatures required to cure epoxy and phenolic resins adequately, tubes made from this metal presently offer the widest range of lining possibilities.

Closure fold seals may consist of unmodified vinyl resin or plasticized cellulose and resin, with or without added color.

Collapsible tubes are available in many combinations of diameters, lengths, openings, and caps. Custom-use tips for ophthalmic, nasal, mastitis, and rectal applications also are available. Only a limited number of internal liners and closure seals are available for tubes fitted with these special-use tips.

Lined tubes from different manufacturers are not necessarily interchangeable. Although some converted resin liners may be composed of the same base resin, the actual liner may have been modified to achieve better adhesion, flow properties, drying qualities, or flexibility. These modifications may have been necessitated by the method of applying the liner, the curing procedure, or, finally, the nature of the liner itself.

CLOSURES

The closures for the formulations also must be studied as a portion of the overall stability program. Although the closure must form an effective seal for the container, the closure must not react chemically or physically with the product. It must not absorb materials from the formulation or leach its ingredients into the contents.

The integrity of the seal between the closure and container depends on the geometry of the two, the materials used in their construction, the composition of the cap liner, and the tightness with which the cap has been applied. Torque is a measure of the circular force, measured in inch-pounds, which must be applied to open or close a container. When pharmaceutical products are set up on a stability study, the formulation must be in the proposed market package. Thus, they should be capped with essentially the same torque to be used in the manufacturing step.

Rubber is a common component of stoppers, cap liners, and parts of dropper assemblies. Sorption of the active ingredient, preservative, or other formulation ingredients into the rubber and the extraction of one or more components of the rubber into the formulation are common problems.

The application of an epoxy and polytetrafluoroethylene lining to the rubber closure reduces the amount of leached

extractives but essentially has no effect on the sorption of the preservative from the solution. Polytetrafluoroethylene-coated rubber stoppers may prevent most of the sorption and leaching.

REFERENCES

1. Connors KA, et al. *Chemical Stability of Pharmaceuticals*, 2nd edn. New York: Wiley, 1986.
2. Carstensen JT. *Drug Stability Principles and Practices*. New York: Marcel Dekker, 1990.
3. Allen Loyd V. Dosage Form Design and Development. *Clinical Therapeutics* 2008; 30(11): 2102–2111.
4. United States Pharmacopeia, Inc. *United States Pharmacopeia 34 – National Formulary 29*. Rockville, MD: US Pharmacopeia Convention, Inc. 2011; 742–745.
5. US Food and Drug Administration. *Current Good Manufacturing Practice for Finished Pharmaceuticals, 21CFR211*.
6. United States Pharmacopeia, Inc. *United States Pharmacopeia 34 – National Formulary 29*, Rockville, MD: US Pharmacopeia Convention, Inc. 2011.
7. Health Products and Food Branch Inspectorate Good Manufacturing Practices (GMP) Guidelines – 2009 edn, GUI-001, Health Canada, May 2009; 62.
8. EudraLex. The Rules Governing Medicinal Products in the European Union, Volume 4. EU Guidelines to Good Manufacturing Practice Medicinal Products for Human and Veterinary Use, Part I, Chapter 1, Quality Management, 2008; 5.
9. United States Pharmacopeia, Inc. *United States Pharmacopeia 34 – National Formulary 29*, Rockville, MD: US Pharmacopeia Convention, Inc. 2011; 11–12.
10. *Q1A (R2) Stability Testing of New Drug Substances and Products*. International Conference on Harmonization (ICH), FDA CDER CBER, 2003.
11. *Q1B Stability Testing: Photostability Testing of New Drug Substances and Products*. International Conference on Harmonization (ICH), FDA CDER CBER, 1996.
12. *Q1C Stability Testing for New Dosage Forms*. International Conference on Harmonization (ICH), FDA CDER CBER, 1996.
13. *Q1D Bracketing and Matrixing Designs for Stability Testing of New Drug Substances and Products*. International Conference on Harmonization (ICH), FDA CDER CBER, 2003.
14. *Q1E Evaluation of Stability Data*. International Conference on Harmonization (ICH), FDA CDER CBER, 2004.
15. *Q1F Stability Data Package for Registration Applications in Climatic Zones III and IV*. International Conference on Harmonization (ICH), FDA CDER CBER, 2003.
16. Kennon L. Use of models in determining chemical pharmaceutical stability. *J Pharm Sci* 1964; 53: 815.
17. McMinn CS, Lintner CJ. (Oral presentation), APHA Acad Pharm Sci Mtg Ind Pharm Tech Sec. Chicago, May 1973. Please check reference for oral presentation.
18. Bentley DL. Statistical Techniques in Predicting Thermal Stability. *J Pharm Sci* 1970; 59: 464.

Chapter 5

Bioavailability and Bioequivalence Testing

Steven B. Johnson, PharmD

INTRODUCTION 51

DISSOLUTION 51

ABSORPTION FACTORS 52

BIOEQUIVALENCE TESTING 53

METHODS FOR DETERMINING BIOEQUIVALENCE 53

MINIMIZING THE NEED FOR BIOEQUIVALENCE STUDIES 54

EVALUATION OF BIOEQUIVALENCE DATA 54

BIOEQUIVALENCE ASSESSMENT AND DATA EVALUATION 54

CRITERIA FOR BIOEQUIVALENCE 56

STUDY DESIGN 58

THERAPEUTIC EQUIVALENCE EVALUATIONS 60

INTRODUCTION

Understanding the concepts of bioavailability and bioequivalence testing is essential in the drug development process because they create the foundation for regulatory decision making when evaluating formulation changes and lot-to-lot consistency in innovator products. They also serve as the primary components to demonstrate therapeutic equivalence between generic products and the reference innovator product.

Changes in bioavailability can be thought of in terms of changes in exposure to the drug. If these changes are substantial, then they can alter the safety and efficacy profile of the compound in question. The bioavailability of orally administered drugs can be affected by numerous factors. These include food or fed state, differences in drug metabolism, drug–drug interactions, gastrointestinal transit time, and changes in dosage form release characteristics (especially for modified release products).

Bioequivalence is an important consideration in ensuring lot-to-lot consistency, including whenever evaluating changes in a marketed product's formulation, manufacturing process, and dosage strength. Bioequivalence is also critical in regulatory authority decision making when determining whether a generic product is therapeutically equivalent to the original innovator product.

In addition, chemical equivalence, lot-to-lot uniformity of physicochemical characteristics, and stability equivalence are other factors that are important, as they too can affect product quality. In this chapter, bioavailability and bioequivalence topics are emphasized for solid oral dosage forms. However, many of the general concepts can also be applied to other dosage forms, including biologics.

GENERAL CONCEPTS

The terms used in this chapter require careful definition, since, as in any area, some terms have been used in different contexts by different authors.

Bioavailability is a term that indicates measurement of both the rate of drug absorption and the total amount (extent) of drug that reaches the systemic circulation from an administered dosage form. It is specific to the parent or active drug substance as contrasted to metabolites.

Equivalence is more a general and relative term that indicates a comparison of one drug product with another. Equivalence may be defined in several ways:

- **Bioequivalence** indicates that a drug in two or more similar dosage forms reaches the systemic circulation at the same relative rate and extent (i.e., the plasma level

profiles of the drug obtained using the two dosage forms are the same).

- **Chemical equivalence** considers that two or more dosage forms contain the same labeled quantities (within specified limits) of the drug.
- **Clinical equivalence** occurs when the same drug from two or more dosage forms gives identical *in vivo* effects as measured by a pharmacological response or by control of a symptom or disease.
- **Pharmaceutical equivalence** refers to two drug products with the same dosage form and same strength.
- **Therapeutic equivalence** implies that two brands of a drug product are expected to yield the same clinical result. The FDA specifically uses the term “therapeutic equivalence” in the evaluation of multi-source prescription drug products (generic drugs).

Area under the Concentration–Time Curve (AUC) is the integral of the concentration–time curve after administration of a single dose of drug or after achieving a steady state. The calculated area under the serum, blood, or plasma concentration–time curve is reported in amount/volume multiplied by time (e.g., $\mu\text{g/mL} \times \text{h}$ or $\text{g}/100 \text{ mL} \times \text{h}$) and can be considered representative of the amount of drug absorbed. Several variants of AUC exist, including AUC_{0-t} , $\text{AUC}_{0-\infty}$, and $\text{AUC}_{\tau, \text{ss}}$, corresponding to the calculated area from time zero to a truncated time point (e.g., $\text{AUC}_{0-4\text{h}}$), the total area under the curve, and the area when steady state has been achieved.

Peak-height Concentration (C_{max}) is the peak of the blood level–time curve and represents the highest drug concentration achieved after drug administration.

Time of Peak Concentration (T_{max}) is the measured length of time necessary to achieve the maximum concentration (C_{max}) after drug administration.

DISSOLUTION

For a drug to be absorbed, it must first go into solution. In Figure 5-1 the steps in the dissolution and subsequent absorption process of a tablet or capsule dosage form are outlined. Similar profiles could be obtained for any solid or semisolid dosage form, including oral suspensions, parenteral suspensions, and suppositories. The theory and mechanics of drug dissolution rate are described in detail in Chapter 6. The physical characteristics of the drug and the composition of the tablet (dosage form) can have an effect on the rates of disintegration, deaggregation, and dissolution of the drug. As such, these can affect the rate of absorption and resultant blood levels of the drug.

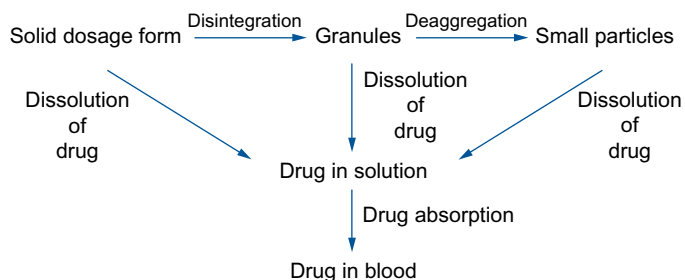


Figure 5-1. Sequence of events involved in the dissolution and absorption of a drug from a solid oral dosage form.

An important aspect of product quality for marketed oral solid dosage forms relates to dissolution testing. The dosage forms actually used by patients will be from lots that have not directly undergone human bioavailability testing. Once adequate product quality has been established by bioavailability testing, subsequent batches manufactured using the same formulation, equipment, and process are likely to be bioequivalent to the original batch tested in humans. This is an important concept in the regulatory control of product quality and is where *in vitro* testing such as assay, content uniformity, tablet hardness, and dissolution are important. Among these several *in vitro* tests, dissolution testing is probably the most important, related to bioavailability. As part of the drug approval process, a dissolution test procedure is established for all oral solid dosage forms. These dissolution tests are incorporated into the United States Pharmacopeia (USP) and apply both to innovator and generic drug products. All marketed batches of these drug products must meet the Abbreviated New Drug Application (ANDA)/New Drug Application (NDA)/USP dissolution tests throughout the shelf-life of the product. Products failing their approved dissolution test and/or a USP dissolution test must be removed from the market.

PROPERTIES OF THE DRUG

The physical characteristics of the drug that can alter bioavailability are discussed in Chapter 39 and consist of the polymorphic crystal form, choice of the salt form, particle size, hydrated or anhydrous form, wettability, and solubility of the drug. These factors should be investigated during product development and should not, therefore, affect the bioavailability of the drug product.

PROPERTIES OF THE DOSAGE FORM

The various components of the solid or semisolid dosage form, other than the active ingredient, are discussed in Chapter 30. Only an overview, for tablet dosage forms, is given here. In addition to the active ingredient, a tablet product usually will contain the following types of inactive ingredients:

- **Glidants** are used to provide a free-flowing powder from the mix of tablet ingredients so that the material will flow when used on a tablet machine.
- **Binders** provide cohesiveness to the tablet. Too little binder will produce tablets that do not maintain their integrity; too much may affect adversely the release (dissolution rate) of the drug from the tablet.
- **Fillers** are used to give the powder bulk so that an acceptably sized tablet is produced. Most commercial tablets weigh from 100 mg to 500 mg, so it is obvious that for many potent drugs the filler constitutes a large portion of the tablet. Binding of drug to the fillers may occur and affect bioavailability.
- **Disintegrants** are used to cause the tablets to disintegrate when exposed to an aqueous environment. Too much will produce tablets that may disintegrate in the bottle because of atmospheric moisture, and too little may be insufficient for disintegration to occur and may therefore

alter the rate and extent of release of the drug from the dosage form.

- **Lubricants** are used to enhance the flow of the powder through the tablet machine and to prevent sticking of the tablet in the die of the tablet machine after the tablet is compressed. Lubricants are usually hydrophobic materials such as stearic acid, magnesium, or calcium stearate. Too little lubricant will not permit satisfactory tablets to be made; too much may produce a tablet with a water-impervious hydrophobic coat, which can inhibit the disintegration of the tablet and dissolution of the drug.

ABSORPTION FACTORS

A significant factor related to drug bioavailability is the fact that many drugs are administered, not as a solution, but as a solid dosage form. Optimal bioavailability might be expected from a solution, since a solid drug must first dissolve to be absorbed, but considerations such as drug stability, unpalatable taste, and the desired duration of action (for controlled-release drug products) may prevent the development of solution-based dosage form.

In the dose titration of any patient the objective is, in conceptual terms, to attain and maintain a blood level that exceeds the minimum effective level required for response but does not exceed the minimum toxic (side-effect) level. This is shown graphically in Figure 5-2. There are several absorption factors that can affect the general shape of this blood-level curve and thus drug response.

DOSE ADMINISTERED

The blood levels will rise and fall in proportion to the dose administered.

AMOUNT OF DRUG ABSORBED

Blood levels achieved are also dependent on the amount of drug absorbed. For example, the effect of having only one-half of the drug absorbed from a dosage form is equivalent to lowering the dose (Figure 5-3).

RATE OF ABSORPTION

If absorption from the dosage form is more rapid than the rate of absorption that gave profile C in Figure 5-4, minimum toxic (side-effect) levels may be exceeded. If absorption from the dosage form is sufficiently slow, minimum effective levels may never be attained, as in profile B in Figure 5-4.

In either instance, the time course and extent of clinical response to the drug may be altered because of changes in dose or rate and extent of absorption.

Both factors, rate and extent of drug absorption, can be affected by the dosage form in which the drug is contained. The

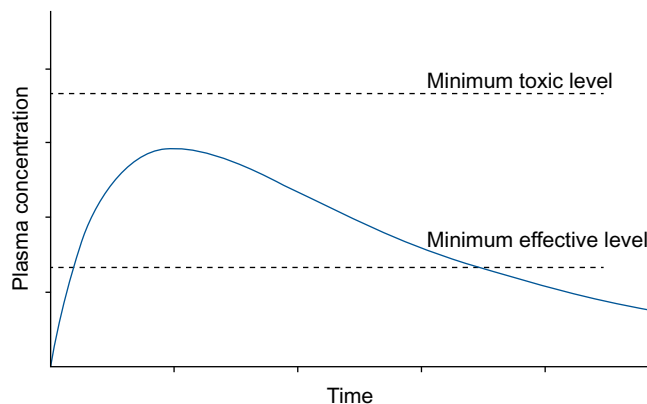


Figure 5-2. Typical plasma-level curve of a drug with effective and toxic (side-effect) profile levels defined.

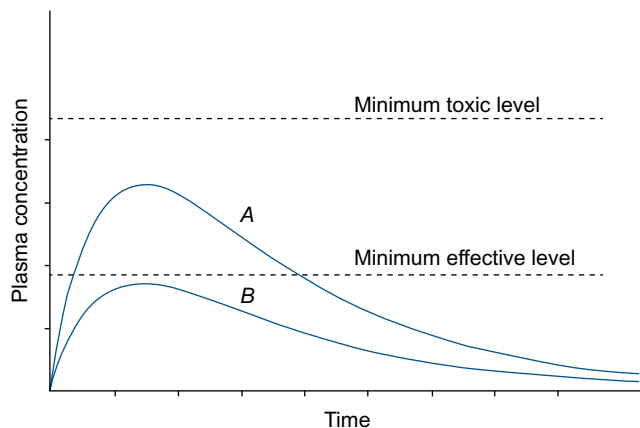


Figure 5-3. Effect of the extent of drug absorption from a dosage form on drug-plasma levels and efficacy. The extent of absorption from dosage form B is 50% of that from dosage form A.

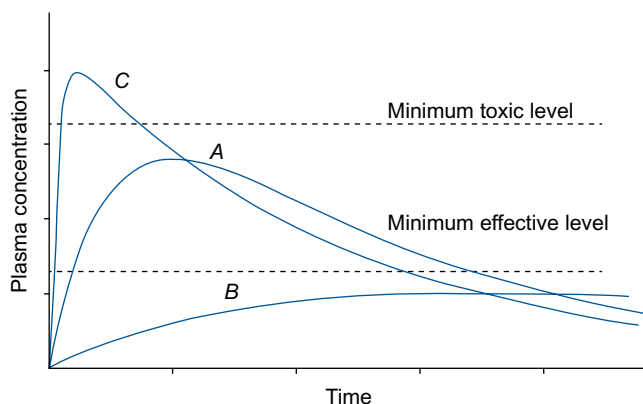


Figure 5-4. Effect of the rate of drug absorption from a dosage form on the plasma-level profile and efficacy. The rates of absorption from dosage forms B and C are 1/10 and 10 times those from dosage form A.

effect of rate of absorption may be intentional, as in controlled-release products, or unintentional, for example, as brought about by a change in the composition and/or method of manufacture of the dosage form.

The choice of the inactive ingredients (excipients) used to prepare a dosage form is up to the individual manufacturer. It is through these changes in composition and manufacturing technique that unintended changes in bioavailability and bioequivalence may occur. Revalidation of bioequivalence may be needed for major changes in the manufacturing process, whereas small changes may not raise significant bioavailability concerns. In situations involving minor changes in the manufacturing process, comparative dissolution testing of the original and reformulated product provides adequate documentation of continued product quality, if the resulting dissolution profiles are similar. These considerations apply to all drug manufacturers, both innovator and generic companies.

BIOEQUIVALENCE TESTING

The awareness of the potential for clinical differences between otherwise chemically equivalent drug products has been brought about by a multiplicity of factors that include, among others:

- better methods for clinical efficacy evaluation
- development of techniques to measure microgram or nanogram quantities of drugs in biological fluids

- improvements in the technology of dosage form formulation and physical testing
- awareness of reported clinical differences from the literature in otherwise similar products
- increased cost of classical clinical evaluation
- objective and quantitative nature of bioavailability tests
- an increase in the number of chemically equivalent products on the market because of patent expirations and the Drug Price Competition and Patent Term Restoration Act of 1984 (Hatch-Waxman Act), which established the generic drug approval procedures that are in place today.

The increase in the number of drugs that are available from multiple sources frequently has placed people involved in the delivery of healthcare in the position of having to select one from among several marketed products. As with any decision, the more data available, the more comfortable one is in arriving at the final decision. The need to make these choices, in light of the potential failure to demonstrate *in vivo* equivalence between products or different batches of the same product, has increased the demand for quantitative data. Bioequivalence testing represents a bridging alternative to clinical testing for efficacy and safety in such cases and is the means by which generic drugs are approved for marketing. In addition, this is also the means by which the quality of all drug products is maintained in situations involving major changes in formulation or manufacturing process.

Requirements for bioequivalence data on drug products should be applied reasonably. The reason for bioequivalence testing should not be overlooked (i.e., it is used as a surrogate, in certain situations, for the clinical evaluation of drug products). In some instances, bioequivalence data cannot reliably be obtained if the bioanalytical methodology is not available. However, in such cases pharmacodynamic data may provide a more sensitive, objective evaluation of a product's therapeutic equivalence and may be explored as an alternative evaluation method in the absence of relevant bioanalytical methodology.

Basic pharmacokinetic evaluation of bioavailability data is not necessary to show bioequivalence of two drug products. Pharmacokinetics has its major utility in the prediction or projection of dosage regimens and/or in providing a better understanding of observed drug reactions or interactions that result from the accumulation of drug in some specific site, tissue, or compartment of the body. The basis for the conclusion that two drug products are bioequivalent must be that the drug concentrations measured in a biological matrix, or alternatively the pharmacological response, for one drug product are essentially the same for the second drug product. The more straightforward decisions in the evaluation of bioequivalence between two drug products are those in which the two products are exactly superimposable (definitely bioequivalent). Those in which the two products differ in their bioequivalence parameters by a large amount, such as 50% or more, are most definitely not bioequivalent. Statistical evaluation of the data is necessary for all situations, particularly for data that exist between these two extremes.

METHODS FOR DETERMINING BIOEQUIVALENCE

Bioequivalence usually involves human testing but sometimes may be demonstrated using an *in vitro* bioequivalence standard, especially when such an *in vitro* test has been correlated with human *in vivo* bioavailability data. In other situations, bioequivalence may be demonstrated through comparative clinical trials or pharmacodynamic studies.

The FDA has categorized (21CFR320.24) various *in vivo* and *in vitro* approaches that may be utilized to establish bioequivalence. These are, in descending order of accuracy, sensitivity and reproducibility,

1. An *in vivo* test in humans in which the active drug substance, as well as active metabolites when appropriate, is measured in plasma.
2. An *in vitro* test that has been correlated with human *in vivo* bioavailability data. This approach is most likely for oral modified release products and is described in detail in FDA Guidance.
3. An *in vivo* test in animals that has been correlated with human bioavailability data.
4. An *in vivo* test in humans, where urinary excretion of the active drug substance, as well as active metabolites when appropriate, is measured.
5. An *in vivo* test in humans in which an appropriate acute pharmacological effect is measured.
6. Well-controlled clinical trials in humans that establish the safety and efficacy of the drug product, for establishing bioavailability. For bioequivalence, comparative clinical trials may be considered. This approach is the least accurate, sensitive, and reproducible approach and should be considered only if other approaches are not feasible.
7. A currently available *in vitro* test, acceptable to FDA, that ensures bioavailability. This approach is intended only when *in vitro* testing is deemed adequate, but no *in vitro*–*in vivo* correlation (IVIVC) has been established. It also can relate to considerations involving the Biopharmaceutics Classification System (BCS).

Most bioequivalence studies involve the direct measurement of the parent drug, as described in item 1 above. Bioequivalence testing in animals is not a recommended approach due to possible differences in metabolism, gastrointestinal physiology, weight, and diet.

MINIMIZING THE NEED FOR BIOEQUIVALENCE STUDIES

If a drug product has been adequately tested and approved for marketing, and if no changes in the manufacturing of the product are made, it is reasonable to assume that all subsequent batches of the product would be expected to be bioequivalent to the original product. If subsequently manufactured batches meet all tests of quality, including the dissolution test, no further human bioequivalence testing is needed.

Depending on the degree of change, bioequivalence may sometimes need to be reconfirmed. Although it is somewhat difficult to categorize such major changes, this issue has been described in a series of FDA guidance documents related to Scale-Up and Post-Approval Changes (SUPAC).

Drug characteristics related to solubility and permeability may allow a reasonable expectation that the drug is unlikely to be subject to significant bioavailability problems. For such drugs, *in vitro* dissolution testing may be adequate, in lieu of *in vivo* testing. These concepts are described in the Biopharmaceutics Classification System (BCS). This classification system provides a scientific framework for classifying drugs based on aqueous solubility and intestinal permeability. In addition, criteria for rapid dissolution are described (not less than 85% dissolved in 30 minutes, using mild agitation and physiological media). The BCS permits waivers of *in vivo* bioequivalence testing for high solubility, high permeability drugs (Class I), which are formulated into immediate release dosage forms having rapid dissolution. The basic tenet behind the BCS is that solutions of drugs are thought to have few bioavailability or bioequivalence issues. Dosage forms containing drugs that are of high solubility and exhibit rapid dissolution behave similarly to a solution. Particularly for such drugs that are, in addition, highly permeable (well absorbed), the likelihood of bioavailability problems is quite small, and consequently, bioequivalence testing for such drugs is thought to be unnecessary. Similarly for oral solutions, bioequivalence testing is not necessary.

EVALUATION OF BIOEQUIVALENCE DATA

The following sections highlight some considerations when evaluating data from bioequivalence studies. The topics discussed are directed specifically toward plasma level evaluations. With minor modifications, the approaches outlined can be used for urinary excretion measurements or for suitable, quantitative, pharmacological response measurements.

Bioequivalence studies are usually conducted in healthy adults under standardized conditions. Most often, single doses of the test and reference product will be evaluated. However, in selected cases, multiple-dose regimens may be used (e.g., when patients are used and they cannot be discontinued from a medication). The goal of the study is to evaluate the *in vivo* performance, as measured by rate and extent of absorption, of the dosage forms under standardized conditions to minimize patient-related and other variability.

The protocol should define the acceptable age and weight range for the subjects to be included in the study, as well as the clinical parameters that will be used to characterize a healthy adult (e.g., physical examination observations, clinical chemistry, and hematological evaluations). The subjects should have been drug-free for at least two weeks prior to testing to eliminate possible drug-induced influences on liver enzyme systems. Normally, the subjects will fast overnight for at least ten hours prior to dosing and will not eat until a standard meal is provided four hours post-dosing. The dosage forms should be given to subjects in a randomized manner, using a suitable crossover design, so that possible daily variations are distributed equally between the dosage forms tested. The protocol should define sample collection times and techniques to collect the biological fluid. The method of sample storage should also be defined.

BIOEQUIVALENCE ASSESSMENT AND DATA EVALUATION

Several parameters are used to provide a general evaluation of the overall rate and extent of absorption of a drug. An analysis of all characteristics is required before one can determine bioequivalence or lack of bioequivalence. It is implicit that the analytical methodology used for analysis of drug in the samples is specific, sensitive, and precise.

In assessing the bioequivalence of drug products, one must quantitate the rate and extent of absorption, which can be determined by evaluating parameters derived from the blood-level concentration–time profile. Three parameters describing a blood-level curve are considered important in evaluating the bioequivalence of two or more formulations of the same drug. These are the peak-height concentration ($C_{\max}$), the time of the peak concentration ($T_{\max}$), and the area under the blood (serum or plasma) concentration–time curve (AUC).

PEAK-HEIGHT CONCENTRATION ($C_{\max}$)

The peak of the blood-level–time curve represents the highest drug concentration achieved after oral administration. It is reported as an amount per volume measurement (e.g., microgram/milliliter ($\mu\text{g/mL}$), unit/mL, or gram/100 mL). The importance of this parameter is illustrated in Figure 5-5, where the blood concentration–time curves of two different formulations of a drug are represented. A line has been drawn across the curve at 4 $\mu\text{g/mL}$. Suppose that the drug is an analgesic, and 4 $\mu\text{g/mL}$ is the minimum effective concentration (MEC) of the drug in blood. If the blood concentration curves in Figure 5-5 represent the blood levels obtained after administration of equal doses of two formulations of the drug and it is known that analgesia would not be produced unless the MEC was achieved or exceeded, it becomes clear that formulation A would be expected to provide pain relief, while formulation B, even though it is well absorbed regarding extent of absorption, might be ineffective in producing analgesia.

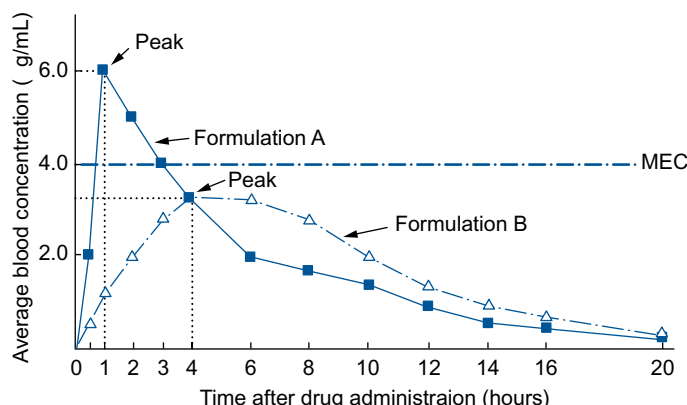


Figure 5-5. Blood concentration–time curves obtained for two different formulations of the same drug, demonstrating relationship of the profiles to the minimum effective concentration (MEC).

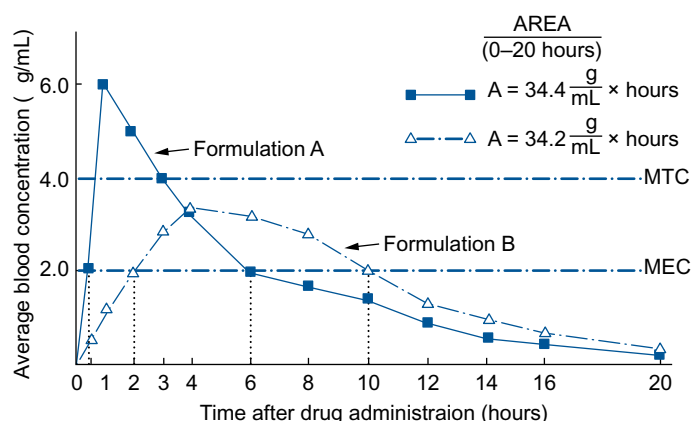


Figure 5-6. Blood concentration–time curves obtained for two different formulations of the same drug, demonstrating relationship of the profiles to the minimum toxic concentration (MTC) and the minimum effective concentration (MEC).

On the other hand, if the two curves represent blood concentrations following equal doses of two different formulations of the same cardiac glycoside, and 4 $\mu\text{g/mL}$ now represents the minimum toxic concentration (MTC) and 2 $\mu\text{g/mL}$ represents the MEC (Figure 5-6), formulation A, although effective, may also present safety concerns, while formulation B produces concentrations well above the MEC but never reaches toxic levels.

Time of Peak Concentration ($T_{\max}$)

The second parameter of importance is the measurement of the length of time necessary to achieve the maximum concentration

after drug administration. This parameter is called the time of peak blood concentration ($T_{\max}$). In Figure 5-5, for formulation A the time necessary to achieve peak blood concentration is 1 h. For formulation B, $T_{\max}$ is 4 h. This parameter is related closely to the rate of absorption of the drug from a formulation and may be used as a simple measure of rate of absorption but is normally not evaluated statistically.

To illustrate the importance of $T_{\max}$, suppose that the two curves in Figure 5-6 now represent two formulations of an analgesic and that in this case the MEC is 2.0 g/mL. Formulation A will achieve the MEC in 30 minutes; formulation B does not achieve that concentration until 2 h. Formulation A would produce analgesia much more rapidly than formulation B and would probably be preferable as an analgesic agent. On the other hand, if one were more interested in the duration of the analgesic effect than on the time of onset, formulation B would present more prolonged activity, maintaining serum concentrations above the MEC for a longer time (8 h) than formulation A (5.5 h).

AREA UNDER THE CONCENTRATION–TIME CURVE (AUC)

The third, and sometimes the most important parameter for evaluation, is the area under the serum, blood, or plasma concentration–time curve (AUC). This area is reported in amount/volume multiplied by time (e.g., $\mu\text{g/mL} \times \text{h}$ or $\text{g}/100 \text{ mL} \times \text{h}$) and can be considered representative of the amount of drug absorbed following administration of a single dose of the drug.

Although several methods exist for calculating the AUC, the trapezoidal rule method is most often used. This method assumes a linear function, $y = bt + a$, and its accuracy increases as the number of appropriate sampling intervals are increased. Table 5-1 and Figure 5-7 describe the process for calculating the AUC using the trapezoidal rule.

Returning to Figure 5-6, the curves, although much different in shape, have approximately the same areas ($A = 34.4 \mu\text{g/mL} \times \text{h}$; $B = 34.2 \mu\text{g/mL} \times \text{h}$), and both formulations can be considered to deliver the same amount of drug to the systemic circulation. Thus, one can see that AUC should not represent the only criterion on which bioequivalence is judged. All the results, as a composite, must be considered in reaching a decision about bioequivalence since no single parameter is adequate to serve this purpose.

The plasma concentration–time curve is the focal point of bioequivalence assessment and is obtained when serial blood samples are analyzed for drug concentration. The concentrations are plotted on the ordinate (y-axis), and the times after drug administration that the samples were obtained are plotted on the abscissa (x-axis).

A drug product is administered orally at time zero, and the plasma drug concentration at this time clearly should be zero. As a product passes through the gastrointestinal (GI) tract, it must undergo a sequence of events depicted in Figure 5-1. As

Table 5-1. Using the Trapezoidal Rule to Calculate Area Under the Concentration–Time Curve.

$AUC_{(0-E)}$ is used for bioequivalence analyses when the $AUC_{(0-t)}$ makes up $\geq 80\%$ of the $AUC_{(0-E)}$. $AUC_{(0-t)}$ is used when the $AUC_{(0-t)}$ makes up $< 80\%$ of the $AUC_{(0-E)}$. When drugs with long half-lives are evaluated, $AUC_{(0-t)}$ may sometimes be used with a truncated time point.

Area under the concentration–time curve from time zero to time t (AUC_{0-t})

1. Plot the concentration–time data for each subject
2. Divide the curve into trapezoids by drawing vertical lines from each datum point to the x-axis. Calculate the area of the trapezoids using the following formula:
3. $AUC_{(t_2-t_1)} = [(C_2 + C_1)(t_2 - t_1)] / 2$
 $AUC_{(0-t)}$ is then calculated by summing the individual areas to the time of the last concentration:
4. $AUC_{(0-t)} = AUC_{(t_2-t_1)} + AUC_{(t_3-t_2)} + \dots + AUC_{(t_n-(t_{n-1}))}$

Area under the concentration–time curve from time zero to infinity (AUC_{0-E})

5. To calculate $AUC_{(0-E)}$, the tail region of the curve must be added to $AUC_{(0-t)}$: $AUC_{(0-E)} = AUC_{(0-t)} + \text{AUC "tail"}$
6. $\text{AUC "tail"} = C_t / \lambda_z$, where: C_t is the last detectable concentration, and λ_z is the terminal elimination rate constant (see Figure 5-9).

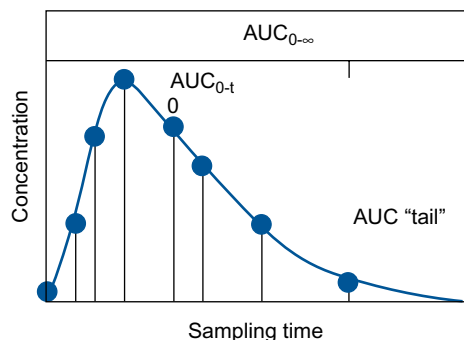


Figure 5-7. Graphical depiction of use of the trapezoidal rule for calculating the area under the concentration-time curve.

the drug is absorbed, increasing concentrations of the drug are observed in successive samples until the maximum concentration is achieved. This point of maximum concentration (C_{max}) is the peak of the concentration-time curve. If a simple model describes the pharmacokinetics of the drug tested, the peak concentration represents approximately the point in time when absorption and elimination of the drug have equalized.

The section of the curve to the left of the peak represents the absorption phase (or absorption and distribution), during which absorption predominates over elimination. The section of the curve to the right of the peak is called the elimination phase, during which elimination predominates over absorption. It should be understood that elimination begins as soon as the drug appears in the bloodstream and continues until the drug has been completely eliminated. Elimination is classically the log-linear portion of the curve. Absorption continues for some period of time into the elimination phase for as long as there is drug (in gradually decreasing amounts) available for absorption in the GI tract.

One must recognize that elimination of the drug includes all processes of elimination of the drug, involving urinary excretion, as well as metabolism by various tissues and organs. The efficiency of metabolism and urinary excretion will determine the shape of the elimination phase of the curve.

Bioequivalence studies normally are performed in healthy, adult volunteers under rigid conditions of fasting and activity because the objective is to obtain quantitative information on the influence of pharmaceutical formulation variables on the drug product's absorption. Drug blood-level profiles, therefore, allow quantification of the rate and extent of drug absorption and are critical in establishing the comparative efficiency of two drug products in delivering the drug to the systemic circulation.

Suggestions that bioequivalence studies should be conducted in a disease-state population are not tenable if the object of the study is to assess drug formulations, unless safety considerations prohibit administration of the drug to healthy volunteers. If, on the other hand, the purpose is to determine the effect of disease on the efficiency of absorption of the drug product, then one must use the disease-state population. The reasoning is obvious. To ensure that any differences observed in the drug blood-level profiles are attributable to formulation factors, as much as possible, one must hold all other variables constant (i.e., food, activity, and state of disease).

One need not be limited to drug blood-level profiles, but in a similar manner, may obtain cumulative urinary drug amount-time profiles. Drug concentration is determined in the urine at specified time intervals, and the amount excreted per interval is calculated by multiplying the concentration by the volume of urine obtained in that interval. The amounts per interval then are combined, and ultimately the total amount excreted in the urine is obtained. This value is analogous to the area under the blood concentration-time curve. However, one limitation to this method is that rate cannot be readily determined. A typical cumulative urinary drug amount-time profile for several nitrofurantoin products is presented in Figure 5-8.

CRITERIA FOR BIOEQUIVALENCE

Under the Drug Price Competition and Patent Term Restoration Act of 1984 (Hatch-Waxman), manufacturers seeking approval to market a generic drug must submit data demonstrating that the drug product is bioequivalent to the pioneer (innovator) drug product. A major premise underlying the 1984 law is that bioequivalent drug products are therapeutically equivalent and, therefore, interchangeable.

The standard bioequivalence study is conducted in a cross-over fashion in a small number of volunteers, usually with 24 to 36 healthy adults. The number of subjects appropriate for a bioequivalence study can be determined on the basis of previous knowledge of the drug's variability. In general, the number of subjects should be sufficient to detect 20% differences in the measured parameters with 80% certainty. Single doses of the test and reference drugs are administered, and blood or plasma levels of the drug are measured over time. Characteristics of these concentration-time curves, such as the area under the curve (AUC) and the peak blood or plasma concentration (C_{max}), are examined by statistical procedures.

Bioequivalence of different formulations of the same drug substance involves equivalence with respect to the rate and extent of drug absorption. Two formulations whose rate and extent of absorption differ by $\pm 20\%$ or less are generally considered bioequivalent. The use of the $\pm 20\%$ criterion is based on a

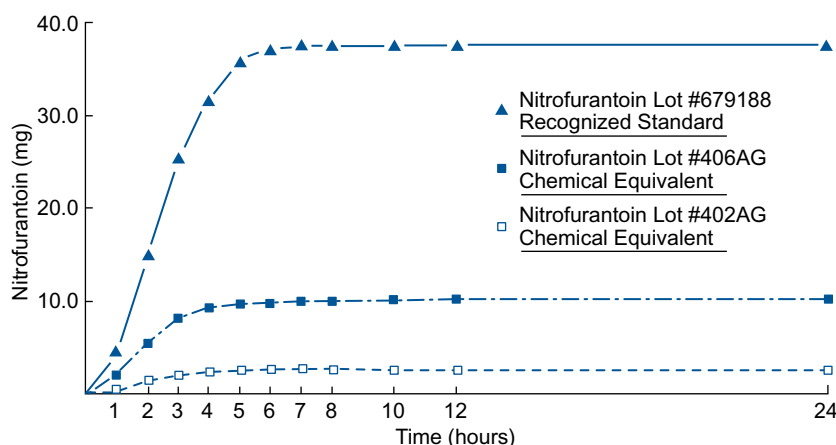


Figure 5-8. Average cumulative amounts of nitrofurantoin excreted from three lots of two commercially available products after a single oral dose of 100 mg of nitrofurantoin.

medical decision that for most drugs, a $\pm 20\%$ difference in the concentration of the active ingredient in blood will not be clinically significant (Figure 5-9).

To verify, for a particular pharmacokinetic parameter, that the $\pm 20\%$ criterion is satisfied, two one-sided statistical tests are carried out using the log-transformed data from the bioequivalence study. In order to interpret the statistical results, the log-transformed data must first be back-transformed. When the log-transformed data are back-transformed, the $\pm 20\%$ now becomes $-20\%/+25\%$. One test is used to verify that the lower bound of the 90% confidence interval of the average response for the generic product is no more than 20% below that of the innovator product. The other test is used to verify that the upper bound of the 90% confidence interval of the average response for the generic product is no more than 25% above that for the innovator product. The current practice is to carry out each of the two one-sided tests at the 0.05 level of significance.

Computationally, the two one-sided tests are carried out by computing a 90% confidence interval. For approval of Abbreviated New Drug Applications (ANDA), in most cases, the generic manufacturer must show that a 90% confidence interval for the ratio of the mean response (usually AUC and C_{max}) of its product to that of the innovator is within the limits of 0.80 and 1.25 after the log-transformed data have been back-transformed. If the true average response of the generic product in the population is near 20% below or 25% above the innovator average, one or both of the confidence limits is likely to fall outside the acceptable range, and the product will fail the bioequivalence test. Thus, an approved product is likely to differ from the innovator by far less than this quantity. These same criteria are applied to other bioequivalence situations, such as post-approval changes in innovator or generic products.

The current practice of carrying out two one-sided tests at the 0.05 level of significance ensures that if the two products truly differ by as much as or more than is allowed by the equivalence criteria, there is no more than a 5% chance that they will be approved as equivalent. This reflects the fact that the primary concern from the regulatory point of view is the protection of the patient against a conclusion of bioequivalence if this does not hold true. The results of a bioequivalence study usually must be acceptable for more than one pharmacokinetic parameter. As such, a generic product that truly differs by $\pm 20\%$ or more from the innovator product with respect to one or more pharmacokinetic parameters would have less than a 5% chance of being approved. Different statistical criteria may be used when bioequivalence is demonstrated through comparative clinical trials, pharmacodynamic studies, or comparative *in vitro* methodology.

Using the two one-sided test procedures, when two drug products differ by more than 12–13% in means, they are unlikely to

pass the bioequivalence confidence interval criteria of 80–125%. A study of more than 200 approved generic drugs indicated that a mean bioavailability difference of only 3.5% existed. Although somewhat larger differences might meet the bioequivalence criteria, the reality is that, for generic drug products approved by FDA, observed differences have been quite small.

FED BIOEQUIVALENCE STUDIES

Food has been shown to alter the bioavailability of some drugs, and this alteration can have a negative impact on the interpretation of bioequivalence results between test and reference products. As a result, bioequivalence studies are usually conducted under fasting conditions. However, in some instances a fasting study may not be reasonable for a particular drug due to safety considerations or perhaps because of the drug's intended clinical indication. In these situations a fed bioequivalence study is sometimes acceptable. A fed bioequivalence study is similar to the standard bioequivalence study except that, following an overnight fast, the test or reference products are administered 30 minutes after the start of a standardized meal. The FDA currently recommends a high-fat, high-calorie meal as described in an FDA Guidance. The composition of this meal is described in Table 5-2. The same statistical criteria as used for the standard bioequivalence study are observed for the resultant fed bioequivalence study data.

AVERAGE BIOEQUIVALENCE

The standard *in vivo* bioequivalence study design is based on administration of the test and reference products on separate occasions to healthy subjects, either in single or multiple doses, with random assignment to the two possible sequences of drug product administration. Samples of plasma or blood are analyzed for drug and/or metabolite(s) concentrations, and pharmacokinetic parameters are obtained from the resulting concentration–time curves. Parameters are analyzed statistically to determine if the test and reference products yield comparable values. Statistical analysis for pharmacokinetic parameters, such as area under the curve (AUC) and peak concentration (C_{max}), is based on the two one-sided tests procedure, which determines whether the average values for pharmacokinetic parameters measured after administration of the test and reference products are comparable (i.e., average bioequivalence). This procedure involves the calculation of a 90% confidence interval for the ratio of the averages of the test and reference products. To establish bioequivalence, the calculated confidence interval must fall within a bioequivalence limit, usually 80–125% for the ratio of the product averages. In addition to this general approach for determining bioequivalence, a FDA Guidance provides specific recommendations for (1) logarithmic transformation of pharmacokinetic data, (2) methods to evaluate sequence effects, and (3) methods to evaluate outlier data.

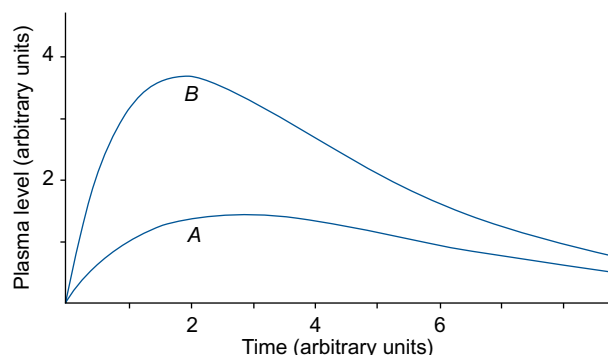


Figure 5-9. Computer simulation of the plasma-level curves for two dosage forms of the same drug, assuming that the rate and extent of drug absorption for dosage form A were 50% and 50%, respectively, of those for dosage form B.

Table 5-2. The FDA Standardized High-Fat Test Meal Composition.

Meal Composition	Energy (kcal)
Protein	150
Carbohydrate	250
Fat	500–600

The example test meal would be two eggs fried in butter, two strips of bacon, two slices of toast with butter, four ounces of hash brown potatoes, and eight ounces of whole milk. Substitutions in this test meal can be made as long as the meal provides a similar amount of calories from protein, carbohydrate, and fat and has a comparable meal volume and viscosity.

POPULATION AND INDIVIDUAL BIOEQUIVALENCE

Statistically, the average bioequivalence approach focuses on the comparison of population averages of a bioavailability metric of interest and not on the variability of the metric for the test and reference products. In addition, average bioequivalence cannot describe a subject-by-formulation interaction, that is, the variation that may be present among individuals in the average test and reference difference. In contrast, population and individual bioequivalence approaches include comparisons of both averages and variability of the study metric. The population bioequivalence approach assesses the total variability of the metric in the population. The individual bioequivalence approach assesses the within-subject variability as well as the subject-by-formulation interaction. However, due to statistical and study design issues with population and individual bioequivalence, respectively, the FDA has deferred recommending these analysis methods.

STUDY DESIGN

AVERAGE OR POPULATION BIOEQUIVALENCE

A conventional, non-replicated crossover design, such as the standard two-formulation, two-period, two-sequence crossover design, may be used to generate data for assessment of population bioequivalence. Replicated-crossover designs or parallel designs also may be used.

INDIVIDUAL BIOEQUIVALENCE

Three important parameters, the within-subject variability for the test metric, within-subject variability for the reference metric and subject-by-formulation interaction are integral components of the individual bioequivalence criterion. A replicated-crossover design of the bioequivalence study should be used to estimate these parameters.

Further information related to current FDA recommendations regarding the design and analysis of bioequivalence studies is available on the internet at <http://www.fda.gov/cder/>, under Regulatory Guidance Documents.

PITFALLS ASSOCIATED WITH CROSS-STUDY COMPARISONS

This is a situation in which the blood concentration–time curve of a drug product in one study is compared with the blood concentration–time curve of that drug product in another study. There are several reasons why such cross-study comparisons are not recommended and may lead to false conclusions. However, if no other data are available, and if important comparisons must be made, cross-study comparison may be informative if one keeps in mind the possible limitations. The following examples, used to illustrate these three points, are taken from actual bioavailability data.

Subject Population

In Figure 5-10, a research lot of potassium phenoxymethyl penicillin was compared with the appropriate reference standard for that product. The research lot drug was found to be bioequivalent, with average peak-serum concentrations differing by 8% and the area differing by only 9%. In another study conducted with a full-manufacture lot of the test product, the same lot of the reference standard potassium phenoxymethyl penicillin was used.

The results of this study are shown in Figure 5-11. Again, the two products were found to be bioequivalent, as the peak and area parameters differed by less than 5%. In these two studies, identical test conditions were used, and the same analytical procedure and laboratory was employed. However, if one compares the plasma levels for the reference standard lot found in Figure 5-10 with the levels for the same lot of tablets in the study in Figure 5-11, sizable differences in blood levels are found, as shown in Figure 5-12.

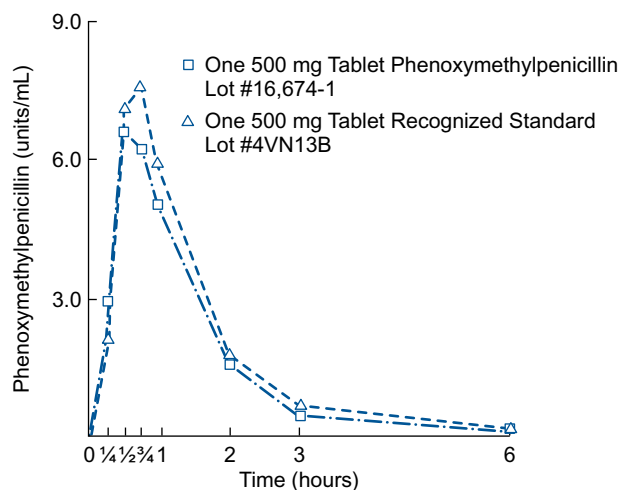


Figure 5-10. Average serum concentration of phenoxymethyl penicillin, following oral administration of 500 mg given as one tablet of recognized standard (Δ) or of test product, research lot ($\square$).

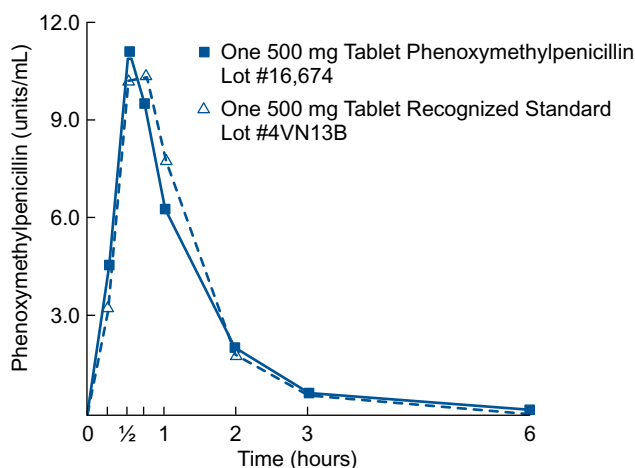


Figure 5-11. Average serum concentration of phenoxymethylpenicillin following oral administration of 500 mg given as one tablet of recognized standard (Δ) or of test product full mfg lot ($\blacksquare$).

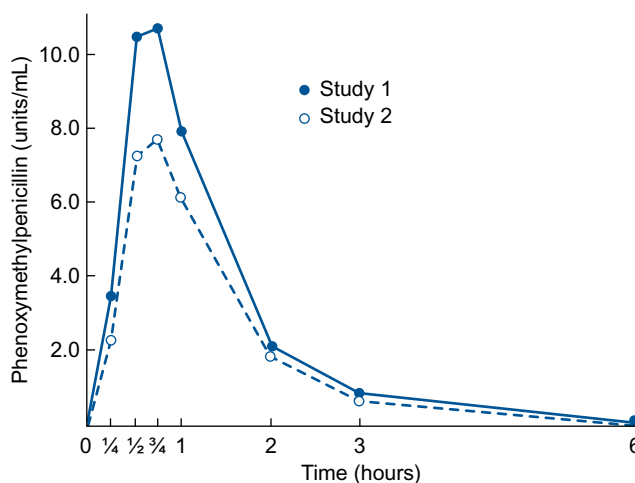


Figure 5-12. Average serum concentration of phenoxymethylpenicillin following a single oral 500 mg dose of recognized standard, in two different subject populations.

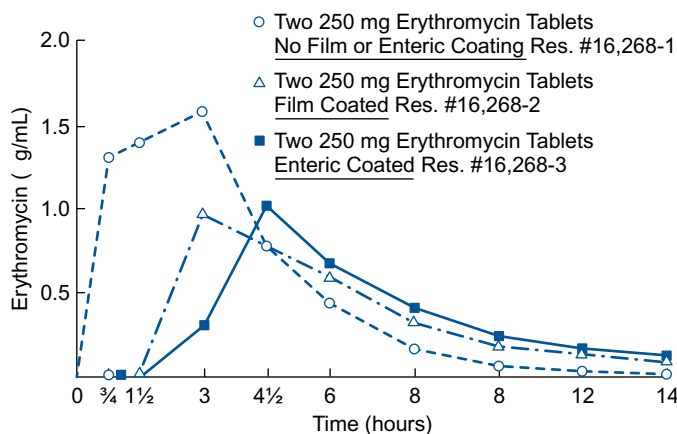


Figure 5-13. Average serum erythromycin concentration following administration in 500 mg doses as three different tablet dosage forms. The results were obtained from 21 healthy adult subjects following an overnight fast of 12 hours before, and 2 hours after, drug administration.

The average peak serum levels for this lot of tablets were found to be 8.5 and 12.5 units/mL in the two respective studies, a difference of approximately 31%. Likewise, the average AUC was found to differ by approximately 21%. Such apparent differences are solely the result of cross-study comparisons and are not due to differences in actual bioavailability.

The same lot of reference standard tablets was used in both studies. Hence, the difference must be due to the experimental variables that occur normally from study to study. The major difference between the two studies was the subject population involved. In the first study, healthy adult male prison volunteers were used, whereas in the second study there were 17 females and 7 males in a hospital clinic, also described as normal, healthy volunteers. An appreciable difference in sex distribution was obvious when comparing these studies. Adjustments for body weight and surface area alone did not correct for the apparent discrepancies in peak concentration or blood-level AUC. It is difficult to determine the exact factors that caused the observed differences. This example should serve as a note of caution in comparing bioavailability values of peak concentration and area under the curve from different studies.

Study Conditions

Parameters such as the food or fluid intake of the subject before, during, and after drug administration can have dramatic effects on the absorption of certain drugs. Figure 5-13 shows the results of a three-way crossover test in which the subjects were fasted 12 hours overnight and 2 hours after drug administration

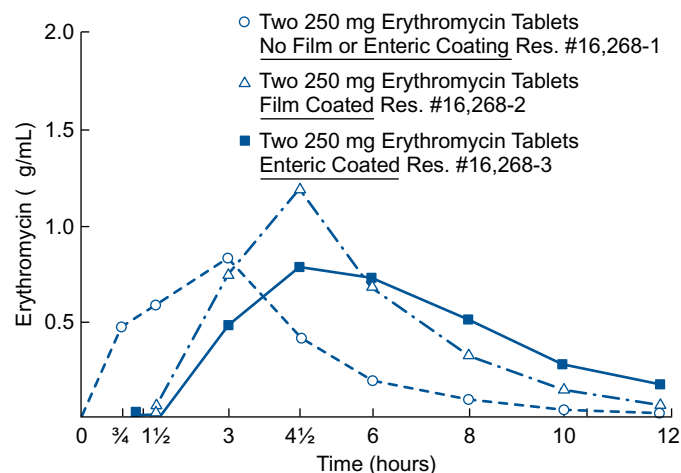


Figure 5-14. Average serum erythromycin concentration following administration in 500 mg doses as three different tablet dosage forms. The results were obtained from 12 healthy adult subjects with only a 2 hour fast before drug administration.

of an uncoated tablet, a film-coated tablet, or an enteric-coated tablet of erythromycin.

The results of this study suggest that the uncoated tablet is superior to both the film-coated and enteric-coated tablets in terms of blood level performance. These results also suggest that neither film coating nor enteric coating is necessary for optimal blood-level performance. Figure 5-14 shows results with the same tablets when the study conditions were changed to only a 2-hour pre-administration fast with a 2-hour post-administration fast. In this case, the blood levels of the uncoated tablet were depressed markedly, while the film-coated and enteric-coated tablets showed relatively little difference in blood levels.

From this second study, it might be concluded that film coating appears to impart the same degree of acid stability as an enteric coating. This might be acceptable if only one dose of the antibiotics was required. However, Figure 5-15 shows the results of a multiple-dose study in which the enteric-coated tablet and the film-coated tablet were administered four times a day, immediately after meals. The results show that the film coating does not impart the degree of acid stability that the enteric coating does when the tablets are administered immediately after food in a typical clinical situation.

Assay Methodology

Depending on the drug under study, there may be more than one assay method available. For example, some steroids can be assayed by a radioimmunoassay, competitive protein-binding,

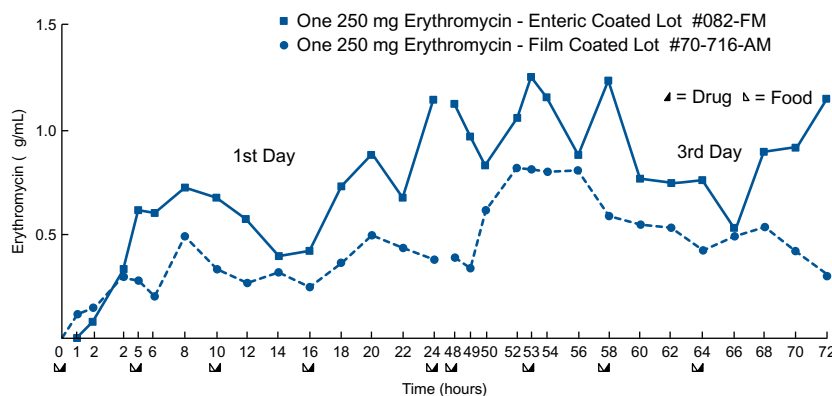


Figure 5-15. Average serum erythromycin concentration-time profiles from drug administered in two different tablet dosage forms. The results were obtained from 24 healthy adult subjects, following administration of 250 mg, four times a day, with meals and at bedtime.

gas-liquid chromatography, or indirectly by a 17-hydroxycorticosteroid assay.

Figures 5-16 and 5-17 show the results of a comparison of prednisone tablets using a competitive protein-binding method and a radioimmunoassay, respectively. The serum concentration-time curves resulting from each method lead to the same conclusion, that the products are bioequivalent. However, Figure 5-18 shows a comparison of the absolute values obtained by the two assay methods with the same product.

Obviously, the wrong conclusion would have been reached if one product had been assayed by one method and the other product by the other method and the results had been compared. Even in cases in which only one assay method is employed, there are numerous modifications with respect to technique among laboratories that make direct comparisons difficult.

The backbone of any bioavailability study involving plasma (or urine) levels of drug, in addition to good study design and subject controls, is the analytical methodology used to determine the levels of a drug. The precision and reliability of the method employed in a given study have been established to a sufficient degree to make the results of the study internally consistent. As demonstrated, major problems arise when, without careful evaluation of the analytical methodology employed, one attempts to compare the data of studies from different laboratories. Even with similar analytical methodology performed by the same laboratory, it would be unreasonable to expect

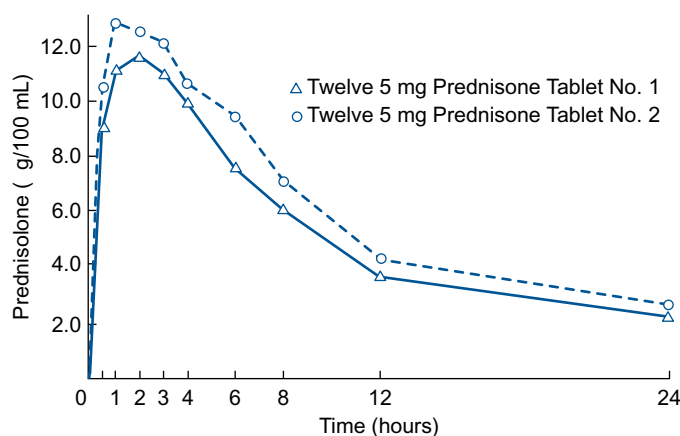


Figure 5-16. Average plasma prednisolone levels following 60 mg of prednisone administered to 24 normal adults as a single oral dose of twelve 5 mg prednisone tablets from two different manufacturers. Plasma levels were determined by a competitive protein-binding assay.

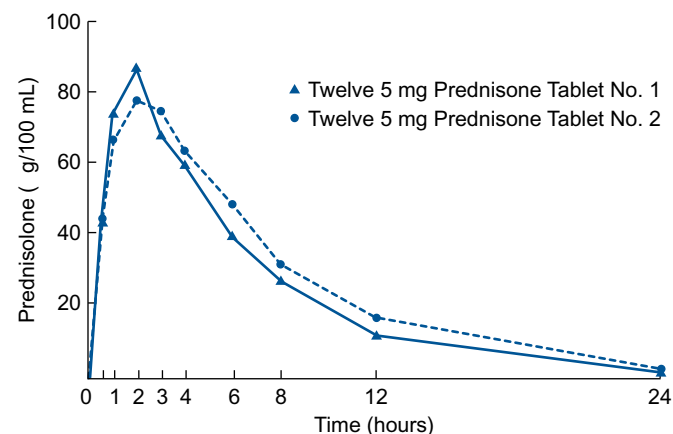


Figure 5-17. Average plasma prednisolone levels, following 60 mg of prednisone administered to 24 normal adults as a single oral dose of twelve 5 mg prednisone tablets from two different manufacturers. Plasma levels were determined by a radioimmunoassay procedure.

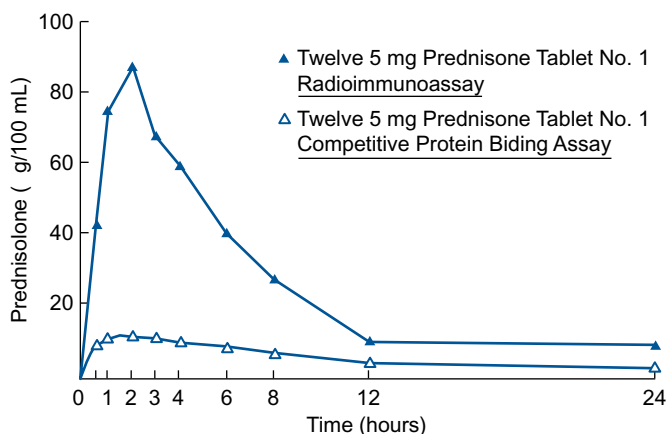


Figure 5-18. Average plasma prednisolone profiles from drug administered as a single 60-mg dose to 24 normal adults. Plasma levels were determined by both a competitive protein-binding assay and a radioimmunoassay.

agreement, using the same dosage form, closer than 20% to 25% for plasma levels from one study to the next.

Under the best conditions, cross-study comparisons are relatively insensitive, and at worst they can be misleading. Cross-study comparisons certainly cannot be used to make decisions or estimate differences in drug products with the generally acceptable sensitivity of difference detection of 20% or less.

THERAPEUTIC EQUIVALENCE EVALUATIONS

The FDA publication *Approved Drug Products with Therapeutic Equivalence Evaluations* identifies drug products approved on the basis of safety and effectiveness. In addition, this list contains therapeutic equivalence evaluations for approved multi-source prescription drug products. These evaluations have been prepared to serve as public information and advice to state health agencies, physicians, and pharmacists to promote public education in the area of drug product selection and to foster containment of healthcare costs.

To help contain drug costs, virtually every state has adopted laws and/or regulations that encourage the substitution of drug products. These state laws generally require either that substitution be limited to drugs on a specific list (the positive formulary approach) or that substitution be permitted for all drugs except those prohibited by a particular list (the negative formulary approach). Because of the number of requests for FDA assistance in preparing both positive and negative formularies, it became apparent that the FDA could not serve the needs of each state on an individual basis. The agency also recognized that providing a single list based on common criteria would be preferable to evaluating drug products on the basis of differing definitions and criteria in various state laws. The therapeutic equivalence evaluations in this publication reflect FDA's application of specific criteria to the approved multi-source prescription drug products.

FDA classifies as therapeutically equivalent those products that meet the following general criteria:

1. They are approved as safe and effective.
2. They are pharmaceutical equivalents in that they (A) contain identical amounts of the same active drug ingredient in the same dosage form and route of administration and (B) meet compendial or other applicable standards of strength, quality, purity, and identity.
3. They are bioequivalent in that (A) they do not present a known or potential bioequivalence problem, and they meet an acceptable *in vitro* standard, or (B) if they do present such a known or potential problem, they are shown to meet an appropriate bioequivalence standard.

4. They are adequately labeled.
5. They are manufactured in compliance with Current Good Manufacturing Practice regulations.

This concept of therapeutic equivalence applies only to drug products containing the same active ingredient(s) and does not encompass a comparison of different therapeutic agents used for the same condition. The FDA considers drug products to be therapeutically equivalent if they meet the criteria outlined above, even though they may differ in certain other characteristics such as shape, scoring configuration, release mechanisms, packaging, excipients, expiration date/time, and minor aspects of labeling (e.g., the presence of specific pharmacokinetic information). The FDA believes that products classified as therapeutically equivalent can be substituted with the full expectation that the substituted product will produce the same clinical effect and safety profile as the prescribed product.

As described in this chapter, the concepts of bioavailability and bioequivalence testing are essential in the drug development process by creating the foundation for regulatory decision for both innovator and generic drug products.

BIBLIOGRAPHY

- Abdou HM. *Dissolution, Bioavailability and Bioequivalence*. Easton, PA: Mack Publishing, 1989.
- Amidon GL, Robinson JR, Williams RL. *Scientific Foundations for Regulating Drug Product Quality*. Alexandria, VA: AAPS Press, 1997.
- Chow S-C. *Design and Analysis of Bioavailability and Bioequivalence Studies*. New York: Dekker, 1992.
- Marston SA, *et al.* Evaluation of direct curve comparison metrics applied to pharmaceutical profiles and relative bioavailability and bioequivalence. *Pharm Res* 1997;14: 1363.
- Schuurmann DJ. A comparison of the two one-sided tests procedure and the power approach for assessing the bioequivalence of average bioavailability. *J Pharmacokinet Biopharm* 1987;15: 657.
- Shah VP, Maibach HI. *Topical Drug Bioavailability, Bioequivalence and Penetration*. New York, NY: Plenum, 1993.
- Shargel L, Yu AB. *Applied Biopharmaceutics and Pharmacokinetics*. Norwalk, CT: Appleton & Lange, 1993.
- Welling PG, *et al.* *Pharmaceutical Bioequivalence*. New York, NY: Dekker, 1991.
- Guidance for Industry - Extended Release Oral Dosage Forms: Development, Evaluation, and Application of In vitro/In vivo Correlations. <http://www.fda.gov/cder/guidance/1306fml.pdf>.
- Guidance for Industry—Immediate Release Solid Oral Dosage Forms—Scale-Up and Postapproval Changes: Chemistry, Manufacturing, and Controls, *in vitro* Dissolution Testing, and *in vivo* Bioequivalence Documentation. 1995. <http://www.fda.gov/cder/guidance/cmc5.pdf>.
- Guidance for Industry—Questions and Answers about SUPACIR. 1997. <http://www.fda.gov/cder/guidance/qaletter.htm>.
- Guidance for Industry—SUPACIR/MR: Immediate Release and Modified Release Solid Oral Dosage Forms Manufacturing Equipment Addendum. 1999. <http://www.fda.gov/cder/guidance/1721fml.pdf>.
- Guidance for Industry—SUPACMR: Modified Release Solid Oral Dosage Forms Scale-Up and Postapproval Changes: Chemistry, Manufacturing, and Controls; *in vitro* Dissolution Testing and *in vivo* Bioequivalence Documentation. 1997. <http://www.fda.gov/cder/guidance/1214fml.pdf>.
- Guidance for Industry—Waiver of *in vivo* Bioavailability and Bioequivalence Studies for Immediate-Release Solid Oral Dosage Forms Based on a Biopharmaceutics Classification System. <http://www.fda.gov/cder/guidance/3618fml.pdf>.
- Guidance for Industry—Food-Effect Bioavailability and Fed Bioequivalence Studies. 2002. <http://www.fda.gov/cder/guidance/5194fml.doc>.
- NTI Letter. Therapeutic Equivalence of Generic Drugs Response to National Association of Boards of Pharmacy. 1997. <http://www.fda.gov/cder/news/ntiletter.htm>.

Chapter 6

Dissolution

Vijai Kumar, MS, MBA and Praveen Hiremath, PhD

DEFINITION OF DISSOLUTION AND THEORETICAL CONCEPTS FOR THE RELEASE OF THE DRUG FROM DOSAGE FORMS 64

MATHEMATICS OF DISSOLUTION 64

DISSOLUTION PROFILE COMPARISONS 78

APPLICATION OF QUALITY BY DESIGN (QBD) PRINCIPLES IN DISSOLUTION TESTING 79

Dissolution is the process by which a solid enters into solution. The earliest reference to dissolution is the 1897 article by Noyes and Whitney, titled as “The Rate of Solution of Solid Substances in Their Own Solution.” The authors suggested that the rate of dissolution of solid substances is determined by the rate of diffusion of a very thin layer of saturated solution that forms instantaneously around the solid particle. They developed the mathematical relationship that correlates the dissolution rate to the solubility gradient of the solid. Their equation is still the basic formula upon which most of the modern mathematical treatments of the dissolution phenomenon revolve.

Interestingly, the work of Noyes and Whitney, together with the studies that followed in the early part of the 20th century, was primarily based on the physicochemical aspects of dissolution applied to chemical substances. The most prominent part of these investigations that deserve recognition are those of Nernst and Brunner, in 1904, for their application of Fick’s law of diffusion to the Noyes-Whitney equation, and those of Hixson and Crowell, in 1931, for their development of famous “Cube Root Law” of dissolution.¹ Hixson-Crowell equation is subsequently provided:

$$Q_0^{1/3} - Q_t^{1/3} = K_{HC}t \quad (1)$$

where Q_t is the amount of drug released in time t , Q_0 is the initial amount of drug in the dosage form/product, and K_{HC} is the rate constant for Hixson-Crowell cube root equation, which describes the surface area-volume relationship. This equation applies to products, such as powders and tablets, where the dissolution occurs in the planes that are parallel to the surface of the dosage form.

Two more alternative explanations were available, as reviewed by William Higuchi (1961),² by the 1950’s. The interfacial barrier model considered that interfacial transport, rather than diffusion through the film, is the limiting step, due to a high activation energy level for the former, first proposed by Wilderman in 1909. Another model was Danckwerts’ model, which appeared in 1951. According to this, constantly renewed macroscopic packets of solvent reach the solid surface and absorb molecules of solute, delivering them to the solution.

By the middle of the 20th century, emphasis started to shift to the examination of the effects of dissolution behavior of drugs on the biological activity of pharmaceutical dosage forms. One of the earliest studies, with this purpose in mind, was conducted by J Edwards, in 1951, on aspirin tablets. He reported, “because of its poor solubility, the analgesic action of aspirin tablets would be controlled by its dissolution rate within the stomach and the intestine.” No *in vivo* studies, however, were conducted by Edwards to support his postulate.

About eight years later, Shenoy and colleagues proved the validity of Edward’s suggestion of the *in vitro/in vivo* correlation by demonstrating a direct relationship between the bioavailability of amphetamine from sustained-release tablets and its *in vitro* dissolution rate. Other studies, especially those reported by Nelson, Levy, and others, confirmed, beyond doubt, the significant effect of the dissolution behavior of drugs on their pharmacological activities. Nelson, in 1957, was the first to explicitly relate the blood levels of orally administered

theophylline salts to their *in vitro* dissolution rates. Due to the importance of these findings, dissolution testing began to emerge as a dominant topic within both the pharmaceutical academia and the drug industry.

During this 20-year period, 1950–1970, a number of studies conducted, especially in the United States, confirmed the significance of dissolution- bioavailability relationship in the pharmaceutical product development. As a result, the basket-stirred-flask test (USP apparatus 1) was adopted as an official dissolution test in 6 monographs of the *United States Pharmacopeia (USP)* and *National Formulary (NF)*, in 1970. Also, due to the sustained interest in the subjects of dissolution and gastrointestinal absorption, an explosion in the number of monographs of the dissolution requirements in subsequent USP/NF editions was noted. Notable developments during this evolution are the adoption of the paddle method (USP apparatus 2) in 1978, the publication of a General Chapter on Drug Release in USP 21 (1985), the presence of 23 monographs for modified-release dosage forms in USP 22-NF 18 (1990), the adoption of the reciprocating cylinder (USP apparatus 3) for extended-release products in 1991, and the adoption of the flow-through cell in (USP apparatus 4) for extended-release products in 1995.

In the late 1960s, dissolution testing became a mandatory requirement for several dosage forms. The role of dissolution in the absorption of drug products, however, is still far from being understood completely. Although considerable efforts were made to establish *in vitro/in vivo* correlations between release of drug from the formulation and drug absorption, the limited knowledge of the complex composition and hydrodynamics of the gastrointestinal fluids remains a real barrier. In spite of the reported success of several *in vitro/in vivo* correlation studies, dissolution cannot be relied upon as a predictor of therapeutic efficiency. Rather, it is a qualitative tool that can provide valuable information about the biological availability of a drug, as well as batch-to-batch consistency. Another area of difficulty is the accuracy and precision of the testing procedure, which is dependent, to a large extent, on the strict observance of so many subtle parameters and detailed operational controls.

In spite of these shortcomings, dissolution is considered, today, as one of the most important quality control procedures performed on pharmaceutical dosage forms, and dissolution studies have become an essential part of drug applications to regulatory bodies worldwide. Whether or not it has been correlated with biological effectiveness, the standard dissolution test is a simple and inexpensive indicator of a product’s physical consistency. If one batch differs from the other in its dissolution characteristics or if the dissolution profiles of the production batches show a consistent trend upwards or downwards, it sounds a sure warning that some factor in the raw material, formulation, or process is out of control.¹ Additionally, dissolution data seems to be a useful tool in the early stages of drug development and molecular manipulation. In the early stages of research, steps may be taken to optimize characteristics that influence subsequent data concerning biological availability. Based on simple dissolution testing, selection of a proper salt for a new drug can be done at an early drug development stage.

DEFINITION OF DISSOLUTION AND THEORETICAL CONCEPTS FOR THE RELEASE OF THE DRUG FROM DOSAGE FORMS

Dissolution is defined as the process by which solid substances enters in solvent to yield a solution. Stated simply, dissolution is the process by which a solid substance dissolves. Fundamentally, it is controlled by the affinity between the solid substance and the solvent.” The physical characteristics of the dosage form, the wettability of the dosage unit, the penetration ability of the dissolution medium, the swelling process, the disintegration, and the deaggregation of the dosage forms are a few of the factors that influence the dissolution characteristics of drugs. Wagner proposed a scheme, depicted in Figure 6-1, for the processes involved in the dissolution of solid dosage forms.

This scheme was later modified to incorporate other factors that precede the dissolution process of solid dosage forms. Carstensen proposed a scheme incorporating the following sequence:

1. Initial mechanical lag
2. Wetting of the dosage form
3. Penetration of the dissolution medium into the dosage form
4. Disintegration
5. Deaggregation of the dosage form and dislodgement of the granules
6. Dissolution and occlusion of some particles of the drug

Carstensen explained that the wetting of the solid dosage form surface controls the liquid access to the solid surface and, many times, is the limiting factor in the dissolution process. The speed of wetting directly depends on the surface tension at the interface (interfacial tension) and upon the contact angle between the solid surface and the liquid. Generally, a contact angle of more than 90° indicates poor wettability. Incorporation of a surfactant, either in the formulation or in the dissolution medium, lowers the contact angle and enhances dissolution. Also, the presence of air in the dissolution medium causes the air bubbles to be entrapped in the tablet pores and act as a barrier at the interface. For capsules, the gelatin shell is extremely hydrophilic, and, therefore, no problems in wettability exist for the dosage itself, although it may exist for the powders inside.

After the solid dosage form disintegrates into granules or aggregates, penetration characteristics play a prime role in the deaggregation process. Hydrophobic lubricants, such as talc and magnesium stearate, commonly employed in tablet and capsule formulations, slow the penetration rate and, hence, the deaggregation process. A large pore size facilitates penetration,

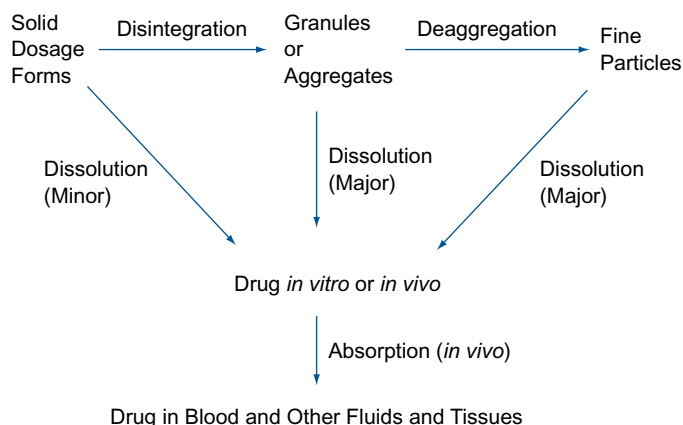


Figure 6-1. Schematic illustration of dissolution process of solid dosage forms.

but, if it is too large, it may inhibit penetration by decreasing the internal strain caused by the swelling of the disintegrant.

After deaggregation and dislodgment occur, the drug particles become exposed to the dissolution medium and dissolution proceeds. Figure 6-2,³ graphically presents the model proposed by Carstensen.

It is apparent from Figure 6-2 that the rate of dissolution of the drug can become a rate-limiting step before it appears in the blood. However, when the dosage form is placed into the gastrointestinal tract in solid form, there are two possibilities for the rate-limiting step. The solid must first dissolve, and the drug in solution must then pass through the gastrointestinal (GI) membrane. Freely water-soluble drugs tend to dissolve rapidly, making the passive diffusion of the drug or the active transport of the drug rate-limiting step for absorption through the GI membrane. Conversely, the rate of absorption of poorly water-soluble drugs will be limited by the rate of dissolution of the undissolved drug or disintegration of dosage form.

The rate of dissolution of drug substance is determined by the rate at which solvent-solute forces of attraction overcome the cohesive forces present in the solid. This process is rate-limiting, when the release of solute into solution is slow and the transport into the bulk solution is fast. In this case, the dissolution is said to be interfacially controlled. Dissolution may also be diffusion controlled, where the solvent-solute interaction is fast, compared to transport of solute into the bulk solution. In diffusion-controlled process, a stationary layer of solute adjacent to the solid/liquid interface is postulated and is commonly referred to as the diffusion layer. The saturation concentration of solute develops at the interface and decreases with distance across the diffusion layer.

MATHEMATICS OF DISSOLUTION

It has long been recognized that the release of the active drug from a drug product may be greatly influenced by the physicochemical properties of the drug, as well as the dosage form.⁴ The availability of the drug is usually determined by the rate of release of the drug from the physical system (dosage form). The release of the drug from its dosage form is usually determined by the rate at which it dissolves in the surrounding medium. The rate of dissolution of a chemical or drug from the solid state is defined as the amount of drug substance that goes into solution per unit time under standardized condition of liquid/solid interface, temperature, and solvent composition. In biopharmaceutics, rate of dissolution usually refers to the rate at which

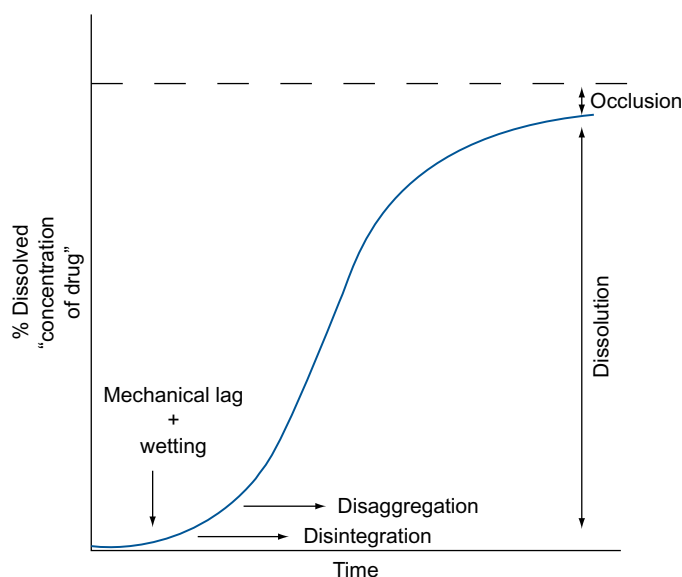


Figure 6-2. The S-shaped dissolution curve of solid dosage forms.

the drug dissolves from an intact dosage form or from fragments or particles from the dosage form during the test.⁵

The following section deals with the introductory concepts on mathematics of dissolution, focusing primarily on intrinsic dissolution.

INTRINSIC DISSOLUTION

The rate of dissolution of a pure pharmaceutical active ingredient, when conditions, such as surface area, temperature, agitation or stirring speed, pH, and ionic strength of the dissolution medium is kept constant, is known as intrinsic dissolution rate. This parameter allows the screening of the drug candidates and aids in understanding their solution behavior under various biophysiological conditions.⁶

Intrinsic Dissolution Rate Constants

The rate at which a substance dissolves in a liquid to form a solution is governed by physical parameters, such as the surface area of the substance at a given time during the process of dissolution, the shape of the substance, the characteristics of the solid/ liquid interface, and the solubility of the substance in the liquid. Hence, dissolution can be considered a specific type of certain heterogeneous reaction, which results in a mass transfer as a net effect between the escape and deposition of solute molecules at a solid surface. Mathematically, the process can be simply described as follows:

$$\frac{dM}{dt} = KA(C_s - C) \quad (2)$$

where M is the mass of the substance remaining to be dissolved, A is the surface area exposed to the dissolution medium, C_s is the saturation concentration referred to as solubility in the dissolution medium, C is the amount dissolved or the concentration of the drug in solution at time t , and K is the intrinsic dissolution rate constant or simply the dissolution rate constant.

The equation expresses the fact that, when C is small, $C < 0.15C_s$, then K is proportional to C_s , since $(C_s - C)$ is large. If this applies, then to a good approximation we may write:

$$\frac{dM}{dt} = KAC_s \quad (3)$$

Equation 3 is commonly referred to as a sink-condition equation, which implies that sink conditions exist during the process of dissolution. It must be noted, however, that A is a constant, except initially, when only very small quantities of solute have dissolved and where there is an amount of solute far in excess of saturation.

When the process of dissolution takes place under sink conditions, a stagnant film of liquid (dissolution medium) is adsorbed onto the solid, the thickness of this film being l cm. The liquid in the film in direct contact with the solid is saturated with drug in solution. The concentration of the drug in solution then drops as the distance from the dissolving solid surface increases. At the end of the film, l cm from the surface, the concentration in the film is the same as that in the bulk solution, C_b . The driving force behind the movement of solute molecules through the stagnant film is the concentration gradient that exists between the saturation concentration of the solute, C_s , in the stagnant layer at the surface of the solid and its concentration on the farthest side of the stagnant film, C_b . A schematic representation of dissolution as a physicochemical phenomenon is shown in Figure 6-3: the greater this difference in concentration, the faster the rate of dissolution.

Applying Fick's first law of diffusion to Equation 2, the flux, J , defined as the rate of flow of material through 1 cm^2 , that is,

$$J = \frac{\frac{dM}{dt}}{A} \quad (4)$$

can be expressed as

$$J = -D \left(\frac{\delta C}{\delta x} \right) \quad (5)$$

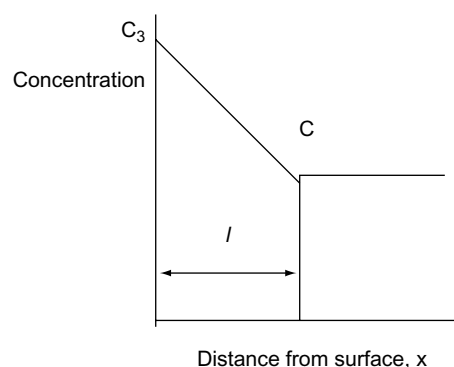
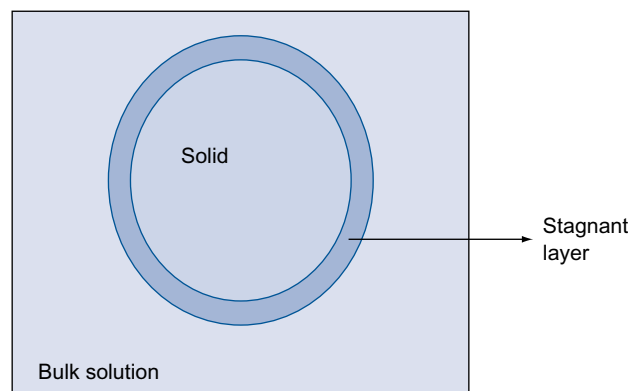


Figure 6-3. Schematic representation of a physical model depicting a dissolution process.

where D is the coefficient of diffusion, x is the distance as shown in the Figure 6-3. If the concentration gradient, $\delta C/\delta x$, is linear, if $C = C_s$ at the surface ($x = 0$), and if $C = C_b$ (the bulk concentration at the interface between the bulk solution and the film, where $x = l$), then

$$\left(\frac{\delta C}{\delta x} \right) = \left(\frac{C_b - C_s}{l} \right) \quad (6)$$

Therefore,

$$\frac{1}{A} \frac{dM}{dt} = -D \frac{C_b - C_s}{l} = -K(C_s - C) \quad (7)$$

Or simply,

$$\frac{dM}{dt} = -KA(C_s - C) \quad (8)$$

If the agitation intensity of the system containing suspended particles is increased, the thickness of the film will decrease progressively. Hence, k is a function of the test as well. Additionally, if the product of $A(C_s - C)$ is maintained constant for many drugs tested by the same test, the relative magnitude of k values will indicate the effective ease of dissolution. In practice, k also includes the dynamics of the shear rate between solid and the solvent, that is, the rate at which fresh solvent contacts the surface of the solid; highly complex processes, including diffusion rate through the boundary layers, depend upon this rate.

The shear rate depends on a multitude of variables that must be controlled, if the test is to be repeatable. Those variables include the flow pattern of solvent in the apparatus, turbulence, viscosity, surface tension, and dissolved gases, all of which are subject to uncontrolled input variables to the system, such as vibrations and system geometry. The theoretical basis for those inputs rests in the realm of chemical engineering's fluid flow and surface boundary theories, which are discussed elsewhere.⁷

The intrinsic dissolution rate has been used as a means to demonstrate the chemical purity and equivalency of the active pharmaceutical ingredient (API). The use of rotating disk system (USP Wood Apparatus), which is similar to USP procedure 1, is most common, although stationary disk systems, vertical diffusion cells, and enhancer cells can also be used to measure the intrinsic dissolution rates.⁸

ROTATING DISK SYSTEM (USP WOOD APPARATUS)

The apparatus consists of steel punch, a die, and a base plate. The base of the die has three threaded holes for the insertion of the screws and the attachment to the base plate. The material is placed in the die cavity, the punch is inserted into the cavity, and the material is compressed. The die is screwed onto the shaft holder, and the shaft holder is mounted on the stirring device. The shaft is a stainless steel rod with hollow die holder. Pellet and die assembly are introduced at once, when the dissolution drive mechanism is lowered. The dissolution is achieved by shear-like motion of the pellet in the dissolution medium. The dissolution vessel is standard curved bottom 1-L flask. Care should be taken that the air bubbles do not form on the surface of the pellet or it will interfere with the dissolution rate (Fig. 6-4).

STATIONARY DISK SYSTEM

The apparatus consists of a steel punch, a die, and a base plate. The die base has three holes for the attachment of the base plate. The three fixed screws on the base plate are inserted through the three holes on the die. Punch is inserted into the die cavity filled with the material, and the material is then compressed. The pellet and die assembly can then be inserted with the pellet side up, into the bottom of the dissolution vessel, which is flat bottomed. The USP Apparatus 2 is the stirring mechanism here. The advantage of this system is that no air bubbles are formed on the pellet surface. There is also no change in the temperature, as the device is small and is totally submerged into the dissolution medium (Fig. 6-5).

CORRELATION BETWEEN THE DISINTEGRATION AND DISSOLUTION

The close correlation between disintegration and dissolution has been studied by many investigators. Both processes exhibit "S"-shaped curves, and a Probit or a Weibull function was suggested to explain the data. In general, however, disintegration has proved a poor indicator of bioavailability, due to the turbulent agitation maintained during the test. Several other factors, such as solubility, particle size, and crystalline structure, among others, have been found to affect the dissolution of the drug substance but have no relevance to disintegration.

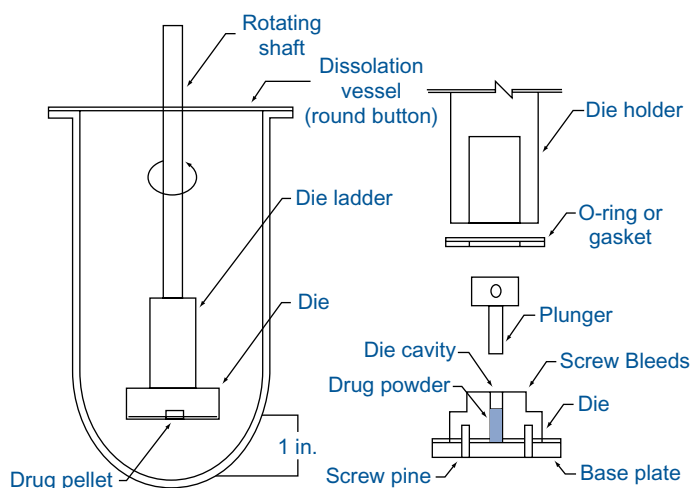


Figure 6-4. Rotating disk system (USP Wood apparatus).

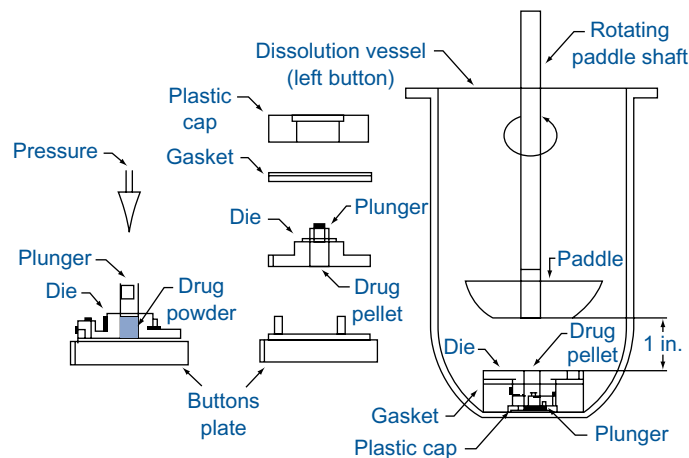


Figure 6-5. Stationary disk system (new apparatus).

FACTORS AFFECTING THE RATE OF DISSOLUTION

The dissolution rate data can be meaningful, only if the results of successive test on the same dosage form are consistent within reason. The dissolution test should yield reproducible result, even when it is performed in different laboratories or with different personnel. To achieve high reproducibility, all variables that influence the test should be clearly understood and possibly controlled.

Factors affecting the dissolution rate of drugs from a dosage form include:

1. Factors related to the physicochemical properties of the drug
2. Factors related to drug product formulation
3. Factors related to dissolution test parameters
4. Miscellaneous factors

FACTORS RELATED TO THE PHYSICOCHEMICAL PROPERTIES OF THE DRUG

Effect of Solubility on Dissolution

The physicochemical properties of the drug substance play a prime role in controlling its dissolution from the dosage form. The modified Noyes and Whitney equation shows that the aqueous solubility of the drug is the major factor determining its dissolution rate. Actually, some studies show that drug-solubility data could be used as a rough predictor of the possibility of any future problems with bioavailability, a factor that should be taken into consideration in the formulation design.

Effect of Particle Size on Dissolution

According to Nernst-Brunner theory, the dissolution rate is directly proportional to the surface area of the drug. Since the surface area increases with the decreasing particle size, higher dissolution rates may be achieved through the reduction of the particle size. This effect has been highlighted by the superior dissolution rate observed after "micronization" of certain sparingly soluble drugs, as opposed to the regularly milled form.

Several investigations have demonstrated an increased absorption rate for griseofulvin after micronization. Similar effects have been reported for chloramphenicol, tetracycline salts, sulfadiazine, and norethisterone acetate. In the case of chloramphenicol, it has been shown that formulations containing smaller particles (50–200 μm) were absorbed faster than formulations containing larger particles (400–800 μm). Figure 6-6 presents the effect of particle-size differences on the dissolution rate of phenacetin and phenobarbital.⁹

However, when employing this technique to enhance dissolution, it is important to recognize the fact that it is the

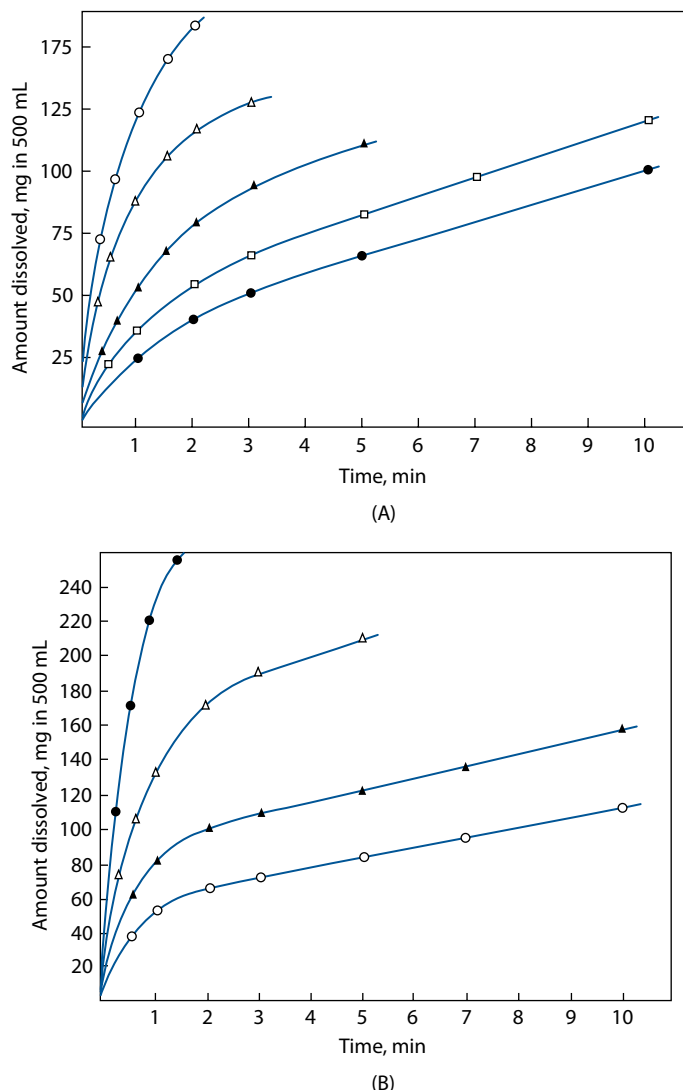


Figure 6-6. Effect of particle size on the dissolution rate of drugs from solid dosage forms.³**A.** Phenacetin: ○ particle size: 0.11-0.15 mm; △ particle size: 0.15-0.21 mm; ▲ particle size: 0.21-0.30 mm; □ particle size: 0.30-0.50 mm; ▤ particle size: 0.50-0.71 mm. **B.** Phenobarbital: ● particle size: 0.07-0.15 mm; △ particle size: 0.15-0.25 mm; ■ particle size: 0.25-0.42 mm; ◆ particle size: 0.42-0.71 mm.

effective surface area that has to be increased. The effective surface area is the surface area available to the dissolution fluid. If the drug is hydrophobic and the dissolution medium has poor wetting properties, reduction of particle size may lead to decreased effective surface area and, hence, a “slower” rate of dissolution.

Physical properties of the drug particles, other than size, also indirectly affect the effective surface area by modifying the shear rate of the fresh solvent that comes in contact with the solid. These properties include the particle shape and the density.

The mechanism by which the reduction in particle size improves dissolution is usually through the enhancement of the drug solubility. It is assumed that the drug solubility is independent of particle size. However, the drug solubility and the surface area can be correlated by the Ostwald-Freundlich equation:

$$\ln S = \frac{2M\gamma}{\rho RT} \frac{1}{r} = \frac{\alpha}{r} \quad (9)$$

where M is the molecular weight, ρ is the density, γ is the interfacial tension or surface free energy of the solid, T is the temperature, R is the gas constant, and r is the radius of the particle.

From this equation,

$$S = S_{\infty} \cdot e^{\alpha/r} \quad (10)$$

The equation shows that the solubility is inversely proportional to particle radius. Therefore, S could be viewed as the solubility of the microparticles, and S_{∞} as the solubility of the macro particles. However, it is obvious that the particle radius has to be reduced to a microlevel, before it can effect a change in solubility.

This extreme reduction in particle size usually cannot be achieved through regular milling or even micronization procedures, and, therefore, other methods have been recommended. One of these involves formation of a solid solution or molecular dispersion, where the molecules of the sparingly soluble drug either are dispersed interstitially in a water-soluble drug or replaced in its crystal lattice.

Another technique, which also produces extremely small particles but still larger than the those produced by solid solution, is by dispersion of the drug into a soluble carrier, such as polyvinylpyrrolidone (PVP) solution. These techniques are employed for the enhancement of dissolution rate of insoluble drugs.

Effect of Solid Phase Characteristics of the Drug on Dissolution

Amorphicity and crystallinity, the two important solid-phase characteristics of drugs affect their dissolution profile. Numerous studies have demonstrated that the amorphous form of a drug usually exhibits greater solubility and higher dissolution rates, as compared to that exhibited by the crystalline form. For example, it was shown that the amorphous form of novobiocin has a greater solubility and higher dissolution rate than the crystalline form. Blood-level studies confirmed such findings, where administration of the amorphous form yielded about three to four times the concentration, compared to the administration of the crystalline form. Similar differences were demonstrated for griseofulvin, phenobarbital, cortisone acetate, and chloramphenicol. Chloramphenicol palmitate is one example that exists in at least two polymorphs. The B form is apparently more bioavailable. The recommendation might be that manufacturers should use polymorph B for maximum absorption. One contradictory example is that of erythromycin estolate, where the dissolution rate of amorphous form is markedly lower than the crystalline form of erythromycin estolate, as exemplified by Figure 6-7. However, a method of controlling and determining crystal form would be necessary in the quality control process.¹⁰

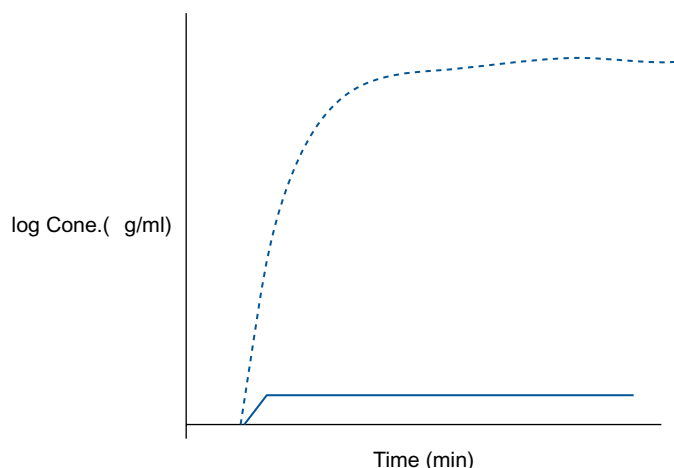


Figure 6-7. Dissolution performance of erythromycin estolate. (Dotted line, Crystalline form; Solid line, Amorphous form.)

Effect of Polymorphism on Dissolution

Polymorphic forms of drugs have been shown to influence changes in solubilizing characteristics and, thus, the dissolution rate of the drug in question. Numerous reports have shown that polymorphism and the state of hydration, solvation, and/or complexation markedly influence the dissolution characteristics of the drug. The drugs that exhibit influence on the dissolution behavior include tolbutamide, chloramphenicol, and others.

FACTORS RELATED TO DRUG PRODUCT FORMULATION

It has been shown that the dissolution rate of a pure drug can be altered significantly when mixed with various excipients during the manufacturing process of solid dosage forms. These excipients are added to satisfy certain pharmaceutical functions, such as diluents (fillers), dyes, binders, granulating agents, disintegrants, and lubricants. Generically identical tablet and capsule products, manufactured by different pharmaceutical manufacturers, were found to exhibit significant differences in dissolution rates for their active ingredients. In certain cases, several studies showed that poor tablet and capsule formulations have been shown to cause a marked decrease in bioavailability and impairment of the clinical response. Such findings, during the 1960s, especially in the case of digoxin and tolbutamide tablets, as well as chloramphenicol and tetracycline HCl (all life-saving drugs), were the triggering factors that compelled the drug-regulatory agencies and compendial authorities to institute the dissolution test as a legal requirement for most solid dosage forms.

Effect of Granulating Agents and Binders

Solvang and Finholt¹¹ have shown that Phenobarbital tablets, granulated with gelatin solution, provide faster dissolution rate in gastric fluid than those prepared using sodium carboxymethylcellulose or polyethylene glycol 6000 as a binder. This observation was attributed to the fact that gelatin imparts hydrophilic characteristics to the hydrophobic drug surface, whereas PEG 6000 forms complex with poor solubility, and sodium carboxymethylcellulose is converted to its less soluble acid form at low pH of the gastric fluid (Fig. 6-8). Even gelatin obtained from various processes and origins has been shown to affect the dissolution rate of dosage forms.¹²

Various studies have been reported in the literature, evaluating the effects of various granulating agents and binders on the dissolution rate of tablets.¹³

Effect of Disintegrants and Diluents

The type and amount of disintegrating agent employed in the formulation significantly controls the overall rate of dissolution of dosage form. Jaminet *et al.*¹² employed several disintegrating agents in manufacturing of Phenobarbital tablets, including Primojel (sodium glycolate of potato starch), Nymcel (polymerized water-soluble brand of sodium carboxymethylcellulose), and Copagel (low viscosity grade of sodium carboxymethylcellulose). The effect on the dissolution rate of tablets by the addition of disintegrants, before and after granulation, was assessed. When added before granulation, Copagel gave tablets with a remarkably slow dissolution rate. However, when added after granulation, Copagel did not result in lowering the dissolution rate. Primojel was not found as effective, particularly on addition after granulation. Levy, in 1963, studied the effect of starch, the most commonly used diluent, on the rate of dissolution of salicylic acid tablets manufactured by the dry, double-compression process (Fig. 6-9).¹⁴ Increasing the starch content from 5% to 20% resulted in a dramatic increase in the dissolution rate, almost three-fold. This was attributed to better and more thorough disintegration. Later, however, Finholt suggested that the hydrophobic drug crystals acquire a surface layer of fine starch particles that imparts a hydrophilic property

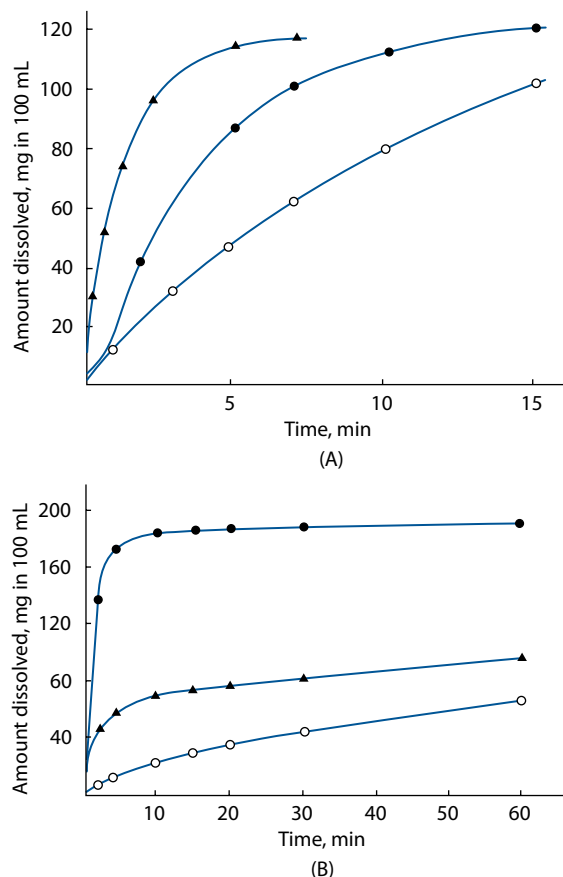


Figure 6-8. Effect of binders and granulating agents on dissolution rate of tablets.⁵ A. Rate of dissolution of phenacetin from powder, granules, and tablets in diluted gastric juice (surface tension 42.7 dynes cm⁻¹, pH 1.85). •, phenacetin powder; ▲, phenacetin granules; ●, phenacetin tablets. B. Dissolution rate of phenobarbital tablets in diluted gastric juice (surface tension 39.4 dynes cm⁻¹, pH 1.50). •, Gelatin binder, <open triangle> CMC, <open circle> Polyethylene glycol 6000.

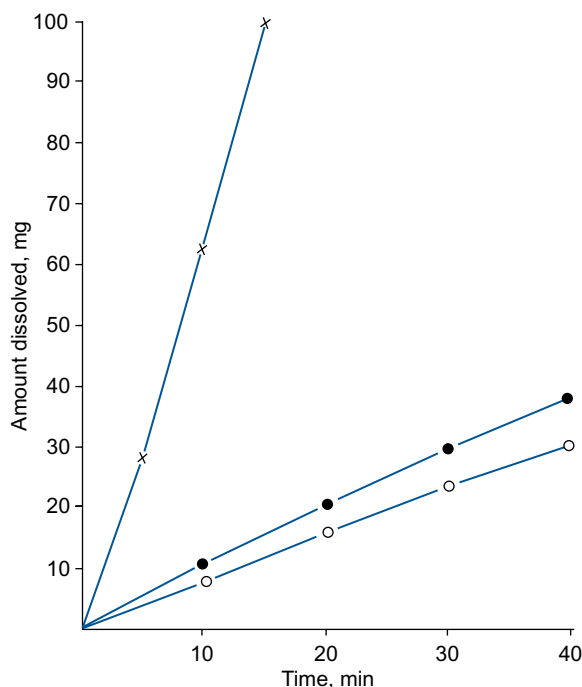


Figure 6-9. Effect of starch content on dissolution rate.⁴ <open circle>, 5%; •, 10%; ×, 20% starch in granules.

to the granular formulation and, thereby, increases the effective surface area and, hence, the dissolution rate (see Fig. 6-9).

Effects of Lubricants

The nature, quality, and quantity of lubricants added can affect the dissolution rate. The effect of various lubricants on dissolution rate of salicylic acid was studied, and it was concluded that magnesium stearate, a hydrophobic lubricant, tends to retard the dissolution rate of salicylic acid tablets, whereas sodium lauryl sulfate enhances dissolution, due to its hydrophilic character combined with surface activity, which increases the microenvironment pH surrounding the weak acid and increases wetting, and better solvent penetration into the tablets¹⁵ (Fig. 6-10) illustrates the effect of lubricants on the dissolution rate of tablets.

Effect of lubricants on the dissolution rate of drugs from dosage form depends on properties of the granules, the lubricant itself, and the amount of lubricant used. If granules are hydrophilic and fast disintegrating, a water-soluble surface-active lubricant will have an insignificant effect on the dissolution. Conversely, if the granules are hydrophobic, the surface-active lubricant will enhance dissolution. It was also found that hydrophobic lubricants, such as magnesium stearate, aluminum stearate, stearic acid, and talc, decrease the effective drug-solvent interfacial area by changing the surface characteristics of the tablets, which results in reducing its wettability, prolonging its disintegration time, and decreasing the area of the interface between the active ingredient and solvent.

FACTORS RELATED TO THE DISSOLUTION TEST PARAMETERS

Method of Granulation

Wet granulation has been shown to improve the dissolution rates of poorly soluble drugs by imparting hydrophilic properties to the surface of the granules. Additionally, the use of

fillers and diluents, such as starch, spray dried lactose, and microcrystalline cellulose, tends to increase the hydrophilicity of the active ingredients and, thus, improve dissolution. Consequently, wet granulation was considered superior to a dry or double-compression procedure. Figure 6-11 shows the effect of different granulation methods on the dissolution rate of tablets.¹⁶

It must be noted that, with the advent of newer tableting machines and materials, it becomes more evident that the critical formulation and proper mixing sequence and time of

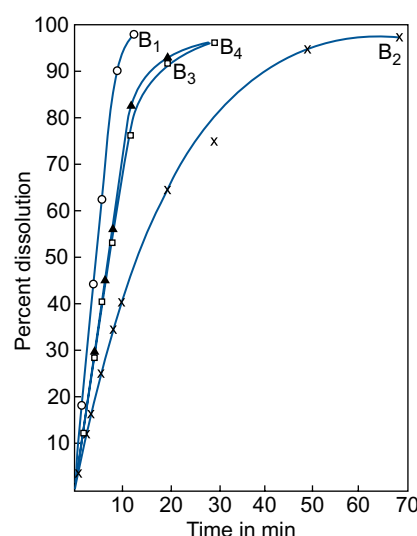


Figure 6-11. Effect of manufacturing process on the dissolution rate of tablets.⁷ B₁, Direct compression with spray-dried lactose. B₂, Wet granulation with ethylcellulose and lactose. B₃, Acacia mucilage and lactose. B₄, Starch paste and lactose.

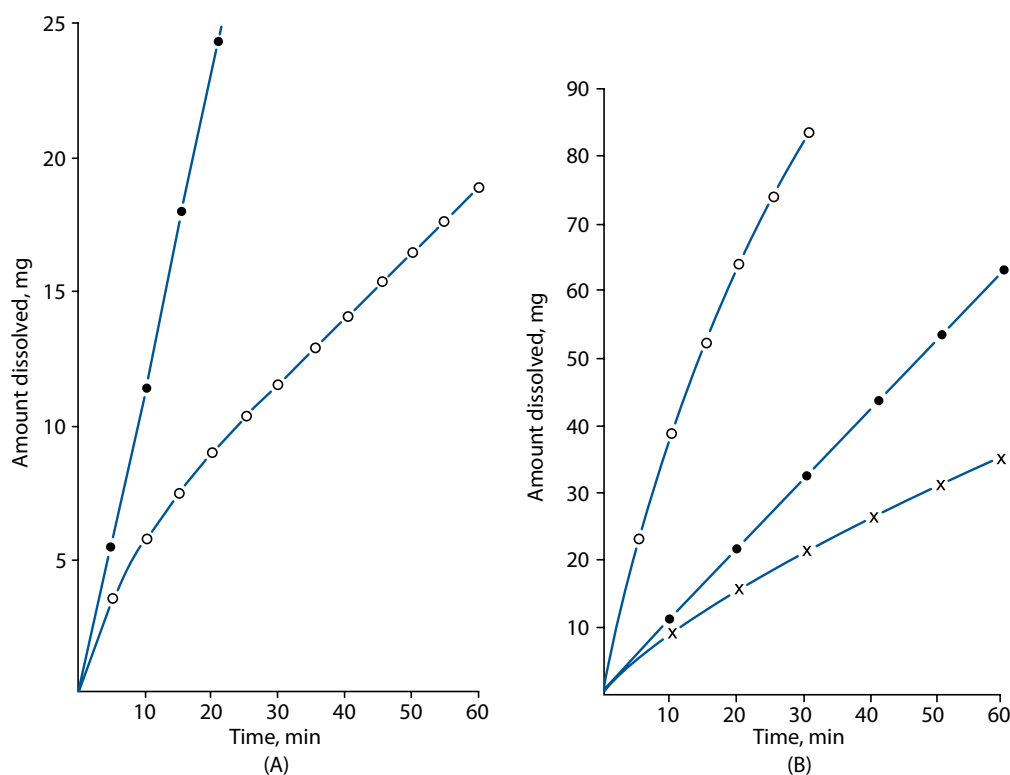


Figure 6-10. Effect of lubricant on the dissolution rate of tablets.⁶ A. Effect of magnesium stearate on dissolution rate of salicylic acid from rotating discs made from fine salicylic acid powder. ○, 3% magnesium stearate; ●, no lubricant added. B. Effect of lubricant on dissolution rate of salicylic acid contained in compressed tablets (formula A). ×, 3% magnesium stearate; ●, no lubricant; ○, 3% sodium lauryl sulfate.

adding the several ingredients are the main criteria that affect the dissolution characteristics of the tablets, not the method of granulation.

Effect of Compression Force on Dissolution Rate

In his early studies of the physics of tablet compression, T Higuchi, in 1953, pointed out the influence of compression force employed in the tableting process on the apparent density, porosity, hardness, disintegration time, and average primary particle size of compressed tablets. There is always a competing relationship between the enhancing effect, due to the increase in surface area, through the crushing effect and the inhibiting effect, due to the increase in particle bonding that causes an increase in density and hardness and, consequently, a decrease in solvent penetrability. The high compression may also inhibit the wettability of the tablet, due to the formation of a firmer and more effective sealing layer by the lubricant under the high pressure and temperature that usually accompanies a strong compressive force (Fig. 6-12).¹⁴ The curve profile of the compressive force of the tablet, versus dissolution rate, can take one of several shapes, as is observed in Figure 6-13.⁹

FACTORS RELATED TO THE DOSAGE FORM

Drug Excipient Interaction

These interactions can occur during any unit operation, such as mixing, blending, drying, and/or granulating, resulting in a change in dissolution pattern of the dosage form in question.

The effect of magnesium stearate on the disintegration time of tablets containing either potato starch or sodium starch glycolate was found to depend on the swelling characteristics of the

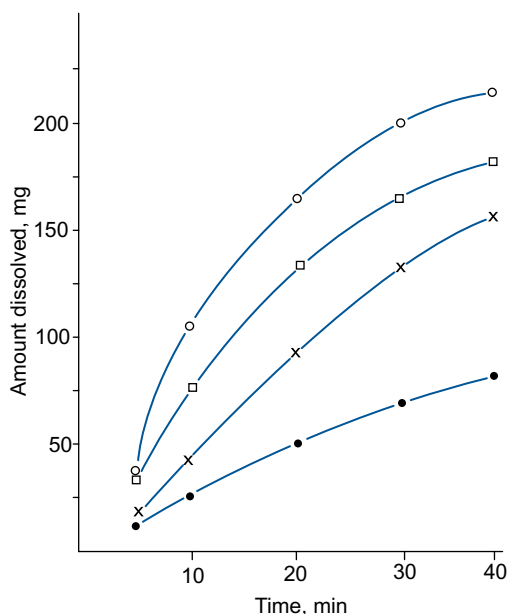


Figure 6-12. Effect of precompression pressure on the dissolution rate of salicylic acid contained in compressed tablets.⁴ ○, 715 kg; ×, 1430 kg; ◆, 2860 kg; ■, 5730 kg pressure per cm.² (Average of five tablets each, formula D.)

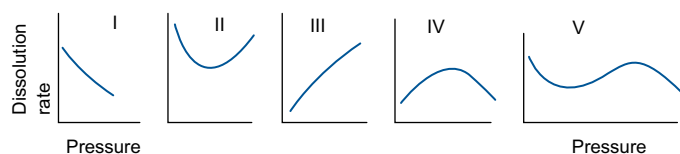


Figure 6-13. Different types of relations between compressional force of tablets and dissolution rate.

disintegrants. These results were attributed to the formation of lubricant film during mixing, which resulted in an increase in disintegration time and, thus, delayed dissolution.¹⁷

It is essential that the formulator have a thorough understanding of these interactions, so that the most appropriate excipients can be selected to enable the formulator to perform optimally. By minimizing, if not eliminating, these interactions, adverse effects on the performance of the final product can be avoided. It must also be noted that better process control is also possible with non-interacting drug-excipient interactions.

Deaggregation

Deaggregation is often a prerequisite for dissolution. In such cases, it can control dissolution. It was reported that two capsule formulations of sodium diphenylhydantoin showed significant deaggregation, dissolution, and thereby absorption rates. The formulation that deaggregated rapidly after the capsule shell was dissolved resulted in exposure of a larger surface area. This resulted in rapid dissolution at neutral pH, but less rapid dissolution when both preparations were exposed to 0.1N hydrochloric acid. Aggregation of other formulation inhibited the conversion of most of its sodium salt to the free acid in acidic medium, whereas such conversion occurred readily with the rapidly disintegrating formulation. As a result, after neutralization of the medium, the latter dissolved and absorbed more readily and rapidly than did the former.

EFFECT OF TEST PARAMETERS ON THE DISSOLUTION RATE

Eccentricity of the Stirring Device

USP 34/NF 29 specifies that the stirring shaft must rotate smoothly without significant wobble. Eccentricity can be measured with a machinist's indicator. It is measured in terms of total indicator reading (TIR), which determines the sum of the distance on both sides (180°) of the axis of rotation.

Guiding the Shaft

One must remember that the shaft of the stirring device extends about 6 in beyond the chuck. An eccentricity of 0.005 in (0.11 mm) at a distance of 1 in (25 mm) from the chuck will be barely perceptible, but, at 6 in, it will amount to 0.30 in (0.75 mm), which is the maximum that can be tolerated.

The shaft mounting should not produce perfect concentricity, but also allow for ease of vertical adjustment. That can best be obtained with a hollow drive shaft and chuck grip on the output with a guide on the other hand. They must be held to close concentricity tolerances with the axis of rotation of the drive tube. The further apart such guides, the better the probability of minimum wobble at the end of the shaft, provided the shaft is straight. The simple trigonometry is illustrated in Figure 6-14.¹

In the lower left view, the distance from the chuck to the basket is approximately 12 in (15 mm). If no guide bushing is used, any inherent eccentricity in the chuck is multiplied 12 times at the basket, which will certainly produce an eccentricity greatly exceeding the acceptable tolerances for eccentricity at the basket or paddle, if not when new, then after the chuck has been in use. In the lower right view, no guide bushing is used, but the shaft is supported at both the ends of the hollow shaft from A to A, and the chuck is brought closer to the flask cover. The inherent eccentricity in the drive (A to A) can be held close at the factory, and the eccentricity of the basket cannot exceed it, if the shaft is straight. In this case, the distance from A to A is about 6 in and matches the distance from the chuck head to the basket, also about 6 in.

Even with a guide bushing, the system shown on the lower left is not recommended, because the chuck may have a twist corrected by the bushing, which might cause a whip in the shaft with attendant vibration. These problems can be minimized by using a resilient grip in the chuck, such as a rubber "O" ring.

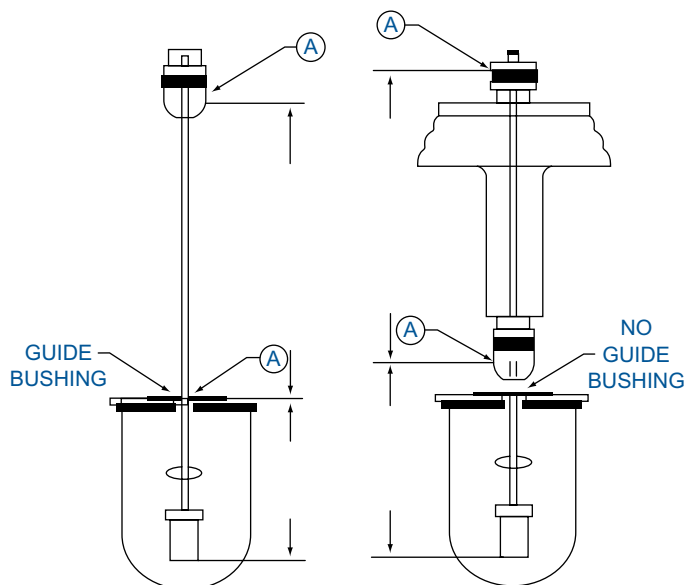


Figure 6-14. The rotating shaft should be supported at two places (A) to minimize wobble, shown by the two arrangements depicted.

Vibration

Vibration is a common variable introduced into the dissolution system from myriad causes. It has the effect of changing the flow patterns of the liquid and of introducing unwanted energy to the dynamic system. Both effects may result in significant changes in the dissolution rates. The speeds of the rotational device selected by official compendium are 50 rpm or 100 rpm. Other speeds are specified for certain drugs. Precise speed control is best obtained with a synchronous motor that locks into the line frequency. Such motors are not only more rugged, but are also far more reliable. Periodic variations in rpm that might result in possible disturbance in rotational devices is commonly referred to as torsional vibration. Such vibration indicates a variation in the velocity of rotation for short periods of time, although the average velocity is well within $\pm 4\%$ of the specified rate.

Alignment of the Stirring Element

There are two important factors to be considered here. These are as follows:

Tilt

USP 34/NF 29 states that the axis of the stirring element shall not deviate more than 0.2 cm from the axis of the dissolution vessel, which defines centering of the stirring shaft to within ± 2 mm. It also constrains tilt. A series of tests suggest that tilt in excess of 1.5 (degrees) may increase dissolution rates using Method 2 from 2% to 25%,¹⁸ which is still a significant variation. The user should be able to adjust the equipment to obtain alignment of the vertical spindles to within 1 (degree) perpendicularly with the base of the drive to which the flasks are mounted. Such alignment cannot be ensured in the factory. Adjustments for perpendicularity must, therefore, be used to bring the equipment into alignment in its final position.¹⁹

Agitation Intensity

The degree of agitation, or the stirring conditions, is one of the most important variables to consider in dissolution. Given the background of various theories of dissolution, it is apparent that agitation conditions can markedly affect diffusion-controlled dissolution, because the thickness of the diffusion layer is inversely proportional to agitation speed. Wurster and Taylor²⁰ employed the empirical relationship

$$K = a(N)^b \quad (11)$$

where N is the agitation rate, K is the reaction (dissolution) rate, and a and b are constants. For diffusion-controlled processes, $b = 1$. Dissolution that is interfacial-reaction-rate-controlled will be independent of agitation intensity, and, thus, $b = 0$.

Agitation intensity within and between various *in vitro* dissolution testing devices can be varied by the dimensions and geometry of the dissolution vessel, volume of dissolution medium, and the degree of agitation or shaking. It is safe to predict that the two dosage forms having particles of differing sizes and densities will not experience identical dissolution system, even though the containers are subjected to the same rate of rotation as oscillation.

Temperature

Because drug solubility is temperature-dependent, careful temperature control during the dissolution process is very important and should be maintained within 0.5° . Generally, a temperature of 37°C is always maintained during dissolution determinations. The effect of temperature variations of the dissolution medium depends mainly on the temperature/solubility curves of the drug and excipients in the formulation (Fig. 6-15).^{21,22}

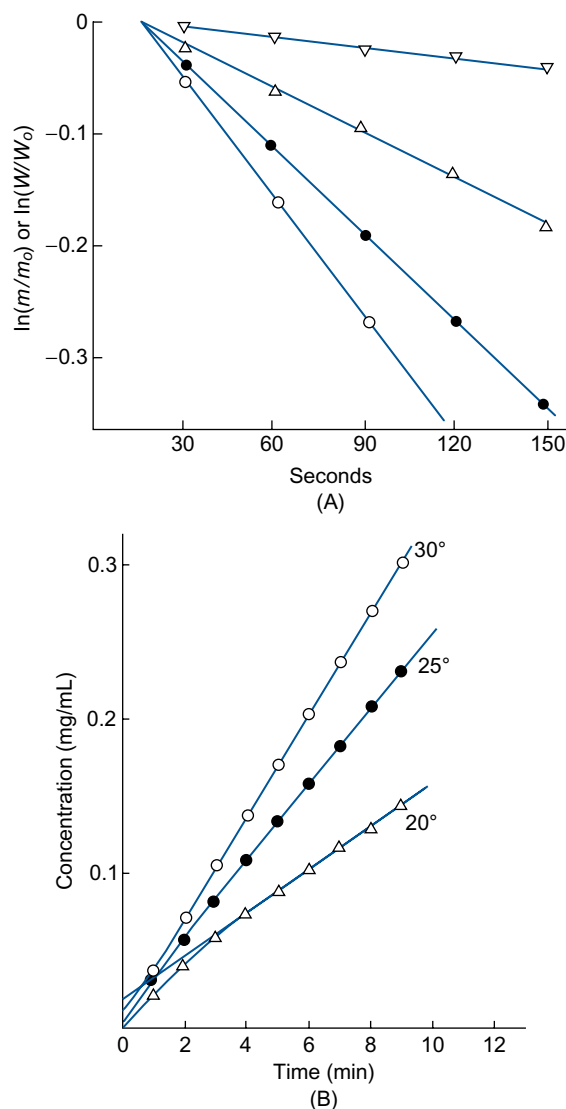


Figure 6-15. Effect of temperature on dissolution and disintegration rates of tablets.^{15,16} A. Dissolution and disintegration curves according to Equations 1 and 2 for position II of the USP basket. Δ , dissolution at 10° ; $\circ$, dissolution at 20° ; $\bullet$, dissolution at 30° ; and, $\circ$ disintegration at 5° .¹⁶ B. Dissolution of phenobarbital anhydrate at various temperatures (at 300 rpm).

For a dissolved molecule, the diffusion coefficient, D , depends on the temperature T , according to the Stokes equation:

$$D = kT/6\pi\eta r \quad (12)$$

where k is the Boltzmann constant and $6\pi\eta r$ is the Stokes force for a spherical molecule (η is the viscosity in cgs or poise units, and r is the radius of the molecule).

DISSOLUTION MEDIUM

Selection of suitable fluid for dissolution testing depends largely on the solubility of the drug, as well as mere economics and practical reasons.

pH of the Dissolution Medium

Great emphasis and effort was first placed on simulating *in vivo* conditions, especially pH, surface tension, viscosity, and sink condition. Most of the early studies were conducted in 0.1N HCl or buffered solutions with a pH close to that of the gastric juice (pH ~ 1.2). The acidic solution tends to disintegrate the tablets slightly faster than water and, thereby, may enhance the dissolution rate by increasing the effective surface area. However, due to the corroding action of the acid on dissolution equipment, currently, it is a general practice to use distilled water, unless investigative studies show a specific need for the acidic solution to generate meaningful dissolution data. Another approach for avoiding the deleterious effects of hydrochloric acid is to replace it with acidic buffers, such as sodium acid phosphate, to maintain the required low pH.

Surface Tension of the Dissolution Medium

Surface tension has been shown to have a significant effect on the dissolution rate of drugs and their release rate from solid dosage forms. Surfactants and wetting agents lower the contact angle and, consequently, improve penetration by the dissolution medium. Measurable enhancement in the dissolution rate of salicylic acid from an inert matrix was reported by Singh and coworkers, when the contact angle, θ , was lowered from 92° (water) to 31°, using 0.01% dioctyl sodium sulfosuccinate (Fig. 6-16).²³ The surface tension was also correspondingly lowered from 60 to 31 dynes/cm. Similar findings were obtained in benzocaine studies when polysorbate 80 was used as the surface active agent (see Fig. 6-16).²³

Other studies conducted on conventional tablet formulations and capsules also showed significant enhancement in the dissolution rate of poorly soluble drugs, when surfactants were added to the dissolution medium, even at a level below the critical micelle concentration, probably by reducing the interfacial tension. Low levels of surfactants were recommended to be included in the dissolution medium, as this seemed to give a better *in vivo* and *in vitro* correlation.

Finholt and Solvang compared the dissolution behavior of phenacetin and phenobarbital tablets in human gastric juice to that in dilute hydrochloric acid with and without various amounts of polysorbate 80 in the dissolution medium. The data showed that both pH and surface tension have significant influence on the dissolution kinetics of the drug studies. For example, they found that not only was the dissolution rate much faster in diluted gastric juice, but that it increased with decreasing particle size, whereas the opposite was the case when 0.1N HCl was used.

Viscosity of the Dissolution Medium

In case of diffusion-controlled dissolution processes, it would be expected that the dissolution rate decreases with an increase in viscosity. In the case of interfacial-controlled dissolution processes, however, viscosity should have little effect. The Stokes-Einstein equation describes diffusion coefficient, D , as a function of viscosity.

Braun and Parrott showed that the dissolution rate of benzoic acid is inversely proportional to the viscosity of the dissolution medium using various concentrations of sucrose and methylcellulose solutions (Fig. 6-17).²⁴

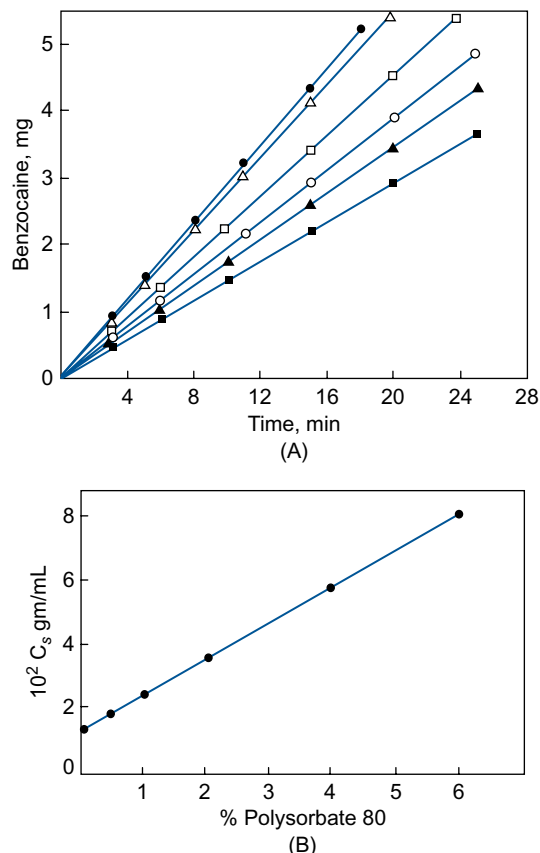


Figure 6-16. Effect of surfactants on dissolution rate.¹⁷ A. Dissolution data for benzocaine in different concentrations of polysorbate 80 using the propeller-driven stirrer apparatus at a stirring speed of 150 rpm. Polysorbate conc: ○, 6%; △, 4%; □, 2%; ●, 1%; ▲, 0.5%; ■, 0%. B. Solubilization data for benzocaine in different concentrations of polysorbate 80.

MISCELLANEOUS FACTORS

In addition to the factors discussed earlier, there are several other factors that can affect the dissolution characteristics of the drug product.

Adsorption

The adsorbent has an influence on the dissolution rate of a slightly soluble solid. It was also reported that the adsorbent is capable of increasing the dissolution rate observed in water under conditions of a decreased concentration gradient applying Nernst-Brunner film theory. Maximum dissolution rate can be obtained when a constant-concentration gradient is maintained. Adsorption isotherms can be employed to calculate the approximate amount of adsorbent required to increase the slower dissolution rate.

Sorption

The effect of water sorption on disintegration and dissolution properties, among other physical properties, of tablets containing microcrystalline cellulose was examined. It was concluded that water sorption from the atmosphere into the tablet containing microcrystalline cellulose is a very rapid first-order process, resulting in substantial changes in the physical properties. These changes are attributed to the breaking of the hydrogen bonds. The relative density of the tablets was found to decrease, resulting in increased disintegration time with increase in water sorption-rate constants. These changes were found irreversible.

Humidity

In relation to the dissolution rate of a drug substance, humidity is usually associated with storage effects. Moisture has shown

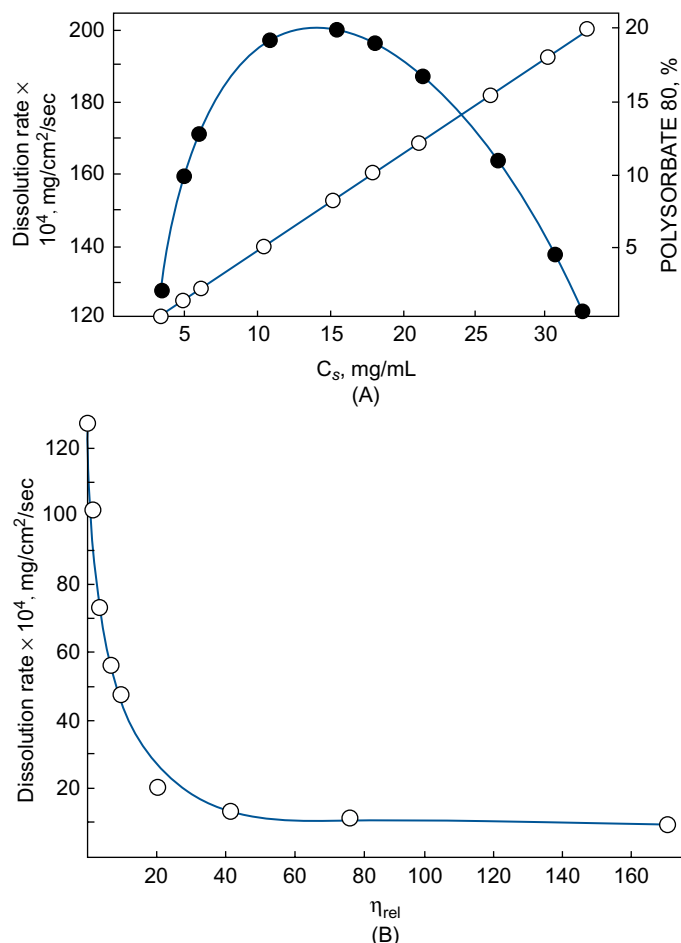


Figure 6-17. Effect of viscosity on dissolution rate.¹⁸ A. Relationship of total solubility (C_s) of benzoic acid at 25° to dissolution rate and concentration of polysorbate 80. •, rate; •, concentration. B. Relationship of viscosity to dissolution rate of benzoic acid in aqueous methylcellulose solutions at 25°.

to influence the dissolution rate of many drugs from solid dosage forms. Environmental conditions to which dosage forms are exposed, moisture in particular, should be rigorously assessed, if reproducible and reliable dissolution data are to be obtained. Additionally, humidity during the manufacture of the dosage forms should be carefully controlled to reproduce the quality of the product from batch to batch.

Detection Errors

Two most common variables leading to inter-laboratory disagreement are the failure to use standards during analysis and external vibration.²⁵ Extreme care must be exercised when laboratory methods are introduced into quality control to ensure that no part of the equipment interferes with sensitive determinations.

Despite the fundamental relationship between bioavailability and dissolution rate, the present evidence suggests that no single dissolution-rate test can be applied to all drugs. The possibility that a single test may be applied to drugs having similar physicochemical properties remains to be established. These observations are attributable, primarily, to the inability to assess and control the many variables affecting the dissolution process of a drug substance.

Sources of Variability

Compendial Methods

When selecting apparatus for dissolution testing, routine quality control, new drug development, or complying with regulatory

requirements, the analyst must follow the latest issue of compendia, including revisions. The modifications introduced in the dissolution testing methods, during recent years, are so numerous that even revisions two or three years old may be outdated.

USP/NF Method 1 (Rotating Basket Method) The USP/NF rotating basket method of dissolution testing essentially consists of a 25.0 ± 3.0 mm diameter $\times$ 37.0 ± 3.0 mm high stainless-steel 40-mesh wire basket, rotated at a constant speed ranging from 25 to 150 rpm. It is immersed in 900 ml of dissolution medium in a vessel of 1000 ml capacity. The medium in the vessel is maintained at a constant temperature of $37 \pm 0.5^\circ\text{C}$ by means of a suitable water bath. The environment in which the apparatus is placed should not contribute significant motion, agitation, or vibration to the assembly. A fitted cover may be used to retard evaporation. The shaft is positioned so that its axis is not more than 2 mm at any point from the vertical axis of the vessel and rotates smoothly without any significant wobble (Fig. 6-18).

The dosage unit is placed in a dry basket at the beginning of each test. Distance between inside bottom of the vessel and the basket is maintained at 25 ± 2 mm during the test.

In case of non-disintegrating dosage forms, this apparatus is superior to Apparatus 2, since it constrains the dosage form in steady state fluid flow. This method may seem inferior for

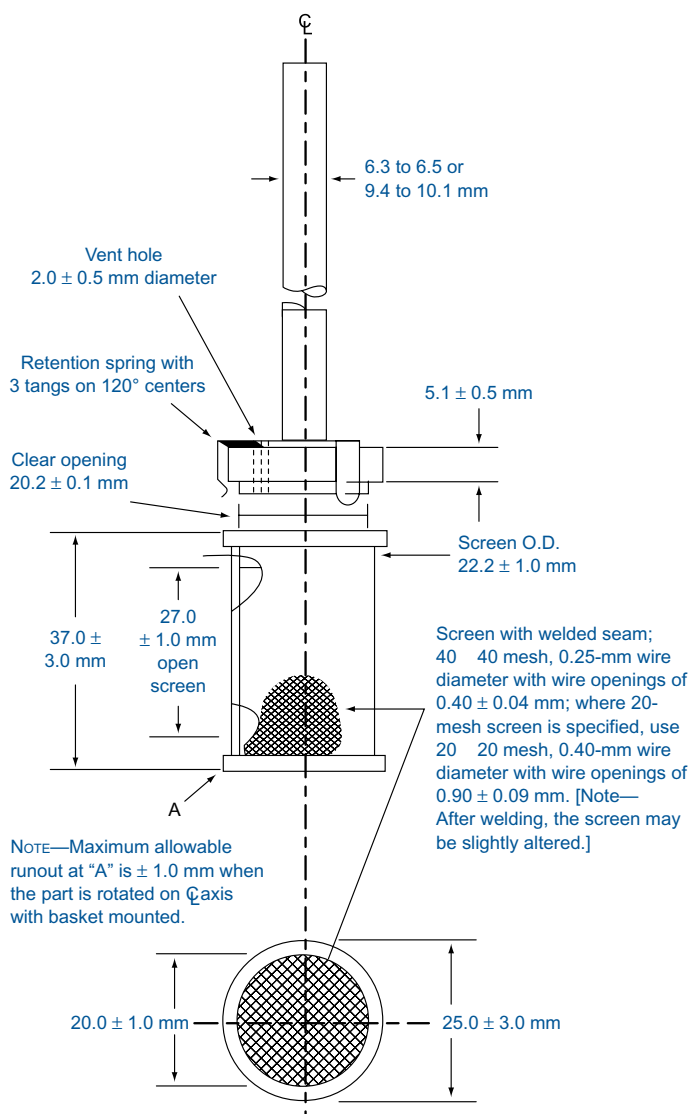


Figure 6-18. USP Apparatus 1.

testing of dosage forms, which contain gums, due to the clogging of screen matrix. In case of floating dosage forms, this method performs well, but care should be taken that excipients do not clog the basket mesh.

USP/NF Method 2 (Rotating Paddle Method)—For all practical purposes, the compendial specifications outlined for this method are identical to method 1, except that the paddle is substituted for the rotating basket.

The metallic or suitably inert, rigid blade and shaft comprise a single entity. The paddle and blade shaft may be coated with suitable inert coating. The dosage form is allowed to sink to the bottom of the vessel before rotation of the blade is started. This apparatus is frequently used for both disintegrating and non-disintegrating dosage form at 50 rpm. Other agitation speeds are acceptable with proper justification.

USP/NF permits variation in the paddle method involving the use of a helix of non-reactive material as a “sinker” for floating dosage forms. Anchoring accomplished by such a device has been severely studied (Fig. 6-19).

In cases of basket and paddle methods, “cone formation” problem has been reported in cases of some of the products. A dosage form containing high amounts of insoluble excipients is expected to form a dense mass at the bottom of the vessel. This cone formation was observed for both poorly and highly soluble drugs, but it has more impact on poorly soluble drugs. The cone formation is less pronounced in cases of basket method, because the dosage is placed in the basket, instead of being dropped at the bottom of vessel, as in paddle method. To eliminate this problem mostly seen in paddle method and,

hence, to improve the reproducibility of the dissolution test, several modifications to the paddle method were suggested and investigated. They are the tilted vessel, the Peak™ vessel with a cone-shape molded into the bottom of the vessel, metal strip, crescent-shaped spindles, mega-paddle, and various propeller shapes. There was discussion regarding use of the peak vessel, because it is not standard equipment. Peak vessel use requires documented justification. It was pointed out that increasing paddle speed often eliminates the cone, but there are limits as to how high the paddle speed can go without losing discriminatory power. Although a comment was made that reducing the amount of excipients in the formulation could reduce or eliminate the cone, it was pointed out that cone formation and dissolution should not be the driving force to modify a formulation. A flow-through cell apparatus was suggested, as an alternative to minimize cone formation.

USP/NF Method 3 (Reciprocating Cylinder)—The assembly consists of a set of cylindrical, flat bottomed glass vessels; a set of glass reciprocating cylinders; stainless steel fittings (type 316 or equivalent) and screens that are made of suitable non-sorbing material and non-reactive material designed to fit the top and bottoms of the reciprocating cylinders; and a motor and drive assembly to reciprocate the cylinders vertically inside the vessels and, if desired, index the reciprocating cylinders horizontally to a different row of vessels. The vessels are immersed in suitable water bath of any size that permits holding the temperature at $37 \pm 0.5^\circ\text{C}$ during the test. The components conform to the specifications, as shown in Figure 6-20, unless otherwise specified in the individual monograph.

One advantage of reciprocating cylinder is that gastrointestinal tract conditions can be easily simulated, as it is easy to make time dependent pH changes. This apparatus is most suitable for non-disintegrating (extended release) or delayed-release (enteric coated) dosage forms.

USP Apparatus 4 (Flow-Through Cell)—The assembly consists of a reservoir and a pump for dissolution medium; a flow-through cell; and a water bath that maintains dissolution medium at $37 \pm 0.5^\circ\text{C}$. The pump forces the dissolution medium upwards through the flow-through cell. The pump has a delivery range between 240 and 960 ml/hr, with the standard flow rates of 4, 8, and 16 ml/min. It must be volumetric to deliver constant flow independent of flow resistance in the filter device; the flow profile is sinusoidal with a pulsation of 120 ± 10 pulses/min.

The components conform to the specifications, as shown in the Figure 6-21, unless otherwise specified in the monograph.

The advantages of flow through cell apparatus most often cited are the ability to test drugs of very low aqueous solubility in the open loop mode and the ability to change the pH conveniently during the test. The disadvantage associated with it might be the operational difficulties of preparing large volumes of medium for operation in the open loop mode and the added time in the system set up and cleaning.

USP Apparatus 5 (Paddle Over Disk)—The Apparatus 2 is used with the addition of a stainless steel disk assembly, designed for holding the transdermal system at the bottom of the vessel. Temperature is maintained at $32 \pm 0.5^\circ\text{C}$. A distance of 25 ± 2 mm between the paddle and blade and the surface of the disk assembly is maintained during the test. The vessel may be covered during the test to minimize evaporation. Disk assembly for holding the transdermal system is designed to minimize any “dead” volume between the disk assembly and the bottom of the vessel. Disk assembly holds the system flat and is positioned such that the release surface is parallel with the bottom of the paddle blade. (For more specifications, refer to Fig. 6-22.)

USP Apparatus 6 (Rotating Cylinder)—The vessel assembly used is the same as Apparatus 1, except the basket and the shaft is replaced with a stainless steel cylinder stirring element to maintain the temperature at $32 \pm 0.5^\circ\text{C}$ during the test. The shaft and cylinder components of the stirring element are fabri-

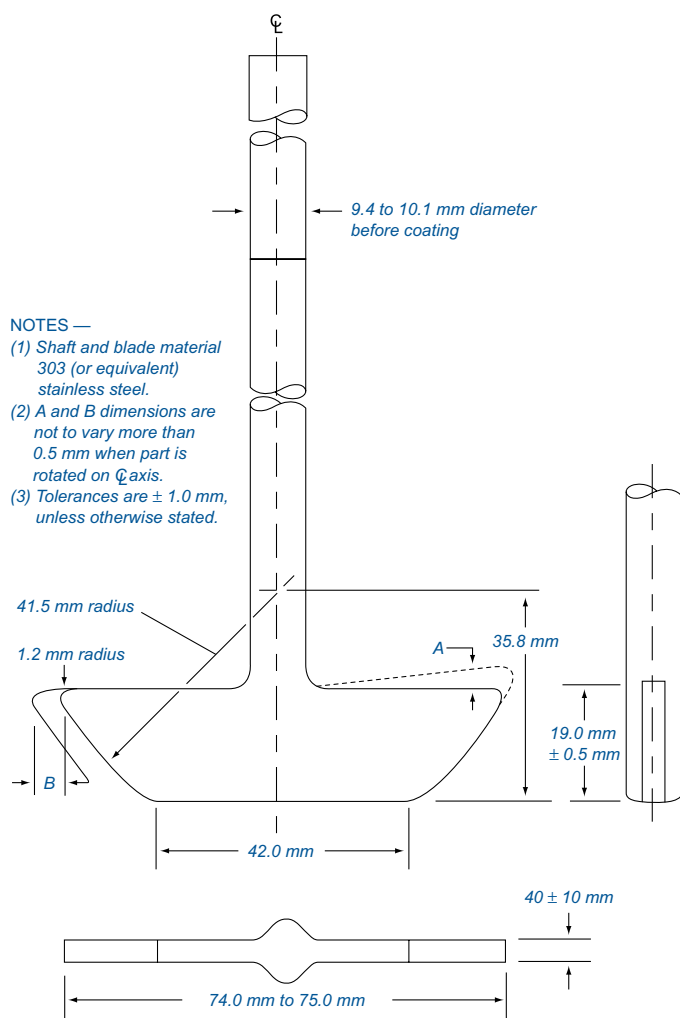


Figure 6-19. USP Apparatus 2.

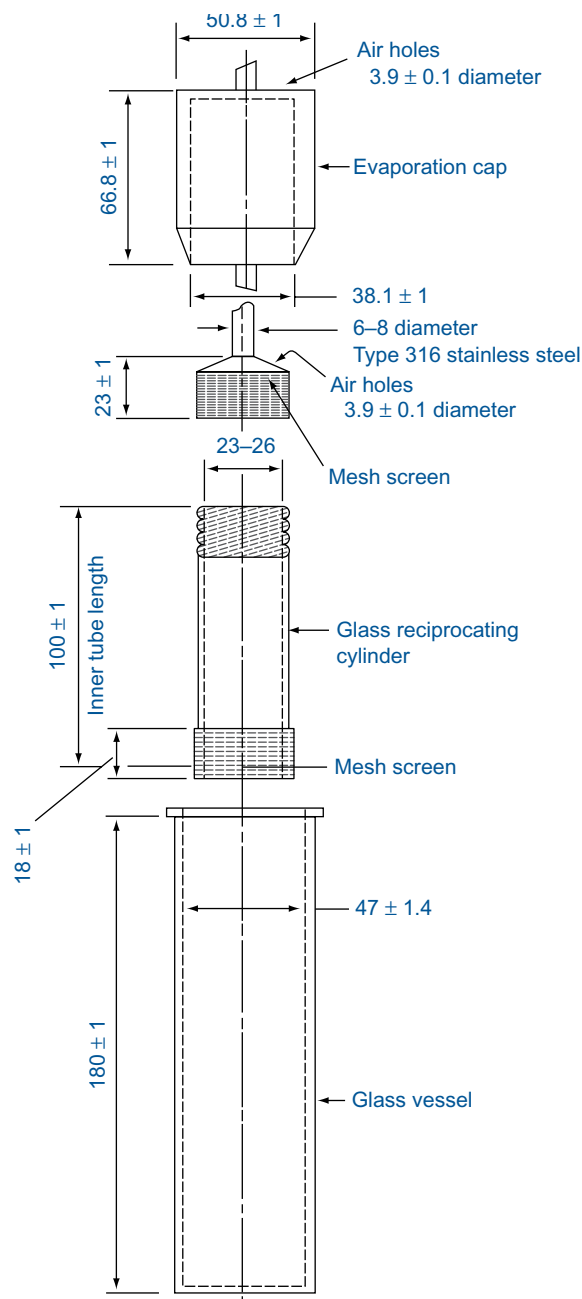


Figure 6-20. USP apparatus 3. (All measurements are expressed in mm unless noted otherwise.)

cated of stainless steel to the specifications, as shown in Figure 6-23. The dosage units are placed on the cylinder at the beginning of each test. The distance between the inside of the vessel and the cylinder is maintained at 25 ± 2 mm during the test.

USP Apparatus 7 (Reciprocating Cylinder)—The assembly consists of a set of volumetrically calibrated or tared solution containers made of glass or other suitable inert material; a motor and drive assembly to reciprocate the system vertically and to index the system horizontally to a different row of vessels automatically, if desired; and a set of suitable sample holders. (For details on specifications, refer to the Fig. 6-24.)

DISSOLUTION OF IMMEDIATE RELEASE SOLID ORAL DOSAGE FORMS

In vitro dissolution tests for immediate release solid oral dosage forms, such as tablets and capsules, are used to (1) assess the lot-to-lot quality of a drug product; (2) guide development

of new formulations; and (3) ensure continuing product quality and performance.

For the drug approval process, it is essential to have the current knowledge about solubility, permeability, dissolution, and pharmacokinetics of a drug product. Based on drug solubility and permeability, the following Biopharmaceutical Classification System (BCS) is recommended in the literature:²⁷

Class 1: High solubility-High permeability drugs

Class 2: Low solubility-High permeability drugs

Class 3: High solubility-Low permeability drugs

Class 4: Low solubility-Low permeability drugs

This classification can be used as a basis for setting *in vitro* dissolution specifications and *in vivo/in vitro* correlation (IVIVC). The BCS suggests that, for high solubility, high permeability (Class 1) drugs and, in some cases for high solubility, low permeability (Class 3) drugs, 85% dissolution in 0.1N HCl in 15 minutes can ensure that bioavailability is not limited by dissolution. In case of low solubility, high permeability drugs (Class 2), drug dissolution may be the rate-limiting step for drug absorption, and an IVIVC may be expected. A dissolution profile in multiple media is recommended for drug products in this category. In case of high solubility, low permeability drugs (Class 3), permeability is the rate controlling step, and a limited IVIVC may be possible, depending on the relative rates of dissolution and intestinal transit. Drugs in low solubility, low permeability (Class 4) present significant problems for oral drug delivery.

The regulatory acceptance of *in vitro* (dissolution) testing as a reliable surrogate for an *in vivo* bioavailability study is commonly referred to as “biowaiver.” Biowaiver may be granted for BCS Class 1 (high solubility, high permeability) drug products for bioequivalence studies, if the drug product is rapidly dissolving. The drug product is considered rapidly dissolving, if not less than 85% of the labeled amount of the drug substance dissolves within 30 minutes, using US Pharmacopeia (USP) Apparatus I at 100 rpm (or Apparatus II at 50 rpm) in a volume of 900 ml or less in each of the following media: (1) 0.1 N HCl or Simulated Gastric Fluid USP without enzymes; (2) a pH 4.5 buffer; and (3) a pH 6.8 buffer or Simulated Intestinal Fluid USP without enzymes.²⁸

DISSOLUTION OF POWDERS

There is no official method in the USP for dissolution testing of powders. The only application of powder dissolution in the USP is the evaluation of the intrinsic dissolution of powders in general chapter <1087> of the USP 34. However, in this method, the powder is pressed into a tablet, like a disk with a defined surface. The dissolution from the surface is evaluated. Dissolution testing of finely divided particles can be performed using paddle method or flow-through cell method.

DOSAGE FORMS FOR ORAL CAVITY

Dosage forms for the oral cavity, such as sublingual tablets, buccal tablets, chewing gums, and chewable tablets, are solid dosage forms placed in the mouth, allowing the active ingredient to dissolve in the saliva and then absorb either via the oral route or by the buccal/sublingual mucosa within the mouth.^{29,30}

- **Chewable tablets:** USP has stated the need to use paddle method for chewable tablets, with the exception of ampicillin chewable tablets where basket method is suggested, and for carbamazepine chewable tablets, where both USP apparatus 2 and 3 are suggested. There have been suggestions to use USP apparatus 3, a reciprocating cylinder, along with glass beads, to create a large amount of agitation within the dissolution medium.
- **Buccal/Sublingual tablets:** In general, the drug release studies from buccal/sublingual tablets has been carried out using USP apparatus 2 (paddle). Modified Franz diffusion cell also has been suggested for these dosage forms.
- **Chewing Gums:** The USP has not yet created an apparatus

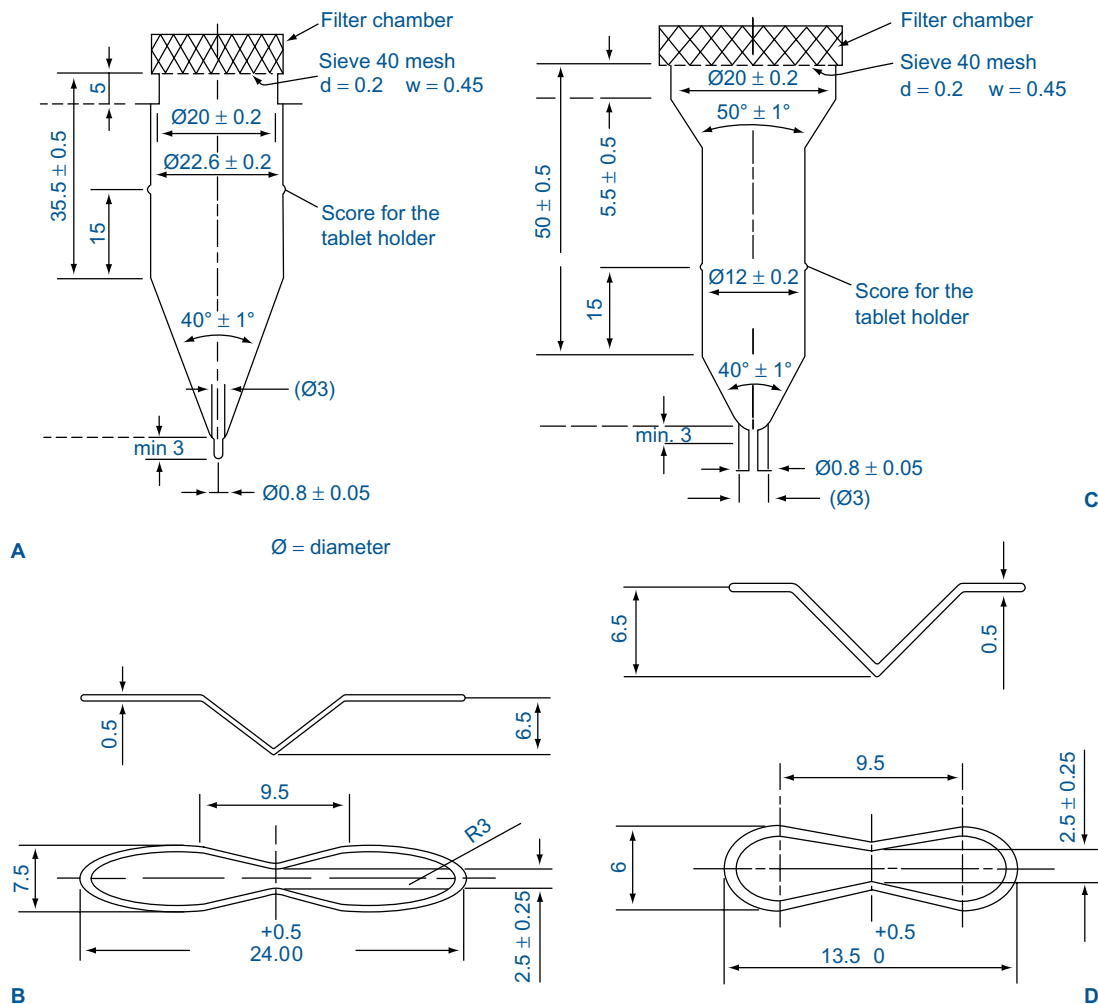


Figure 6-21. USP apparatus . **A.** Large cell for tablets and capsules. All measurements are expressed in mm unless noted otherwise. **B.** Tablet holder for the large cell All measurements are expressed in mm unless noted otherwise. **C.** Small cell for tablets and capsules. All measurements are expressed in mm unless noted otherwise. **D.** Tablet holder for small cell. All measurements are expressed in mm unless noted otherwise.

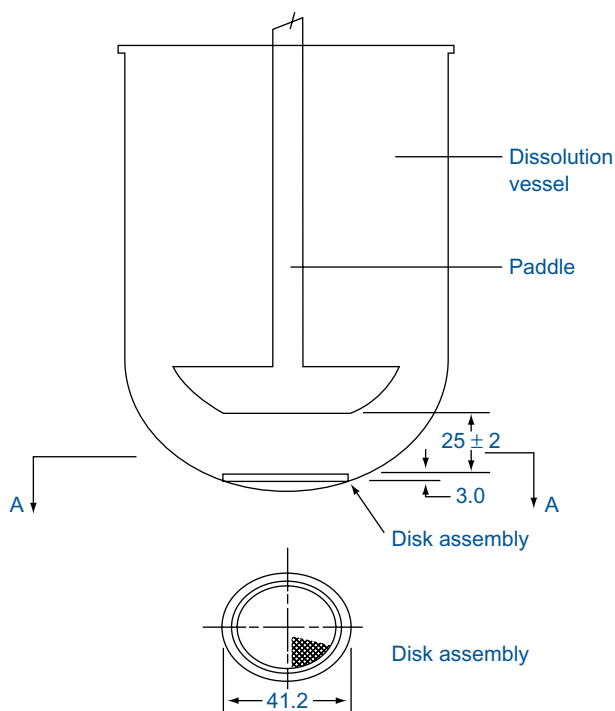


Figure 6-22. USP apparatus 5 (Paddle over disk). (All measurements are expressed in mm unless noted otherwise.)

to test the release of medication from chewing gums. However, the *European Pharmacopoeia* provides a description of a stainless steel 3-piston-apparatus that is required for testing of “medicated chewing gums.” The test is typically operated at 37°C and at 60 cycles/min. Test media with a pH of 6 are commonly used, since this pH corresponds to reported saliva pH values of 6.4 (adults) or 7.3 (children). However, to date, there has been insufficient international experience with this apparatus to draw a firm conclusion about its suitability.

DISSOLUTION OF ORALLY DISINTEGRATING TABLETS³¹

Orally disintegrating tablets (ODT) are solid dosage forms that disintegrate in the oral cavity, leaving an easy to swallow residue. ODT have high porosity, low density, and low hardness. The time for disintegration for ODT is considered less than 1 min. Development of dissolution methods for ODT is comparable to the approach taken for conventional tablets, except when the tablets utilize taste masking. Media that can be used are 0.1 N hydrochloric acid, pH 4.5 and 6.8 buffers. The most commonly used apparatus for running dissolution test for ODT is USP Apparatus 2 (paddle method) with a paddle speed of 50 rpm. USP Apparatus 1 is less frequently used, due to the physical properties of these tablets, as the tablet fragments or disintegrated tablet masses may become trapped in the basket, yielding poorly reproducible dissolution profiles. Since dissolution for ODT is very fast, slower speeds are employed. In

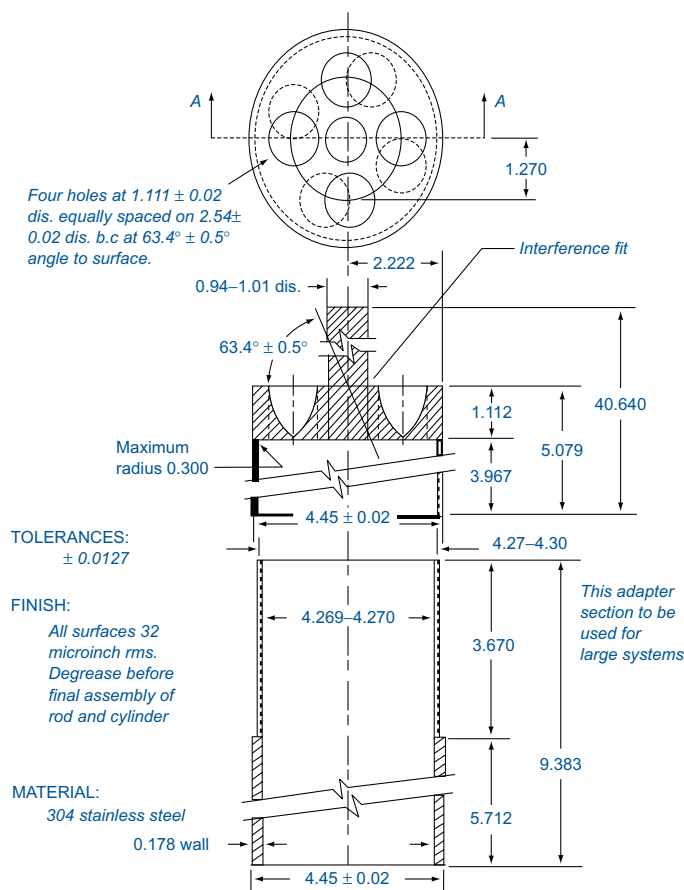


Figure 6-23. USP apparatus 6. (All measurements are expressed in cm unless otherwise noted.)

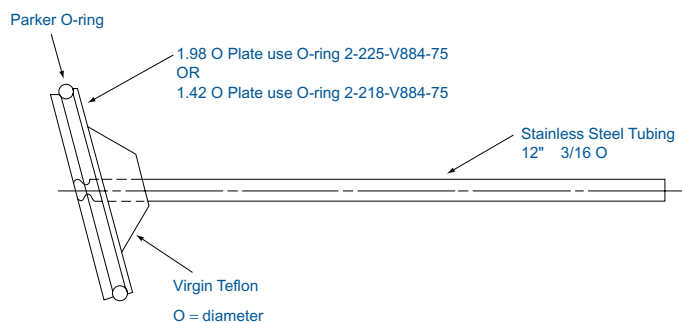


Figure 6-24. USP apparatus 7.

the case of tablets exceeding 1 gram and containing relatively denser particles, larger mounds may be produced on dissolution, which may be prevented by using higher paddle speeds. These two situations expand the suitable range to 25–75 rpm.

DISSOLUTION OF TOPICAL DOSAGE FORMS³²

Drug-release studies from gels, creams, and ointments are becoming an important step, both during the developmental stages of new formulations and as a routine quality control test for assuring the uniformity of the finished product. Also, these studies can often provide useful information on some physicochemical parameters involved in the *in vivo* percutaneous absorption, such as the diffusion coefficient and the solubility of the drug in the specific vehicle used.

Although many investigators have conducted drug release-rate studies from topical dosage forms, it appears that no single apparatus or procedure has emerged as the most favored or accepted as a quasi-standard for others in the field. According to FDA guidelines, the most commonly used method is as follows:

In vitro dissolution method for topical dosage forms is based on an open chamber diffusion cell system, such as a Franz cell system, fitted usually with a synthetic membrane. The test product is placed on the upper side of the membrane, in the open donor chamber of the diffusion cell, and a sampling fluid is placed on the other side of the membrane, in a receptor cell. Diffusion of drug from the topical product to and across the membrane is monitored by assay of sequentially collected samples of the receptor fluid.

Aliquots removed from the receptor phase can be analyzed for drug content by high-pressure liquid chromatography (HPLC) or other analytical methodology.

DISSOLUTION OF SUSPENSIONS

Although most dissolution studies during the last two decades have concentrated on tablets and capsules, some studies have pointed to the importance of the dissolution characteristics of drugs administered in suspension. This is hardly surprising, as suspensions are similar to the disintegrated form of tablets and capsules; if dissolution has become a priority for these formulations, it is logical to extend its concept to suspensions. Indeed, several studies show that the absorption of several poorly soluble drugs administered in suspension formulations are dissolution rate-limited.

Such *in vivo/in vitro* correlation studies have confirmed the importance and the viability of dissolution rate determinations of suspensions, as a discriminative test for rapid screening of new formulations and to control lot-to-lot variability within the same manufacturer and between different commercial manufacturers. In general, most of the dissolution apparatuses described for tablets and capsules could easily be used for suspensions.

The USP Apparatus 2 (paddle method) has frequently been used at a rotation speed between 25 to 50 rpm. However, the rotating filter apparatus by Shah has gained wide acceptance for suspensions, because it provides mild laminar liquid agitation and it functions as an *in situ* non-clogging filter. Sufficient volume of the dissolution medium should be used to maintain sink condition (about 900–1000 mL), and a temperature of 37°C should be maintained.

DISSOLUTION OF SUPPOSITORIES

Although most of the early work on suppositories has been concerned with their physical characteristics, such as softening and liquefaction ranges, homogeneity, smoothness, and neutrality, several reports appeared in the early literature pointing to the direct correlation between their efficacy and the release characteristics of the active ingredients. It has been reported that fatty bases, such as cocoa butter, tend to release hydrophobic drugs, which are highly soluble in the oily base, very slowly. Emulsification of the fatty base significantly improved the drug-release rate. Incorporation of surface-active agents was found to improve the release rate of water-soluble drugs from the fatty suppository base dramatically.

Although many investigators have conducted extensive research on the release of drugs from suppositories, no single method or apparatus design has emerged as the standard procedure for the pharmaceutical laboratory. Many methods for the determination of the dissolution rate of suppositories are based on the dialysis technique, where the suppository is placed in a dialyzing bag made of special membrane or cellophane material. The bag is placed in a beaker or wide-mouth bottle, containing a known volume of distilled water, and then concentration of the drug outside the bag is measured as a function of time.

A slight variation of the basket method of the USP Dissolution Apparatus 1 is also used frequently. Hanson Research markets a basket apparatus for suppository dissolution testing. Hanson's modified basket uses slots, instead of mesh, to provide a suitable porosity. The use of such a basket avoids the blocking of the mesh opening of the regular USP basket, when oil-based

suppositories are used. The system also has the advantage of being capable of testing suppositories that float or have such low specific gravity that it interferes with the flow dynamics of the paddle method.

DISINTEGRATION AND DISSOLUTION OF DIETARY SUPPLEMENTS³³

USP 34/NF 29 contains a section on dietary supplements. One of the specifications appearing in the monographs for some of the supplement dosage forms is Disintegration and Dissolution, 2040. The dissolution procedures for the nutritional supplements use Apparatus 1 and Apparatus 2 and require measurement of one vitamin and folic acid, if applicable, and one mineral, if applicable. Oil-soluble vitamins are exempt from the dissolution requirement.

DISSOLUTION OF MODIFIED RELEASE DOSAGE FORMS³²

Extended Release

In addition to application/compendial release requirements, multipoint dissolution profiles should be obtained in three other media, for example, in water, 0.1N HCl, and USP buffer media at pH 4.5 and 6.8 for the drug product. Adequate sampling should be performed, for example, at 1, 2, and 4 hours and every 2 hours thereafter, until either 80% of the drug from the drug product is released or an asymptote is reached. A surfactant may be used with appropriate justification.

Delayed Release

In addition to application/compendial release requirements, dissolution tests should be performed in 0.1 N HCl for 2 hours (acid stage), followed by testing in USP buffer media, in the range of pH 4.5–7.5 (buffer stage) under standard (application/compendial) test conditions and two additional agitation speeds using the application/ compendial test apparatus (three additional test conditions). If the application/compendial test apparatus is the rotating basket method (Apparatus 1), a rotation speed of 50, 100, and 150 rpm may be used, and, if the application/compendial test apparatus is the rotating paddle method (Apparatus 2), a rotation speed of 50, 75, and 100 rpm may be used. Multipoint dissolution profiles should be obtained during the buffer stage of testing. Adequate sampling should be performed, for example, at 15, 30, 45, 60, and 120 min, following the time from which the dosage form is placed in the buffer, until either 80% of the drug from the drug product is released or an asymptote is reached.

DISSOLUTION PROFILE COMPARISONS

In the presence of minor changes, single point dissolution tests have been employed in evaluating scale-up and post-approval changes.³⁴ For major changes, a dissolution profile comparison performed under identical conditions for the product before and after the change is recommended. Dissolution profile comparison may be carried out using the model dependent or model independent methods. One such model independent approach is subsequently explained.

MODEL INDEPENDENT APPROACH USING A SIMILARITY FACTOR

This approach uses a difference factor (f_1) and a similarity factor (f_2) to compare the dissolution profiles. The difference factor (f_1) calculates the percent (%) difference of the two curves at each time point and is a measurement of the relative error between the two curves:

$$f_1 = \left\{ \left[\sum_{t=1}^n |R_t - T_t| \right] / \left[\sum_{t=1}^n R_t \right] \right\} \cdot 100 \quad (13)$$

where n is the number of time points, R_t is the dissolution value of the reference batch (pre-change) at time t , and T_t is the dissolution value of the test (post-change) batch at time t .

The similarity factor (f_2) is a logarithmic reciprocal square root transformation of the sum of squared error and is a measurement of the similarity in the percent (%) dissolution between the curves:

$$f_2 = 50 \cdot \log \left\{ \left[1 + (1/n) \sum_{t=1}^n |R_t - T_t|^2 \right]^{-0.5} \right\} \cdot 100 \quad (14)$$

To calculate the difference and similarity factor, first, the dissolution profile should be done for 12 units each of the pre-change and the post-change products. The difference factor (f_1) and similarity (f_2) can be calculated using the mean dissolution values from both curves at each time interval. For the curves to be considered similar, f_1 values should be close to 0, and f_2 values should be close to 100. This model independent method is most suitable for dissolution profile comparison, when three or four more dissolution time points are available.

AUTOMATION IN DISSOLUTION TESTING

Due to the large amount of testing required in determining dissolution rate of drugs, automation of the process seemed almost a necessity, not simply a convenience to the analyst. Also, because of modular nature of the dissolution apparatus, automation can be easily accomplished in different ways and by various techniques.

At present, however, the setup of the apparatus, media preparation, and introduction of the dosage forms are mostly done manually. The rest of the process—including the withdrawal of samples, maintenance of a certain pH or of sink conditions, assay performance, and data acquisition and calculations—is, in most cases, fully automated. The automation process not only saves money, time, and effort on the part of the analyst, but, more significantly, it improves the overall reliability and enhances the reproducibility of testing procedures.

Several commercial companies have also introduced semi- and fully-automated dissolution systems. Some of these are the Hanson Research Dissolution System (Northridge, CA; Dissoette and Dissograph apparatuses), Technicon (Tarrytown, NY; Sasdra apparatus), and Applied Analytical (Wilmington, NC).

Millipore's Waters Chromatography Division has introduced a fully automated dissolution system, using a Waters pump, detector, and autosampler, combined with a Hanson Research's dissolution bath and sample transfer system. Samples are analyzed by HPLC, which provides better specificity than ultraviolet (UV) methods of analysis.

Hewlett-Packard manufactures a fully automated dissolution sampling and UV analysis system that can analyze samples from three dissolution baths. One such system is model 2100 C dissolution test system (Fig. 6-25).³⁵ The Model 2100C combines enhanced features with advanced communications to ensure reproducibility and control throughout the dissolution testing process. The operations are technician-friendly. Its convenient vessel layout simplifies manual or automatic sampling. It has built-in height adjustment and permanent centering, which reduces operator errors. It also has chuckless spindles to reduce setup time. The precision control of variables provides most accurate and repeatable results. The unique water bath flow characteristics maintain vessel temperature to better than $\pm 0.1^\circ\text{C}$. It has versatile SystemLink™ for PC communication and printer output and can be configured for use with both basket and paddles.

VALIDATION OF DISSOLUTION METHOD

In general, the approach to validation of a dissolution method is similar to that of any other method. The following discussion briefly summarizes the approach to dissolution assay validation.³⁶

Specificity/Placebo Interference

It is necessary to demonstrate that the results are not unduly affected by placebo constituents, other active drugs, or degradates.



Figure 6-25. Distek dissolution system.

Linearity, Range, Filter Bias, and Recovery Studies

The linearity of the detection method, range, the filter bias, and the recovery of drug from dissolution fluid containing placebo should be determined. System suitability tests for UV-Vis and chromatographic methods are also identified at this stage.

Precision and Ruggedness

Precision testing of the dissolution method should be performed on at least two lots of six tablets each on two days. The average of each run, as well as the standard and relative standard deviations, should be computed. Precision of the dissolution method is usually expressed as the standard deviation for a data set obtained on a single day.

Robustness

The evaluation of robustness, which assesses the effect of making small, deliberate changes to the dissolution conditions, typically, is done later in the development of the drug product. The number of replicates (typically, 3 or 6) is dependent on the intermediate precision.

Effect of Dissolved Gases

Air dissolved in the media may form bubbles that, in turn, could coat the tablets or other dosage form. This is most likely to happen as the medium is heated to test temperature (37°C). The coating can affect the drug release by altering the dissolution and disintegration or dissolution of the tablet. Accordingly, the effect of deaeration on the dissolution rate should be evaluated, or deaerated medium should be specified in the procedure. Effective methods of deaeration include vacuum filtration, helium sparging, hot water placed under a vacuum with or without sonication, and the use of a commercially available medium dispensing device.³⁷

Automation

Validation for automated systems is the same as for manual sampling. A simple experiment should be done to verify the drug does not adsorb to the apparatus tubing and to quantify system carryover.

APPLICATION OF QUALITY BY DESIGN (QbD) PRINCIPLES IN DISSOLUTION TESTING

The main purpose of QbD is to understand and control the critical process parameters (CPP) and critical material attributes (CMA) that have significant influence on the target product quality (TPQ). There has been tremendous interest and discussion surrounding the application of quality by design (QbD) principles in dissolution testing. The future goal is to eliminate

dissolution altogether, with the help of scientifically defined and properly executed experiments. For example, if particle size correlates with dissolution during development, then particle size measurement could be used in lieu of the latter. Similarly, granule size could be used as a future surrogate test for particle size distribution, which could then be used to replace dissolution testing.³⁸

It is not feasible to test various batches produced during process and product establishment in the clinic. Under QbD, variants would be deliberately made at the extremes of the proposed design space to look for variability, rather than trying to make all batches equivalent. It is recognized that the dissolution test can be used as a surrogate for *in vivo* bioavailability studies. The present bioequivalence guidelines and biopharmaceutical classification system (BCS) provide a platform for application of *in vitro* dissolution, as a surrogate for clinical quality (Fig. 6-26). However, to support using a dissolution test as a surrogate for clinical quality in QbD, a much higher level of understanding, which incorporates a link to clinical quality in regulatory documentation, is required. It is, especially, essential to establish a link between *in vitro* tests and the safety and efficacy (volunteer PK). Once this link has been established, the *in vitro* (dissolution) test can be used as a surrogate of clinical performance, and the clinical performance of all variants from the design space establishment can be evaluated (Fig. 6-27). To use a dissolution test as a surrogate tool in QbD, it is very

		Solubility	
		High	Low
Permeability	High	Dissolution accepted 1	Dissolution combined with bioavailability study on most relevant manufacturing/product variables 2
	Low	Dissolution should be acceptable, provided no new excipients affecting transit or permeability 3	Bioequivalence Study Or Follow principles of BCS2 or BCS3 if can demonstrate that compound behaves more like BCS2 or BCS3 in vivo 4

Figure 6-26. Role of BCS in correlating dissolution testing to clinical performance within QbD.

		Solubility	
		High	Low
High	High	Complete dissolution within 30 minutes in most discriminating 'simple' media (physiological pH range). If slower: bioavailability data 1	Limit set based on clinical 'bioavailability' data 2
	Low	Complete dissolution within 15 minutes in most discriminating 'simple' media (physiological pH range). If slower: bioavailability data 3	Limit set on case by case basis 4

Figure 6-27. Importance of BCS class of the drug in setting dissolution specifications.

important to have an *in vitro* dissolution test that has best possible relevancy to *in vivo*. In this direction, several important developments have been made in understanding the dissolution medium composition to establish *in vivo* relevancy, so that it may play a role in QbD in the future. However, modeling of intestinal hydrodynamics, *in vivo* drug precipitation/solid state conversions and conditions for dissolution in the more distal parts of the GI tract still need to be effectively addressed to develop *in vivo* relevant dissolution methods to make QbD application in dissolution testing a reality.³⁹

REFERENCES

- Banakar UV. *Pharmaceutical Dissolution Testing*, 1st edn. New York: Marcel Dekker, 1991.
- Higuchi T. Rate of release of medicaments from ointment bases containing drugs in suspension. *J Pharm Sci* 1961; 50: 874–875.
- Carstensen TJ. *Dissolution—State of the Art 1982. Proceedings of the 2nd Wisconsin Up-date Conference*. Madison WI: Extension Services in Pharmacy, University of Wisconsin, 1982.
- Morrison AB, Campbell JA. Tablet disintegration and physiological availability of drugs. *J Pharm Sci* 1965; 54(1): 1–8.
- Wagner JG. *Biopharmaceutics and Relevant Pharmacokinetics*, 2nd edn. Hamilton: Drug Intelligence Publications, 1971.
- Abdou H. *Dissolution, Bioavailability and Bioequivalence*. Easton PA: Mack, 1989.
- Cox D *et al.* *Pharm Tech* 1978; 2(4): 40–53.
- Viegas TX *et al.* *Pharm Tech* 2001; 44–53.
- Finholt P. In: Leeson LJ, Carstensen TJ, eds. *Dissolution Technology*. Washington DC: APHA, 1974; p 108.
- Aguiar AJ *et al.* Effect of polymorphism on the absorption of chloramphenicol from chloramphenicol palmitate. *J Pharm Sci* 1967; 56(7): 847–853.
- Solvang S, Finholt P. Effect of tablet processing and formulation factors on dissolution rate of the active ingredient in human gastric juice. *J Pharm Sci* 1970; 59(1): 49–52.
- Jaminet F *et al.* [Effect of adjuvants on the speed of solubility in vitro of phenobarbital from different tablet formulae.] *Pharm Acta Helv* 1969; 44(7): 418–432 [in French].
- Alam AS, Parrott EL. Effect of aging on some physical properties of hydrochlorothiazide tablets. *J Pharm Sci* 1971; 60(2): 263–266.
- Levy G *et al.* Effect of certain tablet formulation factors on dissolution rate of the active ingredient. II Granule size, starch concentration, and compression pressure. *J Pharm Sci* 1963; 52: 1047–1051.
- Levy G, Guntow RH. Effect of certain tablet formulation factors on dissolution rate of the active ingredient. III Tablet lubricants. *J Pharm Sci* 1963; 52: 1139.
- Marlowe E, Shangraw R. Dissolution of sodium salicylate from tablet matrices prepared by wet granulation and direct compression. *J Pharm Sci* 1967; 56: 498–504.
- Murthy KS, Samyn JC. Effect of shear mixing on in vitro drug release of capsule formulations containing lubricants. *J Pharm Sci* 1977; 66(9): 1215–1219.
- Hanson W, Hanson R. *Pharm Tech* 1979; 3(3): 42–50.
- Thakker K *et al.* *Pharm Forum* 1980; 6: 177–185.
- Wurster DE, Taylor PW. Dissolution kinetics of certain crystalline forms of prednisolone. *J Pharm Sci* 1965; 54(5): 670–676.
- Nogami H. Dissolution Phenomena of Organic Medicinals involving Simultaneous Phase Changes. *Chem Pharm Bull* 1969; 17: 499–509.
- Carstensen JT *et al.* Time and temperature dependence of disintegration and correlation between dissolution and disintegration rate constants. *J Pharm Sci* 1980; 69: 290–294.
- Singh P *et al.* Mechanistic study of the influence of micelle solubilization and hydrodynamic factors on the dissolution rate of solid drugs. *J Pharm Sci* 1968; 57: 959–965.
- Braun R, Parrott E. Influence of viscosity and solubilization on dissolution rate. *J Pharm Sci* 1972; 61: 175–178.
- Cartwright AC. Sources of variation during collaborative evaluation of in vitro dissolution tests for two solid preparations. *J Pharm Pharmacol* 1979; 31: 434–440.
- Gray V *et al.* The science of USP 1 and 2 dissolution: present challenges and future relevance. *Pharm Res* 2009; 26(6): 1289–1302.
- Amidon GL *et al.* A theoretical basis for a biopharmaceutic drug classification: the correlation of in vitro drug product dissolution and in vivo bioavailability. *Pharm Res* 1995; 12: 413–420.
- Dokoumetzidis A, Macheras P. A century of dissolution research: from Noyes and Whitney to the biopharmaceutics classification system. *Int J Pharm* 2006; 321(1–2): 1–11.
- Azarmi S *et al.* Current perspectives in dissolution testing of conventional and novel dosage forms. *Int J Pharm* 2007; 328(1): 12–21.
- Siewert M *et al.* FIP/AAPS guidelines to dissolution /in vitro release testing of novel/special dosage forms. *AAPS Pharm-SciTech* 2003; 4(1): E7.
- Klancke J. Dissolution Testing of Orally Disintegrating Tablets. *Dissolution Technologies* 2003; 10(2): 6–8.
- US Food and Drug Administration (FDA). (1997). Guidance for Industry. SUPAC-MR: Modified Release Solid Oral Dosage Forms. www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM070640.pdf (accessed 10 April 2012).
- US Pharmacopeia 34/ National Formulary 29
- Moore JW, Flanner HH. *Pharm Tech* 1996; 20(6): 64–74.
- Distek Inc. <http://www.distekinc.com/> (accessed 10 April 2012)
- Skug JW *et al.* *Pharm Tech* 1996; 58:
- Rohrs BR, Stelzer DJ. Deaeration Techniques for Dissolution Media. *Dissolution Technologies* 1995; 2(2): 6–8.
- D'Souza SS *et al.* AAPS Workshop on the Role of Dissolution in QbD and Drug Product Life Cycle: A Commentary. *Dissolution Technologies* 2010; 17(4): 41–5.
- Dickinson PA *et al.* Clinical relevance of dissolution testing in quality by design. *AAPS J* 2008; 10(2): 380–390.

Chapter 7

Modern-Day Drug Discovery and Development

Rick G. Schnatz, PharmD

INTRODUCTION 81

EVOLUTION OF 21ST-CENTURY PHARMACEUTICAL RESEARCH 81

PHARMACEUTICAL RESEARCH ORGANIZATIONS 83

THE SEARCH FOR NEW DRUGS 84

NATURAL PRODUCT SOURCES 85

FUNCTIONS OF RESEARCH SCIENTISTS 85

DRUG DEVELOPMENT 88

INTRODUCTION

Medical and pharmaceutical research provides a basis for the development of new therapeutic approaches to human and animal disease. This process of drug discovery research can be basic (seeking an understanding of biological phenomena that are unknown) or applied (using principles that are known to produce a desired new product or effect). In either case, drug discovery research results from an unmet clinical need, a recognized deficit in treatment options. Drug discovery is the process by which drugs are discovered or designed. In the past most drugs were discovered either by identification of the active ingredient from traditional remedies or by serendipitous discovery. Today we know how disease and infection are controlled at the molecular and physiological level. Drug discovery involves the identification of candidates, synthesis, characterization, screening, and assays for therapeutic efficacy. Drug discovery is still a lengthy, expensive, difficult, and inefficient process with a low rate of new therapeutic discovery. The outcome of a successful drug discovery program is the generation of a therapeutic entity where none previously existed or the replacement of established therapies in favor of a newer modality that is safer and more efficacious.¹ The main function of the pharmaceutical industry is to create products (i.e., drugs that have an impact on health care). Drug development is the process of bringing a new drug to the market once a lead compound has been identified through drug discovery. Drug development includes preclinical research (microorganisms/animal) and clinical trials (human). Products of this type can be foreseen to some extent through knowledge and study and thus are amenable to planned research and development (R&D). For example, if the cause of a disease has been identified as an infection by a microorganism, a search can be undertaken for an agent that will prevent or cure the infection. However, in some instances, the etiology of a disease is unknown despite intensive investigation. In this situation, the pathway to a satisfactory cure or method of prevention cannot be foreseen or forecast. In such cases, products may only be developed after application of careful investigations, from a revolutionary new approach, or perhaps from a serendipitous finding.

Although much of the drug discovery research in the United States is carried out by major pharmaceutical manufacturers and biotechnology companies, this research is dependent on a vast and growing background of scientific knowledge generated by diverse organizations. Universities, private institutes, governmental laboratories, and industrial research all play significant roles in developing new technologies and knowledge that provides the basis for discovery and the ultimate generation of a new product. This new knowledge may involve development of a new technology, improved scientific methodology and instrumentation, or increased understanding of the basic molecular or cell biology underlying a disease.

The major objective of research in the pharmaceutical industry is to produce safe drugs that prevent, cure, or ameliorate disease. Interim research goals that lead to this major objective are to

- Understand the molecular basis of biological mechanisms in health and disease.
- Develop new biological testing procedures relevant to human and veterinary medicine.
- Develop a quantitative understanding of the interaction of drugs with key biological systems, leading to the more rational design of drugs.
- Understand the absorption, transport, and mode of action of drugs.
- Develop drugs of low toxicity, reproducible delivery, and high specificity for a given pathological state or target organ.

This chapter will touch on the above points to illustrate how drug discovery research is used to develop new products that fulfill clinical needs.

EVOLUTION OF 21ST-CENTURY PHARMACEUTICAL RESEARCH

The search for medicines to treat disease began with natural products. Up until the early part of the 20th century, pharmaceuticals were mainly derived from natural products, and the practice of gathering and preparing dried herbs was commonplace. A remedy such as menthol was extracted from peppermint and used for treating coughs and colds. Boneset tea reduced fevers, peppermint relieved an aching tooth or a colicky baby, and foxglove could revive a failing heart. Early challenges associated with natural products were the ability to develop and manufacture drugs of uniform strength and quality. Quality often varied with the raw materials or the skills of the pharmacist. Current challenges for natural products relate more to mining natural biodiversity² and meeting the synthetic challenges.³

Until World War I, most synthetic drugs and chemicals used in the United States were discovered and produced in Europe. When supplies were curtailed by the war, the impetus was provided for the establishment of an independent US chemical and pharmaceutical industry. Accordingly, production of chemicals and drugs was undertaken, which stimulated the development of industrial research. In the following years, the US pharmaceutical industry made major contributions through discovery and development of new drugs, and it assumed a place of leadership in the world.

Toward the mid-20th century, chemical research on the isolation, identification, and synthesis of drugs began to yield many important drug substances. During this time the synthesis and

manufacturing of vitamins was a major focus of companies such as Roche and Merck. Among the major drugs discovered and/or developed in the United States during this period were insulin, sulfonamides, penicillin and broad-spectrum antibiotics, cortisone and other steroid compounds, isoniazid, diuretics, and tranquilizers (now commonly referred to as anxiolytics or antipsychotics). Discovery and development of the sulfonamides, antibiotics, and other anti-infective agents dramatically reduced the death rates from a number of infectious diseases. Use of drugs like isoniazid led to a decline in the United States tuberculosis death rate between 1945 and 1984 from 39 per 100,000 people to 0.7 per 100,000 people. Since a large proportion of the deaths from these diseases had occurred prior to adulthood, these drugs allowed more individuals than ever before to mature and assume productive roles in society.

Since the early 1980s, new classes of drugs that treat hypertension and dyslipidemias have emerged and made inroads in reducing morbidity and mortality from cardiovascular disease. These are now some of the world's most efficacious, safe, and profitable drugs. According to a 2010 update by the American Heart Association (AHA), the death rate for cardiovascular disease fell 28 percent between 1997 and 2007.⁴ Cancer medicines are moving from cytotoxic agents to cytostatic agents, but there is still much work to be done in this complex area of cell signaling and rampant proliferation. New challenges have emerged in areas of dementia, given our expanded lifespan and increased awareness.

In its broadest definition, biotechnology refers to the use of living organisms, such as cells, tissues, or biological molecules such as enzymes or antibodies in the production of products having beneficial use. Biological products include a wide range of agents such as vaccines, blood and blood components, allergenics, somatic cells, gene therapy, tissues, and recombinant therapeutic proteins. In contrast to most drugs that are chemically synthesized and whose structure is known, most biologics are complex mixtures that are not easily identified or characterized. A recent survey found that America's biopharmaceutical research companies have in development 900 biotechnology medicines and vaccines, targeting more than 100 diseases. These include 352 medicines for cancer, 188 for infectious diseases, 69 for autoimmune diseases, and 39 for AIDS/HIV and related conditions.⁵ Many already approved biotechnology medicines treat or help prevent heart attacks, strokes, hepatitis, heart failure, kidney cancer, cystic fibrosis, diabetes, and various types of leukemia.

Of increasing importance as lifespan improves and the world population grows are the classes of drugs that have marked effects on quality of life without significantly affecting longevity. Beginning in the early 1990s, a new term appeared in the evolution of pharmaceuticals. The term *lifestyle drugs* appeared when insurance companies were faced with the decision of paying for products that might improve one's life, function, or appearance, as opposed to a drug one might take to manage an illness or cure a disease.^{6,7} The question was, do insurance companies pay for treatments or happiness? Looking back in time, many consider alcohol to be one of the oldest and most widely used of the lifestyle drugs. Its use can be considered an attempt to increase happiness through pharmacological activity. A turning point in the era of modern lifestyle drugs was the launch of sildenafil citrate (Viagra) in 1998 as a drug of choice for erectile dysfunction.

There are as many different definitions of "lifestyle drugs" as there are people involved in the discussion. It is very difficult to define absolutely. An operational definition of "lifestyle drugs" that has gained some traction is as follows: drugs that could modify or change non-medical or non-health-related goals or conditions at the margins of health and well-being.⁸

The term "lifestyle drugs" generally encompasses pharmaceuticals that do the following:⁶

- Treat conditions resulting from personal responsibility or behavior choices
- Aid in smoking cessation (e.g., bupropion)

- Enhance life or performance
- Treat erectile dysfunction (e.g., sildenafil citrate)
- Treat problems falling outside the medical or social definition of health
- Address male-pattern baldness (e.g., finasteride)
- Straddle medical, social, and environmental domains
- Treat social anxiety disorder (e.g., selective serotonin reuptake inhibitors (SSRIs))

An increasing number of lifestyle drugs can be expected to reach the marketplace in the near future. The debate over who will pay for them will intensify. A looming question for all society is whether pharmaceutical research dollars will be diverted from seeking those medications to treat or mitigate disease to drugs that are designed to increase happiness, pleasure, and a personal sense of well-being. All elements of society—government, industry, health care professionals, patients, and consumers—should participate in these discussions.⁹

The applications of nanotechnology and pharmacogenomics are among the newest evolutions in pharmaceutical research and could have far-reaching and paradigm-shifting implications for drug discovery. Nanotechnology came into its own in 2000 with the birth of the National Nanotechnology Institute. To date more than \$14 billion has been invested by the US government in nanotechnology. Nanotechnology is a science that involves the design and building of structures and devices smaller than 100 nanometers. It is currently being applied to all stages of drug development, from formulation chemistry for targeted drug delivery to diagnostic applications. Nanotechnology is capable of improving drug research and development by providing researchers more information about drugs and targets, allowing high content screening, decreasing volumes of expensive reagents, and enhancing solubility and selective tissue targeting. The largest share of opportunities is emerging in pharmaceutical applications, and this is expected to reach \$18 billion in 2014.¹⁰ The US National Science Foundation predicts that nanotechnology will produce half of the pharmaceutical industry product line by 2015.¹¹

The old adage of the right drug for the right patient was traditionally based on the trial and error method of matching patients with the appropriate treatment protocol. In the evolving scientific principle of pharmacogenomics, health care providers will prospectively analyze a patient's genetic profile and prescribe the best available drug therapy, based on biomarkers that will predict drug response or drug toxicity.

Sometimes referred to as "personalized medicine," pharmacogenomics is the fusing of classical pharmacology with human genetics, using a patient's genotype to optimize drug therapy and minimize toxicity.

Anticipated benefits of this new discipline include the following:¹²

- Pharmaceutical companies will produce entities with a high specificity to defined conditions, based on the proteins, enzymes, and RNA molecules associated with genes and diseases.
- Accurate determination of appropriated drug dosages will replace basing dosages on weight and age. Dosages based on a patient's genetics will maximize the therapy's value and, in all likelihood, will decrease the risk of overdosing.
- Knowledge of an individual's genetic code will allow the person to make lifestyle and environmental changes to avoid or lessen the severity of a genetic disease. This will also allow for careful monitoring and treatments that can be started at the optimal time to maximize patient benefit.
- Drug therapy, from the first dose, will be safer, speeding recovery time. It is likely that pharmacogenomics may reduce the estimated 100,000 deaths and 2 million hospitalizations that occur each year in the United States as the result of adverse drug response.¹³
- Vaccines, made of DNA or RNA, will activate the immune system but will be unable to cause infections.

Pharmacogenomics presents barriers which must be overcome before full benefits can be realized. All healthcare providers will need to gain or develop a deeper understanding of genetics in order to correctly interpret the genetic testing results and their applicability to the patient. The process adds an extra diagnostic step before providers can determine which drug will provide the most appropriate treatment option.

One latest example of personalized medicine is a genetic test that will help the practitioner predict which patient with acute coronary syndrome (ACS) requires a higher dose of clopidogrel. A study published in the *Journal of the American Medical Association* concluded that “among patients with stable cardiovascular disease, tripling the maintenance dose of clopidogrel to 225 mg daily in CYP2C19*2 heterozygotes achieved levels of platelet reactivity similar to that seen with the standard 75-mg dose in noncarriers; in contrast, for CYP2C19*2 homozygotes, doses as high as 300 mg daily did not result in comparable degrees of platelet inhibition.”¹⁴ This finding is but an example of the future that will allow practitioners to treat patients appropriately from the first dose, instead of by trial and error, thus potentially preventing a dangerous and costly negative outcome.

In June 2010 the FDA listed 6 medications for which diagnostic genetic testing was required prior to prescribing. Today there are approximately 30 drugs for which the FDA recommends diagnostics testing and more than 200 medications that contain pharmacogenomics data within the drug's labeling.

PHARMACEUTICAL RESEARCH ORGANIZATIONS

The pharmaceutical and biotechnology industries are leaders among all US industries with respect to support of R&D. The industry finances almost its entire R&D with its own funds; no other industry spends as high a percentage of R&D funds for basic and applied research. A significant portion of every sales dollar is devoted to drug research activities (Table 7-1). For instance, in 2010 US pharmaceutical companies devoted 20.5 percent of their sales revenues to R&D.¹⁵ In 2010 the average R&D spending for the top 20 pharma companies was US \$4.4 billion, or 18.2 percent of sales.¹⁶ This expenditure in part underlies the cost of prescription drugs because only 2 of 10 marketed drugs return revenues that match or exceed R&D costs.¹⁵

Technological advances have led to an explosion of small biotechnology companies that specialize in one or more steps in the preclinical and clinical processes. Often large pharmaceutical companies contract to the smaller outside companies as a way of extending their internal resources, to mitigate capacity fluctuations and to decrease costs. The trend among large

pharmaceutical companies toward outsourcing has led to the growth of many contract research organizations (CROs) that are contracted to produce biopharmaceuticals.¹⁶ Additionally, there is a movement by large pharmaceutical companies to outsource the production of early-phase clinical study drugs to compounding pharmacies. This trend is being driven in part by economics and in part by drug shortages that are affecting patient care worldwide.

The academic community plays a vital role in the development of new drugs. Its role includes, but is not limited to, research on a basic understanding of disease states, development of biochemical or physiological rationales for drug targets, the initial evaluation of new drugs, consultants with companies to use their academic and scientific expertise to guide pharmaceutical research, and, certainly not least, the training of scientists. During the late 1980s and early 1990s, scientists at universities made many basic advances toward identification of enabling technologies, which led to the founding of many biotechnology companies that continue to flourish today. Over time, platform technologies were used to enable the biotechnology-focused¹⁷ company to become a product-focused company, the true endpoint of pharmaceutical research. Examples of platform technology to product conversion are the following: (1) Ligand Pharmaceuticals, founded by Dr. Ron Evans of the Salk Institute, was based on the discovery of novel intracellular receptors and their role in gene transcription, focusing on the identification of agonists and antagonists, primarily of steroid hormone receptors. Ligand and its sister company X-Cepto continue today to work on these important drug target families. Ligand markets ONTAK, Targretin capsules, Targretin gel, and Panretin gel. Ligand's newer product, AVINZA, is a treatment for chronic, moderate-to-severe pain. In addition, Ligand's pharmaceutical partners develop products for men's and women's hormone-related diseases, osteoporosis, metabolic disorders, and cardiovascular and inflammatory diseases. (2) Vertex was founded by Dr. Joshua Boger, using the structure-based design that relies on the high-resolution molecular imaging of the active site of a disease molecule. Vertex has had several major pharmaceutical relationships, including GlaxoSmithKline, for the development and marketing of Agenerase (amprenavir) for HIV, and Kissei for p38 MAP Kinase inhibitors for use in inflammatory disease. (3) In contrast to the target-family focus brought by Ligand and X-Cepto or the structure-based design brought by Vertex, Pharmacopeia, a combinatorial chemistry company founded by Drs. Michael Wigler and Clark Still, employs chemical and biological diversity and high-throughput screening approaches to lead drug discovery. Pharmacopeia has a drug pipeline based on its collaborations with key pharmaceutical partners, such as Schering-Plough, Bristol-Myers Squibb, Celgene, and Cephalon. Thus, partnerships between academia and biotechnology, as well as between biotechnology and large pharma, have propelled both the genesis and evolution of biotechnology. It is likely that change will continue to occur at a faster pace in the smaller and more opportunistic biotechnology areas than within large pharma partners.

Clinicians and clinical scientists often lead the discovery of new uses for drugs and new directions for research based on observations made in the clinical setting. Chlorpromazine was originally synthesized as an antihistamine but was found to be useful as an antipsychotic. The clinical use of this compound, and of other central nervous system drugs, has resulted in a marked reduction in the number of mentally ill patients needing hospitalization.

Research in the academic community has been supported to a major extent by agencies of the US government, such as the Public Health Service (PHS), the National Institutes of Health (NIH), the Centers for Disease Control and Prevention (CDC), and the National Science Foundation (NSF). The pharmaceutical industry also contributes financial support to academic laboratories where research of general or specific interest to the industry is conducted. Institutes established by private

Table 7-1. Leading Pharmaceutical Companies Ranked by R&D Spending (US\$ in Billions) in 2010

Company	R&D Spending	Pharma Sales	R&D as % of Sales
Merck	11	39.8	27.6
Pfizer	9.4	58.5	16.0
Roche	8.6	39.1	22.0
Novartis	7.1	42.0	16.9
GlaxoSmithKline	6.1	36.2	16.9
Sanofi-Aventis	5.1	40.3	12.7
Eli Lilly	4.9	21.1	23.2
Johnson & Johnson	4.4	22.4	19.6
AstraZeneca	4.2	33.3	12.6
Abbott	3.7	19.9	18.6

From Cacciotti J, Clinton P. The Lull between two Storms, *Pharmaceutical Executive*, May 2011. www.oliverwyman.com/media/2011_PharmExec_50.pdf; accessed 20 April 2012.

endowment such as the Sloan-Kettering Institute, Shriners Children's Hospitals, the Institute for the Study of Aging, and the Gates Foundation all pursue basic and applied research in many fields related to the public health. Many hospitals also maintain research clinics and/or privately or publicly endowed foundations to pursue causes and treatment of specific diseases, a related group of diseases, diseases endemic to a certain geographical area, or groups of diseases affecting a certain organ of the body. Because research does not depend on the vending of items or services, it is not immediately self-supportive and must be supported by public as well as private funds.

Interest in pharmacoepidemiological research has prompted the development and need for review criteria in this area. *The Hartzema Guide*¹⁸ makes use of case-controlled and cohort studies as major methodologies in this field. Generally, the evaluation criteria for case-controlled and cohort studies address proper sample-frame definition, compatibility of cases and controls, drug-exposure validations, unintended-effect ascertainment procedures, and related considerations. Although these can be confusing, Hartzema provides interpretation of the statistics used in reporting case-controlled and cohort studies and gives review criteria for meta-analysis, an approach to integrating the pharmacoepidemiological literature.

THE SEARCH FOR NEW DRUGS

Until the early 20th century, most useful drugs, such as morphine, quinine, digitalis, ergot, and atropine, to name a few, were derived from plant sources, and their therapeutic uses were based on serendipitous discoveries. As the science of medicinal chemistry has evolved, screening of natural products has become more methodical. Screening of natural products is based on the concept that evolution favors molecular conservation. However, the future of natural product screening for drug discovery is currently limited by speed and compound diversity.

In the mid-20th century, useful drugs were derived from natural products, chemical syntheses, or combinations of both sources. The approaches used to identify lead molecules that evolved to drugs covered the spectrum from molecular diversity to rational design. Molecular diversity is the differences in physical properties that exist among different molecules. These properties can be expressed as differences in shape or size, polarity and charge, lipophilicity, polarizability, or flexibility.¹⁹

The application of molecular diversity techniques to drug discovery is a multidisciplinary effort requiring the skills of people of varied backgrounds, ranging from computational chemistry to engineering to organic synthesis to molecular biology.²⁰

Rational, or structure-based, drug design refers to a process that begins with a high-resolution map to the active site of a disease target. With an x-ray crystal structure or a nuclear magnetic resonance image, medicinal and protein chemists can engineer molecules to fit, or better fit, the active site. This approach is appealing, has been applied by biotechnology companies in a more high-throughput fashion, and has been successful in the field of HIV protease inhibitors and chemotherapy protein-tyrosine kinases inhibitors.

Structure-based design is not currently applicable to all classes of drug targets, however. Guanine nucleotide-coupled receptors (GPCRs), which have proven to be one of the most feasible classes of drug targets, intertwine between the extra- and intracellular surface seven times. Because of their architecture in the lipid bilayer of the cell membrane, the structure of these "heptahelical receptors" has not been solved. However, using processes described below, many successful drugs have been found that work through GPCRs.

The advent of combinatorial chemistry in the 1980s and 1990s greatly impacted drug discovery. This technology refers to the generation of compounds in sets, or libraries, that are typically chemically related and made by combining sets of reactions, such that the chemical steps are efficiently conducted. Combinatorial chemistry was initially applied to amino acids

and nucleotides by Affymax and NeXagen, respectively. Since then, several companies, predominantly Pharmacoepia and Arqule, have applied this technology to small molecules. Having large numbers of compounds allows for increases in subtle alterations in chemical relatedness as well as chemical diversity. This point is critical since the biological target is the true measure of a successful chemical interaction and use of descriptors to capture chemical diversity does not discriminate in the relevant ways that biology does. This point was well made by Jurgen Drewes²¹ when he wrote, "It is, however, by no means certain to what extent molecular diversity, as viewed by chemists and as calculated by structural descriptors, resembles diversity as 'seen' by a biological target molecule."

Collections of compounds can be designed to be "drug-like" in that they have favorable physicochemical properties in common with known drugs. These properties were first elucidated by Dr. Chris Lipinski of Pfizer and have become known as the "Lipinski Rule of Five." Based on his analysis of drugs developed predominantly at Pfizer, Dr. Lipinski described key traits necessary for a molecule to have suitability as a drug.²² These traits are less than 5 hydrogen-bonds, less than 10 hydrogen-bond acceptors, an octanol-water partition coefficient less than 5, and a molecular weight less than 500. It is important to note, however, that approximately 25 percent of the drugs in the Comprehensive Medical Chemistry (CMC) database do not follow the "Rule of Five." Typically, these exceptions are antibacterials, antineoplastics, or CNS drugs. The large percentage of exceptions would advise against overly strict adherence to these guidelines as rules. Indeed, it is rare to identify a drug in a high-throughput screen (HTS) and best to think of compound collections as sources of leads that can be optimized to drugs by medicinal chemists, biologists, and pharmacologists.

The testing of compound collections in HTSs is another area where key advancement has occurred in the pharmaceutical industry in the last 25 years. HTS allows scientists to devise biochemical assays around a molecular target using new technologies. With more sensitive high-throughput technologies focused on identification of activity at a precise molecular target, the likelihood that a compound will be identified in a high-throughput screen increases. One of the primary impacts of combinatorial chemistry and its HTS counterpart has been more efficient identification of leads from chemical libraries. However, despite its promise and efficiency, over the past 15 years, technology and HTS have not met the expectations of the pharmaceutical industry in producing significantly higher numbers of lead compounds that resulted in new chemical entities or new drug products.

CHEMICAL LIBRARIES AND SAMPLE COLLECTIONS

Prior to the advent of combinatorial chemistry, organic chemists in the pharmaceutical industry synthesized new compounds one at a time. The collection of these compounds was not particularly diverse but led to hundreds of thousands of compounds in a company's sample collection. Combinatorial chemistry has greatly increased the efficiency of compound synthesis, leading to significantly larger compound collections. The explosion in synthesis of chemical libraries necessitated more efficient testing of biological activity.

Initial screening of thousands of compounds is accomplished rapidly by the use of *in vitro* enzymatic or receptor screens. Typically, several unique active lead compounds emerge, which are studied in a variety of secondary assays, either confirming or refuting the original hypothesis.

The small molecular weight of compounds in a chemical library favors the chance of their oral availability. The Lipinski guideline recommends that compounds be about 500 daltons to optimize oral absorption across the gastrointestinal tract.²² Other guidelines include recommendations for polar surface area and hydrophobicity measures.²³ Orally available drugs are highly desirable; thus, small molecule drug discovery remains the focus of pharmaceuticals over their biological counterparts.

Whether a molecular diversity or rational design approach was followed to identify the lead molecule, drug discovery tends to proceed thereafter through an iterative process of chemical modification and biological testing. Teams of scientists improve the characteristics of their lead compound in an optimization process. If successful in building the appropriate characteristics, this process results in a drug candidate.

NATURAL PRODUCT SOURCES

In addition to compound collections, organic chemists and biochemists derive leads from natural product sources. Natural products can be derived from plant and animal sources; in the latter category microbial and marine organisms often are considered separately from ordinary domestic animals. Digitalis glycosides, such as digitalis and digoxin, derive from the foxglove plant and are powerful cardiac stimulants. The poppy plant has provided opium alkaloids (morphine, codeine) used in analgesia; and the belladonna plant provides the belladonna alkaloids (atropine and scopolamine) used as parasympathetic blockers. In addition to the plant alkaloids mentioned earlier, some important natural products include antibiotics, steroid and peptide hormones, vitamins, enzymes, prostaglandins, and pheromones.

Although serendipity plays a relatively large role in the search for natural products, rational biological inputs based on deficiency syndromes, replacement therapy, or known biological effects clearly influence the development of these drugs. Nutritionists, endocrinologists, pharmacologists, microbiologists, biochemists, and physiologists all play a vital role in understanding the underlying biological mechanisms. Antibiotics, steroids, and prostaglandins provided fertile new fields for chemical modification, leading, in all three cases, to drugs that are more useful than the parent compounds. Much research is being undertaken by the NIH and private companies on unique natural products that have anticancer properties. For example, Taxol (paclitaxel), derived from the bark of a Pacific yew tree (*Taxus brevifolia*), was developed for the treatment of ovarian cancer by Bristol-Myers Squibb. As mentioned earlier, current challenges for natural products research relate to mining natural biodiversity² and meeting the synthetic challenges.³

FUNCTIONS OF RESEARCH SCIENTISTS

The pharmaceutical industry is an outstanding example of successful collaboration between scientists of biological and physical science disciplines. Chemists and other physical scientists predominantly have been responsible for synthesis, isolation, and characterization of medicinal agents. However, biological scientists have played an equally essential role in originating meaningful screening and testing models and in the overall evaluation of new agents. Qualified specialists in many fields, including pharmacy, physics, statistics, chemistry, biology, engineering, pharmacology, physiology, medicine, and many others, take part in the tremendous research effort in pharmaceuticals. Cooperation is a major feature of today's scientific investigations. Multidisciplinary teams are essential in industrial research requiring collaboration and effective communication; frequently a hundred or more scientists may be involved in discovering and developing a compound into a useful drug.

Some industrial research laboratories are organized according to scientific disciplines, such as departments of organic chemistry or pharmacology. Other companies may use a project-team style, wherein chemists, biologists, and pharmacologists are organized into a project unit for the purpose of discovering drugs useful for a particular disease state. Frequently the latter organizational approach is focused on therapeutic areas, such as diseases of the cardiovascular, immunological, or central nervous systems. Irrespective of the organizational style, problems in drug discovery and development have become so complex that a multidisciplinary approach to research nearly always is

used. For the sake of simplicity, this section will outline the functions of scientists with particular backgrounds who play leading roles in pharmaceutical research; however, the reader should understand that drug development is a cooperative venture among all scientists.

ORGANIC CHEMISTRY

As noted previously, organic chemists synthesize new drug candidates as well as isolate and characterize natural products, such as alkaloids. In each case, there is interest in the complex relationships between chemical structure and pharmacological action. These structure-activity-relationships (SARs) are fundamental to drug discovery. Once synthesized, compounds are evaluated for numerous types of biological and pharmacological action. Observation of interesting and repeatable biological activity opens pathways for additional chemical research effort in the expansion of the series and often leads to significant new medicinal products. Determination of the pharmacological activity of a compound is an involved process, with very small changes in structure frequently yielding profound changes in the pharmacological effect. Many of the currently used antispasmodics, anticonvulsants, local anesthetics, nonnarcotic analgesics, chemotherapeutic agents, and hypnotics have been products of this approach.

Another research approach is to identify, isolate, and purify compounds from biologically active mixtures. The determination of the structure of a biologically active molecule provides a twofold benefit to pharmacy and medicine. It makes possible research leading to synthesis and modification of the structure. Changes in structure usually are accompanied with changes in biological activity, and occasionally vast improvement is accomplished. For example, our current knowledge of adrenal corticosteroids began with the study of the various components in an extract of the adrenal cortex. The components were characterized structurally, and biological activities were assessed. Eventually, cortisone was synthesized from bile acids. Today, some synthetic analogs of cortisone are available that are superior therapeutically to the naturally occurring steroids.

A second example of slight chemical changes leading to improved biologic activity comes from the tetracyclines, a clinically important group of antibiotics. The first of these, 7-chlorotetracycline, was isolated in 1948 from *Streptomyces aureofaciens*. Shortly thereafter, a group of scientists isolated 5-hydroxytetracycline from *Streptomyces rimosus*, and in 1953 its structure was established. Once the chemical structure of this antibiotic was known, the way was opened for systematic variation of the basic nucleus to obtain new drugs with improved properties. Specifically, the catalytic removal of chlorine from 7-chlorotetracycline gave tetracycline itself, which proved to be superior to either of the above-mentioned antibiotics, and has replaced them to a large extent. Although tetracycline initially had been isolated from a *Streptomyces* species, this useful antibiotic is prepared more readily by semisynthetic methods.

Studies on the structure and synthesis of penicillins led to the development of the semisynthetic penicillins and, later, to cephalosporins and monobactams. These new compounds have made possible major improvements in antibiotic therapy. Total synthesis is made possible by knowledge of chemical structures and, in many instances, is important economically in reducing the cost of the drug. Chloramphenicol, which can be obtained from cultures of *Streptomyces venezuelae*, combats bacteria-produced typhoid dysentery and Rocky Mountain spotted fever. A commercially feasible chemical synthesis has replaced the fermentation process for production of the antibiotic.

MICROBIOLOGY

Since the discovery and development of penicillin during World War II, the search for new antibiotics among the metabolic products of microorganisms has constituted a major research effort in the pharmaceutical industry. The proven clinical

usefulness of antibiotics in treating many bacterial infections has fully justified this effort. Microbiologists have searched among a wide variety of fungi and bacteria, looking for antibiotic substances. In this search, microorganisms from plant tissues, animal sources, the sea, many types of soil, and many other ecological niches have been examined. There are more than 16,500 antibiotics produced by microorganisms that have been reported in the scientific literature, approximately 90 percent of which have been characterized and have had molecular structures assigned.^{24,25} More than 1000 antibiotic substances have been detected and at least partially characterized. A combination of microbiological and chemical methods is required to distinguish the new antibiotics from the host of older ones that have already been discovered.

After a culture has been found to produce a new antibiotic, microbiologists then turn their attention to the biosynthesis of the compound, seeking to improve yields in order to produce quantities of the compound for testing and evaluation. An effort also is made to understand biosynthetic pathways, improve yields further, and facilitate the biosynthetic production of the isotope-labeled antibiotic for pharmacological and toxicological evaluation.

New antibiotics are being evaluated for application in an increasing number of disease conditions. Tests are conducted to determine activity of new antibiotics against a variety of yeasts, molds, and protozoa, as well as against normal and antibiotic-resistant bacterial pathogens. The antibacterial drugs have contributed to major advances in the control of bacterial and other microbial diseases. However, impetus for continued research is provided by problems of drug resistance, patient sensitivity, and the inability to control certain infections.

Microbiologists are concerned not only with the microorganisms that produce antibiotics but also with the microbial pathogens that the antibiotics are expected to control. The mode of transmission of disease and the pathogenicity, virulence, and invasiveness of the infectious microorganisms are under investigation. A serious problem in drug resistance involves the transfer of drug resistance among gram-negative bacteria by means of an episome bearing one or more antibiotic resistance factors. Agents that prevent the emergence of the resistance factor, or that prevent its transfer, have been sought. Current research is being directed toward agents that enhance host resistance.

Integration of microbiological research and organic chemical research resulted in the production of a series of semisynthetic penicillins and cephalosporins. These antibiotics are chemically modified derivatives of biosynthetically produced antibiotics, which possess improved spectra of action or other advantageous chemical and biological properties.

BIOCHEMISTRY, CELL BIOLOGY, AND MOLECULAR BIOLOGY

Pharmaceutical research in biochemistry, cell biology, and molecular biology has exploded in the past 25 years. These areas include investigations of specific mechanisms of action for biologically active compounds. Biochemistry and cell biology are focused on understanding the underlying biochemical and cellular processes that are involved in the wonderfully complex mechanism of living things: the signal transduction processes, the energy-yielding systems, and the synthetic systems for generation of proteins, nucleic acids, and other macromolecules. Normal cellular communication and metabolic patterns are determined, and efforts are made to define the abnormal conditions that occur in various disease states. Biochemists also are involved in the isolation, purification, and characterization of small and large biologically active molecules.

The increasing sophistication of research demands that an understanding of the molecular bases of diseases emerge as a primary goal. This knowledge has strongly influenced both the methodology of testing new drugs and the choice or design of compounds to be tested. Biological targets (i.e., the molecular locations where drugs act) are identified, isolated, and

characterized. Usually this involves the cloning and expression of the target proteins from human tissue as well as from various other species that may serve as model systems in drug testing. Some of the receptor systems for which drugs have been developed include those for catecholamines, opiates, steroids, and various peptide hormones such as bradykinin, angiotensin II, and endothelin. The discovery of the enkephalins, natural brain polypeptides that bind the opiate receptor, has opened new horizons in CNS pharmacology. This information has been useful in acquiring new knowledge of the interaction between drugs and their receptor sites and in understanding the requirements for specific spatial orientation of essential structural features of drugs. Drug design also makes provision for those characteristics that will assure absorption, transport to the receptor site, and elimination of the therapeutic agent.

Biochemists and cell biologists develop the biomedical rationale to guide medicinal chemists in the design of drugs that are more selective for specific aspects of disease. For example, knowledge of the structure and biochemical function of coenzymes stimulated chemists to synthesize a large number of analogs of coenzymes, some of which have proven to be useful compounds in the chemotherapy of cancer; methotrexate is an example of such a compound.

Increasing emphasis is being placed on studies of enzymatic processes, such as those related to the biosynthesis of cholesterol, fatty acids, and triglycerides; regulation and control of protein and nucleic acid synthesis; absorption processes; and biochemical mechanisms in central nervous processes and ischemia. The significance of elevated blood levels of cholesterol and certain other lipids in atherosclerosis has focused attention on drugs affecting cholesterol biosynthesis. Several of these drugs, such as Livalo, Lipitor, Crestor, Zocor, Pravachol, and Mevacor, are now available. These drugs have had a dramatic impact in reducing serum cholesterol, and more cholesterol-reducing drugs may be expected.

Acute problems associated with atherosclerosis often are caused by thrombi. Current antithrombotic approaches are directed at inhibiting platelet aggregation through warfarin (Coumadin); heparins; aspirin; Integrilin (eptifibatide) or Reo-Pro (abciximab)—inhibitors of gplIb/IIIa ; or clopidogrel, prasugrel, and ticagrelor. The desire to inhibit the clotting cascade makes thrombin a good drug target for direct inhibition. This approach involves the investigation of thrombin receptor inhibition. The first direct thrombin inhibitor, Pradaxa (dabigatran etexilate), was approved in 2010 for reducing the risk of stroke and blood clots in people with atrial fibrillation not caused by a heart valve problem. Enzymes that are capable of dissolving a recently formed blood clot, such as streptokinase, tissue plasminogen activator (tPA), and urokinase, have been approved and are useful under specific primary care circumstances in the treatment of stroke.

Major advances have been made in the field of gastrointestinal physiology; many new gastrointestinal polypeptide hormones have been isolated and characterized, and their primary functions have been determined. Evidence for many years pointed to the existence of gastric receptors for histamine in addition to the vascular receptors. More recently, proton-pump inhibitors (PPIs), such as omeprazole, pantoprazole, lansoprazole, esomeprazole, and rabeprazole, have been designed to block specifically the H₂ receptor and have been very effective in the treatment of peptic ulcers when used with appropriate antibiotic treatment.

Molecular biological research has impacted every area of drug discovery. Molecular biology provides insight into the organism's fundamental genetic composition. Genetic testing of a patient may prevent a known toxic effect of a drug, such as abacavir. Abacavir is in a class of medications called nucleoside reverse transcriptase inhibitors (NRTIs) and is used in combination with other medications to treat HIV infection. Hypersensitivity reaction to abacavir is strongly associated with the presence of the HLA-B*5701 allele. Therefore, a prospective HLA-B*5701 screening might spare a patient from a hypersensitivity reaction to the drug.

Of note to pharmaceutical research is the use of recombinant expression as a source of scarce or valuable human proteins, such as growth hormone, antibodies, interferon, and insulin. This science also allows dissection of cell pathways and the generation of reagents for better assays. A major objective of research is the design of satisfactory model systems in animals, cell culture, and other innovative means to give reliable predictions of the safety and efficacy of new drugs in humans. Molecular biology has advanced the reduction in the numbers of animals used in drug research. Tests for biological activity at the molecular level are done first, after which animal tests using standardized, controlled experiments are conducted.

VIROLOGY AND IMMUNOLOGY

The search for antiviral agents, which has depended on the development of methodology for propagation and assaying of viruses in tissue culture, has led to more precise procedures of testing compounds for antiviral activity. Tissue-culture techniques have made possible the production of large quantities of viruses for vaccine manufacture. New and improved vaccines represent a major objective of biologic research. New separation methods developed in biochemistry and physical chemistry have been applied to the isolation and purification of viruses and have led to preparation of highly purified and concentrated vaccines. Such vaccines are more effective and produce markedly fewer side-effects.

The discovery of HIV and its epidemiological implications has opened new avenues of research to develop suitable therapies. In combinations with cocktails of other drugs, HIV protease inhibitors are the main avenue of therapeutic approach for controlling HIV today. The discovery that chemokines and chemokine receptors are involved as coreceptors for HIV defined new pharmaceutical strategies toward small-molecule drug discovery.

New viral threats will surely emerge. Currently, the world is working to understand a new corona virus which has led to the severe acute respiratory syndrome (SARS) threat. New viruses will threaten the world, due to global travel, and could be used in bioterrorism. Advances in virology will add new understanding, disease controls through treatment, and eventually cures.

Recent immunological research has focused attention on a number of important diseases with an autoimmune component, such as rheumatoid arthritis, lupus, inflammatory bowel disease (IBD), and multiple sclerosis. These diseases continue to be poorly treated over time. Suppression of immune functions or induction of immune tolerance may be of great importance. A more detailed knowledge of the molecular basis of B and T differentiation, signal transduction, and leukocyte trafficking is needed to search for drugs that either enhance or inhibit these immune responses.

In addition, a clearer picture of the molecular basis of immune disease is required to improve the probability that additional drugs will be found to alleviate allergic reactions.

Immunological research has also been directed toward cancer. The existence of tumor-specific antigens in both viral- and chemical-induced tumors, as well as new evidence for host reactions to the tumor, increase the possibility of useful immunological approaches against cancer. One of the most important developments in the past decade has been the isolation and production of monoclonal antibodies. These agents can be used to identify tumor-specific antigens and thus serve as powerful *in vitro* diagnostic and therapeutic tools. The technique can be applied to other antigens as well. These substances are being developed as carrier systems for drugs by virtue of their ability to deliver the antibody-drug complex directly to the antigen-producing cell or tissue.

PHARMACOLOGY

The role of pharmacological research in drug discovery continues to evolve. Initially, the pharmacologist was a whole animal biologist who developed animal models of disease to the extent

possible and tested compounds in animals to measure efficacy. Classical pharmacology contributed in two major areas²⁶: (1) the design and operation of animal model for detecting and evaluating the activity of compounds, and (2) determination of the dosage, toxicity, mode of action, metabolism, and fate of a drug candidate in the body. More recently, the molecular pharmacologist is involved in the discovery and validation of new targets for drug discovery, as well as the generation of new assays, both *in vitro* and *in vivo*. Classic pharmacological methods using intact animals, whole organs, and isolated tissues were used more than 15 years ago. These have evolved to more automated and molecular-oriented methods, in which purified or recombinant enzyme and receptor systems are used in initial phases of discovery pharmacology, followed with *in vivo* testing as needed. Potential drug candidates are examined early on for specificity against other unrelated molecules, a means of reducing side-effects in individuals. Safety assessments are done earlier in the discovery process, allowing the physicians and clinical pharmacologists to work together to set the starting points for dosing drugs with minimal side-effects, to monitor what form of toxicity might appear, and to determine contraindications for drug use.

The study of drug absorption, distribution, metabolism, and excretion is frequently referred to as ADME. Drug therapy requires an elaborate and thorough knowledge of the kinetics of these processes after intravenous, oral, topical, sublingual, or intraorifice administration of the drug. Initial studies are often conducted using *in vitro* experimental systems, such as Caco-2 cell permeability, to determine the likelihood of oral absorption, or human liver microsome to determine the likelihood of metabolic stability. Common next steps are to test suitable compounds in animals to determine if the experimental systems are accurate for a particular chemical series and to extend the data set to a whole organism. Experiments are often performed with radioactive forms of the drug to determine the amounts of drug and its metabolites that appear in blood, urine, and tissues. Animals can be used to determine the manner in which a living organism assimilates a drug; however, human pharmacokinetic studies are essential to determine the fate of the compound in humans: Is it accumulated in specific organs, is it excreted into bile or urine, and is it metabolized? To determine the concentration of drugs in biological fluids or tissues requires special separation techniques, as well as sensitive, accurate, and precise instrumental measurements. Accurate quantitation and identification of the drug and its metabolites usually requires the use of chromatographic techniques coupled with mass spectrometry. Sensitive liquid-chromatography mass-spectrometry (LC/MS) methods provide powerful data on early drug candidate molecules, which influence the direction of new chemical synthesis.

TOXICOLOGY

To be certain that a new drug is safe, detailed studies are made of the effects of varying doses and prolonged administration of that drug. The pharmacologist provides acute toxicity data; however, the toxicologist then must refine the acute toxicity measurement in laboratory animals and begin subacute and chronic studies. The latter are conducted in a variety of species, at several dosage levels of the drug, and over periods of time ranging from 3 months up to 30 months. During the test period, animals are observed carefully for all adverse symptoms. At the end of this period, and occasionally during its progress, the animals are sacrificed, and their vital tissues (such as liver, heart, kidney, intestine, or brain) are removed and studied grossly and microscopically by a pathologist.

In addition to gross and microscopic pathology, biochemical and physiological responses are measured as an indication of liver function, kidney function, or endocrine function. During recent years, metabolic investigations have become more sophisticated and have been brought to bear on the comparative effects of drugs on various animals and on humans. In some

instances, the metabolism of drugs or the therapeutic effects of drugs vary from species to species. Such variability can be the basis for differences in toxicity as well as efficacy. For these reasons, increasing emphasis is being given to studies of comparative metabolism in humans and animals to determine which laboratory animal handles the drug in a manner similar to humans. Selection of that species for extensive toxicity testing increases confidence that the toxic reactions that may occur in humans can be predicted by the animal tests.

Reproductive studies to determine the potential effects of the new drug on the reproductive processes and on subsequent generations are performed. Teratological studies are done to determine whether the new drug affects the fetus. Special toxicity tests have been designed to detect specific toxic reactions, such as nerve damage resulting in hearing loss. Ultimately, drugs reaching the marketplace will be assigned an FDA Pregnancy Category (A, B, C, D, X).

Carcinogenicity trials, which are lifelong studies in animals carried out at doses approximating the maximum (tolerated) human dose, provide evidence of a new drug's potential to produce human cancer. Several newer methods of toxicity are evolving in the biotech industry using gene activation methodology whereby one can evaluate if a candidate drug is transcriptionally active at a number of genes involved in liver metabolism or stress responses. These methods are likely to transform toxicological testing in the future.

In 1992, a global effort aimed at establishing uniform standards for toxicology testing of new drugs began, under the auspices of the International Conference on Harmonization (ICH). Guidelines were first published on toxicology testing for drugs and updated in 1998. Biologicals were not covered by those initial guidelines. Product-specific toxicology programs normally are required for biologicals. Toxicological studies are assuming increasing importance in the world of pharmacy and medicine. As knowledge and skills increase, and the ability to measure toxic reactions improves, greater safety and efficacy of new drugs may be ensured.

During the 1990s and continuing into the 2000s, significant progress has been achieved in the concept of replacing animals in toxicology/safety assessment with numerous *in vitro* systems. These are attempts to reduce the number of animals used and to refine the manner in which they are used. A review of available data, from 1973 to 2006, shows a decline in the number of animals used in research, experimentation, teaching, or testing: from just under two million in the 1970s and 1980s to just over one million in 2006.²⁷ Overall, multiple *in vitro* systems have been developed for screening and testing for eye and skin irritation, skin sensitization, teratology, and other endpoints; and a scientific consensus has been reached on the requirements and processes for validation for these *in vitro* tests. However, the use of these newer test systems in place of existing *in vivo* tests is not yet a reality. Much progress and dialog have taken place in the 1990s and 2000s on modification of both US and international requirements and guidelines for testing and for defining an approval process for alternatives and innovations.

Communication between scientists and the literature also have evolved with the explosion of information systems (IS) technologies and access to the Internet. Attention and access to the scientific and patent literature has accelerated. Formerly, the individual scientist subscribed personally to a few journals and depended on a scientific library for coverage of additional new scientific findings. With the tremendous growth of the scientific and patent literature and the emergence of interdisciplinary investigations, individual scientists can conduct desktop searches to stay abreast of the literature. Despite increasing use of various kinds of alerting services and facilities, many of them computer-based, for retrieval or retrospective search of pertinent information, personal perusal of literature remains critical, and the ability to initiate desktop searches is fundamental to independent scientific research.

PHYSICAL CHEMISTRY

Modern research in pharmacy and medicine is supported and expedited by instrumentation. Modern instruments make possible the rapid and accurate measurement of physical and chemical properties of molecules. Separation and characterization of molecules are sometimes possible today in a matter of hours or a few days; only a decade or two ago, such work often required many days, weeks, or even months. Examples of specialized physicochemical and computational methods that are applicable to structural research are electron microscopy, nuclear magnetic resonance (NMR) spectroscopy, mass spectrometry, and crystallography.

NMR spectra identify chemical groups and indicate the nature of neighboring chemical groups in the molecule. Mass spectrometry permits determination of the molecular weight and empirical formula of an organic molecule and of the major fragments of the molecule. With this information, it is often possible to deduce the entire structure of a molecule rapidly and precisely. X-ray crystallographic analysis enables the physical chemist to determine the precise position of each atom of a molecule as it exists in the crystalline form. Structures of both the drug target and a potential drug in the active site of an enzyme have been critical to the discovery of HIV protease inhibitors.

Physicochemical studies are directed at the chemical groups and stereochemical configuration of biologically active molecules; these studies can describe molecules in terms of energy and electron distributions and can approximate the influence of the chemical environment on these distributions. The spatial and electronic conformation of drugs and the changes in conformation that occur in various environments govern the absorption, transport, distribution, and reaction with the receptor site. If description of molecules in these functional terms is achieved, correlation of electronic structure with function may be possible, and the design of safer, specific, and more effective drugs on a rational basis may occur.

INFORMATION SCIENCE

The information systems have spearheaded a great amount of data generation, assimilation, and scientific communication. IS departments are now commonplace in academic, government, and industrial settings. The amount and sophistication of chemical and biological information have led to the critical role of bench-top and desktop computers in assimilating data. Computer-assisted chemistry, computer graphics, and relational databases have added a new dimension in structure and activity relationships. The computer-based monitoring and analysis of animal studies is routine. Online signal processing allows investigators to interact more fully with their experiments. Computer-assisted automation permits collection of more data with increasing accuracy; sophisticated software packages are available commercially or may be developed in-house.

DRUG DEVELOPMENT

Before a new drug candidate can proceed to toxicological or clinical evaluation, considerable analytical chemical development is required to lay the groundwork for subsequent quality control and stability studies. Drug standards are established, and analytical methods for the bulk drug and the proposed final product are devised. Tentative chemical, physical, and biological specifications of the candidate drug are established. Simultaneously with analytical development, pharmaceutical chemists begin formulation studies toward the goal of a stable, highly acceptable product that delivers the correct amount of drug in a reproducible, effective manner. Sometimes a new drug must be modified chemically via esterification to a pro-drug in order to provide a form that is pharmaceutically acceptable and effective. Accelerated and long-term stability studies are started to estimate the conditions in which the product will be stable.

If a compound has desirable activity in an experimental testing system and appears to be safe upon toxicological examination, it becomes a candidate for clinical trial. Two additional tasks must be accomplished before a clinical trial can be undertaken. First, the drug candidate must be in a suitable, stable dosage form, and the candidate compound must be available for absorption and transport to the site of action. The stabilization of a drug candidate must preclude physical or chemical change (discoloration, precipitation, or decomposition). These components, or excipients, must often meet the standards outlined in the US Pharmacopeia/National Formulary (USP/NF), European pharmacopeias, or other national compendia. Because of the many physical forms in which pharmaceuticals are presented, the research necessary is broad in scope and not only involves the principles of physical pharmacy but also requires the application of principles from the allied fields of chemistry and biology.

The second task at this stage is to file an Investigational New Drug (IND) application with the FDA. The IND is, in fact, a document that gives a full description of the new drug, the location and manner of its manufacture, all quality control information and standards, stability, analytical methods, pharmacology, toxicology, documentation of efficacy in animals, qualifications of investigators who will be doing the clinical studies, and the complete protocols of the proposed clinical studies.

A new drug is administered to humans for the first time by a physician or clinical pharmacologist. These Phase I studies are carried out most often in healthy male volunteers in order to study the safety and pharmacokinetics of a new drug. The first trial of a drug in humans is done with great caution and on a very limited basis.

When dosing limits have been established and are found acceptable, the drug is made available to a larger number of practicing specialists for the Phase II study, which principally is concerned with the determination of safety and efficacy in patients having the primary disease for which the drug is to be tested. The minimum effective dose, the maximum tolerated dose, and the dose response (intermediate doses) also must be determined.

If after Phase II the drug still looks promising, it is distributed more widely to selected practicing physicians in the Phase III study. The purpose of the Phase III stage is to secure data from a larger number of patients on efficacy and incidence of side effects.

Finally, before the new drug can be marketed, a New Drug Application (NDA) is filed with the FDA and approval obtained. The NDA contains most of the information included in the IND, revised and updated, as well as all the results of the clinical studies proving safety and efficacy. Almost all clinical, laboratory, and patient history data are processed on computers. These medical data are updated in computerized retrieval systems and are designed to provide timely information during the FDA review. These systems also provide an additional information resource for premarketing and post marketing queries. Only after FDA approval of the NDA can distribution and marketing of the new drug begin.

Depending on the nature of the disease and the clinical endpoints that are monitored, some drugs require long-ranging and expensive clinical trials. Some trials by necessity monitor mortality rates. Clinical trials are carefully designed with the input of statisticians to determine numbers of patients and duration of the studies. Trials cost hundreds of millions of dollars (Figure 7-1) over multiyear periods, and they demand careful monitoring throughout.

Figure 7-1 also depicts the clinical research effort on a new drug and represents the culmination of many years of effort by large numbers of scientists of many disciplines and skills. Clinical research is the proving ground where the intelligence, creativity, and perseverance of laboratory researchers come to fruition. Of the candidate drugs that come to clinical research, only 20 percent survive as safe and efficacious and are added to the portfolio of therapeutics. Indeed, the 2011 report from the Pharmaceutical Manufacturers of America (PhRMA), titled 2011 Profile Pharmaceutical Industry¹⁵ revealed the following information on the drug discovery research process:

- One in 5000 compounds screened is approved for patient use.
- The estimated cost to create one new medicine is \$1.2 to \$1.3 billion.
- It takes an average of 10 to 15 years to develop a new medicine.
- Only 2 of 10 marketed drugs return revenues that meet or exceed the R&D costs (Figure 7-2).

Together with the cost of conducting R&D (Table 7-1), the above data explain the rising costs of prescription drugs. The top 25 selling drugs are listed in the twelfth Annual Pharma

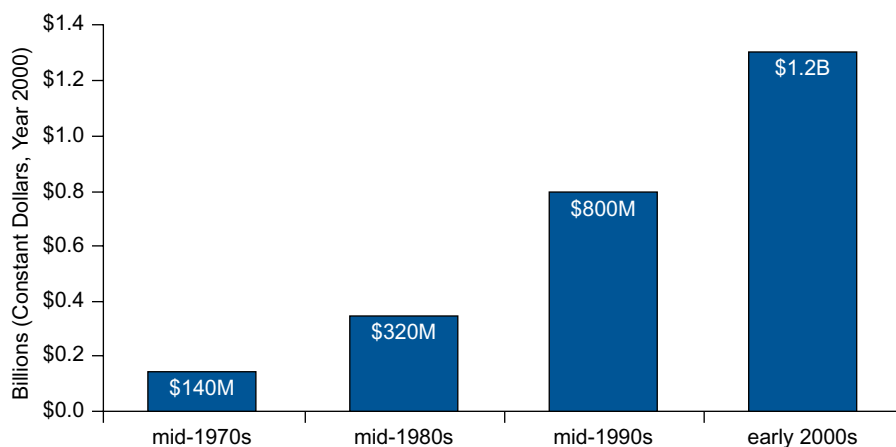


Figure 7-1. The rise in drug development costs. The average cost to develop one new approved drug—including the cost of failures—increased approximately 50% between the late 1990s and the early 2000s. (From Pharmaceutical Research and Manufacturers of America. Chart Pack: Biopharmaceuticals in Perspective. 2010. http://www.phrma.org/sites/default/files/159/phrma_chart_pack.pdf, Accessed 8 May, 2012. Data from: JA DiMasi and HG Grabowski. The cost of biopharmaceutical R&D: is biotech different? *Managerial and Decision Economics* 2007; 28: 469–479; JA DiMasi *et al.* The price of innovation: new estimates of drug development costs, *Journal of Health Economics* 2003; 22:151–185. Note: Data for early 2000s is adjusted to 2000 dollars based on correspondence with study author.)

Developing a new medicine takes an average of 10–15 years

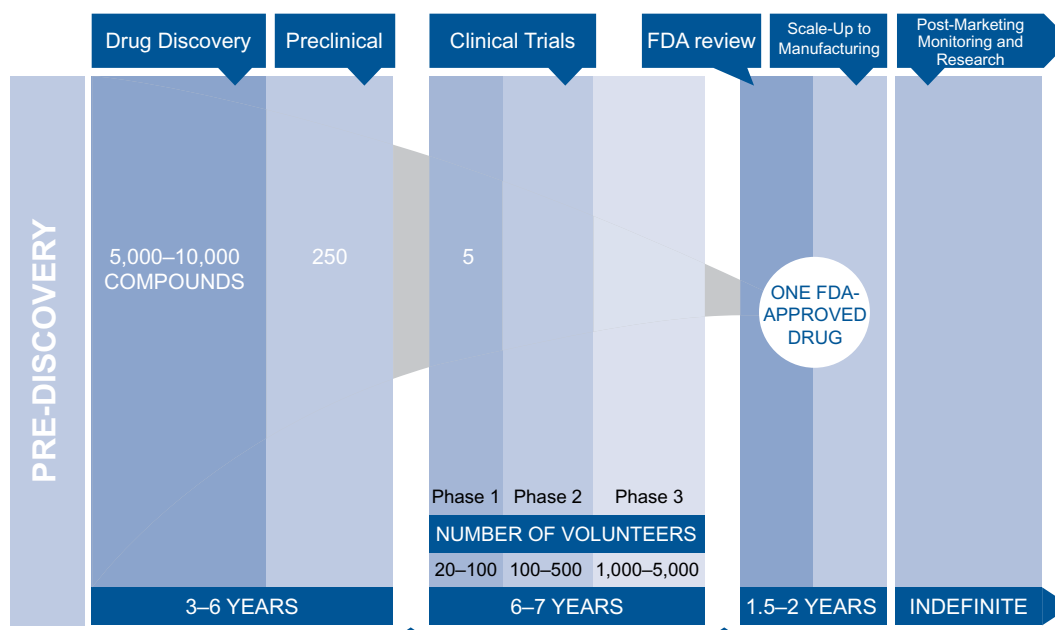


Figure 7-2. The research and development process. (From Pharmaceutical Research and Manufacturers of America, Pharmaceutical Industry Profile 2011. Washington, DC: PhRMA, April 2011.)

Executive report.¹⁶ The top 10 of these are “blockbuster drugs” (i.e., over \$1 billion in sales (Table 7-2)¹⁶) and account for \$48 billion of sales in a \$593 billion market, or 8.1% of the market. Another factor in rising costs that cannot be discounted is the fact that overall research productivity and investment in

innovative research for new approaches and new medicines in the large pharma tier of companies has declined.²⁸ This may in part be attributable to merger and acquisition (M&A) consolidation in the industry. Of the top 10 companies, all are the result of M&A (Table 7-3).¹⁶

Table 7-2. The Top-Selling Drugs in 2010

Rank	Drug	Company	Type of Drug	2010 Sales (US \$ in billions)
1	Lipitor	Pfizer	Cholesterol reducer	7.2
2	Nexium	AstraZeneca	Antiulcer	6.3
3	Plavixq	Bristol-Myers Squibb	Antiplatelet	6.1
4	AdvairDiskus	GlaxoSmithKline	Antiasthmatic	4.7
5	Abilify	Otsuka	Antipsychotic	4.6
6	Seraquel	AstraZeneca	Antipsychotic	4.4
7	Singulair	Merck	Antiasthmatic	4.1
8	Crestor	AstraZeneca	Cholesterol reducer	3.8
9	Actos	Takeda	Antidiabetic	3.5
10	Epogen	Amgen	Erythropoiesis-stimulating agent	3.3

(From Cacciotti J, Clinton P. The Lull between two Storms, *Pharmaceutical Executive*, May 2011. www.oliverwyman.com/media/2011_PharmExec_50.pdf; accessed 20 April 2012.)

Table 7-3. Top 10 Pharmaceutical Companies Worldwide in Prescription Sales (US\$ in billions) for 2010

Rank	Company	Pharma Sales	Increase (%)
1	Pfizer	58.5	28.9
2	Novartis	42.0	9.2
3	Sanofi-Aventis	40.3	-4.1
4	Merck	39.8	58.0
5	Roche	39.1	4.1
6	GlaxoSmithKline	36.2	-4.2
7	AstraZeneca	33.3	1.4
8	Johnson & Johnson	22.4	-0.4
9	Eli Lilly	21.1	5.4
10	Abbott	19.9	27.7

(From Cacciotti J, Clinton P. The Lull between two Storms, *Pharmaceutical Executive*, May 2011. www.oliverwyman.com/media/2011_PharmExec_50.pdf; accessed 20 April 2012.)

The drug discovery research effort represents the culmination of, on average, 10 to 15 years of research and development by many scientists from multiple disciplines. Of the candidate drugs that come to clinical research, only about 20 percent survive as safe and efficacious and are added to the portfolio of therapeutics. The great investment in pharmaceutical research in the early part of the 20th century that has led to advancements in pharmaceutical therapies needs to be remade in the 21st century in support of helping patients to live longer, healthier, and more productive lives.

REFERENCES

- Spilker B. *Guide to Drug Development: A Comprehensive Review and Assessment*. Philadelphia: Lippincott Williams & Wilkins, 2009.
- Tulp M, Bohlin L. Functional versus chemical diversity: is biodiversity important for drug discovery? *Trends Pharmacol Sci* 2002; 23: 225–231.
- Schreiber SL *et al.* Davies K. Diversity-oriented organic synthesis and proteomics. New frontiers for chemistry and biology. *Chem Biol* 2002; 9: 1–2.
- Roger VL *et al.* Heart disease and stroke statistics—2011 update: a report from the American Heart Association. *Circulation* 2011; 123: e18–e209.
- Anon. “Medicines in Development: Biotechnology,” 2011 Report. Washington, DC: Pharmaceutical Research and Manufacturers of America, 2011.
- Wick JY. Lifestyle drugs straddle medical and social domains. *Pharmacy Times* 2006 (March 1).
- Torrey T. Lifestyle drug—the definition of a lifestyle drug. Available at <http://patients.about.com/od/glossary/g/Lifestyle-Drug-The-Definition-Of-A-Lifestyle-Drug.htm> (Accessed 28 April 2012).
- Flower FJ. Lifestyle and non-medical uses of drugs. In: Rang HP *et al.* eds. *Rang and Dale's Pharmacology*, 6th edn, London: Churchill Livingstone, 2007; pp 765–769.
- Lexchin J. Lifestyle drugs: issues for debate. *CMAJ* 2001; 164(10): 1449–1451.
- The Freedomia Group, Nanotechnology in Health Care to 2011 Report; February 2007.
- The Nanotech Report, 4th edn, Lux Research, 2006.
- Human Genome Program. Office of Biological and Environmental Research, US Department of Energy. Available at: http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml (Accessed 28 April 2012).
- Lazarou J *et al.* Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *JAMA* 1998; 279(15): 1200–1205.
- Mega JL *et al.* Based on CYP2C19 genotype and the effect on platelet reactivity in patients with stable cardiovascular disease. *JAMA* 2011; published online November 16, 2011. doi:10.1001/jama.2011.1703.
- Pharmaceutical Research and Manufacturers of America, Pharmaceutical Industry Profile 2011. Washington, DC: PhRMA, April 2011.
- Cacciotti J *et al.* 12th Annual PharmExec 50, The Lull between Two Storms, May 2011.
- McCoy M. *Chem Eng News* 1998 (July): 7.
- Hartzema AG. Guide to interpreting and evaluating the pharmacoeconomic literature. *Ann Pharmacother* 1992; 26(1): 96–98.
- Gordon EM, Kerwin JF. *Combinatorial Chemistry and Molecular Diversity in Drug Discovery*, Malden, MA: Wiley-Liss, Inc., 1998.
- Nuss JM *et al.* Developing a general strategy for the solid supported synthesis of heterocycles: Applications to the generation of molecular diversity and drug discovery. *Pure Appl Chem*. 1997; 69(3): 447–452.
- Drews J. Drug discovery: a historical perspective. *Science* 2000; 287 (5460): 1960–1964.
- Lipinski CA. *Pharmaceutical News* 2002; 9: 195.
- Egan WJ *et al.* Prediction of drug absorption using multivariate statistics. *J Med Chem* 2000; 43: 3867–3877.
- Higton AA, Roberts AD. In Bycroft BW, ed. *Dictionary of Antibiotics: and Related Substances*. New York: Chapman and Hall, 1988; *Dictionary of Natural Products* on CD-ROM. New York: Chapman and Hall, 2006.
- Bérdy J. Bioactive microbial metabolites. *Antibiotics* 2005; 58(C–1).
- Frenkel JK. Choice of animal models for the study of disease processes in man. *Fed Proc* 1969; 28(1): 160–161.
- Kulpa-Eddy J *et al.* A review of trends in animal use in the United States (1972–2006), US Department of Agriculture, Animal and Plant Health Inspection Service, Animal Care, AATFEX 14, Special Issue, 163–165, Proc. 6th World Congress on Alternatives & Animal Use in the Life Sciences, 2007, Tokyo, Japan.
- Bernstein Research, The long view: Pharma R&D productivity—When the cures fail it makes sense to check the diagnosis. Available at: https://www.bcgperspectives.com/content/articles/value_creation_strategy_biopharma_strength_storm_biopharmaceuticals (Accessed 28 April 2012).

Chapter 8

The New Drug Approval Process

Linda A. Felton, PhD and Dennis W. Raisch, PhD, RPh

THE FOOD AND DRUG ADMINISTRATION 93

OVERVIEW OF THE US DRUG APPROVAL PROCESS 93

POST-APPROVAL ACTIVITIES 98

SUMMARY 98

The research and development efforts needed to ensure the safety and efficacy of new drugs are complex, time consuming, and financially risky. Thousands of compounds undergo extensive testing for every one new chemical that receives marketing approval.¹ Research and development costs for each new drug product are estimated at approximately \$1 billion.² It has been reported that only 30% of drugs that reach the marketplace generate sufficient revenue to recover the average cost of development.³ This chapter discusses the stages of new drug development and approval in the United States with a focus on clinical trial design and methodology. Readers are encouraged to refer to specific Food and Drug Administration (FDA) guidance documents for more detailed information (www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm). Note that other countries have similar regulatory authorities that oversee drug approvals.

THE FOOD AND DRUG ADMINISTRATION

The Food and Drug Administration (FDA) oversees the new drug approval process in the United States. The initial legislation, the Pure Food and Drug Act of 1906, was passed in response to Upton Sinclair's exposé of the meat packing industry, *The Jungle*, which described deplorable practices resulting in contamination. Over time, the authority of the agency expanded, but it remained relatively powerless to effectively assure the safety and efficacy of medicines. Subsequently, a sulfanilamide product containing diethylene glycol as a solvent to enhance the aqueous solubility of the drug was developed. Diethylene glycol, however, is a highly toxic agent used in antifreeze solutions, and numerous deaths resulted from its ingestion. Based on these tragic events, Congress passed the Food, Drug, and Cosmetic Act of 1938, which increased the regulatory authority of the FDA to oversee the development of new drug products.⁴ The Act required disclosure of the ingredients and formulation, assay methods, manufacturing processes, and all animal and human testing to the FDA prior to the distribution of drug products.

Although the Act of 1938 required new drug products to be safe, efficacy standards were not established until another tragedy occurred in the early 1960s. Thalidomide, a synthetic sedative/tranquilizer, had been sold in Europe without a prescription and was viewed as a possible alternative to the more toxic barbiturates.⁵ Prior to FDA approval of thalidomide, several incidences of toxicity in Europe were reported. Severe birth defects were noted when the drug was administered to pregnant women, the most common being phocomelia or arrested limb development. These events brought about the Kefauver-Harris Amendments of 1962, which strengthened existing laws and emphasized the need for the safety of approved drugs.⁶ These Amendments required manufacturers to establish both safety and efficacy of new drug products prior to approval and required investigators to file Investigational New Drug Applications (INDs) prior to testing drugs in humans. As a side note, thalidomide is currently approved in the United States for the treatment and prevention of painful skin lesions associated

with erythema nodosum leprosum and multiple myeloma. Other potential uses of this drug under investigation include several types of cancer, Crohn's disease, and autoimmune deficiency-associated diseases.^{5,7}

The role of the FDA is to promote and protect the health of Americans. A multidisciplinary staff, consisting of pharmacists, physicians, pharmacologists, chemists, statisticians, attorneys, and other scientists, as well as administrative personnel, is employed at the FDA to achieve this goal. The FDA consists of several Centers, each designated with specific responsibilities. The Center for Food Safety and Applied Nutrition oversees food and cosmetic products. The Center for Veterinary Medicine is responsible for animal feed and drugs. The Center for Devices and Radiological Health covers the safety and efficacy of medical devices. This Center also oversees radiation-emitting devices, such as lasers, x-ray systems, microwave ovens, and cellular telephones. The Center for Biologics Evaluation and Research (CBER) supervises biologics. Finally, the Center for Drug Evaluation and Research (CDER) is responsible for drugs and drug products. In addition to reviewing the safety and efficacy of all prescription and over-the-counter drug products prior to marketing, the Center for Drug Evaluation and Research is responsible for monitoring drug safety after initial market approval and has the authority to withdraw drugs posing significant health risks from the market. The Center for Drug Evaluation and Research provides health care professionals and consumers with drug-related information and screens television, radio, and print ads for truthfulness and balance.

Other countries have similar regulatory authorities that oversee the approval of new drug products. For example, the European Medicines Agency (EMA) regulates most European markets. The pharmaceutical industry and regulatory authorities, including the FDA, have been working closely with the International Conference on Harmonization (ICH) to develop standardized regulatory processes for the major markets and allow results of international clinical trials to be used to support approvals across countries.

OVERVIEW OF THE US DRUG APPROVAL PROCESS

As mentioned previously, drug discovery and development is a complex and expensive endeavor. The next several sections of this chapter discuss the new drug development and approval processes in the United States. Although focused specifically on the Center for Drug Evaluation and Research and new drugs/drug products, similar requirements exist for biologics. Medical devices are approved in a different manner, depending on how they are used and their invasiveness on human functions. Devices require an approved Investigational Device Exemption (IDE) prior to beginning clinical studies. The drug approval process is divided into two sections: (1) preclinical testing (lead compound selection and animal testing of new chemicals) and (2) clinical testing (administration of new chemicals to humans). Figure 8.1 shows a schematic of these steps.

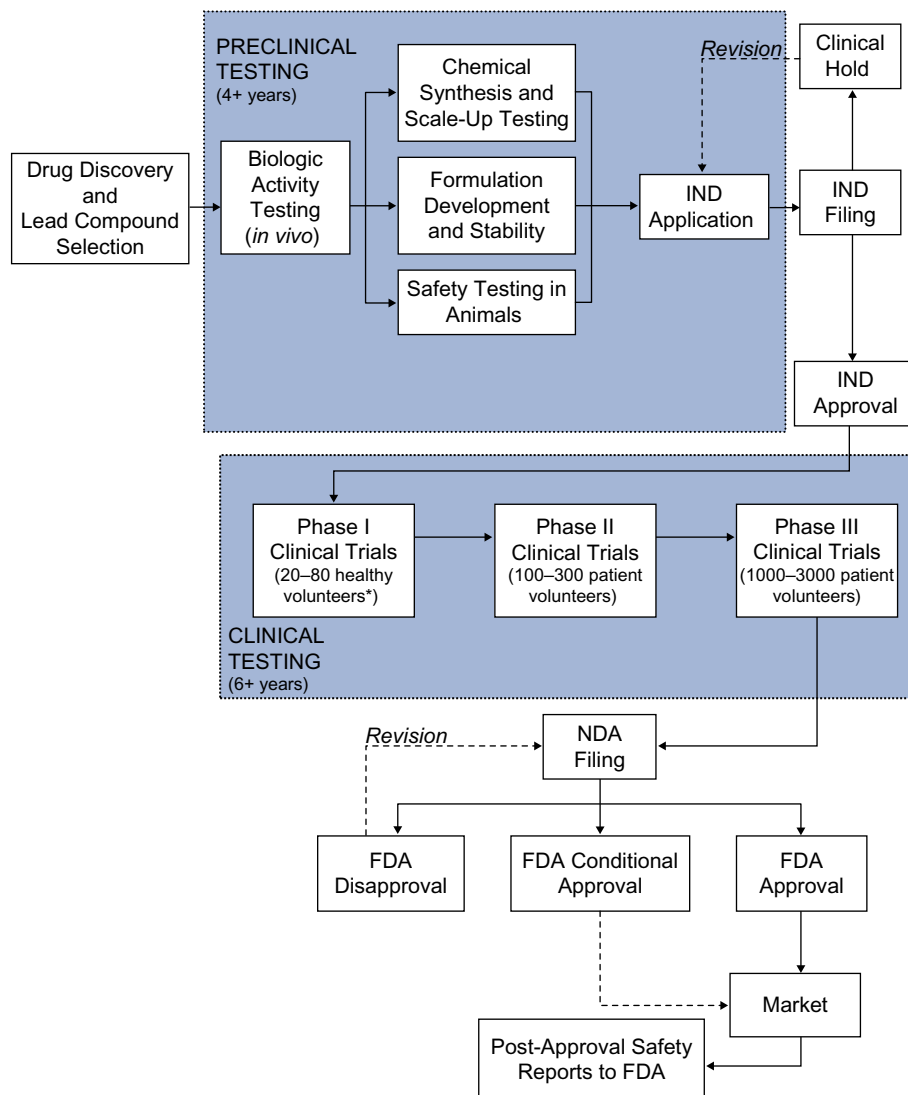


Figure 8-1. Schematic of the new drug approval process in the United States. For life-threatening illnesses such as cancer, patients enrolled in Phase I studies may suffer from the disease.

DRUG DISCOVERY AND LEAD COMPOUND SELECTION

Pharmaceutical companies begin the discovery process by targeting a broad disease category (for example, cancer or cardiovascular disease) or a specific disease state (for example, breast cancer or hypertension). A chemical with potential therapeutic benefit(s), known as a lead compound, must first be identified, and researchers use various high-throughput assay techniques to rapidly screen large numbers of chemicals for biological activity.

Random screening, as the name implies, requires biological testing of a large variety of diverse compounds from existing chemical libraries. Although less up-front financial investment is needed, thousands of compounds may be screened and tested before one agent with significant biochemical activity is identified. A more mechanism-based drug design is targeted synthesis, where researchers focus on one step in a disease process as the target for drug intervention. Although an extensive knowledge of the disease state is required, this more directed approach increases the likelihood of successfully identifying a lead compound. In combinatorial chemistry, one compound is used as a base chemical and various functional groups are randomly added to enhance biological activity. This technique is a more expensive, more complex method of identifying potential lead agents. Another method to enhance biological activity is

drug modeling, where computers are used to manipulate virtual structures and calculate protein binding capabilities. Although initial costs are significant, drug modeling techniques show a great deal of promise for future drug discovery as more research is conducted to identify biochemical pathways. Generally, these discovery techniques are used in combination to identify lead compounds.^{8–10}

PRECLINICAL TESTING

A multidisciplinary team of researchers works to determine many of the lead compound's critical properties. This team might continue to work with the compound throughout the entire development process or the development responsibilities might be transferred to another group of scientists during the clinical testing phase. Preclinical testing includes:

- discovery testing to ensure biological activity *in vivo*
- chemical synthesis and scale-up to ensure adequate quantities of high purity can be made
- formulation development and stability testing to characterize various chemical properties, develop the initial drug delivery system, and determine the stability of the compound
- animal safety testing to ensure limited toxicities of the lead agent.

At this stage of the development process, good laboratory practices (GLPs) are followed. These regulations, in the Code of Federal Regulations (CFR) 21 Part 58, provide standards for the design and conduct of preclinical studies. Qualifications of personnel and requirements for standard operating procedures are specified.

During discovery testing, the specifics of the compound's properties, such as the mechanism of action in animal models, compound specificity, duration of action, and structure-activity relationships, are determined. Adequate quantities of the new chemical compound must be produced at a high level of purity. Impurities present at concentrations greater than 0.1% must be characterized and tested for toxicity. The physicochemical properties of the active compound are determined, and development of the drug delivery system to be used in human testing begins. Animal testing provides initial data regarding the absorption, distribution, metabolism, and excretion (ADME) in a living system. Possible side effects and toxicities are noted. Toxicity studies of at least the same duration as the proposed human testing and a minimum of two weeks must be completed. Active and inactive metabolites must be characterized. Often, the most appropriate animal model to predict human response is not known, thus toxicity studies are conducted in at least 2 animal species, one rodent and one nonrodent, to obtain a comprehensive view of the potential toxicity. Early absorption, distribution, metabolism, and excretion or toxicity problems may be corrected by slight modifications in the chemical structure of the new entity.

Animals should be given the new drug product by the same route intended for humans. Certain dosage forms, such as aerosol, nasal, or buccal delivery systems, might be difficult to administer to animals. In these circumstances, alternative drug delivery routes may be used, and the selected route of administration should ensure sufficient exposure to the new chemical entity. During animal safety testing, dosing studies are conducted, and the highest no-effect dose is determined. In addition to dose, plasma concentrations of the drug are followed, and noted toxicities are correlated to dose and/or plasma concentration.

Generally, once discovery testing shows therapeutic promise, the chemical synthesis, formulation development, and animal safety testing occur concurrently (see Figure 8.1). Although resources may be wasted on earlier failures, successful candidates will be ready for human testing more quickly. The administration of drugs in humans at the earliest time possible ultimately saves valuable resources, as highly toxic compounds can be eliminated and alternative lead compounds can be developed.

Additional preclinical studies may be conducted during clinical testing to support larger trials and, eventually, the marketing of the drug product. Formulation development continues throughout the process, and the data gained from both animal and human testing allow for optimization of the drug delivery system. It is imperative to identify and resolve formulation problems early in the development process, as unresolved problems will surely reemerge later, costing the company both time and money as clinical testing is delayed. More chronic animal exposure experiments are conducted to support further clinical testing.

Pre-IND Meetings

Pre-IND meetings may be held prior to submission of an Investigational New Drug application (IND) and at the request of the sponsor, during these early stages of development to discuss testing plans and data requirements. These meetings are especially useful when a drug has been developed overseas and a great deal of preclinical and clinical data is readily available. During pre-IND meetings, the sponsor and FDA should agree on the acceptable phase of the initial clinical investigation. Clinical data from other countries, if obtained following International Conference on Harmonization requirements, may eliminate the need for Phase I human safety testing in the United States.

FDA guidance documents provide an overview of procedures for requesting formal meetings. These meetings are not intended to replace informal discussions with the FDA.

INVESTIGATIONAL NEW DRUG APPLICATION (IND)

An Investigational New Drug application (IND) must be filed with the FDA and approved prior to administering new drug products to humans. The guidelines for preapproval of all clinical testing are specified in 21 CFR Part 312. FDA guidances for INDs can be found on its website. The name and chemical description of the active components, a list of active and inactive components, and the manufacturers of these components must be provided. The method of preparation and the dosage form to be administered is required. The IND includes all preclinical animal data and the names and locations of the investigators performing the planned clinical trials. Data from clinical trials conducted in other countries should also be included. A major component of the IND is the study protocol(s) implemented to evaluate the drug. The protocol must address all aspects of the trial, including study procedures, informed consent, data collection, and analysis, as well as mechanisms for subject selection, follow-up, and safety monitoring to protect the patient. Upon receipt, an IND number is assigned by the FDA, which is used to track all subsequent communication between the study sponsor and the FDA. The IND is assigned to the appropriate division of the Center for Drug Evaluation and Research, and the contents are thoroughly reviewed. The FDA has 30 days from receipt of the IND to decide if the proposed clinical trial should proceed. Rather than "approving" an IND, the FDA must provide notification if the trial is placed on "clinical hold" pending clarifications or changes to the study protocol. If no further communication is received from the FDA, the sponsor is allowed to begin the study by enrolling patients. Each facility's Institutional Review Board (IRB) must approve the protocol(s) conducted under the IND.

Although the trial may proceed, reviewers at the FDA may place a "clinical hold" on the trial at any time. A "clinical hold" prevents human testing under the IND until FDA concerns have been adequately addressed. Reasons for placing an IND under a clinical hold include unreasonable or significant risk of illness or injury to trial subjects, insufficient information or procedures to assess and minimize patient risks, inadequate qualifications of the clinical investigators, or a misleading, erroneous, or incomplete Investigator's Brochure (a document containing all relevant information about the drug). Revisions to clinical protocols, as well as new protocols or substudies, are submitted to the FDA as amendments to the IND. Progress reports regarding the trial must also be provided annually. The contents of the annual report are specified in the regulations. Furthermore, the sponsor must report any unexpected, serious adverse drug events that occur during the trial must be reported to the FDA and the IRB within specified time periods.

Not all clinical trials require INDs. A sponsor proposing a trial with a commercially available, FDA-approved drug product is exempt, if the trial (1) is not intended to be submitted to the FDA to support labeling changes or a new indication; (2) is not intended to support a major change in advertising; and (3) does not involve a route of administration, dose, or patient population that significantly increases the risk of the drug. An IND is not required, if the trial is exempt according to the above criteria, regardless whether a placebo (inert or inactive treatment) is employed as a control group. Independent investigators, rather than pharmaceutical companies, often conduct these types of clinical trials.

CLINICAL INVESTIGATIONS

Clinical investigations involve the administration of a drug product to humans. This segment of the drug development process requires substantial financial and time commitments.¹¹ Figure 8.2 shows the considerable increase in development costs associated with the initiation of clinical trials. Human

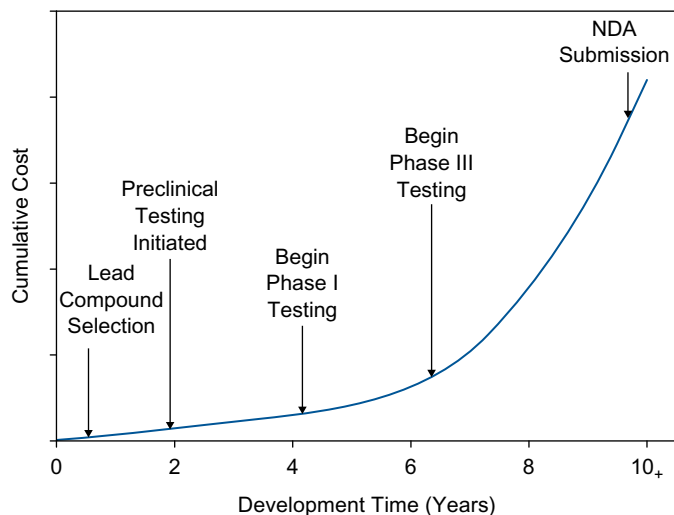


Figure 8-2. Relationship between the time devoted to new drug development and dollars invested.⁶

testing is divided into four phases, each phase having specific objectives. The following sections discuss the various phases of clinical testing.

Phase I Clinical Trials

The first series of experiments performed in humans occurs during Phase I clinical testing. A small number of healthy volunteers (approximately 20–80 people) are exposed to the new drug product in closely monitored trials primarily to assess the compound's safety. For the investigation of drugs to treat life-threatening diseases, such as cancer or Acquired Immune Deficiency Syndrome (AIDs), patients afflicted with the disease may be enrolled.¹² In Phase I trials, the starting dose is low, often one-tenth of the highest no-effect dose in the animal models. After the initial treatment is completed, additional subjects may be recruited, and higher doses may be administered to determine the maximum dose tolerated without significant side effects. During this phase of testing, preliminary absorption, distribution, metabolism, and excretion data of the parent drug and all metabolites should be evaluated. Data regarding pharmacokinetic and pharmacological effects are used in the design of future Phase II trials.

Phase II Clinical Trials

Phase II clinical testing shifts the focus of the trials from safety to efficacy. In comparison to Phase I trials, a larger number of people (100–300 patients) are enrolled, and the majority of these participants suffer from the target illness. Side effects from the new drug product are also investigated. These clinical trials are closely monitored and well-controlled (see Clinical Trial Planning and Design section). Failure during Phase II testing is common, as the human body is more complex than a test tube or animal testing. As with any study conducted in humans, clinical protocols for Phase II trials must be sent to the FDA prior to beginning the trial. They may be included as amendments to the IND approved for a Phase I study.

Phase III Clinical Trials

At the end of Phase II testing, sponsors are encouraged to meet with the FDA personnel who are assigned to the review division for the therapeutic area (for example, Division of Psychiatry Products, Division of Neurology Products, Cardiovascular Renal Products, etc.). These meetings involve review of the acceptability of past trials, the design of future trials, and the general drug development plan. Scientists at the FDA carefully review preclinical and clinical data in evaluating proposed Phase III protocols. Specific areas of the proposed Phase

III trials that are scrutinized include the study objectives, informed consent, inclusion/exclusion criteria, dosing regimens, methods and timing of data collection, duration of treatment and follow-up assessment, blinding of the drug products and plans for maintaining the blind, plans to assess compliance with the protocol, identification of primary outcome variables, and methods to account for dropouts. Addressing these key areas of proposed Phase III protocols is expected to limit the bias of trial results. An overall goal of the meeting is a good-faith agreement between the sponsor and the FDA regarding data required for submission of a New Drug Application (NDA), the final regulatory hurdle before the drug product can be marketed.

Phase III clinical trials are the longest and most comprehensive evaluation of new compounds. Significantly larger numbers of patients (1000–3000) who are afflicted with the target illness are tested. Patients are often recruited, tested, and monitored by several major hospitals and clinics throughout the country. Phase III trials may also be conducted internationally. In addition to determining efficacy, these trials monitor adverse reactions. The new drug may be compared to existing therapeutic regimens (that is, comparator products) or a placebo. The final market formulation for the drug product should be optimized prior to the start of Phase III trials. Compounds that successfully complete Phase III testing have a 95% chance of being approved by the FDA.¹

Prior to the completion of Phase III testing and New Drug Application (NDA) submission, sponsors are encouraged to meet with the appropriate review division of the FDA again. These meetings help establish the format of the submission, so that the review proceeds smoothly and to determine whether additional animal or human trials are necessary. The meeting should be held sufficiently in advance of the tentative New Drug Application (NDA) filing date to allow ample time to incorporate recommended changes or perform additional trials.

Phase IV Clinical Testing

Phase IV trials are post-approval clinical trials designed for one of several reasons. The FDA may mandate Phase IV testing in a specific patient population to further assess efficacy and side effects. Companies may also choose to conduct additional clinical tests to more fully understand how their product compares to other commercially available therapeutic regimens. Since duration of exposure and number of patients treated are often limited during Phase III testing, Phase IV trials may be required to assess long-term safety of the drug.

THE NEW DRUG APPLICATION

Once the Phase III trials have been completed, all preclinical and clinical data are compiled into a New Drug Application (NDA), which is submitted to the FDA for review. The FDA also reviews the product's labeling and package insert. The NDA approval process is the last hurdle prior to marketing. An NDA document typically consists of hundreds of thousands of pages and contains highly detailed information. Regulation guidelines, including the information required for an NDA, are provided in 21 CFR Part 314 Subpart B. Primary items include:

1. safety and efficacy of the drug treatment(s)
2. components of drug product(s)
3. description of methods and controls used in manufacturing the active ingredient and drug delivery system and its packaging
4. proposed labeling.

According to the Center for Drug Evaluation and Research, the time for a standard NDA review has been reduced from a median of 22 months in 1992 to approximately 13 months in 2008 (6 months for priority review).¹³ The faster review times have been attributed to the Prescription Drug User Fee Act (PD-
UFA) of 1992.¹⁴

When an NDA is submitted, relevant sections of the document are distributed to the appropriate reviewers and evaluated first for completeness. If the document is sufficiently complete, the NDA is accepted for review and assigned a priority status. NDAs for new chemical entities are classified as either “P” for priority review or “S” for standard review. A “P” rating is given to new drug products with improved therapeutic effects, safety, and/or side effects in comparison to currently marketed drugs. NDAs assigned a “P” rating are expected to be reviewed in a more timely manner than those assigned an “S” rating. If the NDA is deemed too incomplete to review, it is not filed. The decision to accept the NDA is made within 60 days of the date of submission.

Once the NDA is accepted, detailed evaluation continues, and the FDA has 180 days from submission to complete the review. Each reviewer submits written comments of his or her assigned section and makes a recommendation. The NDA may also be presented to an Advisory Committee for comment. All documents are then compiled and ultimately submitted to the Director of the Office of Drug Evaluation. The FDA may approve the product for market, approve with specific conditions attached (Conditional Approval), or disapprove the drug product. Primary reasons for disapproval include lack of demonstrated safety and efficacy, issues with the manufacturing/processing procedures, or false/misleading labeling. If not approved, a letter is sent to the sponsor detailing deficiencies in the application. If the NDA is approved, an approval letter, along with a draft of the product labeling, is sent to the sponsor. The label is a combination of the draft submitted by the sponsor and revisions provided by the reviewing section of the FDA. Standardized labeling requirements are provided in 21 CFR Part 201.57.

Prior to NDA approval, the FDA conducts an inspection of the sponsor’s facilities to ensure compliance with current Good Manufacturing Practices (cGMPs) as set forth in 21 CFR Parts 210 and 211. These are industry standards to ensure consistent quality of manufactured drug products. Preapproval inspections are conducted within 45 days of the NDA acceptance. If deficiencies are noted, a letter (FDA Form 483) is sent to the sponsor delineating the problems. Once the deficiencies are resolved, the company must provide written certification, and the FDA will then clear the application within 45 days, if the corrections are adequate. As this step is critical in the approval process, companies often hold mock preapproval audits.

The NDA approval process is complicated and challenging, and the process may involve several revisions. In 2008, 70% of priority reviews were approved and 60% of standard reviews were approved as their original submissions (<http://www.fda.gov/downloads/AboutFDA/ReportsManualsForms/Reports/UserFeeReports/PerformanceReports/PDUFA/UCM243358.pdf>).

In response to the FDA Modernization Act of November 1997, the National Institutes of Health (NIH), in collaboration with the FDA, established the Clinical Trials Website (www.clinicaltrials.gov). This website is a searchable database that provides information on clinical trials, including the purpose of the study, recruiting status, criteria for participation, study drug(s), and the location of the trials. Investigators are required by law to provide results on the website once the trial is completed. This process was established to ensure that the results of nonsignificant or negative studies were made public. Most medical journals require proof of registry at clinicaltrials.gov prior to publication of clinical trial results.

THE ABBREVIATED NEW DRUG APPLICATION (ANDA)

In addition to approving new drug products for the United States, the FDA is charged with the approval of generic drug products (21CFR Part 314). This work is accomplished through the Center for Drug Evaluation and Research’s Office of Generic Drugs. A generic drug product must be bioequivalent in comparison to an approved proprietary drug product. The review process for generic drugs is specifically focused on bioequivalence testing rather than safety and efficacy. Thus, conventional

clinical testing is not required. To be considered bioequivalent, both the rate and extent of drug absorption must be within established parameters in comparison to the reference drug. *In vivo* (within a biological system) bioequivalence testing is required for most dosage forms. Applicants may request a waiver from performing *in vivo* bioequivalence studies for certain drug products where bioavailability may be established by submitting (1) a formulation comparison for products whose bioavailability is evident (that is, oral solutions, injectables) or (2) comparative dissolution. The FDA provides guidance on establishing bioequivalence. If any portion of the application is not acceptable, a letter of deficiency that details the insufficiencies and requests additional information and data to resolve these concerns is issued. A tentative approval letter delaying the marketing of the generic product may be issued, if approval of the generic occurs prior to the expiration date of patents or exclusivities of the reference drug product.

RAPID ACCESS TO NEW DRUG PRODUCTS

As a result of the demand for more rapid access to new drug products, the FDA has written several regulations and policies specifically designed for drugs intended to treat severely debilitating or life-threatening illnesses. Subpart E (21 CFR 312.80-.88) regulations expedite the development and approval process. For example, treatment INDs (21 CFR 312.34) are intended to make drugs relatively far along in the development process available to seriously ill patients and are typically made available during Phase III clinical trials.

Accelerated approval is a mechanism whereby products are approved for marketing based on limited data and exposure of small numbers of patients to the product. Approvals may be granted when studies have been performed using surrogate endpoints, such as laboratory finding or physical signs that may not directly measure patient response yet is considered likely to predict therapeutic benefit or delays in disease progression rather than ultimate endpoints, such as mortality. The process is limited to life-threatening indications, such as HIV and cancer. When a drug is approved via accelerated approval, the FDA requires specific post-approval studies be performed within a predetermined amount of time. However, often, these studies are difficult to perform, because patients may not be willing to participate in trials in which the new treatment may not be given. Thus, completion of these post-approval commitments is often delayed.¹⁵ For example, from December 1992 to July 2010, the FDA approved 47 new oncology indications for 35 drugs, but only 26 were converted to regular approval, within a median of 3.9 years. Delays in completing post-marketing studies have been as much as 10.5 years.¹⁵ Lack of completion of confirmatory trials was cited as the reason for not moving to regular approval among 14 of the 21 indications. Furthermore, as with other FDA approved products, the use of accelerated approval products often extends beyond the approved indications, because prescribers are allowed to utilize them in off-label settings.

Orphan Drug Approval

Orphan drugs are drugs used to treat rare diseases or conditions that affect less than 200,000 people in the United States. Orphan drugs go through the same FDA review process previously described. However, the review is expedited, as the majority of orphan drugs are used in the treatment of serious or life-threatening disease. The process by which a company can file an application for orphan drug designation is described in 21 CFR Section 316.20. Due to substantial drug development costs, orphan drugs provide limited opportunities for companies to recoup their investments. The United States federal government, through the Orphan Drug Act of 1983, established tax incentives, reduced user fees, and created exclusivity agreements to encourage research in the orphan diseases.¹⁶ Grants are also available through the FDA to support clinical research, and annual requests for applications may be found in

the Federal Register. Eight new products (new chemical entities and new biologic license applications) for orphan diseases were approved in 2009.¹⁷

OVER-THE-COUNTER DRUG APPROVAL

The approval process for over-the-counter (OTC) drugs is considerably different from prescription medications, and their review is not held to the same standards as an NDA. The first phase of the approval process involves an advisory panel consisting of a multidisciplinary group of scientists that review data provided by manufacturers and other previously published research. The findings are submitted to the FDA, and these reports are subsequently summarized in the Federal Register. Interested parties are given an opportunity to comment. Next, the FDA reviews all statements and publishes a tentative final monograph. The FDA also publishes the nature of the comments received and provides further opportunity for feedback. Then, the final monograph is published in the Federal Register and goes into effect one year after publication. The monographs establish conditions under which over-the-counter drugs are recognized as safe and effective and are not misbranded. By following a monograph, a company can then market an over-the-counter drug without additional FDA approval. For any unsubstantiated claims that a company wishes to make (that is, claims not approved in the monograph), data must be presented to the FDA to justify revision of the monograph or the sponsor may submit a NDA.

POST-APPROVAL ACTIVITIES

SAFETY MONITORING

After an NDA has been granted and marketing is initiated, drug safety is still monitored. Sponsors of the NDA must periodically submit reports of adverse events. For newly approved drugs, these reports are filed quarterly for the first three years, then annually thereafter. For adverse events that are considered serious and unexpected (that is, fatal or life-threatening, permanently disabling, or requiring or prolonging hospitalization), the sponsor must provide a report to the FDA as soon as possible (a written report within 15 days of receipt of the information and a telephone or facsimile report within 7 calendar days). The FDA's MedWatch program (see <http://www.fda.gov/Safety/MedWatch/default.htm>) encourages health-care providers and patients to directly report serious adverse reactions to drugs to the FDA. The program also provides alerts to practitioners regarding actions and recommendations by the FDA. Serious adverse events may require minor labeling changes or the addition of warning or precaution statements. If serious safety concerns arise, the FDA may withdraw approval of the NDA. Another alternative is the addition of a "Boxed Warning" in the product label. Boxed warnings are usually accompanied by a "Dear Health Professional" letter that is sent directly to licensed health providers to increase awareness of the potential problem. The boxed warnings may specify a newly identified risk or provide additional guidelines for use of the products in certain patient groups. Often, an FDA Advisory Committee reviews the NDA in light of the new data prior to an official NDA withdrawal. In some instances, manufacturers have withdrawn drug products prior to FDA action.¹⁸

For certain medications that pose a serious and significant public health concern, the FDA requires the distribution of approved patient medication information. This Medication Guide is intended to ensure patients' safe and effective use of the drug products. The FDA Amendments Act of 2007 further requires a risk evaluation and mitigation strategy (REMS) from manufacturers to ensure that the benefits of a drug or biological product outweigh its risks. A draft guidance on the subject of REMS and Medication Guides was recently issued by the FDA and can be found on its website, along with a list of products subject to a risk evaluation and mitigation strategy.

CHANGES TO AN APPROVED PRODUCT

Any change made to an FDA-approved drug product, including component or composition, chemical synthesis, analytical methods, manufacturing site, manufacturing process, batch size, or labeling, must be submitted to the FDA. Some of the so-called scale-up and post-approval changes (SUPAC) require FDA approval prior to the implementation of the change. Depending on the type of change made and the impact the change may have on the quality of the drug product, notification to the FDA may be provided through annual reports or supplemental new drug applications (SNDA).

SUMMARY

New drug products must be shown as safe and effective before they are approved by the FDA for marketing in the United States. Other countries have regulatory authorities similar to the FDA that oversee new drug approvals. This chapter, focused on drug approval in the United States, outlined the various stages involved in new drug development, including the milestones of Investigational New Drug Application and New Drug Application submission. The costs associated with the development of new drug products are substantial, and the most significant expenditures occur during clinical testing. Thus, the design and conduct of clinical trials are critical to successful drug product development. Several considerations in clinical trial design have been highlighted. The reader is encouraged to refer to specific FDA guidance documents and other referenced materials for further information.

REFERENCES

1. Cato A *et al.* Current challenges and future directions of drug development. In: Cato A *et al.* eds. *Clinical Drug Trials and Tribulations*. New York: Marcel Dekker Inc., 2002: 1–19.
2. Adams CP, Brantner VV. Spending on new drug development. *Health Economics* 2010; 19(2): 130–141.
3. Grabowski HG *et al.* Returns on research and development for 1990s new drug introductions. *Pharmacoeconomics* 2002; 20(Suppl.3): 11–29.
4. Wax PM. Elixirs, diluents, and the passage of the 1938 Federal Food, Drug and Cosmetic Act. *Ann Intern Med* 1995; 122(6): 456–461.
5. Matthews SJ, McCoy C. Thalidomide: a review of approved and investigational uses. *Clin Ther* 2003; 25(2): 342–395.
6. Goyan JE. Science: part of the problem, most of the answer. *Drug Intell Clin Pharm* 1983; 17(7-8): 566–569.
7. Calabrese L, Fleischer AB. Thalidomide: current and potential clinical applications. *Am J Med* 2000; 108(6): 487–495.
8. Craig 3rd, SP, Eakin AE. Structure-based inhibitor design. *Vitam Horm* 2000; 58: 149–169.
9. Lato SM, Ellington AD. Screening chemical libraries for nucleic-acid-binding drugs by in vitro selection: a test case with lividomycin. *Mol Divers* 1996; 2(1-2): 103–110.
10. Lednicher D. Exploration of a chemical series as a drug discovery strategy. *Drug Des Deliv* 1987; 2(1): 69–78.
11. Lakings DB. Nonclinical drug development: pharmacology, drug metabolism, and toxicology. In: Guarino RA, ed. *New Drug Approval Process*. New York: Marcel Dekker Inc., 2000: 17–54.
12. Lu C *et al.* A phase I study of topotecan/paclitaxel alternating with etoposide/cisplatin and thoracic irradiation in patients with limited small cell lung cancer. *Clin Cancer Res* 2003; 9(6): 2085–2091.
13. FDA (2011). *CDER Approval Times for Priority and Standard NMEs and New BLAs* CY 1993 – 2008*. www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/HowDrugsareDevelopedandApproved/DrugandBiologicApprovalReports/UCM123959.pdf (accessed 2 August 2011).
14. FDA (2011). *Prescription Drug User Fee Act (PDUFA)*. www.fda.gov/ForIndustry/UserFees/PrescriptionDrugUserFee/default.htm (accessed 2 August 2011).

15. Johnson J *et al.* Accelerated approval of oncology products: the Food and Drug Administration experience. *Natl Cancer Inst* 2011; 103: 636–644.
16. Shulman SR, Manocchia M. The US orphan drug programme 1983–1995. *Pharmacoeconomics* 1997; 12(3): 312–326.
17. FDA (2010). *CDER New Molecular Entity (NME)/ New BLA Calendar Year Approvals As of December 31, 2009*. www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/HowDrugsareDevelopedandApproved/DrugandBiologicApprovalReports/NMEDrugandNewBiologicApprovals/UCM091096.pdf (accessed 2 August 2011).
18. Chaudhry MU, Simmons DL. Case of the month. Hepatic and renal failure in a patient taking troglitazone and metformin. *J Ark Med Soc* 2001; 98(1): 16–19.

Chapter 9

Metrology and Pharmaceutical Calculations

Roger L. Schnaare, PhD and Shelly J. Prince, PhD

WEIGHTS AND MEASURES 101

WEIGHING AND MEASURING 106

DENSITY AND SPECIFIC GRAVITY 113

PHARMACEUTICAL CALCULATIONS 114

LOGARITHMS 115

PHARMACEUTICAL PROBLEMS 116

ANSWERS TO PROBLEMS 129

APPENDIX A 132

One of the first technical operations that the student of pharmacy must learn is the manipulation of balances, weights, and measures of volume. This entails a study of the various systems of weights and measures, their relationships, and a mastery of the mathematics involved. This chapter considers the fundamental principles of metrology underlying the testing, manufacturing, and compounding of pharmaceutical preparations:

- **Weights and measures**—An accumulation of facts concerning the various systems, with tables of conversion factors and practical equivalents. The relationships among the various systems of weights and measures are clarified.
- **Weighing and measuring**—A discussion of the various types of balances, particularly prescription balances and methods of using, testing, and protecting them; also of various devices and methods for measuring large or small volumes of fluids.
- **Density and specific gravity**—A consideration of the mass/volume ratio of a substance (density), and the ratio of the weight (mass) of one substance to the weight (mass) of another substance taken as the standard (specific gravity).
- **Pharmaceutical calculations**—A review of basic mathematical principles and their use in solving pharmaceutical problems.

WEIGHTS AND MEASURES

Weight is a measure of the gravitational force acting on a body; weight is directly proportional to the body's mass. The latter, being a constant based on inertia, never varies, whereas weight varies slightly with latitude, altitude, temperature, and pressure. The effect of these factors is usually not considered unless very precise weighing and large quantities are involved.

Measure is the determination of the volume or extent of a body. Temperature and pressure have a pronounced effect, especially on gases or liquids. These factors, therefore, are considered when making precise measurements.

All standard weights and measures in the US are derived from or based on the United States National Prototype Standards of the Meter and the Kilogram. The standards are made of platinum-iridium, and are in the custody of the National Institute of Standards and Technology (NIST) in Washington, DC.

HISTORY

A brief outline of the origin of the many systems of weights and measures may help clarify the essential distinctions between them. The sense of the weight of a body cannot be conveyed intelligibly to the mind unless a means of comparison is chosen. As weight is the measure of the gravitational force of a body, this force is expressed in terms of standards of resistance, which exactly balance the body and keep it in equilibrium when used with a mechanical device constructed for this specific purpose. Such standards are termed *weights* and the mechanical devices are called *balances* or *scales*.

The standards that have been chosen by various nations are arbitrary, and instances are common where different standards are in use at the same time in the same country. Many of the ancient standards clearly are referable to variable parts of the human body, such as nail, foot, span, pace, cubit (length of the forearm), and fathom or faethm (stretch of the arms). In the history of metrology three periods may be traced:

- The *Ancient* period, during which the old classical standards originated, terminated with the decline of the Roman Empire. The unit of distance used by all nations for maritime measurements, the *nautical* or *meridian* mile (1/60 of a degree of the earth's equatorial circumference) is exactly equal to 1000 Egyptian fathoms or 4000 Egyptian cubits. These Egyptian measurements, which have persisted for more than 4000 years, were based on astronomical or meridian measurements that were recorded imperishably in the great Pyramid at Ghizeh, whose perimeter is exactly 500 of these fathoms, or 1/2 nautical mile.
- The *Medieval* period extended to the sixteenth century. During this period the old standards were lost, but their names were preserved, and European nations adopted various independent standards.
- The *Modern* period extends from the sixteenth century to the present. Since the seventeenth century, the efforts of most enlightened nations have been directed toward scientific accuracy and simplicity, and during the present century toward international uniformity.

Historical metrology, also referred to as *documentary metrology*, is concerned with the study of monuments and records of ancient periods. *Inductive metrology* is concerned with the accumulation of data concerning the measurement of large numbers of objects that have been referred to as standards but which have no exact measure except by statutory regulation.

The English Systems

In Great Britain, in 1266, the 51st Act of the reign of Henry III declared:

that by the consent of the whole realm of England the measure of the King was made—that is to say, that an English silver penny called the sterling, round and without clipping, shall weigh *thirty-two grains of wheat*, well dried and gathered out of the middle of the ear; and twenty pence (pennyweights) do make an ounce and twelve ounces a pound, and eight pounds do make a gallon of wine, and eight wine gallons do make a bushel, which is the eighth of a quarter.

The 16-ounce pound (*avoirdupois pound*), undoubtedly of Roman origin, was introduced at the time of the first civilization of the British island. However, according to Gray, the word

“haberdepois” was first used in English laws in 1303. A statute of Edward I (AD 1304) states:

that every *pound* of money or of *medicines* is of *twenty shillings weight*, but the pound of all other things is *twenty-five shillings weight*. The *ounce of medicines* consists of *twenty pence*, and the *pound* contains *twelve ounces* [the troy pound], but in other things the pound contains *fifteen ounces*, in both cases the ounce weighing twenty pence.

These laws unfold the theory of the ancient weights and measures of Great Britain, and reveal the standards (i.e., a natural object, grains of wheat). A difference existed then between the troy and the avoirdupois pound, but the weights now in use are 1/16 heavier than those of Edward I, due to the change subsequently made in the value of the coin by the sovereign. In addition, the true pennyweight standard was lost, and, in the next revision of the weights and measures, the present troy and avoirdupois standards were adopted.

The troy *weight* is of still earlier origin. The great fairs of the eighth and ninth centuries were held at several French cities, including Troyes, the gathering place of traders from all countries. Coins were frequently mutilated, so they were sold by weight, and the standard weight of Troyes for selling coin was adopted for precious metals and medicines in all parts of Europe. The troy ounce and the avoirdupois ounce were originally intended to have the same weight, but after the revision it was found that the avoirdupois ounce was lighter by 42½ gr (grains) than the troy ounce. The subsequent adoption of troy weight by the London College of Physicians in 1618, on the recommendation of Sir Theodore Turquet de la Mayerne who compiled their first pharmacopeia, has entailed upon all apothecaries who are governed by British customs to this day the very great inconvenience of buying and selling medicines by one system of weights (the *avoirdupois*) and compounding them by another (the *apothecary* or *troy*).

In the next century efforts were made toward reforming the standards, and in 1736 the Royal Society began the work that ended in the preparation, by Mr. Bird under the direction of the House of Commons, of the standard *yard* and standard *pound* troy in 1760. Copies of these were prepared and no intentional deviation has been made since.

The growing popularity of the French metric system—and the desirability of securing a standard that could be recovered easily in case of loss or destruction, and that should be commensurable with a simple unit—prompted steps in England to secure these advantages in 1816. The labors of English scientists led to the adoption of the *imperial* measures and standards, which were legalized on January 1, 1826; imperial standards are now in use in Great Britain, thus introducing another element of confusion into an already complicated subject. In this system the *yard* is equivalent to 36 inches, and its length was determined by comparison with a pendulum beating seconds of mean time, in a vacuum, at a temperature of 62°F at the level of the sea in the latitude of London, a length that was found to be 39.1393 inches. The *pound troy* (containing 5760 gr) was determined by comparison with a given measure of distilled water under specified conditions. Thus, a cubic inch of distilled water was weighed with brass weights in air at 62°F, the barometer at 30 inches, and it weighed 252.458 gr. The standard for measures of capacity in Great Britain (either dry or liquid) is the *imperial gallon*, which contains 10 lb avoirdupois (each 7000 gr) of distilled water weighed in air at 62°F, the barometer standing at 30 inches. The *bushel* contains eight such gallons.

George Washington, in his first annual message to Congress, in January 1790, recommended the establishment of uniformity in currency, weights, and measures. Action was taken with reference to the currency and recommendations were made by Thomas Jefferson, then the Secretary of State, for the adoption of either the currently used English systems or a decimal

system. However, nothing was accomplished until 1819–1820, when efforts again were made in the United States to secure uniformity in the standards that were in use by the several states. Finally, after a lengthy investigation, on June 14, 1836, the Secretary of the Treasury was directed by Congress to furnish each state in the Union with a complete set of the revised standards, and thus the *troy pound* (5760 gr), the *avoirdupois pound* (7000 gr), and the *yard* (36 inches) are all identical with the British standards. However, the *US gallon* is quite different; the old wine gallon of 231 inch³—containing 58 372.2 gr of distilled water at its maximum density, weighed in air at 62°F, the barometer standing at 30 inches—was retained. The bushel contained 77.274 lb of water under the same conditions, thus making the dry quart about 16 percent greater in volume than the liquid quart.

In 1864 the use of the metric measures was legalized in Great Britain, but was not made compulsory, and in 1866 the United States followed the same course. By the US law of July 28, 1866, all lengths, areas, and cubic measures are derived from the international meter equivalent to 39.37 inches. Since 1893 the US Office of Standard Weights and Measures has been authorized to derive the yard from the meter: one yard equals 3600/3937 m, and the customary weights are referred to the kilogram by an Executive Order approved April 5, 1893. Capacities were to be based on the equivalent; dm³ equals one liter, the decimeter being equal to 3.937 inches. The gallon still remains at 231 inch³ and the bushel contains 2150.42 inch³. This makes the liquid quart equal to 0.946 liter and the dry quart equal to 1.1013 liter, whereas the imperial quart is 1.1359 liter. The customary weights are derived from the international kilogram, based on the value that one avoirdupois lb equals 453.5924277 g and that 5760/7000 avoirdupois lb equals one troy lb.

Avoirdupois weight is used in general in the United States for commercial purposes, including the buying and selling of drugs on a large scale and occasionally on prescription orders.

The Metric System

The idea of adopting a scientific standard for the basis of metrology that could be reverified accurately was suggested by a number of individuals after the Renaissance. Jean Picard, the seventeenth-century French astronomer, proposed that the length of a pendulum beating one second of time at sea level, at latitude 45° should be taken as a unit.

In 1783, the Scottish inventor James Watt first suggested the application of decimal notation, and the commensurability of weight, length, and volume. The French National Assembly in 1790 appointed a committee to decide the preferability of the pendulum standard or a terrestrial measure of some kind as a basis for the new system. The committee reported in 1791 in favor of the latter, and commissions were appointed to measure an arc of meridian and to perfect the details of the commensurability of the units and of nomenclature. However, certain inaccuracies were inherent in the early standards, so they do not bear the intended exact relationships to each other. The present accepted standards are defined in publications of the National Institute of Standards and Technology (NIST).

In its original conception, the meter was the fundamental unit of the metric system, and all units of length and capacity were to be derived directly from the meter, which was intended to be equal to one ten-millionth of the earth's quadrant. Furthermore, it originally was planned that the unit of mass, the kilogram, should be identical with the mass of a cubic decimeter of water at its maximum density. At present, however, the units of length and mass are defined independently of these conceptions.

For all practical purposes, calibration of length standards in industry and scientific laboratories is accomplished by comparison with the material standard of length: the distance between two engraved lines on a platinum-iridium bar, the *International Prototype Meter*, which is kept at the International Bureau of Weights and Measures.

The *kilogram* is defined independently as the mass of a definite platinum-iridium standard, the *International Prototype Kilogram*, which also is kept at the International Bureau of Weights and Measures. The *liter* is defined as the volume of a kilogram of water, at standard atmospheric pressure, and at the temperature of its maximum density, approximately 4°C. The *meter* is thus the fundamental unit on which are based all metric standards and measurements of length and area and of volumes derived from linear measurements.

Of basic scientific interest is that on October 14, 1960, the 11th General Conference on Weights and Measures, meeting in Paris, adopted a new international definition for the standard of length: the meter is now defined as the length equal to 1,650,763.73 wavelengths of the orange-red light of the krypton-86 isotope. This standard will be used in actual measurements only when extreme accuracy is needed.

The kilogram is the fundamental unit on which all metric standards of mass are based. The liter is a secondary or derived unit of capacity or volume. The liter is larger by about 27 ppm (parts per million) than the cube of the tenth of the meter (the cubic decimeter): 1 liter = 1.000027 dm³.

The conversion tables in this publication that involve the relative length of the yard and meter are based on the relation: 1 m = 39.37 inch, contained in the Act of Congress of 1866. From this relation it follows that 1 inch = 25.40005 mm (nearly).

In recent years engineering and industrial interests the world over have urged the adoption of the simpler relation, 1 inch = 25.4 mm exactly, which differs from the preceding value by only 5 ppm. This simpler relation has not as yet been adopted officially by either Great Britain or the United States but is in wide industrial use.

In the United States, the abbreviation "cc" (for cubic centimeter) still persists in general use and is taken as synonymous for the more correct milliliter. The US Pharmacopeia (USP) IX and National Formulary (NF) IV adopted the term *milliliter* with its abbreviated form *mil*, but it proved so unpopular in practice that the following pharmacopeial convention directed the return to the older term cubic centimeter (cc). However, in 1955, USP XV and NF X once again adopted the term *milliliter* with the abbreviation mL.

National jealousies and the natural antipathy to changing established customs interfered greatly with the adoption of the metric system during the early part of the nineteenth century. At present the metric system is in use in every major country of the world. In the United States and Great Britain it is legalized for reference to and definition of other standards, and it is in exclusive use by nearly all scientists and by increasing segments of industry and the public. In the United States the metric system was legalized in 1866, but not made compulsory; in the same year the international prototype meter and kilogram were adopted as fundamental standards. The US silver coinage was based upon the metric system, the half dollar being exactly 12.5 g and the quarter and the dime being of the proportionate weights.

As corporations became more international, the need for a universal standard increased. Since 1875 there has been established and maintained an International Bureau of Weights and Measures, with headquarters in Paris. This Bureau is managed by an international committee that enjoys universal representation. One objective of the committee is to make and provide prototypes of the meter and kilogram for the subscribing nations; approximately 40 such copies have been prepared.

The US prototype standards of both the meter and the kilogram mass, constructed of a platinum-iridium alloy, were brought from Paris in 1890 and are now in the custody of the NIST in Washington, DC. They have been reproduced and distributed by the US government to the various states having bureaus needing such replicas. The original US prototype meter was taken back to Paris in 1957 for reverification and was found to have altered only 3 parts in 100,000,000 after 67 years of use. Thus, there was no demonstrable change within the limits

of experimental error. With the adoption of the krypton-86 wavelength of light definition for the meter the different countries have the means to check their prototype meter bars without returning them to Paris at periodic intervals for comparison with the international meter bar.

ORTHOGRAPHY AND READING

Orthography

There are two methods of orthography of the metric units in use. In the original French, the units are spelled *metre*, *litre*, and *gramme*; in the method proposed by the American Metric Bureau, the units are spelled *meter*, *liter*, and *gram*. For three decades after the original adoption of the metric system, the USP and NF adopted *meter* and *liter*, but used the French *gramme*. Now these official compendia use the spelling *gram*.

Reading

Some difficulty is usually experienced by those unfamiliar with the metric system in reading the quantities. In the linear measures in pharmacy, centimeters and millimeters are used almost exclusively; thus, 0.05 m would not be read five hundredths of a meter, but rather five centimeters (5 cm); if the millimeter column contains a unit, as in 0.055 m, it is read 55 millimeters (55 mm) in preference to fifty-five thousandths of a meter.

Fractions of a millimeter must be read decimally, as 0.0555 m, fifty-five and five-tenths millimeter (55.5 mm). In measures of capacity, cubic centimeters (cc) or milliliters (mL) are used exclusively for quantities of less than a liter. The terms half-liter, quarter-liter, 100 milliliters, and one milliliter are denoted by 500 mL, 250 mL, 100 mL, and 1 mL respectively.

In weight, when the quantity is relatively large and in commercial transactions, the *kilogram* is abbreviated to *kilo*. When less than a *kilogram* and not less than a *gram*, the quantity is read with the *gram* for the unit. Thus, 2000 g would be read either as 2000 grams or as 2 kilos, and 543 g would be read as 543 grams; 2543 g is sometimes read as 2 kilos and 543 grams, although 2543 grams usually is preferred.

For quantities below the *gram*, decigram and centigram usually are not used, but rather *milligram* has been regarded as the most convenient unit. With the increase in the use of extremely small doses of very potent drugs and the wide application of more delicate analytical procedures, the term *microgram* (mcg, µg, or γ), for thousandths of a milligram, is frequently used to designate quantities up to 999 µg (less than 1.000 mg).

Both the metric and English systems of weights and measures are in use in the United States. Even though the metric system has nearly replaced the English system, the pharmacist must have a practical knowledge of both.

WEIGHTS

The Metric System

The USP of 1890 adopted the metric system of weights and measures to the exclusion of all others except for equivalent dosage statements, and the British Pharmacopoeia of 1914 did likewise. In 1944 the Council on Pharmacy and Chemistry of the American Medical Association adopted the metric system exclusively. The advantages of the metric or decimal system, and its simplicity, brevity, and adaptability to everyday needs are now conceded universally.

FRACTIONAL AND MULTIPLE PREFIXES

In many experimental procedures, including some in the pharmaceutical sciences, very small (and occasionally very large) quantities of weight, length, volume, time, or radioactivity are measured. To avoid the use of numbers with many zeros in such cases, the NIST recognizes prefixes to be used to express fractions or multiples of the International System of Units (SI), which was established in 1960 by the General Conference on

Weights and Measures (see the foregoing discussion). The recognized prefixes, which in use are adjoined to an appropriate unit (as, for example, in such quantities as nanogram, picomole, microcurie, microsecond, or megavolt) are defined in Table 9-1.

Table 9-2 lists some metric weights. The prefixes, which indicate multiples, are of Greek derivation: deka, 10; hecto, 100; kilo, 1000. Fractions of the units are expressed by Latin prefixes: deci, 1/10; centi, 1/100; milli, 1/1000.

Only a few of the most convenient denominations are employed in practical work. Whole numbers from one to 1000 are usually expressed in terms of grams, while the kilogram is used as the unit for larger quantities. Quantities between one milligram and one gram are usually referred to in terms of milligrams; microgram (μg or mcg) is used in quantitative analysis, biological studies, and for minute dosage statements.

The English Systems

In the United States, both the avoirdupois and apothecary systems of weight measurement sometimes are used in handling medicines. It must be emphasized that *pharmacists may buy their drugs by avoirdupois weight*. These two systems differ:

1 pound avoirdupois = 7000 gr and is abbreviated lb.

1 pound apothecary = 5760 gr and is abbreviated tb.

1 ounce avoirdupois = 437.5 gr and is abbreviated oz.

1 ounce apothecary = 480 gr and is abbreviated $\mathfrak{z}$.

The *grain avoirdupois* is exactly the same as the *grain apothecary*. The apothecary pound is therefore 1240 gr lighter than the avoirdupois pound, and the apothecary ounce is therefore 42.5 gr heavier than the avoirdupois ounce.

The abbreviations of the denominations of apothecary weight are represented by the signs ounce, $\mathfrak{z}$; dram, $\mathfrak{d}$; scruple, $\mathfrak{s}$; and grain, gr. These have been in use for a long time but possibly may be mistaken for one another in rapid or careless writing. The abbreviations or signs of avoirdupois weight differ from those of apothecary weight, and care should be used not to confound them; they are lb (sometimes written #), pound; oz, ounce; gr, grain. Tables 9-3–9-5 show three English systems of weight.

Table 9-1. Prefixes for Fractions and Multiples of SI Units

Fraction	Prefix	Symbol	Multiple	Prefix	Symbol
10^{-1}	deci	d	10	deka	da
10^{-2}	centi	c	10^2	hecto	h
10^{-3}	milli	m	10^3	kilo	k
10^{-6}	micro	μ	10^6	mega	M
10^{-9}	nano	n	10^9	giga	G
10^{-12}	pico	p	10^{12}	tera	T
10^{-15}	femto	f	10^{15}	peta	P
10^{-18}	atto	a	10^{18}	exa	E

Table 9-2. Metric Weight

1 microgram μg	= 0.000001 g
1 milligram mg	= 0.001 g
1 centigram cg	= 0.01 g
1 decigram dg	= 0.1 g
1 gram g	= 1 g
1 dekagram dag	= 10 g
1 hectogram hg	= 100 g
1 kilogram kg	= 1000 g

Note: The abbreviation μg or mcg is used for microgram in pharmacy, rather than gamma (γ) as in biology.

Table 9-3. Avoirdupois Weight

Pounds	Ounces	Grains
1 =	16 =	7000
	1 =	437.5

Note: 2000 lb = 1 ton, and 2240 lb = 1 long ton.

Table 9-4. Apothecary Weight

Pounds	Ounces	Drams	Scruples	Grains
1 =	12 =	96 =	288 =	5760
	1 =	8 =	24 =	480
		1 =	3 =	60
			1 =	20

Table 9-5. Troy Weight

Pounds	Ounces	Pennyweights	Grains
1 =	12 =	240 =	5760
	1 =	20 =	480
		1 =	24

Jewelers evaluate precious stones with troy weight, which is very similar to apothecary weight. The apothecary and troy grain, ounce, and pound are identical, but the ounces are subdivided differently. The *carat*, used by jewelers, is equal to 3.168 troy grains or four carat grains. When used to express the fineness of gold, one carat signifies 1/24 part. A 14-carat ring is 14/24 pure gold.

As indicated in the footnote to Table 9-6, a number of special metric system units are used in various pharmacopeial and nonofficial descriptions, tests, and assays of drugs and other substances to express linear measurements of very small dimension. These units and their symbols or abbreviations are listed in Table 9-7, together with their equivalents in terms of the other metric units and the inch.

Table 9-6. Metric Linear Measure

1 nanometer (nm)	= 0.000000001 m (0.001 μm : 10^{-9} m: 10 Å)
1 micrometer (μm)	= 0.000001 m (0.001 mm: 10^{-6} m: 10 000 Å)
1 millimeter (mm)	= 0.001 m
1 centimeter (cm)	= 0.01 m
1 decimeter (dm)	= 0.1 m
1 meter (m)	= 1 m
1 dekameter (dam)	= 10 m
1 hectometer (hm)	= 100 m
1 kilometer (km)	= 1000 m

Although the meter (m) is observed to be the initial unit, it is seldom necessary to use it in pharmaceutical practice, and the same holds true for a number of the above measures. The micrometer (μm), millimeter (mm), and centimeter (cm) are employed in the description of many official drugs. Measurements pertaining to spectrometric and colorimetric tests and assays of many official drugs are recorded in micrometers (μm) or reciprocal centimeters (cm^{-1}) for infrared and in nanometers (nm) for ultraviolet and visible wavelengths of light, respectively.

Table 9-7. Equivalent Linear Measurements

Unit	Inches	mm	μm	nm	$\text{\AA}$
1 inch	1	25.4	25 400	2.54×10^7	2.54×10^8
1 mm (millimeter)	0.0394	1	1000	10^6	10^7
1 μm (micrometer)	3.94×10^{-5}	10^{-3}	1	1000	10 000
1 nm (nanometer)	3.94×10^{-8}	10^{-6}	10^{-3}	1	10
1 $\text{\AA}$ (angstrom unit)	3.94×10^{-9}	10^{-7}	10^{-4}	0.1	1

MEASURES

Systems

Two systems of linear measure are used in the United States: English and metric. Two systems of liquid measure are used: apothecary (also called the wine measure or US liquid measure) and metric. The units of the English system of linear measure (inch, foot, yard, and mile) are well known, and need not be described here. The units of the metric systems of linear and liquid measure, and of the apothecary (wine, US liquid) system of liquid measure, with their respective equivalents, are given in Tables 9-7, 9-8, and 9-9.

Pharmacists who fill Canadian or British prescriptions should also be familiar with the substantially different British imperial liquid measure system; the units, with their equivalents, are given in Table 9-10.

The following facts concerning the US system of liquid measure (see Table 9-9) should be noted:

- The apothecary fluid ounce (f ℥) of distilled water weighs 455 gr at 25°C.
- The apothecary pint contains 16 f ℥.
- The US gallon contains 128 f ℥ or 231 inch³. One gallon of distilled water weighs 8.337 avoirdupois lb at 62°F. The US pint therefore weighs 1.04 avoirdupois lb and the pound of distilled water measures only 0.96 pt. *One pound does not measure one pt.*

The following facts concerning the imperial system (see Table 9-9) should be noted:

- The imperial fluid ounce of distilled water weighs 437.5 gr at 15.6°C (60°F). It therefore weighs one avoirdupois oz.
- The imperial pint contains 20 f ℥.
- The imperial gallon contains 160 f ℥. One gal of distilled water weighs 10 avoirdupois lb; 16 f ℥ in this system therefore weighs one avoirdupois lb.

From the above, one can deduce the following:

- The US fluid ounce and minim are larger than the imperial fluid ounce and minim (<minim>). One US minim or fluid ounce equals 1.04 imperial minims or fluid ounces.
- The imperial pint and gallon are much larger than the US pint and gallon.

It is, therefore, inaccurate to use measuring devices calibrated in the US system in measuring quantities directed in English prescriptions when the imperial measure is intended. Conversely, devices calibrated in the imperial system should not be used to measure quantities directed in US prescriptions when the US measure is intended. For example, Canadian pharmacists using American graduated cylinders should calculate percentage solutions on the basis of 454.6 gr of distilled water to the fluid ounce. This is one more argument in favor of adoption internationally by all pharmacists of the metric system of weights and measures.

THE RELATIONSHIPS OF WEIGHTS AND MEASURES

When the systems of weights and measures in use in the United States are examined, the lack of close relation between the different units is appreciated at once. Nevertheless, if the following points are used carefully, many pharmaceutical problems will be greatly simplified.

- Pharmacists may weigh themselves, buy merchandise, sell over the counter, and calculate postage, etc., using avoirdupois weight, which contains 437.5 gr in 1 oz.
- Pharmacists may compound formulas by apothecary weight, which contains 480 gr in 1 f ℥.
- One apothecary fluid ounce of water weighs 455 gr at 25°C. Since 480 minim weigh 455 gr, 1 minim weighs $455/480 = 0.95$ gr. 1 minim does *not* weigh 1 gr. 1 f ℥ does *not* weigh 1 f ℥.

Practical Equivalents

Tables of weights and measures and a table of practical equivalents should be kept in a conspicuous and convenient place

Table 9-8. Metric Liquid Measures

1 microliter (μL)	= 0.000001 L
1 milliliter (mL)	= 0.001 L
1 centiliter (cL)	= 0.01 L
1 deciliter (dL)	= 0.1 L
1 liter (L)	= 1 L
1 dekaliter (daL)	= 10 L
1 hectoliter (hL)	= 100 L
1 kiloliter (kL)	= 1000 L

Note: The standard of capacity is the liter, which is the volume of 1 kg of distilled water at its maximum density (approx. 4°C). Microliters (μL) are used to measure volumes of solutions used in chromatographic procedures for the separation and quantitative determination of some official drugs.

Table 9-9. Apothecary or Wine Measures (US)

Gallon	Pints	Fluid ounces	Fluidrams	Minims
1	8	128	1024	61 440
	1	16	128	7 680
		1	8	480
			1	60

Table 9-10. Imperial Measure (British)

Gallon	Pints	Fluid ounces	Fluidrams	Minims
1	8	160	1280	76 800
	1	20	160	9 600
		1	8	480
			1	60

in the prescription department, and the following equivalents, which are given with practical accuracy, should be committed to memory. Other equivalents may be calculated from these.

LINEAR MEASURE

1 meter = 39.4 inches
 1 inch = 2.54 cm = 25.4 mm
 1 micrometer = 1/1000 mm = 10⁻⁶ m = 1/25, 400 inch

LIQUID MEASURE

1 milliliter = 16.2 minim
 1 fluid ounce = 29.6 mL
 1 pint = 473 mL
 1 gallon = 3790 mL

WEIGHT

1 kilogram = 2.20 lb avoird
 1 pound avoird = 454 g
 1 ounce avoird = 28.4 g
 1 ounce apothecary = 31.1 g
 1 pound apothecary = 373 g
 1 gram = 15.4 gr
 1 grain = 64.8 mg

A table of metric doses with approximate apothecary equivalents is given in Appendix 9A.

Approximate Measures

In apportioning doses for a patient, the practitioner is usually compelled to order the liquid medicine to be administered in certain quantities that have been established by custom, and estimated as shown in Table 9-11.

In almost all cases, careful tests have found that modern teaspoons, tablespoons, dessertspoons, and teaspoons average 25 percent greater capacity than the theoretical quantities given in Table 9-11. The physician and the pharmacist therefore should recommend the use of accurately graduated medicine droppers, teaspoons, and calibrated measuring devices, which may be procured at a small cost (Figure 9-1).

Approximate Dose Equivalents

For many years the apothecaries' system of weights and measures was used widely by physicians and pharmacists when considering the doses of medicinal substances, and it was customary to translate these apothecary doses into relatively exact amounts when the metric equivalents were mentioned. Today, however, doses are established primarily in the metric system without considering the relation of these metric figures to the corresponding quantities in any other system of weights and measures.

It should be emphasized that exact alternative formulas in the avoirdupois system of weights and measures are not

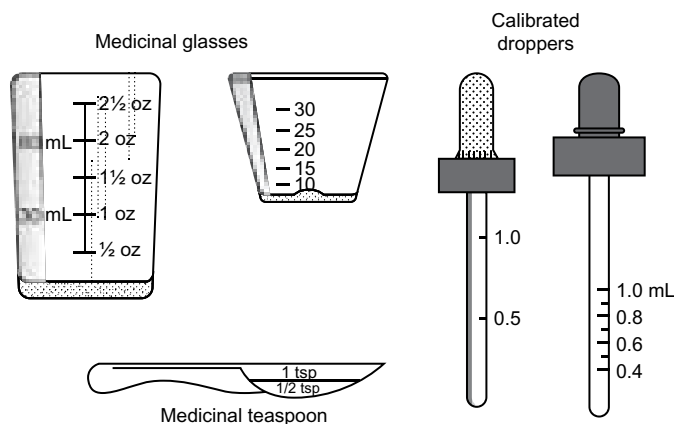


Figure 9-1. Examples of measuring devices.

obtained by using approximate equivalents but, for the purpose of compounding, should be calculated with the use of practical equivalents.

WEIGHING AND MEASURING

Having studied the several systems of weights and measures, students may now learn to apply their knowledge to the *weighing* and *measuring* of pharmaceuticals. The former process requires the use of the *balance*, or, for manufacturing purposes, *scales*, and the latter process requires the use of the *measure*, the *graduate*, and the *pipet*. The successful performance of many of the operations in pharmacy depends on a thorough knowledge of the principles of the balance and a correct understanding of its care and use; because weighing is nearly always the preliminary step in any compounding, it will be discussed first.

There is a relativity of accuracy in weighing (or measuring) that must not be overlooked, as illustrated by the following graded list: coal, salt, sugar, epsom salt, penicillin G, morphine, digoxin, vitamin B₁₂, and radium. One of the most important things for the pharmacist to learn is the degree of tolerance or error permissible in weighing or measuring any particular ingredient. Obviously, the final item on the list, radium, must be measured with much greater precision and accuracy than coal, the first item.

The empiric weighing and measuring methods of the kitchen, embodied in such concepts as a handful, a pinch, or "sweeten to suit your taste," have no place in pharmacy. Accurate work can be accomplished only by means of suitable apparatus.

WEIGHING

In pharmacy, weighing usually refers to ascertaining a definite weight of material to be used in compounding a prescription or manufacturing a dosage form.

The *balance* may be defined as an instrument for determining the relative weights of substances. It should be selected correctly for the specific task at hand, used skillfully, protected from damage, and checked periodically, if accurate results are to be obtained. Of even greater importance is its construction. Standards for balances are given by the NIST.¹

Construction of the Balance

For systematic consideration pharmaceutical balances may be classified as follows: single-beam (equal-arm or unequal arm), compound lever, torsion, and electronic.

SINGLE-BEAM EQUAL-ARM BALANCES

The principle on which single-beam equal-arm balances (or scales) operate is clearly evident in the construction of the

Table 9-11. Approximate Measures Established by Custom

Household Measurement	Apothecary Notation	Metric Volume
1 tumblerful	f $\frac{3}{4}$ viii	240 mL
1 teacupful	f $\frac{3}{4}$ iv	120 mL
1 wineglassful	f $\frac{3}{4}$ ii	60 mL
2 tablespoonfuls	f $\frac{3}{4}$ i	30 mL
1 tablespoonful	f $\frac{3}{4}$ iii or $\frac{3}{4}$ ss	15 mL
1 dessertspoonful	f $\frac{3}{4}$ ii	8 mL
2 teaspoonful	f $\frac{3}{4}$ i	5 mL
1/2 teaspoonful	f $\frac{3}{4}$ ss	2.5 mL

Note: one drop is often considered to be one minim, but this is incorrect, as drops are variable.

classical two-pan analytical balance. This type has a metallic lever or beam, divided into two equal arms at the center by a knife-edge, on which it is supported. At exactly equal distances from this point of support, and situated in the same plane, are placed the end knife-edges; these suspend the pans, which carry the substances to be weighed. A properly constructed balance of this type should meet the following requirements:

- *When the beam is in a horizontal position, the center of gravity should be slightly below the point of support, or central knife-edge, and perpendicular to it.* The relative sensitivity of the balance depends on the fulfillment of this principle, which may be illustrated roughly by forcing a pin through the center of a circular piece of pasteboard. If the edge of the pasteboard is touched slightly, it does not oscillate at all, but rather revolves around the center to a degree corresponding to the impulse given it. In this position it illustrates neutral equilibrium. If the pin is removed and inserted at a very short distance above the center, and the edge of the pasteboard touched as before, it will oscillate slowly, corresponding to a very sensitive beam, the point of support being slightly above the center of gravity as in the balance. If the pin is removed again and inserted far above the center, and the same impulse imparted to the edge, it will oscillate quickly, illustrating stable equilibrium characteristic of a beam that comes to rest quickly and is not particularly sensitive. Unstable equilibrium may be illustrated by balancing the disc so that the point of support is below the center. The slightest touch then causes it to reverse its position completely and finally come to rest with the center of gravity below the point of support.
- *The end knife-edges must be exactly equal distances from the central knife-edge; they all must be in the same plane and the edges absolutely parallel to each other.* It is apparent that the conditions of a good prescription balance cannot be satisfied if there is inequality in the length of the arms of the beam. The distance from the central knife-edge to the one on the left must be exactly the same as the distance from the central knife-edge to the one on the right, otherwise unequal weights would be required to establish equilibrium. If the central knife-edge is placed either above or below a line drawn so that it connects the end knife-edges, the loading of the pans either will cause the beam to cease oscillating or diminish the sensitivity in proportion to the load. If the knife-edges are not parallel, the weight of a body will not be constant upon every part of the pan, but will be greater if placed near the edge on one side, and correspondingly less at a point directly opposite.
- *The beam should be inflexible, but as light in weight as possible, and the knife-edges in fine balances should bear upon agate plates.* The rigidity of the beam is necessary because any serious deflection caused by a loading of the pans would lower the end knife-edges and thus accuracy in weighing would be impossible. The beam should not be heavier than necessary because the sensitivity of the balance thereby would be lessened; to diminish friction, which constantly increases with the age and use of a balance, the bearings of the knife-edges should be agate plates—polished flat pieces of the very hard mineral called agate. A single-beam equal-arm balance with the rider beam graduated to 28 g in increments of 0.2 g, and to 1 oz in increments of 0.01 oz, is shown in Figure 9-2.

UNEQUAL-ARM BALANCES

The unequal-arm balance is the type preferred for laboratory work when large amounts are to be weighed (Figure 9-3). The



Figure 9-2. Single-beam equal-arm balance. (Courtesy of OHAUS Corporation.)

lever principle on which these scales are constructed is based on the law of physics that at equilibrium the force applied at one end of the lever multiplied by the length of the arm (distance from the fulcrum to the point where the force is applied) must be equal to the product of the force acting at the opposite end of the lever and the length of the other arm. The inequality in the length of the arms of this beam permits the convenient use of movable weights upon the graduated longer arm of the beam, thus dispensing with the use of small weights, which are liable to be lost. This scale is of great advantage in laboratory or manufacturing work because it is particularly adapted for weighing liquids; a sliding tare is set on one beam for the weight of the container, and other sliding weights can be adjusted to the weight of liquid desired. These are available with the beams graduated either in the avoirdupois or metric system.

COMPOUND-LEVER BALANCES

The principle of the compound lever was first applied in the construction of balances by Robervahl of Paris, in about AD 1660. It was skillfully adapted for both prescription balances and the general counter and platform scales. The principal objection to this type of scale, when compared with single-beam balances, consists in the multiplicity of points of contact and suspension, thus necessarily increasing friction and the liability to disarrangement; however, their general convenience has made them popular.



Figure 9-3. Manufacturing laboratory scale and weights. (Courtesy of OHAUS Corporation.)

TORSION BALANCES

A simple illustration of the principle of torsion is afforded by tying a stout piece of cord to a firm support and inserting a lead pencil in the middle of the cord between the strands, at right angles to it. If the free end of the cord is stretched tightly, resistance is offered to any effort to turn the lead pencil over; if the pencil is released, it at once flies back to its original position. *Torsion* is the term applied to this method of twisting. The principle of supporting the beam of a balance on a tightly stretched wire, with the view of doing away with knife-edges and diminishing friction, occupied the attention of inventors for years.

In 1882 Professor Roeder and Dr. Springer contrived an ingenious torsion balance that gave promise of valuable results. Two illustrations of this original balance were shown on page 54 of the first edition of *Remington's Practice of Pharmacy* in 1885. Improvements have increased its efficiency greatly. The most important difficulty in applying the principle of torsion resistance was overcome by placing a weight just above the center of gravity. Torsional resistance tends to keep the beam in a horizontal position, while the elevation of a weight above the center of gravity, by its tendency to produce unstable equilibrium, exercises an opposite effect—the beam is inclined to be top heavy and, therefore, to tip on either side. If now the weight is made adjustable by mounting it upon a perpendicular screw so that it can be raised or lowered, it is possible to arrange these opposite forces so that one exactly neutralizes the other. In this manner sensitivity is obtained.

The torsion principle has been applied to prescription balances, as well as analytical balances and scales designed to carry heavier loads. In the torsion prescription balance two beams are used, supported on three frames, each of the latter having a flattened metallic band stretched tightly over its edge.

The torsion balance, which has a rider beam graduated upon the upper edge from 1/8 to 15 gr and on its lower edge from 0.01 to 1.0 g, furnishes a very convenient means of weighing small quantities without having to use small weights. Most modern balances have a direct-reading dial instead of a rider beam, with the metric scale on the upper scale and the apothecary scale on the lower.

The prescription balance may be placed upon a base containing a drawer that can be used for holding weights or powder papers.

ELECTRONIC BALANCES

Electronic balances are single-pan balances with digital or direct-reading features (Figure 9-4). Taring a weighing paper,



Figure 9-4. Electronic single-pan balance. (Courtesy of OHAUS Corporation.)

weigh boat, or beakers is done automatically with the push of a button or lever without the need of external balancing weights. These balances are much more sensitive than the traditional prescription balance, are easier and quicker to use, but are usually more expensive than a torsion balance.

Prescription Balances

The most common type of prescription balance uses the taut-wire frame or torsion principle. Such balances, manufactured to meet the requirements of the NIST Class III balances, have a maximum maintenance sensitivity of 6 mg with no load and with full load (i.e., addition of the 6 mg weight to one pan causes the indicator or the rest point to be shifted not less than one division on the index plate). The Class III balance is used to weigh quantities up to 60 g, depending on the stated capacity and subject to the physical limit of the amount of the material that can be placed on the pan. Electronic balances typically have a sensitivity less than 10 mg (easily meeting standards for a Class III balance) and can weigh small quantities of drug more accurately than a torsion balance. All prescription departments must have a Class III balance.

REQUIREMENTS

A prescription balance should meet the following general requirements:

- It should be constructed so as to support its full capacity without developing undue stresses, and should not be thrown out of adjustment by repeated weighings of the capacity load. (The capacity of the balance will be seen on the metal plate attached to it.) If the capacity is not stated, it is assumed to be at least 15 g (½ oz). The Class III balances usually have a capacity of 60 g (2 oz).
- The removable pans of a torsion prescription balance should be of equal weight. If the pans show any difference in weight, they should be adjusted by leveling the balance or using small pieces of paper. Pans with any appreciable corrosion or wear should be refinished or replaced.
- A prescription balance should have a leveling device, usually leveling feet or screws, so that the balance can be adjusted to a level position. A balance that does not have these is not entitled to be designated as a prescription balance.
- The balance that has a rider or graduated dial should have, at the end of the graduation, a stop that halts the rider or dial at the zero reading. The reading edge of the rider should be parallel to the graduations on the beam.
- The indicator points, when there are two on the balance, should be sharp, and their ends should not be separated by more than 1 mm (0.04 inch) when the scale is in balance. The distance from the face of the index plate to the indicator pointer or pointers should be small (1 mm or less) to protect the operator against making errors resulting from parallax, because it is unlikely that the eye of the operator will be exactly in line with the indicator and the division on the index plate. The indicating elements as well as the lever system of the balance should be protected against drafts. The balance should have a lid that allows a weighing to be made when the lid is closed.
- A torsion prescription balance must have a mechanical means for arresting the oscillation of the mechanism.

TESTING

Certain tests may be used to satisfy the user regarding the construction and character of a torsion balance when its origin, history, or condition is in doubt. Additional tests are carried

out by the NIST, manufacturers, and local and state testing agencies.

A Class III torsion prescription balance meets the following basic tests. Use a set of *test weights* and keep the rider or graduated dial at zero unless directed to change its position.

- Sensitivity requirement: Level the balance, determine the rest point, place a 10 mg weight on one of the empty pans, and again determine the rest point. Repeat the operation with a 10 mg weight in the center of each pan. The rest point is shifted not less than one division of the index plate each time the 10 mg weight is added. The sensitivity requirement for an electronic balance is supplied by the manufacturer.
- Arm ratio test: This test is designed to check the equality of length of both arms of the balance. Determine the rest point of the balance with no weight on the pans. Place 30 g of test weights in the center of each pan and determine the rest point. If the second rest point is not the same as the first, place a 20 mg weight on the lighter side; the rest point should move back to the original place on the index plate scale or farther.
- Shift tests: These tests are designed to check the arm and lever components of the balance.
 - a. Determine the rest point of the indicator without any weights on the pans.
 - b. Place one of the 10 g weights in the center of the left pan, and place the other 10 g weight successively toward the right, left, front, and back side of the right pan, noting the rest point in each case. If in any case the rest point differs from the rest point determined in (a), add the 10 mg weight to the lighter side; this should cause the rest point to shift back to the rest point determined in (a) or farther.
 - c. Place a 10 g weight in the center of the right pan, and place a 10 g weight successively toward the right, left, front, and back sides of the left pan, noting the rest point in each case. If in any case the rest point is different from that obtained with no weights on the pans, this difference should be overcome by addition of the 10 mg weight to the lighter side.

A balance that does not measure up to these tests *must* be corrected.

- Rider- and graduated-dial tests: Determine the rest point for the balance with no weight on the pans. Now place on the left pan the 500 mg test weight and move the rider to the 500 mg point on the beam. Now determine the rest point. If it is different from the zero rest point, add a 10 mg weight to the lighter side. This should bring the rest point back to its original position or farther. Repeat this test, using the 1 g test weight and moving the rider or graduated dial to the 1 g division. If the rest point is different it should be brought back at least to the zero rest point position by the addition of 10 mg to the lighter pan. If the balance does not meet this test, the graduated beam or the rider must be corrected. For balances equipped with a dial scale, the dial must be corrected.

PROTECTION

The necessity for protecting the delicate mechanism of a balance is overlooked frequently, notwithstanding the possibility of having a precision apparatus irretrievably ruined by lack of care in using or cleaning it or in protecting it while at rest. The position chosen for the balance or scales should be on a level and firm counter, desk, or table, where it will be subjected to little risk of damage from dampness, dust, or corrosive vapors and where the knife-edges will not be liable to become dulled by jarring or other vibrations.

In the analytical class of balances, protection is afforded by enclosing them in glass cases having sash doors in the front, sides, or back. They are protected against damage from vibration by a lever for elevating or locking the beam, so that the knife-edges are not in contact with any surface when not in use. To prevent damage from jarring while the balance is in use, from a weight falling on the pan, or other accident, the finest balances are provided with pan supports, which break the fall and serve the additional purpose of quickly arresting the beam, thus saving time while weighing.

In using a prescription balance, neither the weights nor the substance that is to be weighed should be placed on the balance pans while the beam is free to oscillate. The desired weight should be placed upon one pan (usually the one on the right-hand side) and an amount of the substance to be weighed, approximately the desired weight, upon the opposite pan. The beam should be released by means of the lever, and if the substance is in excess, the beam should be locked and a small portion removed and the beam again released and the oscillations observed. This procedure should be repeated until the correct amount is obtained. In case of a deficiency of the substance to be weighed, the reverse procedure is followed until the correct amount is obtained. With practice this can be done very deftly and very quickly and the sensitivity of the balance retained for years.

Substances that react with metals, such as iodine, and those that are adhesive, such as the extracts, should not be weighed directly upon the pans, but rather upon counterpoised watch crystals, or upon glazed paper, care being taken to balance the papers before weighing the substance. In cleaning the balances, great care should be exercised; polishing powders should be used sparingly, as a portion is very apt to find its way into crevices and elude detection until an attempt is made to adjust the balances, when the increased weight of one of the sides of the beam leads to its discovery. Frequent cleaning with soft leather generally is sufficient to keep a balance in good order, but once neglect makes it necessary to use more active measures, some simple polishing powder for the metal work, soap-suds for the nickel plate, and simple brushing for the lacquered brass are all that is necessary.

As the pans are subjected to more wear and tear than any other part of the balance, it is economical to use *solid* rather than *plated* pans because constant friction wears off the plating and the additional cost for replating soon absorbs the difference in price. Equipped in this way, and with agate bearings, a prescription balance is durable and really inexpensive because it will remain fully equal to the most exacting demands for a long time.

Weights Used in Pharmacy

The weights used by the pharmacist are very important, and care in their selection and examination is necessary. False economy must be avoided, as the use of cheap, inaccurate weights ultimately leads to serious consequences. Official inspectors have found pharmacies using prescription weights that were so worn that the characters on their faces had disappeared; also, weights have been found with bits of hardened extract and dirt almost entirely obscuring their characters. An unused set of standard weights should be kept on hand so that at least once a year the weights in daily use can be tested and adjusted or rejected if necessary. The standard weights should be used also when the balance is tested. The set should contain the following weights in a well-fitted box with forceps: one 50 g, two 20 g, one 10 g, one 5 g, two 2 g, one 1 g, one 500 mg, two 200 mg, one 100 mg, one 50 mg, two 20 mg, and one 10 mg, all adjusted to NIST tolerances for analytical or Class P weights.

METRIC WEIGHTS

For weighing larger quantities, janned iron metric weights are available. They are preferably hexagonal, to distinguish them from the round avoirdupois weights. Sets of brass weights,

usually in the range of 10 g to 1000 g, fitted into holes of appropriate size in a block of plastic (*block weights*), are especially convenient for many weighing operations. For prescription compounding, accurate sets of weights ranging from 10 mg to 50 g are available.

For analytical purposes, metric weights are used exclusively; usually, the highest weight is 100 g, the lowest 1 mg. The weights from 1 g upward are of finely lacquered brass or of non-magnetic stainless steel or rhodium-plated bronze. The smaller weights are made of squares of platinum or aluminum foil, with one edge turned up to permit them to be handled easily with the forceps. Fractions of a milligram are weighed by means of the rider on the graduated beam of the balance.

In analytical work and in using the Class III balance in prescription work, the weights should never be handled with the fingers but always with the forceps which accompany an accurate set of weights. In the more expensive sets of weights the forceps are tipped with bone, ivory, or plastic to prevent the wearing away of the weights during handling. With proper care the accuracy of a fine set of weights may be maintained for years.

COMMON AVOIRDUPOIS WEIGHTS

Avoirdupois weights are usually made of iron, and are flat and circular and japanned to prevent rusting. These weights form a pyramidal pile, and range from ½ oz to 4 lb; if found to be incorrect, they may be adjusted by adding to or diminishing the amount of lead that is hammered into a depression in the base of each weight. They are sometimes made of brass in this form, and sometimes of zinc (the latter, however, are brittle and unserviceable). For general use in the pharmacy, the cylindrical weights, known technically as block weights, are preferable. The advantages of block weights are that the gaps left by missing weights are readily noticeable, and the greater part of the surface of the weight is protected from the action of corrosive vapors when the weights are not in use.

APOTHECARY WEIGHTS

Apothecary weights may be obtained either as *block weights* or in the less-desirable *flat* forms. The round, flat, brass *dram* weights, which have the denomination stamped on their faces in raised characters, are still used but should be replaced. With flat weights, the denomination is often only faintly stamped on the face and thus is liable to be obliterated by constant use or by corrosive contact.

Undoubtedly, the best grain weights are the aluminum wire weights. The wire weights are less susceptible to corrosive action than are the brass weights. Also, the wire weights are more easily and quickly distinguished from one another than are other weight forms, so there is less likelihood of dangerous mistakes: the number of sides in the wire weights at once gives the denomination.

Aluminum grain weights, which are cut out of aluminum plates, are also less liable to be corroded. They usually can be more accurately adjusted than brass weights. The corners of the aluminum weights are clipped, and each weight is usually pressed into a curved form so that it may be picked up easily (Figure 9-5).

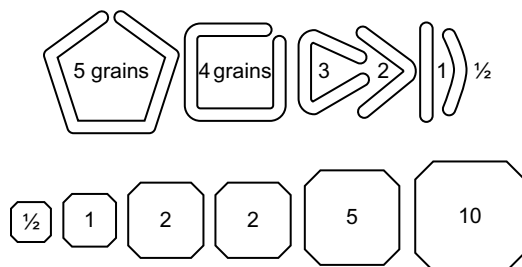


Figure 9-5. Aluminum wire and aluminum grain weights.

The need for apothecary weights in modern practice is decreasing. Apothecary weights can easily be converted to the corresponding metric weights, which are easier to use and less prone to error. In addition, electronic balances do not use external weights. Typically, an electronic balance can display weights in several systems at the discretion of the operator.

Minimum Weighable Quantity

All of the balances described must be used within a degree of error that can be tolerated in prescription compounding and in pharmaceutical manufacturing. The USP allows a maximum error of 5 percent in a single weighing operation. Since the sensitivity requirement of a balance represents the absolute error in using that balance, the percent error will depend on the amount of drug weighed and will increase as the amount of drug decreases.

The minimum weighable quantity (MWQ) with no more than 5 percent error can be calculated for any balance knowing the sensitivity requirement (SR) (i.e., the absolute error) from the following:

$$\text{MWQ} = \text{SR} \frac{100\%}{5\%}$$

EXAMPLES

Calculate the MWQ with no more than a 5 percent error for a balance with a sensitivity requirement of 10 mg.

$$\text{MWQ} = 10 \text{ mg} \frac{100\%}{5\%} = 200 \text{ mg}$$

Calculate the MWQ with no more than a 3 percent error for a balance with a sensitivity requirement of 2.5 mg.

$$\text{MWQ} = 2.5 \text{ mg} \frac{100\%}{3\%} = 83.3 \text{ mg}$$

MEASURING

In pharmacy, *measuring* usually refers to the exact determination of a definite volume of liquid. Many types of apparatus are used in this operation, depending on the kind and quantity of liquid to be measured and the degree of accuracy required. (The NIST has requirements for graduates.¹)

Large Quantities

Glass measures are preferred for measuring liquids. Although glass measures are subject to breakage, they can indicate volume more accurately because of the transparency of glass.

THE MENISCUS

When an aqueous or alcoholic liquid is poured into a graduate, surface forces cause its surface to become concave—the portion in contact with the vessel is drawn upward. This phenomenon is known as the formation of a *meniscus* (Figure 9-6), and in determining the volume of a liquid the *reading must be made at the bottom of this meniscus*. This regulation has been established by the NIST, and all glass measuring vessels are graduated on this basis. Liquids with large contact angles, such as mercury, form an *inverted meniscus*, and the reading then is made at the top of the curved surface.

PROCEDURE

Pharmaceutical manufacturers package liquid preparations in glass or plastic containers equipped with a plastic screw-cap. These containers serve as a stock bottle from which liquids may be poured directly into a graduate. The procedure for pouring liquid from screw-capped containers is as follows:

1. Remove the cap and place it on the counter while the transfer of liquid is made.

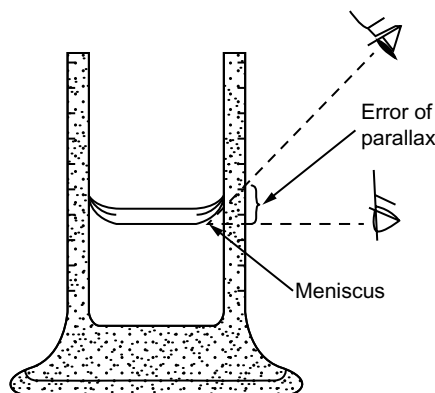


Figure 9-6. Error of measurement due to parallax.

2. While holding the graduate in the left hand, grasp the original container with the label in such a position that any excess of liquid will not soil the label if it should run down the side of the bottle.
3. Raise the graduate and hold it so that the graduation point to be read is on a level with the eye, and measure the liquid. (The extension of the graduating mark into a circle that passes entirely around the graduate is an improvement that obviates the necessity of placing the graduate upon a level place, as the corresponding mark upon the opposite side may be seen through the glass and the graduate easily leveled even when held in the hand.)
4. Replace the cap, and return the bottle to the counter or shelf.
5. Pour the liquid into the bottle or mortar for dispensing or compounding.

METALLIC MEASURES

Metallic measures are nearly cylindrical in shape, but are slightly wider at the bottom. These are generally used for measuring liquids when the quantity is over a pint. A set usually consists of five (gallon, half-gallon, quart, pint, and half-pint) of these measures. Measures made of tinned iron, or of the enameled sheet iron called agateware, are greatly inferior to those made of *tinned copper* or *stainless steel*; tinned-iron measures soon become rusty, and particles of enameling can chip off, leaving the exposed iron to contaminate the measured liquids.

The initial cost of copper or stainless-steel measures is greater than tinned iron, but they are far more durable. Care must be taken to protect them from blows that will cause dents as these may be serious enough to detract from their accuracy. Cylindrical metric measures, usually made of monel metal or stainless steel and having a diameter just half their height, are available in various sizes. Such containers are relatively expensive, but their resistance to corrosion and wear is a tremendous advantage. Copper, of course, should not be used where it is likely to catalyze oxidation.

GRADUATED GLASS MEASURES

Graduated glass measures are nearly always used for quantities of 500 mL or 1 pt or less. There are of two forms, *conical* and *cylindrical* (Figures 9-7 and 9-8). The conical graduate is suitable for some measurements because of the greater ease with which it can be handled, but cylindrical measures are more accurate because of their uniform and smaller average diameter. In a graduated cylinder, the error in volume caused by a deviation of ± 1 mm in reading the meniscus remains constant along the height of the uniform column; the same deviation causes a progressively larger error in a conical graduate because the diameter, and thus the volume of the 1 mm column, increases along its vertical axis. It is safe to assume that practically all good-grade modern graduates comply with the NIST requirements for internal diameters at stated volumes.



Figure 9-7. Class conical graduate. (Courtesy of Kimble Glass.)

A study has indicated that, to improve accuracy, the lower portions of graduates should not be used, and therefore should not be marked.² A composite tabulation (Table 9-12) shows the calculated and the assigned blank portions of graduates. The elimination of the lower markings on graduates was suggested, and in 1955 the NIST specifications for graduates used this principle.¹ The NIST *Handbook* states, "A graduate shall have an initial interval that is not subdivided, equal to not less than one-fifth and not more than one-fourth of the capacity of the graduate." For accurate measurement of volumes less than 1.5 mL, a graduated pipet or a graduated dropper could be used.

EFFECT OF LIQUID AND CONTAINER

It is difficult to measure accurately when pouring from a completely filled bottle because of the uneven flow of the liquid. After the first portion of the liquid is removed, the shape of the bottle does not influence the ease of pouring to any appreciable extent unless the neck is extremely narrow.

Viscous liquids pour slowly, but their accurate measurement is not difficult. Experiments showed that when glycerin is poured into a graduate without letting the liquid run down the inside surface, the precision of measurement can be very high. Naturally, the chance of hitting the inner surface is greater with smaller than with larger graduates. The increase in possible deviation then is caused by the slow movement of the viscous liquid to the desired mark.



Figure 9-8. Glass cylindrical graduate. (Courtesy of Kimble Glass.)

Table 9-12. Unmarked (Unreliable) Portions of Graduates

Capacity of graduate (mL)	Calculated blanks (1951)		NIST blanks (mL)
	2.5% ^a allowed (mL)	5% ^a allowed (mL)	
5	3.0	1.5	1
10	4.4	2.2	2
25	11.8	5.9	5
50	15.8	7.9	10
100	20.9	10.5	20
250	36.3	18.2	50
500	66.5	33.2	100
1000	—	—	200

^a Calculations by Goldstein and Mattocks² based on deviation of ± 1 mm from graduation mark and allowable errors of 2.5% and 5%.

Viscous liquids introduce another factor: drainage time. Graduates are calibrated to contain or deliver indicated volumes within specified limits. Aqueous, alcoholic, and hydroalcoholic liquids can be drained from a graduate in 30 seconds so completely that the delivered and contained volumes are fairly close. When 25 mL of glycerin was measured in the same cleaned and dried cylinders, the received volume measured 23.7 mL after the same time period. Silicone-treated glassware, which now is used frequently, drains completely in a few seconds.

The viscosity factor might be altered when another liquid is to be mixed with the glycerin by measuring and mixing both liquids in a suitable graduate.

Small Quantities

For measuring smaller quantities of liquids, graduated glass tubes of small diameter should be used. The narrower bore permits greater distances between the graduations on the apparatus, thus allowing greater accuracy in making the reading. For example, with a buret the pharmaceutical chemist can estimate volumes to the nearest 1/100 mL.

Pipets and similar apparatus are more accurate and convenient than very small graduates. The graduations on very small graduates are necessarily in the very small, lowest portion of a comparatively tall measure. To measure 1 mL of a volatile oil in a graduate, the surface that the oil must traverse when this measure is inverted is so great that probably 20 percent of the oil will be left adhering to the measure. In liquid preparations in which the smaller liquid is miscible with the larger quantity of diluting liquid, the graduate may be rinsed and this loss recovered, but inconveniences are largely overcome and greater accuracy secured by using a pipet.

In administering small quantities of liquids, the very convenient "drop" is almost always used. It should be emphasized that *one drop is not equivalent to 1 <minim>* and that *60 drops are not equivalent to 1 f℥*. This impression doubtless arose because 60 ordinary drops of water are about equal to 1 f℥, but the volume of a drop of fluid depends on many factors, including density, temperature, viscosity, surface tension, and the size and nature of the orifice from which it is dropped. Thick, viscous liquids, such as the mucilages and the syrups, necessarily produce large drops because the drop adheres to the surface of the glass as long as its weight does not overcome its power of adhesion, whereas chloroform, a mobile liquid that has very little adhesion to the dropping surface, produces very small drops. The greater the surface tension, the larger the drop, and the greater the extent of surface to which the drop adheres, the larger, proportionally, the drop.

A *normal or standard drop measure* was recommended by the Brussels Conference of 1902 for international adoption. This dropper is recognized in the USP.

MEDICINE DROPPER³

The pharmacopeial medicine dropper consists of a tube made of glass or other suitable transparent material that generally is fitted with a collapsible bulb and, while varying in capacity, is constricted at the delivery end to a round opening having an external diameter of about 3 mm. The dropper, when held vertically, delivers water in drops, each of which weighs between 45 mg and 55 mg.

When drops are specified on a prescription, the usual custom has been to employ an *eyedropper*, but now the standard dropper should be supplied. When accuracy is required, it is particularly important to use a specially calibrated dropper for administering potent medicines. The volume error incurred in measuring any liquid by means of a calibrated dropper should not exceed 15 percent under normal conditions.³

TEASPOON

For household purposes, an American Standard Teaspoon has been established by the American National Standards Institute as containing 4.93 ± 0.24 mL. In view of the almost universal practice of employing teaspoons ordinarily available in the household for the administration of medicine, the teaspoon may be regarded as representing 5 mL and is so accepted by the USP.

It must be kept in mind that the actual volume delivered by a teaspoon of any given liquid is related to the latter's viscosity and surface tension, among other influencing factors.

The Human Factor

The *human factor of carefulness* is of paramount importance in every pharmaceutical operation in which accuracy is essential. Accurate measurement of liquids requires accurate equipment, careful manipulation, good vision, and a steady hand.

DENSITY AND SPECIFIC GRAVITY

Several terms are used to express the mass (weight) of equal volumes of different substances:

- *Absolute density* is the ratio of the mass of an object, determined in or referred to a vacuum, at a specified temperature, to the volume of the object at the same temperature. This relationship is expressed mathematically as:

$$\text{Absolute density} = \frac{\text{Mass in grams (in a vacuum)}}{\text{Volume in millimeters}}$$

- *Apparent density* differs from absolute density only in that the mass of the object is determined in air; the mass is influenced by the difference in the buoyant effect of air on the object being weighed, and on the standard masses (weights) used for comparison. If the object and masses are made of the same material, or have the same density, there will be no difference in the buoyant effect, and the apparent density will be identical with the absolute density.
- *Relative density* is an expression sometimes employed to indicate the mass of 1 mL (not cc, which is very slightly different) of a standard substance, such as water, at a specified temperature, relative to water at 4°C taken as unity. Thus, at 4°C the relative density of water is 1.0000, whereas its absolute density at the same temperature is 0.999973. Water attains its maximum absolute density of 0.999973 at 3.98°C. To convert a relative density of water to absolute density, the former should be multiplied by 0.999973.
- *Specific gravity* may be defined as the ratio of the mass of a substance to the mass of an equal volume of another substance taken as the standard. For gases, the standard may be hydrogen or air; for liquids and solids, it is water.

From what has been stated, it is obvious that in a determination of specific gravity there will be, in general, a difference in the result if the masses (weights) are determined in air or in vacuum. If the masses are determined in, or referred to, a vacuum, the result is a *true specific gravity* (sometimes called *absolute specific gravity*); if the masses are determined in air, the calculated result is an *apparent specific gravity*. The difference between these specific gravities is, as a rule, very small.

A very important variable in specific gravity determinations is temperature, and this is doubly important because both the temperature of the substance under examination and the temperature of the standard may be different. The temperatures are commonly shown as a ratio, with the temperature of the water always being indicated in the denominator. The common practice with regard to the determination of specific gravity is that defined by the USP: "Unless otherwise stated, the specific gravity basis is 25°/25°, i.e., the ratio of the weight of a substance in air at 25° to that of an equal volume of water at the same temperature."

But it is not always convenient, or desirable, to determine the weight of both the substance and the water at 25°C, or even to determine the weight of the substance at the same temperature as that at which the water is weighed. Thus, the substance may be weighed at 25°C and compared with the weight of an equal volume of water at 4°C, in which case the specific gravity is reported as being on a 25°C/4°C basis. In the case of theobroma oil, which is solid at 25°C, the specific gravity is determined on a 100°C/25°C basis; for alcohol, it is determined on a 15.56°C/15.56°C basis because many years ago the US government adopted 60°F (15.56°C) as the temperature at which alcoholometric measurements are to be made for government control of alcoholic liquids.

It is apparent that a completely informative statement of specific gravity must indicate the temperature of the substance under examination, as well as that of the equal volume of water. Furthermore, it should be stated whether the determinations of mass (weight) were made on an in vacuum or in air basis; in the latter case, the material of construction of the weights also should be indicated (as the buoyant effect of air on weights depends on their volume).

CALCULATIONS

The principle underlying the determination of the specific gravity of either a liquid or a solid is the same: to find the ratio of the mass (weight) of the substance to that of an equal volume of water. This may be expressed by a simple relationship:

$$\text{Specific gravity} = \frac{W_s}{W_w}$$

where W_s is the weight of the substance, and W_w the weight of an equal volume of water.

DENSITY

Density is defined as the mass of a substance per unit volume. It has the units of mass over volume. *Specific gravity* is the ratio of the weight of a substance in air to that of an equal volume of water. In the metric system both density and specific gravity may be numerically equal, although the density figure has units. In the English system, density and specific gravity are not numerically equal; for example, the density of water is 62.4 lb/ft³ and the specific gravity is 1. This shows the convenience of the metric system. The equations for calculating density, weight, and volume are:

$$\text{Density} = \frac{\text{Weight}}{\text{Volume}}$$

$$\text{Weight} = \text{Density} \times \text{Volume}$$

$$\text{Volume} = \frac{\text{Weight}}{\text{Density}}$$

Given any two variables, the third one can be calculated.

Examples

1. A pharmacist weighs out 2 kg of glycerin (density 1.25 g/mL). What is the volume of the glycerin?

$$\text{Volume} = \frac{2000 \text{ g}}{1.25 \text{ g/mL}} = 1600 \text{ mL}$$

2. What is the weight of 60 mL of oil whose density is 0.9624 g/mL?

$$\text{Weight} = 60 \text{ mL} \times 0.9624 \text{ g/mL} = 57.7 \text{ g}$$

3. Calculate the weight of 30 mL of sulfuric acid (density 1.8 g/mL).

$$\text{Weight} = 30 \text{ mL} \times 1.8 \text{ g/mL} = 54 \text{ g}$$

4. If a prescription order requires 25 g of concentrated hydrochloric acid (density 1.18 g/mL), what volume should the pharmacist measure?

$$\text{Volume} = \frac{25 \text{ g}}{1.18 \text{ g/mL}} = 21.2 \text{ mL}$$

PROBLEMS

(Please see answers at the end of the chapter)

1. What is the weight in grams of 1 L of alcohol (density 0.816 g/mL)?
2. What is the volume (mL) of 1 lb (avoir) of glycerin (density 1.25 g/mL)?
3. What is the volume (mL) of 65 g of an acid whose density is 1.2 g/mL?

PHARMACEUTICAL CALCULATIONS

Pharmaceutical dispensing and compounding calculations use simple arithmetic. The errors that may arise often are due to carelessness, as in improper placing of decimal points, incorrect conversion from one system of measurement to another, or uncertainty over the system of measurement to be used. Before proceeding with any calculation, it is imperative that the problem presented (in a prescription, chart order, formula, etc.) be read carefully, that the information given and required be identified, and that the procedure to be used in the calculation be selected.

Before students read this part of the chapter and attempt to solve the problems, the information in the preceding part of this chapter must be understood thoroughly. Often, several steps are necessary to solve problems. Shortcuts should not be taken unless one is certain they are proper. Many problems can be solved by more than one procedure, such as by ratio and proportion or by dimensional analysis. If students find a procedure that is more logical to them and gives the correct answer, it should be used. Thus, the solutions to sample problems used here generally should be considered suggestions, rather than the only way to solve a given type of problem.

MATHEMATICAL PRINCIPLES

A few mathematical principles (e.g., common decimal fractions, exponents, powers and roots, significant figures, and logarithms) will be reviewed, as these are areas where students often become careless or have forgotten skills. Following this, various types of practical pharmaceutical problems that the pharmacist may be required to solve are discussed and solutions are given. Where practical, rules for solving these problems are given. No attempt is made to elaborate on any mathematical theory.

The problems generally consist of determining the quantity or quantities of material(s) required to compound prescriptions properly and make products used to aid the compounding of prescriptions. The materials used to compound prescription orders may be pure or mixtures of substances in varying strengths. The strengths of mixtures may be denoted in different ways. Conversions may be necessary between systems of varying strengths or between different measuring systems. At the end of each section, sample problems are given for the student to solve, the answers to which appear at the end of the chapter.

Because of the decreasing importance of the apothecary system, the metric system is emphasized here. Chemicals and preparations most likely will be purchased using the avoirdupois or metric systems. Prescription orders are filled in the system indicated on the order, usually the apothecary or metric systems.

The student should become familiar with the terminology used in writing prescription orders, such as Latin words and abbreviations used in giving directions to the pharmacist and patient. The prescriber occasionally may use Roman numerals instead of Arabic numerals, so students must be familiar with these (even if the practice is declining).

Significant Figures

Weighing and measuring can be carried out with only a certain maximum degree of accuracy; the result always is approximate due to the many sources of error such as temperature,

limitations of the instruments employed, personal factors, and so on. Pharmacists must achieve the greatest accuracy possible with their equipment, but it would be erroneous to claim that they have weighed 1 mg of a solid on a Class III prescription balance, which has a sensitivity requirement of 10 mg, or that they have measured 76.32 mL of a liquid in a 100 mL graduate, which can be read only to 1 mL. When quantities are written, the numbers should contain only those digits that are *significant* within the precision of the instrument.

Significant figures are digits that have practical meaning. In some instances zeros are significant; in other instances they merely indicate the order of magnitude of the other digits by locating the decimal point. For example, in the measurement 473 mL all the digits are significant, but in the measurement 4730 mL the zero may or may not be significant. In the weight 0.0316 g the zeros are not significant but only locate the decimal point. In any result the last significant figure is only approximate, but all preceding figures are accurate.

When 473 mL is recorded, it is understood that the measurement had been made within ± 0.5 mL or somewhere between 472.5 and 473.5 mL. The student should stop to consider the full implications of this specifically that the measurement is subject to a maximum error of:

$$\frac{0.5}{473} \times 100 = (\text{approx}) 0.1\% \text{ or } 1 \text{ part in } 1000$$

A zero in a quantity such as 473.0 mL is a significant figure and implies that the measurement has been made within the limits 472.95 mL and 473.05 mL or with a possible error of:

$$\frac{0.05}{473} \times 100 = (\text{approx}) 0.01\% \text{ or } 1 \text{ part in } 10,000$$

Thus, 473 is correct to the nearest milliliter, and 473.0 is correct to the nearest 0.1 mL.

Rules

1. When adding or subtracting, retain in the sum or remainder no more decimal places than the least number entering into the calculations. For example,

11.5 g	11.50 g
2.65 g	2.65 g
3.49 g	3.49 g
17.64 g	17.64 g

Answer : 17.6 g Answer : 17.64 g

In the first column 11.5 g was weighed to 0.1 g or with an accuracy of ± 0.05 g. Although the other two weighings were made with an accuracy of ± 0.005 g, the sum can be expressed properly only to one decimal place.

In the second column 11.50 g was weighed to the nearest 0.01 g or with an accuracy of ± 0.005 g. Since all weighings were made with this degree of accuracy, the sum may be stated as in the example, 17.64 g.

Retain all figures possible until all the calculations are completed and then retain only the significant figures for the answer. Additions or subtractions involving both large and small quantities, each expressed with maximum significance, are often useless. For example, if one were to add 1.2 and 0.041 g, the physical sum would be 1.2 g, regardless of the fact that the two numbers add numerically to 1.241. To express the physical sum as 1.241 g would convey an erroneous degree of accuracy with which the quantity was known.

2. When multiplying or dividing, retain in the answer no more significant figures than the least number entering

Table 9-13. Different Equivalent Values and the Number of Significant Figures to Which the Answer is Correct

Weight (g)	Equivalent weight (gr/g)	Equivalent weight (gr)	Significant figures
4.522 ×	15.432 =	69.78	4
4.522 ×	15.43 =	69.77	4
4.522 ×	15.4 =	69.6	3
4.522 ×	15 =	68	2

into the calculation. The meaning of this rule may be illustrated by the use of equivalents during conversions from one measuring system to another. Table 9-13 gives different equivalent values and the number of significant figures to which the answer is correct. Always use an equivalent that will give the desired degree of accuracy. Repeated multiplication of an approximation increases the error progressively; therefore, retain all figures during calculations and drop insignificant figures as the final step.

FRACTIONS

COMMON FRACTIONS

An example of a common fraction is $3/8$. It is read as “three-eighths” and indicates three parts divided by eight parts of the same thing. The units with both numbers must be the same. Pharmacists measure $3/8$ of a fluid ounce into a graduate, they measure 3 fluidrams, out of 8 fluidrams (a fluid ounce contains 8 fluidrams).

The following principles should be applied when using common fractions:

1. The value of a fraction is not altered by multiplying or dividing both numerator and denominator by the same number.
2. Multiplying the numerator or dividing the denominator by a number multiplies the fraction by that number.
3. Dividing the numerator or multiplying the denominator by a number divides the fraction by that number.
4. To add or subtract fractions, form fractions with the *lowest common denominator*, perform the arithmetical operation, and reduce to the lowest common denominator.
5. To multiply fractions, multiply all numbers above the line to form the new numerator and multiply all numbers below the line to form the new denominator. Cancel if possible to simplify and reduce to the lowest common denominator.
6. To divide by a fraction, multiply by the reciprocal of the fraction.

DECIMAL FRACTIONS

Fractions with the power of 10 as the denominator are known as *decimal fractions* and are written by omitting the denominator and inserting a decimal point in the numerator as many places from the last number on the right as there are ciphers of 10 in the denominator.

The following principles should be applied when using decimal fractions:

- When adding or subtracting decimals, align the decimal points under each other.
- When multiplying decimals, proceed as with whole numbers, then place the decimal point in the product as many places from the first number on the right as the sum of the decimal places in the multiplier and the multiplicand.

- When dividing by a decimal fraction, move the decimal point to the right, in both divisor and dividend, as many places as it is to the left in the divisor to form a whole number in the divisor; proceed as with whole numbers. The decimal point in the quotient should be placed immediately above the decimal point in the dividend.
- When converting a common fraction into a decimal fraction, divide the numerator by the denominator and place the decimal point in the correct place.
- When converting a decimal fraction into a common fraction, place the entire number, as the numerator, over the power of 10 containing the same number of ciphers of 10 as there are decimal places. Cancel, if possible, to simplify.

EXPONENTS, POWERS, AND ROOTS

In the expression $2^4 = 16$, the following names are given to the terms: 16 is called the *power* of the *base* 2 and 4 is the *exponent* of the power. If the exponent is 1, it usually is omitted. The following laws should be recalled:

- The product of two or more powers of the same base is equal to that base with an exponent equal to the sum of the exponents of the powers; e.g., $2^5 \times 2^3 = 2^8$.
- The quotient of two powers of the same base is equal to that base with an exponent equal to the exponent of the dividend minus the exponent of the divisor; e.g., $2^8 \div 2^3 = 2^5$.
- The power of a power is found by multiplying the exponents; e.g., $(2^8)^3 = 2^{24}$.
- The power of a product equals the product of the powers of the factors; e.g., $(2 \times 3 \times 4)^2 = 2^2 \times 3^2 \times 4^2$.
- The power of a fraction equals the power of the numerator divided by the power of the denominator; e.g.,

$$\left(\frac{2}{3}\right)^2 = \frac{2^2}{3^2}$$

- The root of a power is found by dividing the exponent of the power by the index of the root; e.g.,

$$\sqrt[3]{3^6} = 3^{\frac{6}{3}} = 3^2$$

- Any number other than 0 with an exponent 0 equals 1; e.g., $2^0 = 1$. A number with a negative exponent equals 1 divided by the number with a positive exponent equal in numerical value to the negative exponent; e.g.,

$$2^{-4} = \frac{1}{2^4}$$

LOGARITHMS

Logarithms (logs) were invented to facilitate the solution of involved and lengthy problems. Many calculations that are difficult by ordinary arithmetical processes are performed rapidly and easily with the aid of logs; the advent of modern calculators and computer spreadsheet programs has made this use of logs obsolete. Logs still appear, however, in many chemical and pharmacokinetic equations.

The log of a number is the exponent of the power to which a given base must be raised in order to equal that number.

$$Y = a^x$$

$$\log_a Y = x$$

John Napier, of Scotland, who discovered logs over three centuries ago, used the natural log number, 2.71828+, as the base. Henry Briggs, using Napier's discovery a few years later, introduced 10 as the base, which is the most convenient for practical purposes. Napier's system is called *natural logs*, and Briggs' system is called *common logs*. In this latter system the natural numbers are regarded as powers of the base 10 and the corresponding exponents are the logs; e.g.,

$$6 = 10^{0.7782}$$

$$\log_{10} 6 = 0.7782$$

For natural logs,

$$6 = e^{1.792}$$

$$\ln_e 6 = 1.792$$

LAWS AND RULES

The following laws, governing the use of logs, are based on the laws of exponents, and hence hold for any log system.

- The log of a product equals the *sum* of the log of the component numbers; for example, for 25×2 :

$$\log(25 \times 2) = \log 25 + \log 2 = 1.3979 + 0.3010 = 1.6989$$

- The log of a quotient equals the log of the numerator minus the log of the denominator; for example, for $25 \div 2$:

$$\begin{aligned}\log(25/2) &= \log 25 - \log 2 = \log 10^{1.3979} - \log 10^{0.3010} \\ &= 1.3979 - 0.3010 = 1.0969\end{aligned}$$

- The log of a power of a number equals the log of the number multiplied by the exponent of the power; for example, for $(25)^{12}$:

$$\log(25)^{12} = 12 \log 25 = 12 \times 1.3979 = 16.7748$$

- The log of a root of a number equals the log of the number divided by the index of the root; for example, for $\sqrt{25}$:

$$\log \sqrt{25} = \log 25^{1/2} = \frac{\log 25}{2} = \frac{1.3979}{2} = 0.6990$$

- The log of a negative power of a number equals the reciprocal of the number multiplied by the exponent of the power; for example, $(5)^{-2}$:

$$\log(5)^{-2} = -2 \log 5 = -2 \times 0.6990 = -1.398$$

The Log of a Number

The logarithm of a number can be easily obtained from a calculator or computer spreadsheet program.

- Find the logarithm of 273:

$$\begin{aligned}\log 273 &= 2.4362 \\ \ln 273 &= 5.6095\end{aligned}$$

- Find the logarithm of 0.08206:

$$\begin{aligned}\log 0.08206 &= -1.08587 \\ \ln 0.08206 &= -2.5003\end{aligned}$$

The Antilog of a Number

To find the number corresponding to a given log (or antilog), the reverse procedure of that discussed above is employed (i.e., the appropriate numerical base is raised to the exponent expressed by the logarithm).

- Find the number corresponding to the antilog 3.8357:

$$\begin{aligned}\log X &= 3.8357 \\ X &= 10^{3.8357} = 6850\end{aligned}$$

- Find the number corresponding to the natural log 0.4351:

$$\begin{aligned}\ln X &= 0.4351 \\ X &= e^{0.4351} = 2.71828^{0.4351} = 1.5451\end{aligned}$$

- Using the Henderson–Hasselbalch equation for an acidic substance, find the ratio of ionized to un-ionized drug at a pH of 3.0. The pK_a of the drug is 7.4.

PHARMACEUTICAL PROBLEMS

The student who knows algebra, has studied the previous sections of this chapter, and recognizes the Roman numerals and Latin abbreviations used on prescription orders (for directions to the pharmacist and patient by the prescriber) should have sufficient knowledge to solve the routine problems encountered in a pharmacy. The various symbols and abbreviations and their meanings must be well understood. Explanations of some practical problems, representative of those faced in practice, are presented below. Practice problems follow each section, and the answers to these problems are found at the end of this chapter. To solve each problem properly, the following procedure is suggested:

- Analyze the problem carefully so that all data are clearly fixed in the mind; determine what is given and what is asked.
- Select the most direct method of solving the problem. Not all problems can be solved properly in one step. Look up doses, equivalents, and abbreviations when you are not sure.
- Prove or check the result. Many problems encountered in pharmacy still utilize the apothecary and avoirdupois systems; however, solving these problems in contemporary practice is based on converting these systems into metric units prior to solving the problem mathematically. This approach will be followed in this text. Methods for the mathematical manipulation of apothecary and avoirdupois units and direct problem solving in these systems can be found in previous editions of *Remington*.

ADDITION

Review weighing and measuring systems discussed earlier in this chapter.

Rules

- Add like quantities. Using the metric system, if the quantities are not alike, change them to a common unit.
- When adding decimals, keep the decimal points directly under each other.
- When adding fractions, reduce to the lowest common denominator (LCD), add the resulting numerators, and reduce the fraction, if possible, by canceling.

Examples

- Add 3 kg, 33 g, and 433 mg.
Convert to a common unit. The gram is convenient because it is the unit of weight.

$$\begin{aligned}3 \text{ kg} \times \frac{1000 \text{ g}}{\text{kg}} &= 3000 \text{ g} \\ 33 \text{ g} &= 33 \text{ g} \\ 433 \text{ mg} \times \frac{1 \text{ g}}{1000 \text{ mg}} &= 0.433 \text{ g}\end{aligned}$$

Answer: 3033.433 g

PROBLEMS

1. Add 25 mg, 25 g, 210 mg, 2 kg, 1.75 g, 215 mg, 454 g, and 30 mg.
2. The following quantities of a drug were removed from a container: 31 g, 225 g, 855.6 g, and 45.4 g. What is the total weight removed from the container?

SUBTRACTION**RULES**

1. Subtract only like quantities. If the quantities are not alike, change to a common unit.
2. Treat common and decimal fractions as indicated in the section on addition.

EXAMPLES

1. Subtract 285 mL from 1 L. Convert to a common unit.

$$\begin{array}{r} 1000 \text{ mL} \\ - 285 \text{ mL} \\ \hline 715 \text{ mL} \end{array}$$

Answer: 715 mL

PROBLEMS

1. How much is left in a 5 L container after the removal of 895 mL?
2. A pharmacist buys 5 g of a potent drug and at different times dispenses 0.2 g, 0.85 g, 90 mg, and 150 mg on prescription orders. How much of the drug remains?

MULTIPLICATION**RULES**

1. The product has the same denomination as the multiplicand.
2. If the multiplicand is composed of different denominations in the metric system, form a common unit before multiplying and reduce the product to measurable units.
3. Multiply fractions and decimals as in any arithmetic problem, and reduce fractional quantities to measurable or weighable units.

EXAMPLES

1. What will be the total weight of the ingredients in a prescription order for 25 units, each unit containing 0.4 g of Solid F, 0.01 g of Solid G, and 5 mg of Solid H? First, convert to a common unit such as grams.

$$0.4 \text{ g} + 0.01 \text{ g} + 0.005 \text{ g} = 0.415 \text{ g total weight of one unit}$$

$$0.415 \text{ g/unit} \quad 25 \text{ units} = 10.375 \text{ g total weight of all units}$$

2. Multiply 22.4 mL by 2.65.

$$\begin{array}{r} 22.4 \text{ mL} \\ \times 2.65 \\ \hline 59.36 \text{ mL} \end{array}$$

PROBLEMS

1. Multiply 48.5 mL by 3.24.
2. A certain preparation is to contain 0.0325 g of a chemical in each milliliter of solution. How much must be weighed out to make 5 L of the solution?

3. How much cod liver oil is necessary to make 2500 capsules, each containing 0.33 mL?
4. How many milligrams are used to make 1500 units, each of which contains 250 µg of a drug?

DIVISION**RULES**

1. The quotient always has the same denomination as the dividend.
2. If the dividend is composed of different denominations, form a common unit in the metric system before dividing and reduce the quotient to weighable or measurable quantities.
3. Treat fractions and decimals as explained in the multiplication section.

EXAMPLE

1. Divide 3 L by 25.

$$\frac{3 \text{ L}}{25} = 0.120 \text{ L or } 120 \text{ mL}$$

PROBLEMS

1. How many 65 mg capsules can be made from 50 g of a drug?
2. The dose of a drug is 0.1 mg. How many doses are contained in 15 mg of the drug?
3. How many 325 mg capsules of a drug can be filled from a 454 g amount?

CONVERSION

As long as the student knows the interrelationships of the various units within the different weighing and measuring systems (e.g., 20 gr = 1 ℥, 3 ℥ = 1 lb; 1000 mg = 1 g), there are only three conversions necessary to memorize in order to convert between the apoth, avoird, and metric systems. These are:

$$\begin{aligned} 1 \text{ gr (avoird)} &= 1 \text{ gr (apoth)} \\ 1 \text{ gr} &= 64.8 \text{ mg} \\ 1 \text{ f℥} &= 29.6 \text{ mL} \end{aligned}$$

Learn them!

With these three conversions the student is able to derive all other necessary conversions.

Apothecary Conversions

Various equalities within the apothecary system may be calculated.

1. The number of grains in a dram, grains in a pound, and so on may be calculated using the following steps.

$$\begin{aligned} \frac{20 \text{ gr}}{\cancel{\text{℥}}} \times \frac{3 \cancel{\text{℥}}}{3} &= \frac{60 \text{ gr}}{3} \\ \frac{60 \text{ gr}}{3} \times \frac{8 \cancel{\text{℥}}}{3} \times \frac{12 \cancel{\text{℥}}}{\cancel{\text{lb}}} &= \frac{5760 \text{ gr}}{\cancel{\text{lb}}} \end{aligned}$$

Cancel the units. If they do not cancel properly, something has been omitted.

2. Convert 1 ℥ (apoth) to weighable quantities in the avoird system:

$$1 \text{ gr (apoth)} = 1 \text{ gr (avoird)}$$

Since 1 gr (apoth) = 1 gr (avoir), the number of grains in one system equals the number of grains in the other system; e.g., 480 gr (apoth) = 480 gr (avoir).

$$\begin{array}{r} \frac{20 \text{ gr}}{\text{M}} \times \frac{3 \text{ M}}{3} \times \frac{83}{\text{ʒ}} = \frac{480 \text{ gr}}{\text{ʒ (apoth)}} \\ 480 \text{ gr (apoth)} = 480 \text{ gr (avoir)} \\ 437.5 \text{ gr} = 1 \text{ oz avoir} \\ \frac{480 \text{ gr}}{-437.5 \text{ gr}} \\ \hline 42.5 \text{ gr} \end{array}$$

Answer: 1 ʒ (apoth) = 1 oz, 42.5 gr (avoir)

3. Conversions in the metric system are made in the same manner. Convert 1 g to mg.

$$1 \text{ g} \times \frac{1000 \text{ mg}}{\text{g}} = 1000 \text{ mg}$$

Convert 1 g to kg.

$$1 \text{ g} \times \frac{1 \text{ kg}}{1000 \text{ g}} = 0.001 \text{ kg}$$

The same procedure is valid for volume measurements in the metric system.

4. Conversions between the apothecary and metric weight systems can be based on the conversion factor 15.4 gr = 1 g, which may be restated as 15.4 gr/g or 1 g/15.4 gr.

- a. How many milligrams equal 1 gr?

$$\frac{1 \text{ g}}{15.4 \text{ gr}} = 0.0648 \text{ g/gr} = 64.8 \text{ mg/gr or } 64.8 \text{ mg} = 1 \text{ gr}$$

- b. How many grams are in 1 ʒ?

$$\frac{1 \text{ g}}{15.4 \text{ gr}} \times \frac{480 \text{ gr}}{\text{ʒ}} = \frac{311 \text{ g}}{\text{ʒ}}$$

- c. How many grams are in 1 oz (avoir)? *Remember: 1 gr (apoth) = 1 gr (avoir).*

$$\frac{1.000 \text{ g}}{15.4 \text{ gr}} \times \frac{437.5 \text{ gr}}{\text{oz}} = \frac{28.4 \text{ gr}}{\text{oz}}$$

Other weight conversions are then found in a similar manner.

5. Conversions between the apothecary and metric measuring systems can be based on the conversion factor 1 f ʒ = 29.6 mL, which may be restated as 1 f ʒ/29.6 mL or 29.6 mL/f ʒ. How many minims are in 1 mL?

$$\frac{480 \text{ m}}{\text{f ʒ}} \times \frac{1 \text{ f ʒ}}{29.6 \text{ mL}} = \frac{16.2 \text{ m}}{\text{mL}}$$

RULES

1. The USP states that for prescription compounding one uses practical equivalents, defined as exact equivalents rounded to three (3) significant figures.
2. To calculate quantities required in pharmaceutical formulas, the USP directs the use of practical equivalents.
3. In converting doses the USP uses approximate equivalents. Use USP tables wherever possible.

EXAMPLES

1. Convert 1 pt, 4 f ʒ into milliliters. First, convert into fluid ounces.

$$\frac{16 \text{ f ʒ}}{\text{pint}} \times 1 \text{ pint} + 4 \text{ f ʒ} = 20 \text{ f ʒ}$$

Second, convert fluid ounces to milliliters.

$$20 \text{ f ʒ} \times \frac{29.6 \text{ mL}}{\text{f ʒ}} = 592 \text{ mL}$$

Answer: 1 pt, 4 f ʒ = 592 mL.

2. What is the weight of 1200 g in the apothecary system?

$$1200 \text{ g} \times \frac{15.4 \text{ gr}}{\text{g}} = 18,480 \text{ gr}$$

Or:

$$1200 \text{ g} \times \frac{1 \text{ lb}}{373 \text{ g}} = 3.22 \text{ lb}$$

3. Convert 1 pound (apoth) into grams.

$$\frac{1 \text{ g}}{15.4 \text{ gr}} \times \frac{480 \text{ gr}}{\text{ʒ}} \times \frac{12 \text{ ʒ}}{\text{lb}} \times 1 \text{ lb} = 374 \text{ g}$$

4. Convert 25 gr to grams.

$$25 \text{ gr} \times \frac{1 \text{ g}}{15.4 \text{ gr}} = 1.62 \text{ g}$$

5. Convert 50 g to grains.

$$50 \text{ g} \times \frac{15.4 \text{ gr}}{\text{g}} = 770 \text{ gr}$$

PROBLEMS

1. Convert:
 - a. 6.50 grains into milligrams.
 - b. 3/10 grain into milligrams.
 - c. 3½ apoth ounces into grams.
 - d. 2 ʒ into milligrams.
 - e. 3½ avoirdupois ounces into grams.
 - f. 1 lb avoirdupois into grams.
2. Convert:
 - A. 550 g into weighable quantities in the avoirdupois system.
 - B. 450 mg into grains.
 - C. 550 g into weighable quantities in the apoth system.
 - D. 100 µg into grains.
 - E. 1 kg into lb (avoir).
3. Convert the following doses into metric weights:
 - a. 1/100 gr.
 - b. 1/320 gr.
 - c. 1/6 gr.
 - d. 5 gr.
 - e. 20 gr.
4. Convert:
 - a. 200 minim into mL.
 - b. 3 ʒ into mL.
 - c. 8 ʒ into mL.
 - d. 1 pt into mL.
 - e. 5 <minim> into mL.
 - f. 0.1 mg into gr.
 - g. 5 mg into gr.
5. Answer the following questions.
 - a. How many gr are in 1 ʒ?
 - b. How many drams are in 1 ʒ?

- c. How many grains are in 1 oz (avoir)?
- d. How many gr are in ½ lb (apoth)?
- e. Convert 250 gr to weighable quantities in the apothecary system.

HOUSEHOLD EQUIVALENTS

Common household equivalents are found in Table 9-11. These are used to interpret the prescriber's instructions to the patient. The teaspoonful usually is indicated by the symbol $\text{f}\overline{3}$ or 5 mL, although 1 $\text{f}\overline{3}$ does not equal 5 mL. The problem of "the teaspoonful" has been discussed by Morrell and Ordway⁴ and by Madlon-Kay and Mosch.⁵ For practical purposes, a teaspoonful is equal to 5 mL, and 1 $\text{f}\overline{3}$ in the directions to the patient on the prescription means 1 teaspoonful.

For purposes of solving most compounding and dispensing problems, the exact equivalents rounded to three significant places should be used.

DOSAGE CALCULATIONS

In the past various rules for calculating infants' and children's dosages have been devised. All of them give only approximate dosages because they erroneously assume that the child is a small adult; some of them are still used because as yet no absolute method of calculating an infant or child's dose has been found. Children are sometimes more susceptible than adults to certain drugs. Doses for infants and children, where they are known, may be found in the USP Drug Information, the *Pediatric Dosage Handbook* published by APhA, and textbooks on pediatrics.⁶⁻⁸ Doses should not be calculated when it is possible to obtain the actual infant's or child's dose.

RULES FOR APPROXIMATE DOSES FOR INFANTS AND CHILDREN

1. *Young's Rule* (for children 2 years old and older)

$$\frac{\text{Age (years)}}{\text{Age (years)} + 12} \text{ Adult dose} = \text{child's dose (approx)}$$

2. *Clark's Rule*

$$\frac{\text{Weight (lb)}}{150} \times \text{Adult dose} = \text{Child's dose (approx)}$$

3. *Fried's Rule* (for infants up to 2 years old)

$$\frac{\text{Age (months)}}{150} \text{ Adult dose} = \text{Child's dose (approx)}$$

4. *The Square Meter Surface Area Method* relates the surface area of individuals to dose. It is thought that this is a more realistic way of relating dosages.

$$\frac{\text{Body surface area of child}}{\text{Body surface area of adult}} \text{ Adult dose} = \text{Child's dose (approx)}$$

The average body surface area for an adult has been given as 1.73 m²; hence,

$$\frac{\text{Body surface area of child (m}^2\text{)}}{1.73 \text{ m}^2} \text{ Adult dose} = \text{Child's dose (approx)}$$

Calculating Doses for Individuals—of Any Age or Size

Many drugs have doses stated as the amount of *drug/m² body surface area* and may be calculated as follows:

$$\frac{\text{Dose of drug}}{\text{m}^2 \text{ body surface area}} \text{ Body surface area (m}^2\text{)} = \text{dose}$$

Many physiological functions are proportional to body surface area, such as metabolic rate and kidney function.

Drug doses are often stated in *mg/kg body weight* and may be calculated as follows:

$$\frac{\text{Dose of drug}}{\text{kg body weight}} \text{ Body weight (kg)} = \text{Dose}$$

This is the most common way of determining children's doses. Drug doses also may be stated in *units*, as with vitamins A and D, penicillin, and hormones. This means that a certain quantity of biological activity of that drug is called 1 unit. When the term unit is used in connection with a drug, the calculations involved are the same as those for more familiar weight or volume notations. The USP often standardizes the unit for such drugs, so the expression "USP Units" is used. This means the units are calculated based on a USP assay procedure and reference standard.

EXAMPLES

1. The adult dose of a drug is 325 mg. What is the dose for a 3-year-old child?

Use Young's Rule:

$$\text{Child's dose (approx)} = \frac{3}{3 + 12} \times 325 \text{ mg} = 65 \text{ mg}$$

2. What is the dose for a 40 lb child if the average adult dose of the medication is 10 mg?

Use Clark's Rule:

$$\text{Child's dose (approx)} = \frac{40}{150} \times 10 \text{ mg} = 2.67 \text{ mg}$$

3. What is the dose for an 8-month-old infant if the average adult dose of a drug is 250 mg?

Use Fried's Rule:

$$\text{Infant's dose (approx)} = \frac{8}{150} \times 250 \text{ mg} = 13.3 \text{ mg}$$

4. If the average adult dose of a drug is 50 mg, what is the dose for a child who has a body surface area equal to 0.57 m²?

$$\text{Child's dose (approx)} = \frac{0.57}{1.73} \times 50 \text{ mg} = 16.5 \text{ mg}$$

PROBLEMS

1. What is the dose of a drug for a 9-month-old infant if the average adult dose is 25 mg?
2. What is the dose of a drug for a 6-year-old child if the average adult dose is 98 mg?
3. What is the dose of a drug for a child who weighs 28 lb if the average adult dose is 100 mg?
4. What is the dose of a drug for an individual who has a 1.21 m² body surface area? The average adult dose is 400 000 units.
5. What is the dose of a medicament for a child who weighs 66 lb if the dose is stated as 2.5 mg/kg body weight?
6. What is the dose of a drug for an average adult patient if the dose of the drug is 45 mg/m²?

PROBLEM-SOLVING METHODOLOGY

The problem-solving method illustrated in solving pharmaceutical problems is *dimensional analysis* (which is based on *ratio*

and proportion). Dimensional analysis is widely used in many scientific disciplines and offers a consistent way to solve problems. Dimensional analysis also overcomes many difficulties students and pharmacy practitioners have in problem interpretation and provides a well-defined, consistent starting point in the solution of pharmaceutical problems.

Dimensional Analysis

The basis for dimensional analysis is the formation of relationships between quantities, multiplication and canceling units until only the units of the desired answer remain. As an example, if 100 g of a drug cost US\$1.80, how much will 25 g cost? Begin by collecting all of the information in the problem and identify all relationships with units and labels. In this problem, we know the following:

$$\frac{\$1.80}{100 \text{ g drug}}, 25 \text{ g drug}$$

Identify the units you want for the answer (US\$). Identify a relationship from the problem that contains the unit(s) desired for the answer, forming the skeleton of the process.

$$\frac{\$1.80}{100 \text{ g drug}} \quad ? = \$$$

Complete the process by using terms from the problem (or equivalents) necessary to cancel out units until only the unit(s) of the answer remain on the right side.

$$\frac{\$1.80}{100 \text{ g drug}} \quad 25 \text{ g drug} = \$$$

Solve mathematically.

$$\text{Answer} = \$0.45$$

Dimensional analysis can be used to solve most pharmaceutical problems, regardless of complexity, using a consistent procedure:

1. Collect all the information and relationships in the problem complete with units and labels.
2. Identify the unit(s) and label of the answer.
3. Select a starting point corresponding to the unit(s) and label of the answer in the numerator.
4. Complete the process using relationships in the problem and known conversions to cancel units.
5. Solve the problem mathematically.

More complex problems use the same basic procedure: e.g., if 100 g of a drug cost US\$1.80, what would be the cost of the drug to prepare 4 f3 of a solution containing 5 g of the drug per teaspoonful?

Step 1: Collect all information and relationships:

$$\frac{\text{US\$1.80}}{100 \text{ g drug}}, \frac{5 \text{ g drug}}{1 \text{ tsp}}, 4 \text{ f3}$$

Step 2

$$= \text{US\$}$$

Step 3

$$\frac{\text{US\$1.80}}{100 \text{ g drug}} \quad ? = \text{US\$}$$

Step 4

$$\frac{\text{US\$1.80}}{100 \text{ g drug}} \times \frac{5 \text{ g drug}}{1 \text{ tsp}} \times \frac{1 \text{ tsp}}{5 \text{ mL}} \times \frac{29.6 \text{ mL}}{1 \text{ f3}} \times 4 \text{ f3} = \text{US\$}$$

(The 3rd and 4th terms are known definitions and equivalents needed to cancel units.)

Step 5

$$\text{Answer} = \text{US\$0.53}$$

With practice, steps 2 through 4 can be written in one operation.

EXAMPLES

1. Determine the amount of each ingredient contained in one dose of the following prescription.

Solid A	300 mg
Solid B	150 mg
Solid C	200 mg

M ft capsules, D.T.D. No. 12

The directions to the pharmacist are to mix and make 12 capsules each containing the three solids in the amounts indicated. Thus, the dose of each ingredient is as stated in the prescription.

2. How much of each ingredient is contained in one dose of the following prescription?

Solid E	7.2 g
Solid F	0.24 g
Solid G	1.2 g

M div capsules, No. 24

In this prescription the prescriber requests that 24 capsules be made from the three ingredients. The amounts of the ingredients requested are considerable, and drugs usually do not have doses of 7.2 g or 1.2 g, so division of the amounts by the number of doses (24) is required. The pharmacist should check a textbook or compendium to confirm the average adult dose.

$$\text{Drug E: } \frac{7.2 \text{ g}}{24 \text{ capsules}} \times 1 \text{ capsule} = 0.300 \text{ g}$$

$$\text{Drug F: } \frac{0.24 \text{ g}}{24 \text{ capsules}} \times 1 \text{ capsule} = 0.010 \text{ g}$$

$$\text{Drug G: } \frac{1.2 \text{ g}}{24 \text{ capsules}} \times 1 \text{ capsule} = 0.050 \text{ g}$$

3. A prescription calls for 10 units of a drug to be taken 3 times a day. How much will the patient have taken after 7 days?

$$\frac{10 \text{ units}}{\text{dose}} \times \frac{3 \text{ doses}}{\text{day}} \quad 7 \text{ days} = 210 \text{ units}$$

4. If 250 units of an antibiotic weigh 1 mg, how many units are in 15 mg?

$$\frac{250 \text{ units}}{\text{mg}} \quad 15 \text{ mg} = 3750 \text{ units}$$

5. If the dose of a drug is 0.5 mg/kg of body weight/24 h, how many grams will a 33-lb infant receive per 24 h and per week?

$$\frac{1 \text{ g}}{1000 \text{ mg}} \times \frac{0.5 \text{ mg}}{\text{kg} \times 24 \text{ hours}} \times \frac{1 \text{ kg}}{2.2 \text{ lb}} \times 33 \text{ lb} \times 24 \text{ hours} = 0.00750 \text{ g}$$

$$\frac{0.00750 \text{ g}}{\text{day}} \times \frac{7 \text{ days}}{\text{week}} \times 1 \text{ week} = 0.0525 \text{ g}$$

6. A patient is to receive 260 µg of a drug 4 times a day for 14 days. How many 1/250 gr tablets must be dispensed?

$$\frac{1 \text{ tablet}}{\frac{1}{250} \text{ gr}} \times \frac{1 \text{ gr}}{64.8 \text{ mg}} \times \frac{1 \text{ mg}}{1000 \text{ g}} \times \frac{260 \text{ g}}{\text{dose}} \times \frac{4 \text{ doses}}{\text{day}} \times 14 \text{ days} =$$

$$56.2 \text{ tablets} = 57 \text{ tablets}$$

7. An antibiotic is available as an injection containing 10 mg antibiotic/mL. How many milliliters are needed for an infant weighing 8 kg, the dose being 1.4 mg/kg of body weight?

$$\frac{1 \text{ mL}}{10 \text{ mg}} \times \frac{1.4 \text{ mg}}{\text{kg}} \times 8 \text{ kg} = 1.12 \text{ mL}$$

8. A preparation for coughs contains 1.5 g of an expectorant per 100 mL. How many grains of the expectorant are there in a teaspoonful?

$$\frac{15.4 \text{ gr}}{1 \text{ g}} \times \frac{1.5 \text{ g}}{100 \text{ mL}} \times \frac{5 \text{ mL}}{1 \text{ tsp}} \times 1 \text{ tsp} = 1.16 \text{ gr}$$

PROBLEMS

1. Calculate the dose for each ingredient in the following prescription.

Chemical J	10 mg
Chemical K	50 mg
Chemical L	300 mg

M ft capsules, D.T.D. No. 14

2. Calculate the dose of each ingredient in the following prescription.

Drug Q	10.5 g
Drug R	6.3 g

M div 21 doses

3. An f3 prescription contains 6 f3 of a tincture. If 1 teaspoonful 4 times a day is prescribed, how much tincture does the patient take per dose and how much is taken daily?

4. How many 0.3 mL doses are contained in 15 mL of a solution?

5. If 1 mg of a hormone equals 22.5 units, how many milligrams are required to obtain 1 unit?

6. If a bottle contains 80 units of a drug/mL, how many milliliters must the patient take to get a 60 unit dose? If the bottle contains 10 mL total volume of the drug solution, how many days' supply will patients have if they use 60 units a day?

7. A 10 mL ampule contains a 2.5 percent solution of a drug. How many milliliters are needed to give a dose of 150 mg?

8. The dose of an antibiotic is 75 mg for a child. How much of a flavored suspension containing 125 mg antibiotic/5 mL must be given to the child per dose?

9. How many milligrams of a drug are there in each teaspoonful of a syrup that contains 0.5 percent of the drug?

REDUCING AND ENLARGING FORMULAS

Determine the total weight or volume of ingredients and convert, if necessary, to the system of the quantities desired. The quantities in the original and new formulas will have the same ratio.

EXAMPLES

1. The formula for a syrup is:

Drug M	140 g
Sucrose	450 g
Purified water qs	1000 mL

- a. Find the quantities required for 100 mL.

$$\text{Drug M: } \frac{140 \text{ g}}{1000 \text{ mL}} \times 100 \text{ mL} = 14.0 \text{ g}$$

$$\text{Sucrose: } \frac{450 \text{ g}}{1000 \text{ mL}} \times 100 \text{ mL} = 45.0 \text{ g}$$

Purified water: to make 100 mL

- b. What quantities are required to compound 60 mL of the syrup?

$$\text{Drug M: } \frac{140 \text{ g}}{1000 \text{ mL}} \times 60 \text{ mL} = 8.40 \text{ g}$$

$$\text{Sucrose: } \frac{450 \text{ g}}{1000 \text{ mL}} \times 60 \text{ mL} = 27.0 \text{ g}$$

Purified water: to make 60 mL

2. Calculate the amounts needed for 100 g of antiseptic powder, as follows:

Solid A 2 g

Solid B 1 g

Solid C 7 g

Solid D 25 g

Solid E $\frac{115 \text{ g}}{150 \text{ g}}$

$$\text{factor} = \frac{100 \text{ g}}{150 \text{ g}} = 0.667$$

Solid A: 2 g $\times 0.667 = 1.33 \text{ g}$

Solid B: 1 g $\times 0.667 = 0.667 \text{ g}$

Solid C: 7 g $\times 0.667 = 4.67 \text{ g}$

Solid D: 25 g $\times 0.667 = 16.7 \text{ g}$

Solid E: 115 g $\times 0.667 = 76.7 \text{ g}$

3. Prescriptions where the instruction to the pharmacist calls for making a certain number of doses of an ingredient or mixture of several ingredients are a type of formula enlargement. The expression usually used is "D.T.D.," which means "let such doses be given." Occasionally the prescriber will not use this expression, but inspection of amounts of the ingredients indicates that this is what is desired. For example,

Solid H	50 mg
Solid K	150 mg
Liquid N	0.2 mL

M ft capsules, D.T.D. No. 24

The pharmacist checked the individual doses of the ingredients and found them to be slightly below the average adult dose, confirming that the prescriber wanted the quantities listed to be multiplied by 24.

$$\text{Solid H: } \frac{50 \text{ mg}}{\text{capsule}} \times 24 \text{ capsule} = 1200 \text{ mg or } 1.2 \text{ g}$$

$$\text{Solid K: } \frac{150 \text{ mg}}{\text{capsule}} \times 24 \text{ capsule} = 3600 \text{ mg or } 3.6 \text{ g}$$

$$\text{Solid N: } \frac{0.2 \text{ mL}}{\text{capsule}} \times 24 \text{ capsule} = 4.8 \text{ mL}$$

PROBLEMS

1. The formula for a liquid preparation is

Liquid C	35 mL
Solid B	9 g
Liquid R	2.5 mL
Liquid P	20 mL
Purified water, sufficient to make	100 mL

Calculate the quantities of the ingredients to make 2.5 L.

2. The formula for an ointment is:

Solid G	1
Liquid D	30

Solid M	3
Ointment base, sufficient to make	100

Calculate quantities of the ingredients for 2 lb (apoth).

3. How much of each of the three solids and how much purified water are needed to properly compound the following prescription order?

Solid N	0.1 mg
Solid Q	2.5 mg
Solid R	150.0 mg
Purified water, qs	5 mL

M ft solution, D.T.D. No. 48

4. How much of each ingredient is required to compound 90 mL of the following product?

Solid S	7.5 g
Solid T	25 g
Oil C	350 mL
Alcohol	250 mL
Purified water, qs	1000 mL

PERCENTAGE

Percent means per hundred. Fifteen percent is written 15% and means 15/100, 0.15, or 15 parts in a total of 100 parts. Percent is a type of ratio and has units of parts per 100 parts. Thus, 10% of 1500 tablets is $10/100 \times 1500$ tablets = 150 tablets.

To change percent to a fraction, the percent number becomes the numerator and 100 is the denominator. To change a fraction to percent, put the fraction in a form having 100 as its denominator; multiply by 100 so that the numerator becomes the percent.

$$\frac{1}{2} = \frac{50}{100}; \frac{50}{100} \times 100 = 50\%$$

$$\frac{1}{8} = \frac{12.5}{100}; \frac{12.5}{100} \times 100 = 12.5\%$$

Calculations involving percentages are encountered continually by pharmacists. They must be familiar not only with the making a certain number of doses of an ingredient or mixture of arithmetical principles, but also with certain compendial interpretations of the different type percentages involving solutions and mixtures. The USP states:

Percentage concentrations of solutions are expressed as follows:

- *Percent weight in weight (w/w)* expresses the number of g of a constituent in 100 g of product.
- *Percent weight in volume (w/v)* expresses the number of g of a constituent in 100 mL of product, and is used regardless of whether water or another liquid is the solvent.
- *Percent volume in volume (v/v)* expresses the number of mL of a constituent in 100 mL of product.

The term *percent* used without qualification means, for mixtures of solids, percent weight in weight; for solutions or suspensions of solids in liquids, percent weight in volume; for solutions of liquids in liquids, percent volume in volume; and for solutions of gases in liquids, percent weight in volume. For example, a one percent solution is prepared by dissolving 1 g of a solid or 1 mL of a liquid in sufficient of the solvent to make 100 mL of the solution.

Ratio Strength

Ratio strength is another manner of expressing concentration. Such phrases as “1 in 10” are understood to mean that one part of a substance is to be diluted with a diluent to make 10 parts

of the finished product. For example, a 1 : 10 solution means 1 mL of a liquid or 1 g of a solid dissolved in sufficient solvent to make 10 mL of solution. Ratio strength can be converted to percent by:

$$\frac{1 \text{ g substance}}{10 \text{ mL solution}} \times 100 \text{ mL solution} = 10 \text{ g substance}$$

$$\frac{10 \text{ g substance}}{100 \text{ mL solution}} = 10\%$$

The expression “parts per thousand” (e.g., 1 : 5000) always means parts weight in volume when dealing with solutions of solids in liquids and is similar to the above expression. A 1 : 5000 solution means 1 g of solute in sufficient solvent to make 5000 mL of solution. This can be converted to percent by:

$$\frac{1 \text{ g substance}}{500 \text{ mL solution}} \times 100 \text{ mL solution} = 0.02 \text{ g substance}$$

$$\frac{0.02 \text{ g substance}}{100 \text{ mL solution}} = 0.02\%$$

The expression “trituration” has two different meanings in pharmacy. One refers to the process of particle-size reduction, commonly by grinding or rubbing in a mortar with the aid of a pestle. The other meaning refers to a dilution of a potent powdered drug with a suitable powdered diluent in a definite proportion by weight. It is the second meaning that is used in this chapter.

When pharmacists refer to a “1 in 10 trituration” they mean a mixture of solids composed of 1 g of drug plus sufficient diluent (another solid) to make 10 g of mixture or *dilution*. In this case the “1 in 10 trituration” is actually a solid dilution of a drug with an inert solid. The strength of a trituration may also be stated as percent w/w. Thus, the term trituration has come to mean a solid dilution of a potent drug with a chemically and physiologically inert solid.

The meanings implied by the USP statements in the section on percentage are illustrated below with a few examples of the three types of percentages.

Weight-in-Volume Percentages

This is the type of percent problem most often encountered on prescriptions. The volume occupied by the solute and the volume of the solvent are *not* known because sufficient solvent is added to make a given or known final volume.

EXAMPLES

1. Prepare 1 f $\bar{3}$ of a 10% solution. Since this is a solution of a solid in a liquid, this is a w/v solution.

$$\frac{10 \text{ g drug}}{100 \text{ mL soln}} \times \frac{29.6 \text{ mL}}{1 \text{ f}\bar{3}} \times 1 \text{ f}\bar{3} = 2.96 \text{ g drug}$$

2.96 g is dissolved in sufficient purified water to make 29.6 mL of solution.

2. How much of a drug is required to compound 4 f $\bar{3}$ of a 3% solution in alcohol?

$$\frac{3 \text{ g drug}}{100 \text{ mL soln}} \times \frac{29.6 \text{ mL}}{1 \text{ f}\bar{3}} \times 4 \text{ f}\bar{3} = 3.55 \text{ g drug}$$

3. How much 0.9% solution of sodium chloride can be made from $\frac{1}{2}$ f $\bar{3}$ of NaCl?

$$\frac{100 \text{ mL soln}}{0.9 \text{ g NaCl}} \times \frac{31.1 \text{ g}}{1 \text{ f}\bar{3}} \times 0.5 \text{ f}\bar{3} = 1730 \text{ mL soln}$$

4. How many grams of a drug are required to make 120 mL of a 25% solution?

$$\frac{25 \text{ g drug}}{100 \text{ mL soln}} \quad 120 \text{ mL} = 30 \text{ g drug}$$

5. How would you prepare 480 mL of a 1 in 750 solution of an antiseptic? *Remember:* Percent w/v is indicated. One in 750 means 1 g of the antiseptic dissolved in sufficient solvent to make 750 mL solution.

$$\frac{1 \text{ g drug}}{750 \text{ mL soln}} \quad 480 \text{ mL} = 0.64 \text{ g drug}$$

Dissolve 0.64 g of antiseptic in sufficient solvent to make 480 mL solution.

6. How much of a substance is needed to prepare 1 L of a 1 : 10 000 solution? The ratio 1 : 10 000 means 1 g of a substance in 10 000 mL of solution.

$$\frac{1 \text{ g substance}}{10,000 \text{ mL soln}} \quad \frac{1000 \text{ mL}}{1 \text{ L}} \quad 1 \text{ L} = 0.1 \text{ g substance}$$

7. How would you prepare 120 mL of 0.25% solution of neomycin sulfate? The source of neomycin sulfate is a solution which contains 1 g neomycin sulfate/10 mL.

$$\frac{10 \text{ mL stock soln}}{1 \text{ g drug}} \quad \frac{0.25 \text{ g drug}}{100 \text{ mL soln}} \quad 120 \text{ mL soln} = 3 \text{ mL stock soln}$$

Add sufficient purified water to 3 mL of stock solution to make 120 mL.

PROBLEMS

- How would you make 3 f $\bar{3}$ of a 12.5% solution?
- How many liters of a 4% solution can be made from 4 $\bar{3}$ of a solid?
- How many liters of an 8% solution can be made from 500 g of a solid?
- How many grams of a drug are needed to make 4 L of a 1 in 500 solution?

Weight-in-Weight Percentages

Density must be considered in some of these problems. If a weight-in-weight solution is requested on a prescription, both the solute and solvent must be weighed, or the solute and the solvent may be measured if their densities are taken into consideration in determining the volumes. Since the solutions are made to a given weight, a given volume is not always obtainable.

EXAMPLES

1. What weights of solute and solvent are required to make 2 $\bar{3}$ of a 3% w/w solution of a drug in 90% alcohol?

$$\frac{3 \text{ g solute}}{100 \text{ g soln}} \times \frac{31.1 \text{ g soln}}{1 \bar{3} \text{ soln}} \times 2 \bar{3} \text{ soln} = 1.87 \text{ g solute}$$

$$\frac{31.1 \text{ g soln}}{1 \bar{3} \text{ soln}} \times 2 \bar{3} \text{ soln} = 62.2 \text{ g soln}$$

$$62.2 \text{ g soln} - 1.87 \text{ g solute} = 60.3 \text{ g solvent}$$

2. The solubility of boric acid is 1 g in 18 mL of water at 25°C. What is the percentage strength, w/w, of a saturated solution? One gram of boric acid in 18 mL of water makes a saturated solution, and 18 mL of water weighs 18 g, hence the weight of solution is 19 g. The amount of boric acid present is 1 g in 19 g of solution; therefore, the following relationship can be set up:

$$\frac{1 \text{ g drug}}{19 \text{ g soln}} \times 100 \text{ g soln} = 5.26 \text{ g drug}$$

$$\frac{5.26 \text{ g drug}}{100 \text{ g soln}} = 5.26\%$$

3. How many grams of a chemical are needed to prepare 200 g of a 10% w/w solution? Ten percent w/w means 10 g of solute in 100 g total solution. The following relationship may be set up:

$$\frac{10 \text{ g solute}}{100 \text{ g soln}} \quad 200 \text{ g soln} = 20 \text{ g solute}$$

4. How would one make a 2% w/w solution of a drug in 240 mL of alcohol? The density of alcohol is 0.816 g/mL.

- a. First, convert 240 mL to weight. *Remember:* alcohol is the solvent and it has a density different from that of water.

$$\frac{0.816 \text{ g alcohol}}{1 \text{ mL alcohol}} \quad 240 \text{ mL alcohol} = 195.8 \text{ g (196 g) alcohol}$$

- b. Two percent w/w means 2 g solute in 100 g solution. In this problem the final weight of solution is not known; 240 mL (196 g) of alcohol represents the solvent only. The solvent is 98% w/w of the total solution, so the following relationship may be set up:

$$\frac{2 \text{ g solute}}{98 \text{ g alcohol}} \quad 196 \text{ g alcohol} = 4.00 \text{ g solute}$$

- c. Dissolve 4.00 g of the drug in 240 mL alcohol. The resulting solution will be 2% w/w and have a volume slightly larger than 240 mL because of the volume displacement of the drug.
5. How much of a 5% w/w solution can be made from 28.4 g of a chemical?

$$\frac{100 \text{ g soln}}{5 \text{ g chemical}} \quad 28.4 \text{ g chemical} = 568 \text{ g soln}$$

6. How many milliliters of a 70% w/w solution having a density of 1.2 g/mL will be needed to prepare 600 mL of a 10% w/v solution?

- a. Drug needed:

$$\frac{10 \text{ g drug}}{100 \text{ mL soln (10%)}} \times 600 \text{ mL soln (10%)} = 60 \text{ g drug}$$

- b. Weight of 70% solution needed.

$$\frac{100 \text{ g soln (70%)}}{70 \text{ g drug}} \quad 60 \text{ g drug} = 85.7 \text{ g soln (70%)}$$

- c. Volume of 70% solution needed.

$$\frac{1 \text{ mL soln (70%)}}{1.2 \text{ g soln (70%)}} \quad 85.7 \text{ g soln (70%)} = 71.4 \text{ mL soln (70%)}$$

Compounding problems involving solid preparations (such as mixtures of powder) and semisolid preparations (such as ointments, creams, and suppositories) are also percent w/w. The following is an example of this.

1. How much drug is required to make 2 $\bar{3}$ of a 10% ointment?

$$\frac{10 \text{ g drug}}{100 \text{ g oint}} \times \frac{31.1 \text{ g oint}}{1 \text{ } \frac{1}{2} \text{ oint}} \times 2 \frac{1}{2} \text{ oint} = 6.22 \text{ g drug}$$

The same procedure could be used for such mixtures as powders and suppository masses. Instead of using units in the various measuring systems, quantities can be indicated "by parts." The term "parts" then can mean any unit in any measuring system, as long as the units are kept constant.

2. How many grams of each of the following three ingredients are required to make 30 g of the product?

Solid A	0.5 part
Powder B	3.0 parts
Powder C, qs	30.0 parts

Since the product is a mixture of powders, percent w/w is indicated. In the above prescription order the total product is 30 parts because powder C is used to "qs" or "make up to" 30 parts. Therefore, 0.5 g of solid A and 3.0 g of powder B are needed.

$$\frac{1 \text{ g drug}}{19 \text{ g soln}} \times 100 \text{ g soln} = 52.6 \text{ g drug}$$

$$\frac{52.6 \text{ g drug}}{100 \text{ g soln}} = 52.6\%$$

3. How much of each of the following ingredients is needed to make 60 g of the ointment?

R _x Solid D	3.0 parts
Solid E	6.0 parts
ointment Base Q	$\frac{30.0 \text{ parts}}{39.0 \text{ parts total}}$
$\frac{3.0 \text{ g solid D}}{39.0 \text{ g oint}} \times 60 \text{ g oint}$	$= 4.26 \text{ g solid D}$
$\frac{6.0 \text{ g solid E}}{39.0 \text{ g oint}} \times 60 \text{ g oint}$	$= 92.3 \text{ g solid E}$
$\frac{30.0 \text{ g base Q}}{39.0 \text{ g oint}} \times 60 \text{ g oint}$	$= 46.2 \text{ g base Q}$

4. What is the percent strength of a salt solution obtained by diluting 100 g of a 5% solution to 200 g? Assign the 5% solution as soln 1. Assign the final solution as soln 2.

$$\frac{5 \text{ g salt}}{100 \text{ g soln 1}} \times \frac{100 \text{ g soln 1}}{200 \text{ g soln 2}} \times 100 \text{ g soln 2} = 2.5 \text{ salt}$$

$$\frac{2.5 \text{ g salt}}{100 \text{ g soln 2}} = 2.5\% \text{ w/w}$$

PROBLEMS

1. How much of the drug and solvent are needed to compound the following prescription?

R _x Compound A	6% w/w
Solvent, qs	4 } \frac{1}{2}

2. How many grams of solute are needed to prepare 240 g of a 12% w/w solution?
3. How many kilograms of a 20% w/w solution can be made from 1 kg of the solute?
4. How would you prepare, using 120 mL of glycerin (density, 1.25 g/mL), a solution that is 3% w/w with respect to a drug?
5. How much of each substance is needed to prepare a total of 24 g of the following suppository mass?

Compound K	0.3 g
Solid H	0.15 g
Suppository base, qs	2.0 g

6. How would one prepare 500 mL of a 15% w/w aqueous solution?
7. How much of each of the ingredients is required to make 1 kg of the following mixture?

Powder P	1 part
Powder Q	8 part
Powder R	12 part
Powder S	15 part
Total	$\frac{36 \text{ part}}{\text{---}}$

8. How much of each ingredient is required to prepare the following ointment?

R _x Coal Tar Solution	10%
Hydrophilic Ointment, qs	30 g

Volume-in-Volume Percentages

A direct calculation of percentage from the total volume is made. Volumes, unlike weights, may not be additive. However, this does not present a problem because the final solution is made up to the desired volume with the diluent.

In the case of potent substances, a properly prepared stock solution permits the pharmacist to obtain accurately a quantity of solid that might otherwise be difficult to weigh. In the case of frequently prescribed salt solutions, a stock solution readily provides the required amount of salt without the necessity of weighing and dissolving it every time.

Stock solutions may be of various concentrations depending on the requirements for use. The stock solutions should be labeled properly and fractional parts needed to make various strengths also may be listed as a further convenience.

There is a type of compounding and dispensing problem that involves the concept of stock solutions. This involves the patient diluting a dose from the prescription order to a given volume to obtain a solution of desired concentration.

For example, how many grams of a salt are required to make 90 mL of a stock solution, 5 mL of which makes a 1 : 3000 solution when diluted to 500 mL? Assign the stock solution as soln 1. Assign the final dilution as soln 2.

$$\frac{1 \text{ g salt}}{3000 \text{ mL soln 2}} \times \frac{500 \text{ mL soln 2}}{5 \text{ mL soln 1}} \times 90 \text{ mL soln 1} = 3.0 \text{ g salt}$$

EXAMPLES

1. How many minims of a liquid are needed to make 6 } \frac{1}{2} of a hand lotion containing 0.5% v/v of the liquid?

$$\frac{16.2 \text{ mL liq}}{1 \text{ mL liq}} \times \frac{0.5 \text{ mL liq}}{100 \text{ mL lotion}} \times \frac{29.6 \text{ mL lotion}}{1 \text{ } \frac{1}{2} \text{ lotion}} \times 6 \frac{1}{2} \text{ lotion} = 14.4 \text{ mL liq}$$

Add sufficient lotion to 14.4 minim of the liquid to make 6 } \frac{1}{2} of the product.

2. How much 90% alcohol is required to compound 500 mL of a 10% alcohol mixture?

$$\frac{100 \text{ mL (90\%)}}{90 \text{ mL alcohol}} \times \frac{10 \text{ mL alcohol}}{100 \text{ mL (10\%)}} \times 500 \text{ mL (10\%)} = 55.5 \text{ mL (90\%)}$$

PROBLEMS

1. How many minims of a liquid are needed to make 4 } \frac{3}{4} of a 12.5% v/v solution?
2. What volume of 50% v/v alcohol could be prepared from 1 L of 95% v/v alcohol?
3. What is the percentage strength, weight in weight, of a liquid made by dissolving 16 g of a salt in 30 mL of water?

- How much drug will be required to prepare 1℥ of a 2.5% solution?
- What is the percentage, weight in weight, of sugar in a syrup made by dissolving 5 kg of sugar in 8 kg of water?
- How many grams of a drug are required to prepare 120 mL of a 12.5% aqueous solution?
- How much drug is needed to compound a liter of a 1 : 2500 aqueous solution?
- A solution contains 37% of active ingredient. How much of this solution is needed to prepare 480 mL of an aqueous solution containing 2.5% of the active ingredient?
- How much of a drug is required to make 2 qt of a 1 : 1200 solution?

STOCK SOLUTIONS

To facilitate the dispensing of certain soluble substances, the pharmacist frequently prepares or purchases solutions of high concentration. Portions of these concentrated solutions are diluted to give required solutions of lesser strength. These concentrated solutions are known as *stock solutions*. This procedure is satisfactory if the substances are stable in solution or if the solutions are to be used before they decompose.

In the case of potent substances, a properly prepared stock solution permits the pharmacist to obtain accurately a quantity of solid that might otherwise be difficult to weigh. In the case of frequently prescribed salt solutions, a stock solution readily provides the required amount of salt without the necessity of weighing and dissolving it every time.

Stock solutions may be of various concentrations depending on the requirements for use. The stock solutions should be labeled properly and fractional parts needed to make various strengths may also be listed as a further convenience.

There is a type of compounding and dispensing problem that involves the concept of stock solutions. This involves the patient diluting a dose from the prescription order to a given volume to obtain a solution of desired concentration.

For example, how many grams of a salt are required to make 90 mL of a stock solution, 5 mL of which makes a 1 : 3000 solution when diluted to 500 mL? Assign the stock solution as soln 1. Assign the final dilution as soln 2.

$$\frac{1 \text{ g salt}}{3000 \text{ mL soln 2}} \times \frac{500 \text{ mL soln 2}}{5 \text{ mL soln 1}} \times 90 \text{ mL soln 1} = 3.0 \text{ g salt}$$

PROBLEMS

- How much of a drug is needed to compound 120 mL of a prescription order such that when 1 teaspoonful of the solution is diluted to 1 qt, a 1 : 750 solution results?
- How many grams of a drug are needed to make 240 mL of a solution of such strength that when 5 mL is diluted to 2 qt, a 1 : 2500 solution results?
- An ampule of solution of an anti-inflammatory drug contains 4 mg of drug per mL. What volume of the solution is needed to prepare a liter of solution that contains 2 µg of the drug per mL?

PARTS PER MILLION

An expression that is occasionally used in compounding prescriptions is *parts per million* (ppm). This is another way of expressing concentration, particularly concentrations of very dilute preparations. A 1 percent solution may be expressed as 1 part/100; a 0.1 percent solution is 0.1 parts/100 or 1 part/1000. A 1 ppm solution contains 1 part of solute/1 million parts of solution; 5 ppm is 5 parts solute/1 million parts solution, and so on. Remember that the two parts must have the same units, except in the metric system where 1 g = 1 mL of water.

Sodium fluoride is a drug that may be prescribed by a dentist as a preventative for tooth decay in children. It is used only in very dilute solutions due to the drug's toxicity and because only minute quantities are needed. For example, how much sodium fluoride would be needed to prepare the following prescription?

R_x Sod Fluoride, qs
Purified water, qs 60 mL
Make soln such that when 1 f℥ is diluted to 1 glassful of water a 2 ppm soln results.
sig: 1 f℥ in a glassful of water a day.

The mathematics to solve this compounding problem are easy once the steps for calculating the answer are outlined. This problem should be worked "backward."

- The amount of NaF needed is not known.
- One glassful of water has a volume of 240 mL. The concentration of NaF in 240 mL is 2 ppm.
- The NaF solution poured into the glass came from a teaspoonful dose (1 f℥), which is equal to 5 mL.
- The 5 mL dose came from the prescription order bottle containing a NaF solution.

$$\frac{2 \text{ g NaF}}{1,000,000 \text{ mL dilution}} \times \frac{240 \text{ mL dilution}}{5 \text{ mL R}_x} \times 60 \text{ mL R}_x = 0.00576 \text{ g NaF}$$

The pharmacist would weigh out 5.76 mg (actually, one would weigh out a larger quantity and take an aliquot part) and qs to 60 mL.

Another variation of this problem is the prescriber requesting the concentration in terms of fluoride ion (F⁻). In this case the atomic weight of F⁻ and molecular weight of NaF are used in the calculation. If the request called for 2 ppm fluoride, the initial calculations would be the same as above, and an additional step would be added at the end. The 5.76 mg would now represent the weight of fluoride ion needed. This must be converted to weight of NaF. The molecular weight of NaF is 42 and the atomic weight of fluorine is 19. The following proportion can be set up.

PROBLEM

- How many milligrams of NaF are needed in the following prescription?

R_x Sodium fluoride
Purified water, qs to 90 mL
M ft solution such that when 1 f℥ is diluted to 1 glassful of water a 3 ppm NaF soln results.

DILUTION AND CONCENTRATION

Stock solutions can be diluted to make a product that has a lower concentration; also mixtures of powders or semisolids (e.g., ointments) can be diluted to give a product of lower concentration of the drug(s). The diluent is an inert solid or semisolid or base that does not contain any active ingredients.

Mixtures also may be concentrated by adding pure drug or mixing with a product containing a higher concentration of the drug. For example, how much of a diluent must be added to 50 g of a 10 percent ointment to make it a 5 percent ointment?

- How many grams of active ingredient are in 50 g of 10% ointment?

$$\frac{10 \text{ g drug}}{100 \text{ g oint (10\%)}} \quad 50 \text{ g oint (10\%)} = 5 \text{ g drug}$$

2. How many grams of a 5% ointment can be made from 5 g of active ingredient?

$$\frac{100 \text{ g oint (5\%)}}{5 \text{ g drug}} \times 5 \text{ g drug} = 100 \text{ g oint (5\%)}$$

3. How many grams of base must be added to the 50 g of the original 10% ointment?

$$100 \text{ g oint (5\%)} - 50 \text{ g oint (10\%)} = 50 \text{ g base}$$

The term *trituration* was used previously to mean a dilute powder mixture of a drug. It is often necessary to dilute this mixture further to obtain the required amount of drug.

1. How much of a 1 in 10 trituration of a potent drug contains 200 mg of the drug? A 1 in 10 trituration means 1 g of drug in 10 g of mixture or 1 g of drug plus 9 g diluent. *Remember:* Mixtures of solids are percent w/w.

$$\frac{10 \text{ g trituration}}{1 \text{ g drug}} \quad \frac{1 \text{ g drug}}{1000 \text{ mg drug}} \quad 200 \text{ mg drug} = 2 \text{ g trituration}$$

2. How much diluent must be added to 10 g of a 1 : 100 trituration to make a mixture that contains 1 mg of drug in each 10 g of the final mixture?
- a. Determine the amount of drug in 10 g of trituration.

$$\frac{1 \text{ g drug}}{100 \text{ g trituration}} \times 10 \text{ g trituration} = 0.1 \text{ g drug}$$

- b. Determine the amount of mixture that can be made from 0.1 g (100 mg) of drug.

$$\frac{10 \text{ g mixture}}{1 \text{ mg drug}} \quad \frac{1000 \text{ mg drug}}{1 \text{ g drug}} \quad 0.1 \text{ g drug} = 1000 \text{ g mixture}$$

- c. Determine the amount of diluent needed.

$$1000 \text{ g mixture} - 10 \text{ g trituration} = 990 \text{ g diluent}$$

PROBLEMS

1. The following prescription order was received in a pharmacy. If the only *R* cream available is a 10% concentration, how much of the 10% cream and how much diluent are required to compound the prescription?

$$\bar{R}_x \quad \bar{R} \text{ Cream } 3\% \dots 30 \text{ g}$$

2. How many grams of a 1 : 100 trituration contain 100 µg of the active ingredient?
3. How many grams of a 1 : 1000 dilution can be made from 1 g of a 1 : 25 trituration?

MIXING DIFFERENT STRENGTHS

RULES

1. The sum of the products obtained by multiplying a series of quantities by their respective concentrations equals the product obtained by multiplying a concentration by the sum of the quantities. For example, the sum of the products—obtained by multiplying the individual weights or volumes of a series of preparations by the concentration of a given ingredient contained in each preparation—is equal to the product obtained by multiplying the total weight of the series of preparations by the percentage of the given ingredient resulting

from a homogeneous mixture of the same series of preparations.

2. When mixing products of varying strengths, the units and type of percent (w/w, w/v, v/v) must be kept constant.

EXAMPLES

1. What is the percent of alcohol in a mixture made by mixing 5 L of 25%, 1 L of 50%, and 1 L of 95% alcohol?
- a. Determine the total amount of alcohol in the three solutions and the total amount of solution (1 L = 1000 mL). Assume additivity of volumes on mixing.

$$\frac{25 \text{ mL alcohol}}{100 \text{ mL (25\%)}} \times 5000 \text{ mL (25\%)} = 1250 \text{ mL alcohol}$$

$$\frac{50 \text{ mL alcohol}}{100 \text{ mL (50\%)}} \times 1000 \text{ mL (50\%)} = 500 \text{ mL alcohol}$$

$$\frac{95 \text{ mL alcohol}}{100 \text{ mL (95\%)}} \times 1000 \text{ mL (95\%)} = 950 \text{ mL alcohol}$$

- b. Determine the percent of alcohol in the mixture. There is a total of 2700 mL of alcohol in 7000 mL of total solution.

$$\frac{2700 \text{ mL alcohol}}{7000 \text{ mL mixture}} \times 100 \text{ mL mixture} = 38.6 \text{ mL alcohol}$$

$$\frac{38.6 \text{ mL alcohol}}{100 \text{ mL mixture}} = 38.5\%$$

2. What is the strength of a mixture obtained by mixing 50 g of a 5%, 100 g of a 7.5%, and 40 g of a 10% ointment?

$$\frac{5 \text{ g drug}}{100 \text{ g oint (5\%)}} \times 50 \text{ g oint (5\%)} = 2.5 \text{ g drug}$$

$$\frac{7.5 \text{ g drug}}{100 \text{ g oint (7.5\%)}} \times 100 \text{ g oint (7.5\%)} = 7.5 \text{ g drug}$$

$$\frac{10 \text{ g drug}}{100 \text{ g oint (10\%)}} \times 40 \text{ g oint (10\%)} = 4.0 \text{ g drug}$$

There is a total of 14.0 g of active ingredient in 190 g of total mixture.

$$\frac{14.0 \text{ g drug}}{190 \text{ g mixture}} \times 100 \text{ g mixture} = 7.37 \text{ g drug}$$

$$\frac{7.37 \text{ g drug}}{100 \text{ g mixture}} = 7.37\%$$

PROBLEMS

1. What percent of a drug is contained in a mixture of powder consisting of 0.5 kg, containing 0.038% of a drug, and 10 kg, containing 0.043% of a drug?
2. What is the strength of a mixture produced by combining the following lots of alcohol: 2 L of 95%, 2 L of 50%, and 7 L of 60%?
3. What is the percent of drug content in the following mixture: 2 kg of 3%, 300 g of 2.5%, and 500 g of 4.2% resin?

ALLIGATION ALTERNATE

Alligation is a rapid method of calculation that is useful to the pharmacist. The name is derived from the Latin *alligatio*,

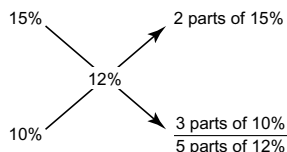
meaning the act of attaching, and it refers to lines drawn during calculation to bind quantities together. This method is used to find the proportions in which substances of different strengths or concentrations must be mixed to yield a mixture of desired strength or concentration. When the proportion is found, a calculation may be performed to find the exact amounts of the substances required.

RULES

1. Line up the concentrations of all the starting materials in a vertical column in order of concentration, traditionally from high to low. Pure drugs are defined as being 100%; solvents or vehicles are designated as 0%.
2. Place the concentration of the desired product in a second column such that it is bracketed by concentrations of starting materials. With two starting materials, the desired product simply falls between the two.
3. Cross subtract the two columns to give a parts formula that can be used to calculate specific amounts of each starting material.

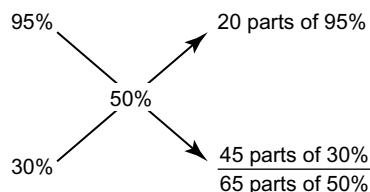
EXAMPLES AND PROCEDURE

1. In what proportion must a preparation containing 10% of drug be mixed with one containing 15% of drug to produce a mixture of 12% drug strength? Applying the above rules gives:



The concentrations of the starting material are lined up in the first column in decreasing or increasing order and the desired percent or concentration is placed in the center column. The third column is obtained by cross-subtracting as indicated by the arrows and gives a parts formula for mixing the two starting materials. Thus, mixing 2 parts of 15% drug preparation with 3 parts of 10% drug preparation will produce 5 parts of a drug mixture of the desired 12% strength.

2. In what proportion must 30% alcohol and 95% alcohol be mixed to make 500 mL of 50% alcohol? Set up the problem in the following manner:



In a total of 65 parts, 20 parts of 95% alcohol 45 parts of 30% alcohol are needed. Since the total is proportional to 500 mL, the following can be calculated:

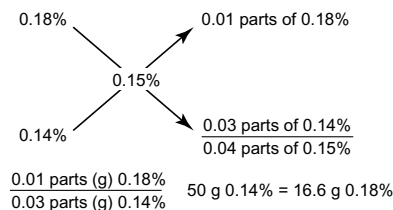
$$\frac{20 \text{ parts (mL) } 95\%}{65 \text{ parts (mL) } 50\%} \times 500 \text{ (mL) } 50\% = 154 \text{ mL } 95\%$$

$$\frac{45 \text{ parts (mL) } 30\%}{65 \text{ parts (mL) } 50\%} \times 500 \text{ (mL) } 50\% = 346 \text{ mL } 30\%$$

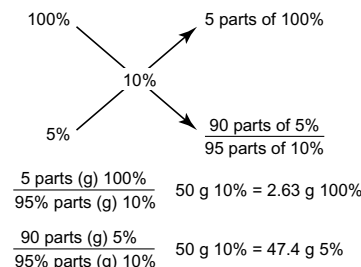
Since volumes are not additive, sufficient water may be needed to make 500 mL.

3. How many grams of an ointment containing 0.18% of active ingredient must be mixed with 50 g of an ointment

containing 0.14% of active ingredient to make a product containing 0.15% of active ingredient?



4. Occasionally it is necessary for a pharmacist to increase the strength of a product. For example, a prescription calls for 50 g of a 10% ointment. The pharmacist only has a 5% ointment and the pure ingredient available. How much of the 5% ointment and the pure ingredient are needed to compound the prescription?



PROBLEMS

1. How much ointment containing 12% drug and how much ointment containing 16% drug must be used to make 1 kg of a product containing 12.5% drug?
2. In what proportion should 50% alcohol and purified water be mixed to make a 35% alcohol solution? (The purified water is 0% alcohol.) *Note:* This problem may be solved by a method other than alligation as was shown above.
3. How many grams of 28% w/w ammonia water should be added to 500 g of 5% w/w ammonia water to produce a 10% w/w ammonia concentration?
4. How many milliliters of 20% dextrose in water and how many milliliters of 50% dextrose in water are needed to make 1 L of 35% dextrose in water?

PROOF SPIRIT

For tax purposes, the US government calculates the strength of pure or absolute alcohol (herein referred to as $\text{C}_2\text{H}_5\text{OH}$) by means of *proof degrees*. This means that 100 proof spirit contains 50% (by volume) or 42.49% (by weight) of $\text{C}_2\text{H}_5\text{OH}$, and its specific gravity is 0.93426 at 60°F. Thus, 2 proof degrees equals 1% (by volume) of $\text{C}_2\text{H}_5\text{OH}$. One proof gallon is one gal of 50% (by volume) of $\text{C}_2\text{H}_5\text{OH}$ at 15.56°C (60°F). In other words, a proof gallon is a gallon that contains $\frac{1}{2}$ gal of $\text{C}_2\text{H}_5\text{OH}$. A proof gallon is 100 proof.

The term *10 degrees under proof* (10° up) signifies that 100 volumes of the spirit contains 90 volumes of proof spirit plus 10 volumes of water, and *30 degrees over proof* (30° op) indicates that 100 volumes diluted with water yields 130 volumes of proof spirit. To prepare proof spirit, 50 volumes of $\text{C}_2\text{H}_5\text{OH}$ are mixed with 53.71 volumes of water to allow for the contraction that occurs to yield 100 volumes of product.

The terms *proof strength*, *proof gallon*, and *proof spirit* are used so that the tax is levied only on the actual quantity of $\text{C}_2\text{H}_5\text{OH}$ contained in any mixture. Therefore, it is sometimes necessary for the pharmacist to convert alcohol purchased

to proof strength to compute tax refunds or convert proof strengths to percent for compounding purposes.

A quantity of solution that contains $\frac{1}{2}$ gal of C_2H_5OH is said to contain one proof gal. Proof gallons may be calculated by the following two equations:

$$\text{Proof gal} = \frac{\text{gal v/v strength}}{50\% \text{ v/v}}$$

$$\text{Proof gal} = \frac{\text{gal proof strength}}{100 \text{ proof}}$$

The second equation is the same as the first because proof strength is always twice the % v/v strength. With these equations, given any two variables the third can be calculated.

EXAMPLES

1. What is the taxable alcohol in 1 pt of alcohol USP?

$$1 \text{ pt} = \frac{1}{8} \text{ gal} \quad (8 \text{ pt} = 1 \text{ gal})$$

Alcohol USP is 95%v/v; therefore,

$$\text{proof gal} = \frac{\text{gal \%strength}}{50\%} = \frac{\frac{1}{8} \text{ gal } 95\%}{50\%} = 0.2375 \text{ proof gal}$$

2. How much Diluted Alcohol USP can be made from 1 qt of alcohol labeled $\frac{1}{2}$ proof gallon? Diluted alcohol USP is 49%v/v; therefore,

$$\begin{aligned} \text{Proof gal} &= \frac{\text{gal \%strength}}{50\%} \\ \text{gal} &= \frac{0.5 \text{ proof gal } 50\%}{49\%} = 0.510 \text{ gal} \end{aligned}$$

PROBLEMS

1. How many proof gallons are there in 1 qt of a preparation that is labeled 75%v/v alcohol?
2. How many proof gallons are there in a pint of an elixir that contains 14% alcohol?
3. How much Diluted Alcohol USP can be made from 1 gal of 190 proof alcohol?

SATURATED SOLUTIONS

Occasionally, it is necessary for a pharmacist to make saturated solutions. Solubility in the USP/NF is expressed as the number of milliliters of a solvent that will dissolve 1 g of a solid; for example, 1 g dissolves in 0.5 mL of water. In other words, if 1 g of a solid is dissolved in 0.5 mL of water, a saturated solution results. An example will illustrate this.

1. How much of a drug is needed to make 120 mL of a saturated solution if 1 g of the drug dissolves in 7.5 mL of water?

Calculate the amount of drug that can be dissolved in 120 mL of water.

$$\frac{1 \text{ g drug}}{7.5 \text{ mL water}} \times 120 \text{ mL water} = 16 \text{ g drug}$$

When 16 g of the drug are dissolved in 120 mL of water, a saturated solution results that has a volume greater than 120 mL because the solid will take up a certain volume. Only 120 mL would be dispensed.

2. What is the % w/w of the above solution?

$$\begin{aligned} 120 \text{ g(mL) water} + 16 \text{ g drug} &= 136 \text{ g solution} \\ \frac{16 \text{ g drug}}{136 \text{ g solution}} \times 100 \text{ g solution} &= 11.8 \text{ g drug} \\ \frac{11.8 \text{ g drug}}{100 \text{ g solution}} &= 11.8\% \text{ w/w} \end{aligned}$$

PROBLEMS

1. What is the solubility of a chemical if a saturated aqueous solution is 0.5% w/w?
2. How many grams are needed to make 500 mL of a saturated solution if 1 g of the solute is soluble in 14 mL of solvent?

MILLIEQUIVALENTS

The quantities of electrolytes administered to patients are usually expressed by the term *milliequivalents* (mEq). The reason that weight units (mg, g) are not used is because the electrical activity of the ions, which in this instance is important, may be best expressed as milliequivalents. (See Chapter 15 "Ionic Solutions and Electrolytic Equilibria.")

A milliequivalent is 1/1000 of an *equivalent* (Eq). An equivalent is the weight of a substance that combines with or replaces one gram-atomic weight (g-at wt) of hydrogen. In pharmacy, the terms equivalent and equivalent weight (Eq wt) have been used interchangeably. For problem solving it is convenient to identify the molar weight in terms of mg per mmol and the number of mEq per mmol as follows:

$$\begin{aligned} \text{Molecular weight} &= \frac{\text{g}}{\text{mole}} = \frac{\text{mg}}{\text{mmol}} \\ \frac{\text{mEq}}{\text{mol}} &= \text{valence} \end{aligned}$$

For example, KCl has a molecular weight of 74.5; the above parameters would be 74.5 mg/mmol and 1 mEq/mmol.

Water of hydration contributes to the molecular weight (mol wt) of a compound but *not* to the valence, and the total mol wt is used to calculate mEq.

EXAMPLES

1. Calcium (Ca^{2+}) has a gram-atomic weight of 40.08. Determine the number of mEq/mmol. As the valence of the calcium ion is 2, there are 2 mEq/mmol.
2. A solution (100 mL) that contains 409.5 mg of NaCl/100 mL has how many mEq of Na^+ and Cl^- ? The molecular weight of NaCl is 58.5.

$$\text{There is } \frac{1 \text{ mEq } Cl^-}{\text{mmol NaCl}} \text{ and } \frac{1 \text{ mEq } Na^+}{\text{mmol NaCl}}$$

$$\frac{1 \text{ mEq } Cl^-}{\text{mmol NaCl}} \times \frac{1 \text{ mmol NaCl}}{58.5 \text{ mg NaCl}} \times \frac{409.5 \text{ mg NaCl}}{100} \text{ mL} \times 100 \text{ mL} = 7.0 \text{ mEq } Cl^-$$

Since NaCl is a 1:1 electrolyte, the solution contains 7.0 mEq of Cl^- and 7.0 mEq of Na^+ .

3. A prescription order calls for a 500 mL solution of potassium chloride to be made so that it will contain 400 mEq of K^+ . How many grams of KCl (mol wt: 74.5) are needed?

$$\begin{aligned} \frac{1 \text{ mEq}}{\text{mmol}} \text{ and } \frac{74.5 \text{ mg}}{\text{mmol}} \\ \frac{1 \text{ g KCl}}{1000 \text{ mg KCl}} \times \frac{74.5 \text{ mg KCl}}{\text{mmol KCl}} \times \frac{1 \text{ mmol KCl}}{\text{mEq } K^+} \times 400 \text{ mEq } K^+ = 29.8 \text{ g KCl} \end{aligned}$$

4. How many mEq of K^+ are in a 250 mg tablet of potassium phenoxymethyl penicillin (mol wt: 388.5; valence: 1)?

$$\frac{1 \text{ mEq } K^+}{\text{mmol Pen}} \text{ and } \frac{388.5 \text{ mg Pen}}{\text{mmol Pen}}$$

$$\frac{1 \text{ mEq } K^+}{\text{mmol Pen}} \times \frac{1 \text{ mmol Pen}}{388.5 \text{ mg Pen}} \times \frac{250 \text{ mg Pen}}{\text{Tab}} \times 1 \text{ Tab} = 0.644 \text{ mEq } K^+$$

5. How many mEq of Mg are there in 10 mL of a 50% magnesium sulfate injection? The mol wt of $MgSO_4 \cdot 7H_2O$ is 246.

$$\frac{2 \text{ mEq } Mg^{2+}}{\text{mmol drug}} \text{ and } \frac{246 \text{ mg drug}}{\text{mmol drug}}$$

$$\frac{2 \text{ mEq } Mg^{2+}}{\text{mmol drug}} \times \frac{1 \text{ mmol drug}}{246 \text{ mg drug}} \times \frac{1000 \text{ mg drug}}{\text{g drug}} \times \frac{50 \text{ g drug}}{100 \text{ mL}} \times 10 \text{ mL} = 40.7 \text{ mEq } Mg^{2+}$$

6. A vial of sodium chloride injection contains 3 mEq/mL. What is the percentage strength of this solution? The mol wt of NaCl is 58.5.

$$\frac{1 \text{ mEq}}{\text{mmol}} \text{ and } \frac{58.5 \text{ mg}}{\text{mmol}}$$

$$\frac{1 \text{ g}}{1000 \text{ mg}} \times \frac{58.5 \text{ mg}}{\text{mmol}} \times \frac{1 \text{ mmol}}{1 \text{ mEq}} \times \frac{3 \text{ mEq}}{\text{mL}} \times 100 \text{ mL} = 17.6 \text{ g}$$

$$\frac{17.6 \text{ g}}{100 \text{ mL}} = 17.6\%$$

PROBLEMS

- What is the mEq wt of ferrous ion (Fe^{2+}) which has an atomic weight of 55.85 g?
- What is the mEq wt of sodium phosphate ($Na_2HPO_4 \cdot 7H_2O$)?
- How many mEq of Na^+ are in 60 mL of a 5% solution of sodium saccharin (mol wt: 241; valence: 1)?
- How many mEq of Ca^{2+} are there in a 600 mg calcium lactate pentahydrate (mol wt: 308.30) tablet?
- How many mEq of sodium are there in a 5 gr sodium bicarbonate tablet? The mol wt of $NaHCO_3$ is 84 and the valence is 1.
- How many mEq of Na are there in 500 mL of $\frac{1}{2}$ normal saline solution? Normal saline solution contains 9 g NaCl/L; mol wt NaCl is 58.5.
- How much KCl is needed to make a pint of syrup that contains 10 mEq of K^+ in each tablespoonful? The mol wt of KCl is 74.5.

TEMPERATURE

RULES

The relationship of Celsius (C) and Fahrenheit (F) degrees is:

$$9(^{\circ}C) = 5(^{\circ}F) - 160$$

where $^{\circ}C$ is the number of degrees Celsius and $^{\circ}F$ is the number of degrees Fahrenheit.

EXAMPLES

- Convert $77^{\circ}F$ to $^{\circ}C$.

$$9(^{\circ}C) = 5(77) - 160$$

$$^{\circ}C = \frac{385 - 160}{9} = 25^{\circ}C$$

- Convert $10^{\circ}C$ to $^{\circ}F$.

$$9(10) = 5(^{\circ}F) - 160$$

$$^{\circ}F = \frac{90 + 160}{5} = 50^{\circ}F$$

PROBLEM

- Convert:
 - $30^{\circ}C$ to $^{\circ}F$
 - $100^{\circ}C$ to $^{\circ}F$
 - $37^{\circ}C$ to $^{\circ}F$
 - $120^{\circ}F$ to $^{\circ}C$

REFERENCES

- National Institute of Standards and Technology (NIST). *Specifications, Tolerances, and Other Technical Requirements for Weighing and Measuring Devices*. NBS Handbook 44. Washington, DC: US Department of Commerce, NBS, USGPO, 1989.
- Goldstein SW, Mattocks AM. *Professional Equilibrium and Compounding Accuracy* (pamphlet). Washington, DC: APhA, 1967. USP XXVI, 2003.
- Morrell CA, Ordway EM. The capacity and variability of teaspoons. *Drug Std* 1954; 22: 216.
- Madlon-Kay DF, Mosch FS. Liquid medication dosing errors. *J Family Pract* 2000; 49(8): 741-744.
- Shirkey HC. Dosage (posology). In: Shirkey HC, ed. *Pediatric Therapy*, 5th edn. St Louis: Mosby, 1975: 19.
- Benitz WE, Tatro DS. *The Pediatric Drug Handbook*, 3rd edn. St Louis: Mosby, 1995.
- Nelson JD. *Pocketbook of Pediatric Antimicrobial Therapy*, 4th edn. Dallas: Jodane, 1981.

ANSWERS TO PROBLEMS

DENSITY

- 816 g
- 363 mL
- 54.2 mL

ADDITION

- 2480 g or 2.48 kg
- 1160 g or 1.16 kg

SUBTRACTION

- 4100 mL or 4.11 L
- 3.71 g

MULTIPLICATION

- 157 mL
- 163 g
- 825 mL
- 375 mg

DIVISION

- 769 capsules 15 mg remainder
- 150 doses
- 1396 capsules 300 mg remainder

CONVERSIONS

- 422 mg
 - 19.4 mg
 - 109 g
 - 7780 mg
 - 99.4 g
 - 454 g

2. a. lb, 3 oz, 173 gr
b. 6.94 gr
c. 1 lb, 5 $\frac{3}{4}$, 5 $\frac{3}{4}$, 26 gr
d. 0.00154 gr
e. 2.2 lb
3. a. 0.648 mg
b. 0.203 mg
c. 10.8 mg
d. 0.325 or 0.324 g
e. 1.299 or 1.296 g
4. a. 12.3 mL
b. 11.1 mL
c. 237 mL
d. 473 mL
e. 0.309 mL
f. 0.00154 gr
g. 0.0772 gr
5. a. 480 gr
b. 8 $\frac{3}{4}$
c. 437 $\frac{1}{2}$ gr
d. 2880 gr
e. 4 $\frac{3}{4}$, 10 gr

DOSAGE CALCULATION

1. 1.5 mg
2. 32.7 mg
3. 18.7 mg
4. 280 000 units
5. 75 mg
6. 77.9 mg

PROBLEM-SOLVING METHODOLOGY

1. D.T.D. No. 14 means dispense 14 such doses. Assuming the doses have been checked, they are for chemicals J, K, and L (10 mg, 50 mg, and 300 mg, respectively)
2. Drug Q: 0.5 g, Drug R: 0.3 g
3. 0.469 mL/dose, 1.88 mL/day
4. 50 doses
5. 0.0444 mg
6. 0.75 mL contains 60 units: 13 $\frac{1}{2}$ days' supply
7. 6 mL
8. 3 mL
9. 25 mg

REDUCING AND ENLARGING

1. Liquid C 875 mL
Solid B 225 g
Liquid R 62.5 mL
Liquid P 500 mL
2. Solid G 7.46 g
Liquid D 224 g
Solid M 22.4 g
Base 492 g
3. Solid N 4.8 mg
Solid Q 120 mg
Solid R 7.2 g
Add sufficient purified water to make 240 mL solution.
4. Solid S 0.675 g
Solid T 2.25 g
Oil C 31.5 mL
Alcohol 22.5 mL

PERCENTAGE**Weight-in-volume percentages**

1. Dissolve 11.1 g in sufficient solvent to make 3 $\frac{3}{4}$.
2. 2.84 L
3. 6.25 L
4. 8 g

Weight-in-weight percentages

5. Compound A 7.46 g, Solvent 117 g
6. 28.8 g
7. 5 kg
8. Dissolve 4.64 g of drug in 120 mL (150 g) of glycerin.
9. Compound K 3.6 g, Solid H 1.8 g, Base 18.6 g
10. Dissolve 88.2 g of the solute in 500 mL of purified water. Dispense 500 mL.
11. Powder P 27.8 g, Powder Q 222 g, Powder R 333 g, Powder S 416 g
12. 3 g of coal tar solution, 27 g of hydrophilic ointment

Volume-in-volume percentages

1. 240 minim
2. 1900 mL
3. 34.8% w/w
4. 0.740 gr
5. 38.5% w/w
6. 15 g
7. 0.4 g
8. 32.4 mL of a 37% solution
9. 1.58 g

STOCK SOLUTIONS

1. 30.3 g
2. 36.3 g
3. 0.5 mL

PARTS PER MILLION

1. 13 mg

DILUTION AND CONCENTRATION

1. g of 10% cream and 21 g of diluent (base)
2. 0.01 g
3. 40 g

MIXING DIFFERENT STRENGTHS

1. 0.0428%
2. 64.5%
3. 3.16%

ALLIGATION ALTERNATE

1. 875 g of 12% ointment and 125 g of 16% ointment
2. 35 parts of 50% alcohol and 15 parts of purified water
3. 139 g of 28% ammonia water
4. 500 mL each of the 20% and 50% solutions are needed

PROOF SPIRIT

1. 0.375 proof gal
2. 0.035 proof gal
3. 1.94 gal

SATURATED SOLUTIONS

1. 1 g in 199 mL
2. 35.7 g of solute—dispense 500 mL

MILLIEQUIVALENTS

1. 27.9 mg/mEq
2. 134 mg/mEq
3. 12.5 mEq
4. 3.89 mEq
5. 3.86 mEq Na
6. 38.5 mEq Na
7. 23.5 g

TEMPERATURE

1. a. 86°F
b. 212°F
c. 98.6°F
d. 48.9°C

APPENDIX A: METRIC DOSES WITH APPROXIMATE APOTHECARY EQUIVALENTS

Liquid Measure			
Metric	Approximate apothecary equivalents	Metric	Approximate apothecary equivalents
1000 mL	1 quart	3 mL	45 minims
750 mL	1½ pints	2 mL	30 minims
500 mL	1 pint	1 mL	15 minims
250 mL	8 fluid ounces	0.75 mL	12 minims
200 mL	7 fluid ounces	0.6 mL	10 minims
100 mL	3½ fluid ounces	0.5 mL	8 minims
50 mL	1¾ fluid ounces	0.3 mL	5 minims
30 mL	1 fluid ounce	0.25 mL	4 minims
15 mL	4 fluid drams	0.2 mL	3 minims
10 mL	2½ fluid drams	0.1 mL	1½ minims
8 mL	2 fluid drams	0.06 mL	1 minims
5 mL	1¼ fluid drams	0.05 mL	¾ minims
4 mL	1 fluid dram	0.03 mL	½ minims

Weight			
Metric	Approximate apothecary equivalents	Metric	Approximate apothecary equivalents
30 g	1 ounce	20 mg	½ grain
15 g	4 drams	25 mg	3/8 grain
10 g	2½ drams	20 mg	1/3 grain
7.5 g	2 drams	15 mg	¼ grain
6 g	90 grains	12 mg	1/5 grain
5 g	75 grains	10 mg	1/6 grain
4 g	60 grains (1 dram)	8 mg	1/8 grain
3 g	45 grains	6 mg	1/10 grain
2 g	30 grains (½ dram)	5 mg	1/12 grain
1.5 g	22 grains	4 mg	1/15 grain
1 g	15 grains	3 mg	1/20 grain
750 mg	12 grains	2 mg	1/30 grain
600 mg	10 grains	1.5 mg	1/40 grain
500 mg	7½ grains	1.2 mg	1/50 grain
400 mg	6 grains	1 mg	1/60 grain
300 mg	5 grains	800 µg	1/80 grain
250 mg	4 grains	600 µg	1/100 grain
200 mg	3 grains	500 µg	1/120 grain
150 mg	2½ grains	400 µg	1/150 grain
125 mg	2 grains	300 µg	1/200 grain
100 mg	1½ grains	250 µg	1/250 grain
75 mg	1¼ grains	200 µg	1/300 grain
60 mg	1 grain	150 µg	1/400 grain
50 mg	¾ grain	130 µg	1/500 grain
40 mg	⅔ grain	100 µg	1/600 grain

Chapter 10

Statistics

Sanford Bolton, PhD and Richard Hirsch, PhD

INTRODUCTION 133

LOCATION PARAMETERS 134

DESIGN OF EXPERIMENTS AND COLLECTION OF DATA 135

DESIGN AND CONDUCT OF CLINICAL TRIALS 137

THE BINOMIAL AND NORMAL PROBABILITY DISTRIBUTIONS 138

ESTIMATION AND STATISTICAL INFERENCE 141

DATA TRANSFORMATIONS 173

INTRODUCTION

The laws of physics in connection with mathematical models and tools are quite capable to deterministically describe natural phenomena. But the more complex the situation to be described gets, i.e., the more influences need to be considered if we wanted to describe them accurately, the more we find our results to be erroneous. And in a scale where quantum effects also play a role, we have to deal with true uncertainty.

Statistics gives us a means to deal with probabilities or errors in measurements. With the help of statistics, we can (among others) describe data, calculate estimates of its distribution, and decide whether to reject a hypothesis in a rational and reproducible manner.

DATA

Data may come in discrete steps, as for instance, the number of patients that benefit from a certain therapy, or they may be continuous, meaning that the values can have infinitely many manifestations, even in a finite interval. An example for continuous data are, for instance, the masses of tablets.

In reality, however, all our measuring devices have a limited measuring accuracy, so in principle all experimental data comes in discrete quantities. This is important when statistical methods demand continuous data as a prerequisite, like the Mann-Whitney U test. We have to deal carefully with values that occur multiple times – statisticians call them “ties”. (Ties are not supposed to occur with continuous data because there will always be a difference, however small it may be. But in gathering real data, ties become real, too, because of the limited precision of the measurement.)

DATA VISUALISATION

In statistics we want to explore data and make inferences from it. An important first step is to visualise the data. The human brain has developed extraordinary capabilities for pattern recognition, and thus we can grasp important statistical parameters like location and spread of the data, or spot correlations, clusters, outliers, and so on, with a single glance.

Stem-and-Leaf Plot

A simple means to display data is the stem-and-leaf plot. It puts the data in order and provides visual information on the location, spread, and form of the data. It retains a least two significant digits of each data point.

A stem-and-leaf plot is constructed as follows:¹

1. Split each score or value into two sets of digits. The first or leading set of digits is the stem, and the second, or trailing, set is the leaf.
2. Draw a vertical line, and list all possible stem digits left to the line from lowest to highest.

3. For each data point write the leaf values on the line labeled by the appropriate stem number.

If appropriate, you can list each stem digit twice and put leaves starting with digits 0 to 4 on the first line, and leaves starting with digits 5 to 9 on the second.

Example 1. Example of a stem-and-leaf plot

For example consider the heights of the students of my statistics class: Their heights in centimetres are as follows: 195, 191, 198, 185, 158, 170, 160, 158, 172, 165, 185, 169, 187, 180, 178, 172, 180, 173, 168, 168, 172, 174, 160, 184, and 171.

The corresponding stem-and-leaf plot is presented in Table 10-1.

Table 10-1. Stem-and-Leaf Plot of Students' Heights

15	88
16	005889
17	01222348
18	004557
19	158

SAMPLES AND POPULATIONS

Many experiments have as an objective the definition or comparison of two or more groups of data. For example, one may wish to compare the efficacy of two antihypertensive agents or a new antipsychotic drug versus a placebo. Or it may be desired to estimate the average drug content and variability of a batch of tablets. In virtually all such experiments, it is not realistic to observe all possible experimental units. In fact, sometimes the entire population of conceivable observations cannot be identified completely. The potential experimental material for a clinical study comparing an antipsychotic drug to a placebo would include not only patients but also persons with the disease who are not yet diagnosed. All of these people are the population or universe. Clearly, one would not perform an experiment that included the entire population for many reasons:

- All of these people could not be identified.
- The time or money to conduct such a huge experiment is not available.
- To include so many people in such an experiment could be dangerous or unethical.

It is not necessary to run such a large experiment to arrive at a fair conclusion regarding the efficacy of the drug. In fact, in

most cases, the test consists of a relatively small *sample* taken from a relatively large *population*.

Another more concrete example is the process of sampling in quality control. It may be of interest to estimate the proportion of defective tablets or the average drug content and uniformity of tablets in a production batch. Certainly in the latter case every tablet in the batch would not be examined because the test is destructive, i.e., the tablet is destroyed during the analysis for drug content. Rather, a sample of 20 tablets would be chosen to estimate the average drug content of the more than 1 million tablets in the batch.

Thus, in typical experiments in the pharmaceutical sciences, a small sample from the population is examined in order to make inferences about the large population.

SUMMARY NUMBERS

In the next step we try to reduce the amount of data. We can try to specify a distribution by a mathematical model (e.g., normal distribution) and a finite set of parameters of that distribution. A sentence like, "The data follow a normal distribution with mean μ and variance σ^2 " contains more information than a thousand or more data points. This is because any normal distribution is completely determined by its mean and its variance (or its standard deviation, the square root of the variance).

LOCATION PARAMETERS

The single most important parameter for any distribution of data from ordinal or interval scales is its location parameter or central value.

Mode

For data from nominal scales, we can only name a mode, i.e., the value that occurs most frequently. A data set or distribution may have more than one mode; it is then called bimodal or multimodal.

Median

The median is a location parameter for data on ordinal or interval scales. At most, one half of the data are lower than the median, and at most, one half of the data is higher. If the data set consists of an odd number of observations (n say), the median equals the $(n+1)/2$ observation in the ordered data set. For an even number of observations, the median can be either stated as the average of the $(n/2)$ th and the $(n/2 + 1)$ th value (for data from an interval scale) or as the $(n/2)$ th or $(n/2 + 1)$ th values themselves, then referred to as "lower median" and "upper median."

For a complete population the median is identical to the 50% quantile.

Mean

The arithmetic mean can only be stated for data from an interval (or ratio) scale. The population mean is commonly denoted by the Greek letter μ , whereas the mean of a sample data set is referred to as $\bar{x}$. The sample mean $\bar{x}$ is an unbiased estimator for the population mean μ . The mean is defined as the sum of the observed values divided by the number of measurements:

$\bar{x} = \left(\sum_{i=1}^n x_i \right) / n$ which makes it the average of the observed values.

Spread Parameters

Other important information about a set of data is the variability or spread of the data.

RANGE

The difference between the largest and smallest value in a data set is known as the range.

VARIANCE

Another measure for the variability of the data that depends on all values is the standard deviation σ or its square, the variance

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

For independent random variables X and Y , the variances are additive: $Var(X + Y) = Var(X) + Var(Y)$. For instance, the content of a drug in the parts of a divisible tablet varies by the variance Var_{tbl} of the content of the whole tablet, plus the variance Var_{div} introduced by the breaking: $Var_{tot} = Var_{tbl} + Var_{div}$. The additive property of variances is the key to analysis of variance (ANOVA, see below).

Covariance

When considering data in more than one dimension, we are usually interested in the extend of the relationship of any two dimensions. For example, if we take the masses, compaction forces, hardness, and friability of a set of tablets, we might want to know whether compaction forces increase or decrease with the mass of the granules in the die, or how hardness or friability change with the compaction force. The answers might seem trivial in this example, but the statistical concept of covariance can give a quantitative answer to this question. The covariance is a generalisation of the variance. If we rewrite the formula for the calculation of the variance as:

$$Var(x) = \frac{\sum (x_i - \bar{x})(x_i - \bar{x})}{n - 1}$$

we find that the covariance is quite similar:

$$Cov(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

We have just replaced the second factor in the numerator by the difference of value and mean in another dimension. (Thus, if we calculate the covariance of one dimension with itself, we get the variance.) From the equation above it is also obvious that $Cov(x, y) = Cov(y, x)$.

The covariance $Cov(x, y)$ tells us whether the two variables x and y are correlated or not: If the covariance is positive, y increases as x increases; if it is negative, y gets smaller with increasing x ; if the covariance is zero, x and y are independent of each other.

FREQUENCY DISTRIBUTIONS

A *frequency distribution* of a data set can be constructed by counting the number of data points falling into a series of intervals (usually of equal size). The frequency distribution and its corresponding graph, a *histogram* or *bar chart*, show the distribution of the data, its central value (e.g., mean or median), and variability (e.g., SD or range). Example 2 shows the weights of 50 weanling rats to be used in an experiment.

Example 2. The weights in gram of 50 rats at weaning were as follows

30	47	37	29	38
32	42	32	30	34
34	32	33	37	36
39	33	45	40	35
43	41	35	32	41
36	27	28	35	30
38	28	41	37	34
41	36	32	30	37
31	31	35	28	25
26	49	34	34	33

Table 10-2 is a frequency distribution with 13 intervals derived from the data given in Example 2. A rule of thumb is to use 8 to 20 intervals, depending on the quantity and spread of the data. The histogram or bar chart of these data is shown in Figure 10-1.

BIAS, PRECISION, AND ACCURACY

Precision refers to the reproducibility of a series of measurements. If the values are very close to each other, the measurements are said to be precise. *Accuracy* refers to the closeness of measurements to the true value. For example, if a tablet contains exactly 200 mg of drug, and three analyses show a drug content of 205, 205, and 206 mg, it might be concluded that the analysis is precise, but not accurate. *Bias* refers to a systematic difference from the true value. Figure 10-2 illustrates these concepts.

The three assays observed above seem to be biased on the high side, i.e., errors in the assay procedure result in too-high values. Figure 10-2 shows that “precise” data need not be accurate. In fact, there need not be any relationship or correlation between the qualities of precision and accuracy. Note that biased data cannot be accurate but can be precise.

In addition to the concept of bias in the area of experimental measurements, it appears also in the field of experimental design. Bias can be introduced into an experiment, not because of an error in an experimental measurement, but because of poor judgment. For example, consider an experiment where the efficacy of oral and sublingual nitroglycerin are to be compared by administering both products to 20 patients on two different occasions and measuring the time to incidence of an angina attack in a treadmill test. Each of 20 patients will receive both the oral and buccal forms. If each patient receives the buccal drug on Monday and the oral drug on the following Sunday, a bias may be observed in the experimental results even if the measurements are not biased. This could be due to either the day of the week when the test was given (gloomy Monday versus a holiday weekend day) or an order effect where there is a

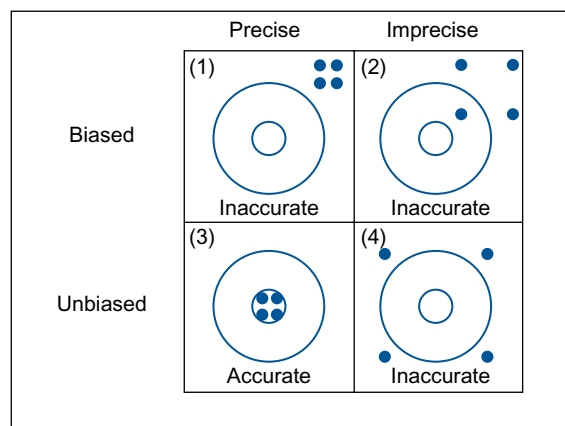


Figure 10-2. Diagram illustrating bias, precision, and accuracy. The shots on targets 1 and 2 are biased; in both cases the shots cluster away from the bull's-eye. The clusters on targets 3 and 4 both are unbiased; the center of each cluster is on the bull's-eye. The shots on targets 1 and 3 are precise; both sets are bunched together. The shots on targets 2 and 4 are scattered widely, hence imprecise. Only the shots on target 3 are accurate—precise and unbiased.

different effect, depending on which drug is given first. For example, there may be psychological factors causing the response to drug taken first to be systematically better (or worse) than that taken second, or the weather may be such as to cause more positive results on the first occasion. In the latter case, differences between the two dosage forms would be exaggerated (biased) in favor of the drug administered first, the buccal drug. To obviate this potential bias, we would give ten of the patients the oral drug first (Monday) and the buccal drug second (Sunday). The other ten patients would receive the products in opposite order. Perhaps, an improvement in this design would be to test the drugs on the same day of the week, e.g., Monday.

Table 10-2. Frequency Distribution of Rat Weights

Weight Group (g)	Frequency	Weight Group (g)	Frequency
24–25	1	38–39	3
26–27	2	40–41	5
28–29	4	42–43	2
30–31	6	44–45	1
32–33	8	46–47	1
34–35	9	48–49	1
36–37	7		

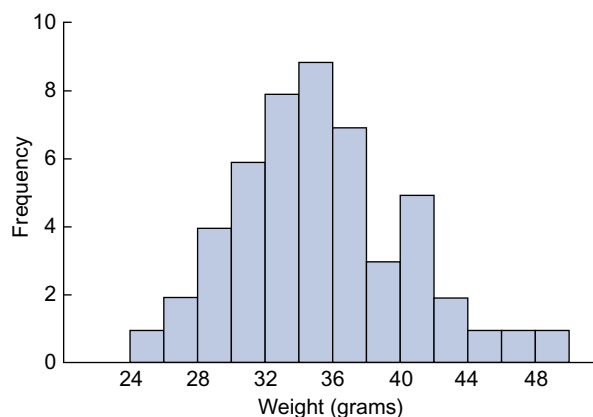


Figure 10-1. Bar chart showing frequency distribution of weights of 50 weanling rats (data in Example 2).

DESIGN OF EXPERIMENTS AND COLLECTION OF DATA

The application of statistics in the analysis of data is optimal when the data are collected in a planned or designed manner. If data are analyzed after the fact (retrospective analysis), great care should be taken to examine the data for possible bias. For example, prescription-volume data gathered for the years 1970 to 1980 may be available only from cities with populations greater than 500 000 or from cities in the Western States. Clearly, conclusions from such data should not be applied indiscriminately to the entire country. Also, the information may have been gathered on a voluntary basis; without knowledge of the characteristics of those who did and did not supply the information, the conclusions could be tainted.

The manner in which data are collected is connected to the planning and design of experiments. In the collection of data, a small sample generally is taken from a large population or universe. Sometimes a sample is taken inadvertently when the original intention was to obtain data from the population. For example, when a questionnaire is sent to every pharmacist in the state, there always will be some people who do not respond to the questionnaire, and anything less than 100% response constitutes a sample. A variety of examples of sampling methods is illustrated below.

SAMPLING BY QUESTIONNAIRE

Suppose that questionnaires on the sales of certain drugs were sent to all pharmacists in a state and only 50% were returned. In this type of survey, the results tabulated from such a sample

probably would be biased because those who did not return the questionnaire would not be represented in the sample.

It has been shown that persons who respond may have different characteristics from those who do not respond. In this hypothetical example, perhaps unanswered questionnaires were represented largely by pharmacists who had large drug sales and were too busy to answer. In another community a pharmacist may have little or no sales of the drugs, resulting in a nonresponse. The reason for each unanswered questionnaire is unknown. These unreturned questionnaires cause a bias, the direction and magnitude of which is unknown.

Other potential errors in this type of response that may introduce bias include the way in which the question is asked, the order in which questions are asked, and the psychological interaction between the interviewer and respondent. Questionnaire and survey techniques that can be employed to reduce or eliminate bias in the sample of responses have been proposed by mathematical statisticians.²

For example, public opinion polls use certain statistical sampling techniques that not only reduce bias but also optimize the information gathered. The Census Bureau has information about the percentages of men, women, and children in the US in various income and nationality groups, in addition to many other detailed categorizations. A sample may be designed to contain the same proportion of particular group(s) as that in the population. Instead of mailing questionnaires, interviewers may be recruited and assigned quotas of the types of people to interview. The interviewers fill out the questionnaires for each respondent during the interview, ensuring a complete response.

It is not possible to elaborate fully on the various methods of sampling here. One should be aware of problems in sampling and aware that a sampling design can be used that will give the limits of error of the resulting compilation for any given cost.²

SAMPLING IN THE CHEMICAL LABORATORY

The procedure for gathering data in the laboratory differs from that of the questionnaire. Different kinds of sampling processes include the sampling of material to be assayed chemically or physically, sampling of analytical reagents and instruments when multiple instruments are available, and sampling of analysts, i.e., the chemists who will perform the assay.

By way of illustration, several samples may be taken from a large lot of digitalis leaves for the chemical determination of acid-insoluble ash, or drug may be analyzed in samples taken from a blend. For the sample to be representative of the lot, the samples should be taken from different parts of the lot to ensure that every part of the lot is represented. Determinations from five samples taken from the same part of a lot (e.g., the top of a container) probably will have values closer together than five samples taken from different parts of the lot (e.g., the top, top-middle, middle, low-middle, and bottom of a container). Despite the good precision, the former five samples may give a biased estimate of the average value for the lot. The more heterogeneous the lot, the more effort should be expended in being sure that every part of the lot is represented by a sample. It might be that the granulation having the most drug is in the bottom of the lot; samples all taken from the top would give an estimate of average drug content that is considerably lower than the true value.

Another aspect of sampling in a chemical determination is the sampling of the chemists who perform the chemical analysis.

If a single chemist makes several determinations on portions taken from the same sample of thoroughly mixed material, one expects the results to be more precise than if several chemists made these determinations. Probably the true reproducibility of a method can be indicated only in terms of how closely an analyst at one laboratory can check an analyst at another laboratory on exactly the same material. Thus, due to slight differences in technique, one chemist always might obtain higher results than another chemist. Thus, the technique of chemists will have an effect on the results and the reproducibility of the method.

SAMPLING IN BIOLOGICAL AND CLINICAL EXPERIMENTS

A typical animal experiment might involve determining the temperature response of rabbits to pyrogens. The results of such an experiment constitute a sample of all possible results that could be obtained from the population of all possible rabbits, laboratories, and technicians. Using different rabbits, laboratories, and technicians will give different results, all contributing to the variability or error in the experiment. The differences between results from two or more laboratories are usually greater than differences between results obtained by two or more technicians in the same laboratory.

Concurrent conditions – such as season of the year, temperature, and humidity – sometimes can contribute to the experimental variability. In biological experiments differences between animals are relatively large, so experiments repeated in the same laboratory with different animals but under otherwise identical conditions will give different results. The use of statistical procedures gives an estimate of the amount of variation to be expected due to animal differences. The same can be said of clinical studies where more than one clinical site is needed to produce reliable, unbiased results.

Appropriate statistical designs and procedures will eliminate or account for potential bias in experiments. This point may be illustrated by an extreme example, an illustration of what not to do: A technician wishes to compare two drugs as to their effects on the growth of rats. Thirty rats from a single cage are used; the first 15 rats caught are put on Drug 1, and the last 15 caught are put on Drug 2. The first 15 rats caught are less lively than the last 15, and because they are less lively, they very likely differ in size and temperament from the last 15 rats. Thus, the results were biased from the very beginning, and one drug was favored merely because of the method of choosing the animals used for each drug.

Obviously, some method entirely free from subjective influences (unconscious or conscious) should be used. A table of random numbers³ or computer-generated random numbers commonly is used to assign animals or patients to treatments. Table 10-3 is a short table of random numbers.

In a biological assay, it often is possible to design a dosage schedule that takes advantage of the reduced within-animal variation compared to between-animal variation. Because more than one dose may be given to a single animal, the order of dosing also must be designed to account for possible trends in response to consecutive doses caused by changes in the animal with time or due to site of application.

Note that in a Latin square, each letter occurs only once in each row and each column of the square. A Latin square design was applied to an assay involving two levels of doses of the standard (high and low doses, S_H and S_L , respectively), and

Table 10-3. A Short Table of Random Numbers

39	61	09	51	68	81	26	30	52	20	61	41	25
89	35	48	61	72	10	84	34	10	44	72	94	77
37	98	37	56	40	30	70	31	75	03	68	32	15
20	55	68	05	53	73	60	28	96	48	91	81	18

Example 3. The use of the random number table

Suppose that 10 patients are to be assigned to two treatment groups, 5 in each group. Table 10-3 can be used to assign patients randomly to groups. Patients first are numbered from 1 to 10. One way of assigning treatments to patients is to read across Table 10-3, and the first five distinct numbers will be assigned to the first treatment. The remaining patients are assigned to the second treatment. A zero will correspond to patient number 10. The first five numbers are 3, 9, 6, 1, and 0. (Note that if a number repeats itself, we skip the number and proceed to the next one.) Therefore, patients numbered 3, 9, 6, 1, and 10 are assigned to the first treatment. If 100 patients are to be assigned to the two treatments, two-digit numbers would be used: reading across, the 50 patients assigned to the first group would be numbered 39, 61, 9, 51, and so on.

There are many ways of using random numbers to ensure randomness in statistical experiments. The number of ways is limited only by the ingenuity of the experimenter. For example, random assignment could be accomplished by assigning patients to Group 1 or 2 as they enter the study, according to the appearance of an odd or even number in the random-number table.

Example 4. The use of a Latin square design

This can be illustrated by an epinephrine assay (see *Remington's Practice of Pharmacy*, 14th ed., p 633), where a single dog is given 16 consecutive doses, the order of which is determined by a Latin square design, illustrated by:

A	D	B	C
D	C	A	B
B	A	C	D
C	B	D	A

two levels of doses of the unknown (U_H and U_L), where the four doses correspond to the letters, A, B, C, and D. The dosage schedule is given in Table 10-4. In this type of design, each dose occurs once in each order of administration (e.g., each of the four preparations are represented once in each group). In such an assay equal doses of epinephrine elicit a smaller and smaller rise in blood pressure with each succeeding dose. Therefore, order is important.

In all biological experimentation the design should be planned so that differences in treatment do not coincide with factors that could influence the outcome, such as differences in age, weight, sex, dates of administration, and so forth. This is known as *confounding* in statistical jargon. For example, if males are given a control treatment and females are given a comparative active treatment, the differences between treatments are said to be confounded by sex. That is, it cannot be determined if the outcomes observed are due to treatment, sex, or a combination of these factors.

Animals or patients should be assigned to doses or treatments at random, taking advantage of the availability of optimal experimental designs. Fisher⁴ has written an excellent book on planning or designing experiments, which explains fully the various types of designs mentioned here. Cochran and Cox⁵ detail useful experimental designs and provide complete directions for

Table 10-4. Typical Dosage Schedule for an Epinephrine Assay Using a Latin Square Design

	First Dose	Second Dose	Third Dose	Fourth Dose
First group	U_L	S_H	U_H	S_L
Second group	S_H	S_L	U_L	U_H
Third group	U_H	U_L	S_L	S_H
Fourth group	S_L	U_H	S_H	U_L

the analysis of data using these designs. Another book by Cox⁶ is less mathematically oriented and comprehended more easily.

DESIGN AND CONDUCT OF CLINICAL TRIALS

Proof of the efficacy and safety of new drugs or treatments requires testing in human subjects. This is best achieved by carrying out *controlled clinical trials*. The use of a placebo treatment or an established treatment as a *control*, a basis of comparison, usually is necessary. Thus, the effects of treatment with those of a concurrently tested control or placebo are compared. The trial includes an adequate number of patients to allow a reliable projection of the results to future patients. Theoretically, the results cannot be projected beyond the types of severity of disease or the ages and sex of the patients included in the trial, although in practice this is not always the case.

The distribution of variables such as age, sex, differences in diagnosis, and initial severity of disease among treatments may be controlled by stratification. Usually patients are assigned to treatments at random, and allowances are made for the effects of the variables by using suitable statistical methods. A restricted randomization procedure is useful if it is desired to ensure that about an equal number of patients enter the trial on each treatment. Table 10-5 illustrates a completely randomized design in which 15 patients are allocated at random, five to each of three treatments.

Note that the individual patients in each triad (the groups of three) are assigned randomly to one of the three treatments. Here the randomization is restricted in that each set of three patients, when entered, must be assigned to treatments A, B, and C. The patients are assigned to treatments as they enter the trial. The first patient (#1) gets treatment B. This scheme prevents runs in the randomization wherein a long consecutive number of patients are assigned to the same treatment. Another example is shown in Table 10-6 for a simple crossover design, in which the individual patients take both treatments consecutively and are assigned randomly to one of two treatment order groups.

The latter design may be more efficient than a completely randomized design because each patient acts as his or her own control, thus eliminating patient-to-patient variability in the statistical analysis. However, this advantage may be offset if drug carryover effects are present, or if the severity of the disease wanes in the second period to the point at which treatment differences no longer can be demonstrated.

To be certain that the random allocation is followed strictly, and to remove subjective bias on the part of both the patient

Table 10-5. Allocation of Patients in Randomized Design

A	B	C	A	B	C
3	1	2	12	10	11
6	5	4	14	13	15
8	9	7			

Table 10-6. Crossover Design

Group	Patient	Period 1	Period 2
I	1	A	B
	4		
	5		
	7		
	10		
II	2	B	A
	3		
	6		
	8		
	9		

and the clinical investigator in assessing the effects of the treatments, the clinical trial should be carried out blind. A *double-blind trial* is one in which neither the patient nor the investigator is made aware of the nature of treatment administered.

To ensure that the study remains blinded, all treatments must be packaged as identical-appearing dosage forms. This may require a great deal of ingenuity on the part of the packaging pharmacist, especially with respect to the taste of orally administered liquid products, the color and shape of tablets, and so on. In some cases, the characteristic side-effects of the drugs make it difficult to keep a study blind. In these situations, one must rely more heavily on objective measures of response and less on subjective measures. However, the *placebo effect* may also result in changes in so-called *objective* measure of response.

THE BINOMIAL AND NORMAL PROBABILITY DISTRIBUTIONS

Statistical conclusions are based on *probability*. The process of *statistical inference* first considers an assumption about the distribution of the population data. If the observed data from the sample collected do not conform reasonably to the assumed distribution, the results are viewed as *significant*, i.e., the sample data show significant differences from the assumed distribution. For example, it may be assumed or hypothesized that an antibiotic will cure 80% of the patients treated. If three of six patients are cured with the drug, this is the question: What is the probability that three or fewer of six patients treated will be cured if the probability of a single patient being cured is 80%? If this calculated probability is small, the probability of a cure is probably not 80% but rather some lesser value.

To compute these probabilities, the properties of the assumed probability distribution must be known. Two important and often-used distributions in statistical theory are the binomial and normal distributions, which are examples of a discrete and continuous probability distribution, respectively. The experiment discussed in the preceding paragraph, which related to the cure of patients treated with an antibiotic, is an example of one application of the binomial distribution.

THE BINOMIAL DISTRIBUTION

The binomial distribution is applicable to data from which one of two mutually exclusive and independent outcomes are possible as a result of a single observation or experimental trial. A patient may be cured or not cured. Only one of these two mutually exclusive events can occur at the time of observation. *Independence*, in this context, means that the probability of a cure for any given patient is 80%, regardless of the experimental outcome of the other patients in the study.

The problem to be solved is to compute the probability that three (or less) of six patients will be cured if the probability of a cure for an individual patient is 0.8, or 80%. The general solution to this problem uses the binomial distribution. If two

independent and mutually exclusive outcomes are possible as the result of an experimental trial, the probability $p(k)$ of exactly k outcomes of one kind (arbitrarily called *successes*) in n binomial trials (n patients in this example) is:

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

where p is the probability of success.

Exercise 1.

Calculate the probability of exactly three successes (cures) in six trials (patients) if $p=0.8$; i.e., the probability of a success or cure is 0.8.

$$\text{Answer : } p(x) = \binom{6}{3} \cdot 0.8^3 \cdot 0.2^3 = \frac{6 \cdot 5 \cdot 4}{3 \cdot 2 \cdot 1} \cdot 0.512 \cdot 0.008 = 0.082$$

Thus, the probability of exactly three cures in six patients is 0.082. This is interpreted to mean that the chance of observing exactly three successes in six binomial trials with $p=0.8$ is approximately 8 in 100.

There are seven possible outcomes for the treatment of six patients as shown in Table 10-7, an example of a binomial probability distribution defined by $n=6$ and $p=0.8$. It lists all the possible outcomes, with the probability of each outcome. The sum of all the probabilities is equal to 1. This distribution is shown graphically in Figure 10-3. A knowledge of this distribution allows a decision to be made as to whether three or fewer cures in six patients is a probable outcome for patients treated with a drug that has a cure rate of 80%. The probability of observing three or less successes (0, 1, 2, or 3 successes) is $p(0) + p(1) + p(2) + p(3) = 0.0988$, or about 1/10. Is this sufficient evidence to say that the true probability of a cure for the drug is less than 0.8? This question will be discussed in more detail in the section on Statistical Inference.

Table 10-7. Binomial Distribution for $n=6$ and $p=0.8$

Number of Successes	Probability of Outcome	Number of Successes	Probability of Outcome
0	0.0000026	4	0.24576
1	0.001536	5	0.39322
2	0.01536	6	0.262144
3	0.08192		

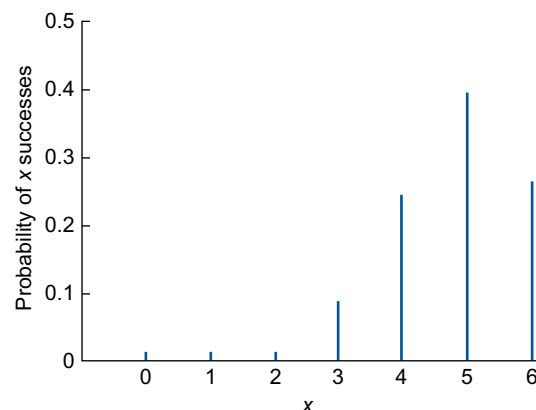
**Figure 10-3.** Binomial distribution for $n = 6$ and $p = 0.8$.

Table 10-8 lists individual probabilities for $p=0.2$, 0.5 , and 0.8 , for n equal to 6 to 10, inclusive.

Exercise 2.

Calculate the probability of four successes in six trials for $p=0.8$.

$$\text{Answer : } p(4) = \binom{6}{4} \cdot 0.8^4 \cdot 0.2^2 = 0.246$$

The mean of the binomial distribution is np . Thus, for the binomial distribution with $p=0.8$ and $n=100$, the mean is $np=80$. That is, if 100 patients were treated with the antibiotic that has a cure rate of 80%, one could expect to see 80% or 80 patients cured of 100 treated on the average. The standard deviation of a binomial distribution is: $\sqrt{np(1-p)}$. The standard deviation of the number of patients cured of 100 treated in the above example is:

$$\sqrt{np(1-p)} = \sqrt{100 \cdot 0.2 \cdot 0.8} = 4$$

This can be interpreted as follows. If 100 patients are treated, it may be expected that 80 are cured on the average, but in any given experiment one probably would not see exactly 80 cured. The number cured will vary around 80, the mean, with a standard deviation equal to 4.

THE NORMAL DISTRIBUTION

The normal distribution can be considered as the underlying foundation of statistical theory and its applications. It is a continuous probability distribution ranging from $-\infty$ to ∞ . A normal

distribution is defined by its mean μ and standard deviation σ . The mean can be any positive or negative value, but the standard deviation is always positive. So there are infinitely many normal distributions. The normal distribution with zero mean and unit standard deviation is called the standard normal distribution. Figure 10-4 shows the graphs of two normal probability curves. The normal distribution is characterized by the symmetry about its mean; most of the data cluster around the mean (68 of the data lies in the interval from $\mu-\sigma$ to $\mu+\sigma$). There are fewer values as the deviation is farther from the mean. The normal distribution is a theoretical probability distribution, not exactly observed in practical situations. However, much data approximate the normal distribution closely enough to make its application useful.

The Central Limit Theorem (CLT) is perhaps the most powerful theorem in statistics. It supports the pervasive use and importance of the normal distribution in statistical analyses. In simple terms, the CLT states that averages or means approach normality as n , the sample size, increases, no matter what the distribution of the individual variables. For data that are close to normal, means from even a small sample size will be approximately normal. For data that have distributions far from normal, larger sample sizes will be needed for the averages to be close to normal. The concept of the CLT is illustrated by the following example⁷

The outcome o of a disease after treatment can be (1) death $o=1$, (2) not cured but continue treatment $o=2$, and (3) cured $o=3$. The probabilities of these three outcomes are 0.1, 0.3, and 0.6, respectively. This distribution is shown in Figure 10-5. This is a discrete distribution (three possible outcomes in a single trial), and it clearly is not normal. Figure 10-6 shows the distribution of means of size 20 ($n=20$). The means are obtained by treating 20 patients, assigning outcomes of 1, 2, or 3, according to the previous definition, and computing the mean. The distribution shown in Figure 10-6 was constructed from a computer simulation representing outcomes that can be expected in

Table 10-8. Short Table of Binomial Probabilities

$p = 0.2$											
Probability of x successes in n trials											
x	0	1	2	3	4	5	6	7	8	9	10
n											
6	0.262	0.393	0.246	0.082	0.015	0.002					
7	0.210	0.367	0.275	0.115	0.029	0.004					
8	0.168	0.336	0.294	0.147	0.046	0.009	0.001				
9	0.134	0.302	0.302	0.176	0.066	0.017	0.003				
10	0.107	0.268	0.302	0.201	0.088	0.026	0.006	0.001			
$p = 0.5$											
Probability of x successes in n trials											
x	0	1	2	3	4	5	6	7	8	9	10
n											
6	0.016	0.094	0.234	0.313	0.234	0.094	0.016				
7	0.008	0.055	0.164	0.273	0.273	0.164	0.055	0.008			
8	0.004	0.031	0.109	0.219	0.273	0.219	0.109	0.031	0.004		
9	0.002	0.018	0.070	0.164	0.246	0.246	0.164	0.070	0.018	0.002	
10	0.001	0.010	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.010	0.001
$p = 0.8$											
Probability of x successes in n trials											
x	0	1	2	3	4	5	6	7	8	9	10
n											
6		0.002	0.015	0.082	0.246	0.393	0.262				
7			0.004	0.029	0.115	0.275	0.367	0.210			
8			0.001	0.009	0.046	0.147	0.294	0.336	0.168		
9				0.003	0.017	0.066	0.176	0.302	0.302	0.134	
10				0.001	0.006	0.026	0.088	0.201	0.302	0.268	0.107

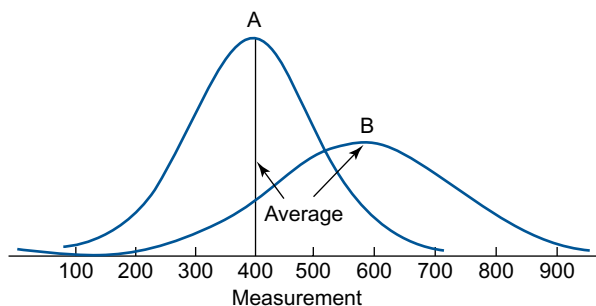


Figure 10-4. Normal probability curves.

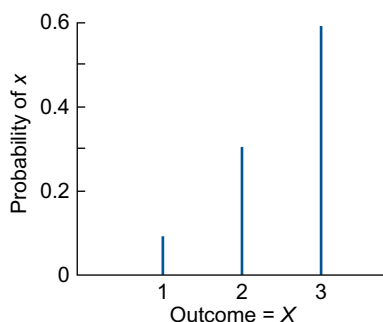


Figure 10-5. Probability distribution of outcomes after drug treatments.

realistic situations. Note that the averages cluster almost symmetrically around 2.5 (the mean) and are beginning to look like a normal distribution. One also should note the small variability of the mean results, most values ranging between approximately 2.2 to 2.7. A single outcome varies from 1 to 3.

The CLT allows the use of statistical methods that assume an underlying normal distribution of the data when dealing with averages of data that do not come from a normal distribution.

COMPUTING PROBABILITIES FROM THE NORMAL DISTRIBUTION

Area represents probability, so the area under the whole normal curve is 1. The probability of observing a single value from a continuous distribution such as the normal is 0. However, one can calculate the probability of observing values in any interval $[x_1, x_2]$ —designated as $P(x_1 \leq x \leq x_2)$ —by computing the area under the curve in that interval. Table 10-9 is a short compilation of cumulative probabilities from the standard normal curve (Fig 10-7) that has a mean of $\mu = 0$ and a standard deviation of $\sigma = 1$. Table 10-9 shows the probability of observing a value less than or equal to Z . For example, the probability of observing a value less than or equal to -1 from the standard normal distribution is 0.16. The symmetry of the normal curve indicates that the probability of a value being greater than or equal to $+1$ is also 0.16. Since the total area under the normal curve is 1 and area represents probability, the area less than or equal to $Z = +1$ is $1 - 0.16 = 0.84$. This relationship is illustrated in Figure 10-7. In general, to calculate the area in any interval, Z_1, Z_2 , look up the cumulative areas corresponding to Z_1 and Z_2 . The difference of the two areas is the area between Z_1 and Z_2 , or the probability of observing a value in that interval.

Exercise 3.

Calculate the probability of a value falling between -1.96 and $+1.28$ for the standard normal curve.

Answer: The area for values less than $Z = -1.96$ is 0.025. The area corresponding to values less than $Z = 1.28$ is 0.90. The difference is 0.875. Thus, the probability of observing a value between -1.96 and $+1.28$ is 0.875.

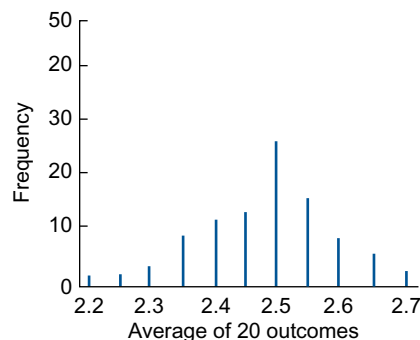


Figure 10-6. Simulation of distribution with average of sample size of 20.

Table 10-9. Short Table of Areas for the Standard Normal Distribution (Area for Values Less Than z)

z	Area	z	Area	z	Area
-3	0.0013	-1.28	0.1003	1.2	0.8849
-2.58	0.0049	-1.2	0.1151	1.28	0.8997
-2.3	0.0107	-1.0	0.1587	1.4	0.9192
-2.2	0.0139	-0.84	0.2005	1.5	0.9332
-2.1	0.0179	-0.8	0.2119	1.6	0.9452
-2.0	0.0228	-0.6	0.2743	1.645	0.950
-1.96	0.025	-0.4	0.3446	1.8	0.9641
-1.9	0.0287	-0.2	0.4207	1.9	0.9713
-1.8	0.0359	0.0	0.500	1.96	0.975
-1.7	0.0446	0.2	0.5793	2.00	0.9772
-1.645	0.050	0.4	0.6554	2.1	0.9821
-1.6	0.0548	0.6	0.7257	2.2	0.9861
-1.5	0.0668	0.8	0.7881	2.3	0.9893
-1.4	0.808	1.0	0.8413	2.58	0.9951
				3.00	0.9987

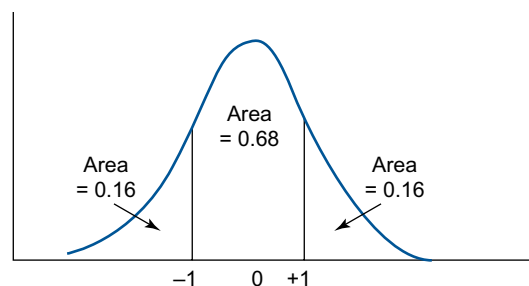


Figure 10-7. Standard normal distribution.

Since there are an infinite number of normal distributions (defined by their means and standard deviations) a reasonable question is, How would one calculate probabilities from a normal distribution that is different from the standard normal distribution? Fortunately, there is a simple transformation called z -transformation that converts data from any normal distribution into the standard normal distribution; Table 10-9 can then be used to compute the probabilities.

On order to do the Z transformation for any x from a normal distribution with mean μ and standard deviation σ , just subtract μ from x and divide the result by σ :

$$z = \frac{x - \mu}{\sigma}$$

The resulting z value follows the standard normal distribution.

Example 5.

What is the probability that a tablet will weigh between 185 and 210mg if tablet weights have an approximately normal distribution with mean 200mg and a standard deviation of 10mg?

Using the Z transformation, we obtain:

$$z_l = \frac{185\text{mg} - 200\text{mg}}{10\text{mg}} = -1.5$$

$$z_u = \frac{210\text{mg} - 200\text{mg}}{10\text{mg}} = 1.0$$

The probabilities of finding values less than $z_l = -1.5$ and $z_u = 1.0$ in the standard normal distribution are found from Table 10-9. They are 0.07 and 0.84, respectively. Therefore, the probability of finding a tablet weighing between 185 and 210mg is $0.84 - 0.07 = 0.77$. Note carefully that the transformation is equivalent to finding a value that is between -1.5 and $+1.0$ standard deviations from the mean (i.e., 185mg is -15 mg from the mean, which is equal to -1.5 standard deviation units).

What is the probability that a tablet will weigh less than 180.4 or more than 219.6mg? The Z transformation results in the values -1.96 and $+1.96$. It is an important property of any normal distribution that 95% of the values are found in this interval.

Several Z values appear frequently when testing for statistical significance:

- 68% of the values are within ± 1 standard deviation of the mean value.
- 80% of the values are within ± 1.28 standard deviations of the mean value.
- 90% of the values are within ± 1.65 standard deviations of the mean value.
- 95% of the values are within ± 1.96 standard deviations of the mean value.
- 99% of the values are within ± 2.58 standard deviations of the mean value.

NORMAL APPROXIMATION TO THE BINOMIAL DISTRIBUTION

The CLT can be applied to the binomial distribution if n , the number of binomial trials, is sufficiently large. As a general rule, if both np_0 and $n(1-p_0)$ are equal to or greater than 5 (p_0 being the true probability of success), the normal approximation can be used. For binomial distributions with p_0 close to 0.5, the approximation is good for values of np_0 and $n(1-p_0)$ smaller than 5. Under these conditions, $(p - p_0)/\sqrt{p_0(1-p_0)/n}$ is approximately normally distributed with mean 0 and standard deviation 1 (the standard normal distribution). This transformation allows easy calculation of binomial probabilities. The approximation is improved if $1/(2n)$ is subtracted from the absolute value of the numerator. This is known as the *Yates continuity correction*.

ESTIMATION AND STATISTICAL INFERENCE

ESTIMATION AND CONFIDENCE INTERVALS

After gathering data from, for example, a survey or an experiment, it is often of interest to estimate the mean value or average of the population. As has been noted, the sample average, $\bar{x}$, is

Example 6.

When inspecting 100 tablets for quality, what is the probability of observing a proportion of defective tablets equal to or greater than 0.10, if the true proportion defective is 0.07? Using the continuity correction

$$z = \frac{\mu - 2n}{\sigma} = \frac{|0.10 - 0.07| - 1/200}{\sqrt{0.07 \cdot 0.93/100}} = 0.98$$

From Table 10-9, the probability of a value less than $z = 0.98$ is approximately 0.84. The probability of a value greater than $z = 0.98$ is $1 - 0.84 = 0.16$. Therefore, the probability of observing a proportion greater than 0.10, when inspecting 100 tablets from this batch, is approximately 0.16.

not exactly equal to the population average, μ , but in a well-designed and implemented experiment, $\bar{x}$ should be an unbiased estimate of the true mean. Thus, the best estimate of the true, but unknown, population average is the sample mean $\bar{x}$.

However, the mean of the sample gives no idea of the precision of this average. If the average assay of 10 tablets is 100mg, it is not known how close this value is to the unknown true value. It would be important to have some estimate of the reliability of the result. *Confidence intervals*, or confidence limits, give an interval that may encompass the mean with a known probability. That is, a 95% confidence interval of 97 to 103mg means that one would give 19 to 1 odds that the true mean is in this interval. It cannot be said for certain that the true mean is in the interval, but if the experiment were repeated many times and a 95% confidence interval constructed each time, then 19 of 20 such intervals would contain the true mean. For any given experiment, there is no way to tell if the true mean is in the interval, but it is known that the chances are 95% that the true mean is in the interval.

In statistical inference, no statements can be made with assurity. Statistical proofs are not like mathematical proofs. Statistical conclusions are couched in terms of probability. A statement such as "The means are significantly different," means that it is believed that the means are different but there is a chance, albeit a small one, that the conclusion is incorrect. However, the probability of making the wrong decision is known.

Symmetric confidence limits are computed as:

$$\bar{x} \pm z \cdot \sigma_{\bar{x}}$$

where z is an appropriate constant, depending on the probability statement (degree of confidence) associated with the confidence interval. For a normally distributed variable with standard deviation known, the value of z is obtained from Table 10-9. For example, to obtain a 95% confidence interval, ± 1.96 standard deviations covers 95% of the area. For a 90% confidence interval, $z = 1.65$; for a 99% confidence interval, $z = 2.58$. Note that if the standard deviation is unknown but estimated from the sample data, the value of z for normally distributed variables is replaced by t , obtained from the t distribution, which will be introduced in the next section.

On rare occasions, a *one-sided* or an *unsymmetrical* confidence interval may be appropriate. One use of a one-sided interval is described under linear regression as applied to stability prediction.

STATISTICAL INFERENCE AND THE T DISTRIBUTION

Statistics are used most often as a decision-making tool. The familiar phrase "the difference is statistically significant" results from the application of statistical inference to experimental data. The procedure allows a probability statement to be made

Example 7.

A drug shows an average blood pressure reduction of 9.8 torr when tested on 100 patients. The standard deviation is known to be 8 torr. A 95% confidence interval for the mean blood pressure reduction in torr is:

$$9.8 \pm 1.96 \cdot \frac{8}{\sqrt{100}} = 9.8 \pm 1.57$$

A 99% confidence interval is:

$$9.8 \pm 2.58 \cdot \frac{8}{\sqrt{100}} = 9.8 \pm 2.06$$

Note that if the interval has a higher probability of containing the true mean, the confidence interval is wider.

Example 8.

A survey of 1000 pharmacists showed that 30% have more than 15 years of experience and 70% have less than 15 years of experience. A 95% confidence interval on the proportion of pharmacists with more than 15 years of experience is:

$$p \pm 1.96 \cdot \sqrt{p(1-p)/n} = 0.3 + 1.96 \cdot \sqrt{0.3 \cdot 0.7/1000} = 0.3 \pm 0.028$$

This means that the true proportion is between 0.272 and 0.328 with 95% probability.

about comparative data. Statements made using this approach cannot be made with absolute certainty. Because experimental results generally come from sample data, one never can be sure of the exact properties of population data. However, decisions can be made with a known probability of error.

t Test

In addition to estimating the mean assay of a batch of 10 tablets, the 10 assay values were obtained to perform a statistical test comparing the average result to that expected based on the labeled potency of 100 mg. If every one of the 3 000 000 tablets in this batch were assayed, the average potency would be known. The random sample of 10 is representative of the entire batch, but it is extremely unlikely that the sample average exactly will equal the batch average. The question to be asked is, in view of the variability of the 10 assays and the average result, can it be ascertained that these 10 tablets came from a population with an average of 100 mg? The solution to this question, an example of statistical inference, is obtained using a simple *t* test. This *t* test consists of the following steps, which can be considered typical in many designed experiments.

Construct a Null Hypothesis

A null hypothesis is an assumption about the parameter under investigation, which is the mean value in this example. The null hypothesis is a statement that assumes that the parameter is equal to some value, usually a null value. That is, the hypothetical value is considered to represent a situation of no change. How to construct the null hypothesis is not always obvious, but a few examples should make this concept clearer.

For the tablet assays, no change means that the population average, μ , is equal to the labeled potency, 100 mg. The null hypothesis is of the following form:

$$H_0: \mu = 100 \text{ mg}$$

The statistical test allows a decision to be made: the sample of tablets are or are not representative of a population with mean 100 mg.

Construct an Alternative Hypothesis

An alternative hypothesis makes an assumption about alternative values of the parameter, usually encompassing complementary values. Thus, if H_0 is $\mu = 100 \text{ mg}$, an alternative could include all values greater than or less than 100 mg. This is a two-sided alternative represented as $H_a: \mu \neq 100 \text{ mg}$. In some cases, a one-sided alternative may be suitable. This may be expressed as

$$i) H_a: \mu > 100 \text{ mg} \text{ or } H_a: \mu < 100 \text{ mg}$$

The process of statistical inference will result in one of two possible decisions: either accept or reject the null hypothesis. *Rejection* means the alternative is accepted. For a two-sided alternative, it is anticipated in advance that if the null hypothesis is not true, that the true average could be either greater or smaller than the hypothetical or assumed value. A one-sided alternative is viable if the alternative only can take on either a lower or higher value than the hypothetical value or if only higher (or lower) values are of interest.

It is not clear always which alternative (one- or two-sided) is correct or appropriate for any given situation. Usually, two-sided alternatives must be considered because, in most situations, smaller and larger values of the parameter are possible and relevant. Some situations where one-sided alternatives may be best will be discussed.

Choose the Level of Significance

The level of significance also is known as the *alpha level* (α) or *error of the first kind*. This is the basis of the well-known statement (e.g., the difference is significant at the 5% level). The α error is set in advance and has the following meaning. The level of significance or α error is the probability of erroneously stating that the difference between the observed value of the parameter (the mean in this example) and the hypothetical value is real or significant.

The α error commonly is chosen as 5%, although this is not obligatory. A more conservative approach would be to choose a level of 1%. This would mean that an error of the first kind, i.e., erroneously declaring a difference, is only 1%. It will be seen that a larger difference is needed for significance if the α error is made smaller. That is, it is more difficult to find a significant difference.

Beta Error and Power

Usually, only the α error is chosen in advance of the experiment. However, it should be understood that there is a second kind of error that should be considered when making statistical decisions. This error, the *beta error* (β), is the probability of declaring no difference between the observed sample value and hypothesized value of the parameter when, in fact, a difference of size delta (δ) exists. The α level, β error, and sample size are related. Sample-size determination, an important topic, is discussed in most elementary statistics books.^{1,8} When it is declared that differences are (or are not) significant, only the α level is considered, and not the β error.

Choose a Sample

The choice of a proper sample and the size of the sample are very important considerations in statistical experimentation and experimental design. The number of objects to be included in the sample is a consequence of the α and β errors. The manner in which samples are chosen will dictate the statistical analysis.

Example 9.

Consider the estimation of the tablet potency of a batch of tablets based on an assay of 10 individual randomly selected tablets. The assay values in mg are:

98.6	99.3	97.9	100.3	99.6
98.0	100.1	97.5	98.4	99.1

The average is 98.88 mg, and the standard deviation is 0.954 mg. In this example an estimate of the mean and standard deviation is obtained from a sample of size 10.

When the standard deviation is unknown but an estimate is available from a relatively small sample, the t distribution is used to describe the distribution of the means. The t distribution may be defined as the distribution of:

$$t = \frac{\bar{x} - \mu}{SD/\sqrt{n}}$$

The t values show a symmetrical distribution centered at 0; i.e., the mean is 0. The t distribution is spread out more than the standard normal distribution. Some commonly used points from the t distribution are shown in Table 10-10. The t distribution is defined by degrees of freedom (DF), which in Example 9 is $n - 1$. Note that when the DF are large (i.e., n is very large), the values in the t table approach the corresponding values from the standard normal-curve table (see Table 10-9). For example, the value below which 97.5% of the area is found is 1.96 when the DF are infinite in the t table.

When the SD is unknown, values from the t table are used to construct confidence intervals in exactly the same manner as was done using Table 10-9. Table 10-10 shows t values that cut off areas of the t distribution in one tail or symmetrically in both tails of the distribution. For example, the two-tailed 5% points cut off 2.5% of the area in each tail. For $DF = 9$, t values below -2.262 and greater than $+2.262$ comprise 5% of the area. Conversely, it can be said that the probability of finding a t value between -2.262 and $+2.262$ is 95% for $DF = 9$.

Table 10-10 gives values for one-tailed probabilities for $p = 0.5\%$ and 2.5% . These values correspond to the two-tailed probabilities of 0.01 (1%) and 0.05 (5%). For example, for 9 DF, the probability of finding a t value greater than $+2.262$ (or smaller than -2.262) is 2.5%. In the current example of tablet assays, a 95% confidence interval can be constructed using the t distribution. The mean is 98.88 mg and the sample SD is 0.954 mg. The t value for 95% of the area for 9 DF is 2.262. The 95% confidence interval in mg is:

$$98.88 \pm 2.262 \cdot \frac{0.954}{\sqrt{10}} = 98.88 \pm 0.68 = 98.20 \text{ to } 99.56$$

This can be interpreted to mean that the probability is 95% that the true mean of the batch lies between 98.20 and 99.56 mg.

Are you surprised by the narrow limits of the interval based on only 10 tablets? The reason for the tight limits is the small standard deviation. Note that this does not guarantee that the true mean, μ , lies in this interval; the probability that it does so is just 0.95. As has been emphasized before, statistical statements and conclusions are probabilistic in nature.

Table 10-10. The Distribution of t Giving Both the Two-Sided or Two-Tailed Probability and the One-Sided or One-Tailed Probability According to Degrees of Freedom

DF	One Tail							
	P = 0A	P = 0.3	P = 0.2	P = 0.1	P = 0.05	P = 0.025	P = 0.01	P = 0.005
	Two Tails							
	P = 0.8	P = 0.6	P = 0.4	P = 0.2	P = 0.1	P = 0.05	P = 0.02	P = 0.01
1	0325	0.727	1.376	3.078	6.314	12.706	31.821	63.657
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106
12	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055

(continued)

Table 10-10. (Continued)

DF	One Tail							
	P = 0A	P = 0.3	P = 0.2	P = 0.1	P = 0.05	P = 0.025	P = 0.01	P = 0.005
	Two Tails							
	P = 0.8	P = 0.6	P = 0.4	P = 0.2	P = 0.1	P = 0.05	P = 0.02	P = 0.01
13	0.259	0.538	0.870	1.350	1.771	2.160	2.650	3.012
14	0.258	0.537	0.868	1.345	1.761	2.145	2.624	2.977
15	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947
16	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.534	0.863	1.333	1.740	2.110	2.567	2.898
18	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819
23	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.531	0.857	1.318	1.711	2.064	2.492	2.797
25	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787
26	0.256	0.531	0.856	1.315	1.706	2.056	2.479	2.779
27	0.256	0.531	0.855	1.314	1.703	2.052	2.473	2.771
28	0.256	0.530	0.855	1.313	1.701	2.048	2.467	2.763
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462	2.756
30	0.256	0.530	0.854	1.310	1.697	2.042	2.457	2.750
40	0.255	0.529	0.851	1.303	1.684	2.021	2.423	2.704
60	0.254	0.527	0.848	1.296	1.671	2.000	2.390	2.660
120	0.254	0.526	0.845	1.289	1.658	1.980	2.358	2.617
∞	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576

In this simple example, the choice of experimental units (tablets to be analyzed) appears to be uncomplicated. However, further thought reveals many alternatives. Ten tablets are to be chosen from 3 000 000. Some possible sampling schemes include (1) take the first 10 tablets from the batch, (2) take the last 10 tablets, (3) take tablets at regular intervals during the run and select 10 of these tablets at random, or (4) take 10 tablets at random from the entire batch. A random sample is one in which each object has an equal probability of being chosen (see under Sampling for more detail). Random samples will assure a valid statistical analysis.

Random sampling can be visualized as a kind of lottery device, in which all of the tablets are mixed and one selected. In many cases, random sampling is not convenient, or the design can be improved by using variations of random sampling schemes. Although the statistical analysis in the present example assumes a random sample, it would not be convenient to implement this procedure for a batch of 3 000 000 tablets. Scheme 3, above, is a more realistic sampling scheme; although it is not truly random, one can proceed as though it were random for this example. In this case, a sample size of 10 tablets is chosen, not for statistical reasons, but because this number has been written into the quality-control procedure. A better procedure would be to base the number of samples on the α and β levels.⁸

Determine Whether the Test Should be One- or Two-Sided

In this example, a two-sided test is chosen because the observed average potency could be either lower or higher than the hypothetical value of 100 mg. That is, there is no reason to

believe, based on the manufacturing process, that the observed value should deviate on one side rather than the other of the labeled potency.

Make Observations and Construct a t Test

Having gathered the tablets and performed the assays, the value of t is computed. This allows one to make the decision, significant or not significant. For a two-sided test, the t value is computed as:

$$t_{n-1} = \frac{|\bar{x} - \mu|}{SD/\sqrt{n}}$$

where μ is the hypothetical mean defined by the null hypothesis. In this example, t is:

$$t_{n-1} = \frac{|\bar{x} - \mu|}{SD/\sqrt{n}} = \frac{|98.88 - 100|}{0.954/\sqrt{10}} \gg 3.71$$

The t value then is compared to the t values in Table 10-10 at the specified α level with $n-1$ DF. For a two-sided test, the absolute value of t is noted, because either small or large values of the difference ($\bar{x} - \mu$) will lead to significance. If the observed value of t is equal to or greater than the value in the table, the difference between the observed and hypothetical values of the parameter, the mean in this example, is declared to be statistically significant. The value of t for a two-sided test at the 5% level for 9 DF is 2.262, the same value used for the 95% confidence interval. This is no coincidence, as will be shown below. Since the observed absolute value of t (3.71) is larger than the value in Table 10-10 (2.262), significance is declared. The true

potency is apt not to be 100 mg, but rather some lower value, based on the observed value of 98.88 mg.

An examination of the equation for t reveals that large differences between the observed and hypothetical mean coupled with a small SD of the mean lead to large values of t . This makes sense, from an intuitive point of view, as large differences with small variability suggest that the difference is real. Also, one should note that had the test showed a non-significant difference, it cannot be said with any assurance that the mean of the batch is 100 mg. In fact, it seems extremely unlikely that this should be true. The data simply do not provide sufficient evidence to show that the mean is different from 100 mg. In this case, the confidence interval would give a region in which the mean probably lies.

If the above test had been performed at the 1% level, it still might have been concluded that the mean of the batch was not 100 mg. The value of t at the 1% level with 9 DF from Table 10-10 is 3.25. Because 3.71 is greater than 3.25, one would declare significance. A test significant at the 1% level gives greater assurance that the true mean differs from 100 mg, compared to a test significant at the 5% level.

There is a relationship between the two-sided t test and the confidence interval. For example, if the 95% confidence interval does not cover the hypothetical value defined by H_0 , the test will show significance and vice versa. This suggests that the true mean is different from the hypothetical mean. In the example discussed above, the 95% confidence interval was calculated as 98.2 to 99.56, which does not cover the hypothetical value 100. Therefore, it may be concluded correctly that the t test will show a significant result at the 5% level. Had the confidence interval in t included 100, the t test would not be significant.

The example described above is known as a one-sample t test. In this test, the experimental design consists of determining the mean value of a random sample from a single population and comparing the mean to some hypothetical value. Thus, it may be of interest to compare the mean tail-flick value of an analgesic compound in rats to some value that represents activity based on previous experience, or to compare the average assay result of 10 tablets to the labeled amount or to a previously accumulated average, as may be available from quality-control records.

TWO INDEPENDENT SAMPLE T TEST

A common design in research involves the comparison of two treatments applied to two independent groups. For example, in a clinical study, a drug is compared to a placebo using 20 patients for the drug treatment and 20 different patients for the placebo treatment. Or the dissolution of tablets prepared by a marketed formulation is compared to the dissolution of tablets prepared from an experimental formulation. Note that this design differs from the one-sample test in that averages are obtained from two groups for purposes of comparison, whereas in the one-sample test, the average of a single group is compared to some hypothetical value.

Three key assumptions are necessary for the two independent sample t test to be valid: (1) each of the two groups are distributed normally, (2) each of the two groups are distributed with the same variance, and (3) the two samples are independent.

The independence assumption is very important. Independent samples mean that the results for any single individual do not influence the results for any other individual. In the case of a clinical trial, independence would mean that the treatment effect for one patient does not influence the result of a treatment for other patients. If one patient discussed the results of his or her medication with another patient in the study, their results would not be independent. If treatments are applied to more than one rat in a cage, their results would not be independent. In the latter case, competition for food and other animal interactions might favor the stronger animal and influence the treatment effect.

Equality of variance also is an important assumption. If the variances are reasonably close, the test should be conducted as usual. As a general rule, if the variances do not differ by more than a factor of four, no special procedure is needed. If the variances differ widely, a modified procedure should be used (the Behrens-Fisher test⁹). The normality assumption is less critical. The CLT results in approximate normality of means of non-normal variables.

This statistical design consists of randomly dividing n objects into two groups of size n_1 and n_2 . Treatment 1 is applied to the first group (n_1) and treatment 2 is applied to the second group (n_2). Optimal treatment allocation in this design is to have an equal number of experimental units ($n/2$) in each group if the primary objective is to compare the means of the two groups. However, if n_1 is not equal to n_2 , the data are analyzed easily, and not much is lost if the two samples are close in size. In animal and human experiments, samples often are lost due to patient dropouts and animal deaths. An experiment that is carried out according to this plan sometimes is called a *parallel design*—two separate groups are treated in parallel.

The test is similar to the one-sample test. In a typical experiment to compare the mean results of the two samples, the null hypothesis is:

$$H_0 : \mu_1 = \mu_2$$

A two-sided alternative has an alternative hypothesis:

$$H_a : \mu_1 \neq \mu_2$$

Once the level (usually 0.05) is specified and data are obtained, a t test is performed. This allows a decision to be made about the equality of the underlying population averages. As in the one-sample case, a value of t is computed as:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

where $\bar{x}_1$ is the mean of first sample of n_1 observations, $\bar{x}_2$ is the mean of second sample of n_2 observations, and

$$s^2 = \frac{\sum x_{1i}^2 - (\sum x_{1i})^2 / n_1 + \sum x_{2i}^2 - (\sum x_{2i})^2 / n_2}{n_1 + n_2 - 2}$$

where $\sum x_{1i}^2$ is the sum of squares of observations in the first sample, $\sum x_{1i}$ is the sum of observations in the first sample, $\sum x_{2i}^2$ is the sum of squares of observations in the second sample, $\sum x_{2i}$ is the sum of observations in the second sample, s^2 is the pooled variance of the two samples.

The examples illustrated so far have used a two-sided test. A one-sided test may be used when the difference can only occur in one direction or when only one direction is relevant.

PAIRED T TEST

In many situations, the scientist is interested in comparing the means of two experimental treatments using a paired-sample design. This differs from the independent two-sample design in that each of two different treatments may be applied to a single group of experimental units (e.g., patients). In a bioavailability study, a generic drug is compared to a standard drug in each of 20 patients. A new analytical method is compared to a previously used method by comparing assay results on different concentrations of the same material divided into two parts.

The paired design has certain advantages over the two independent sample or parallel groups design. It has been noted that significance is determined by the ratio of the difference of the averages divided by the standard error. This ratio can be increased by reducing the standard error. One way is to increase the sample size. Another way of increasing the value of t is to reduce the variability.

In a two independent sample test, the variability is a result of the differences among different experimental units (differences

Example 10.

Suppose one sample of four and one sample of five are taken, respectively, from each of two lots of amobarbital capsules, and the amount of amobarbital is determined in each capsule. It is desired to determine if there is a significant difference between the two samples.

$$H_0: \mu_1 = \mu_2$$

where μ_1 and μ_2 represent the true averages of the two lots of capsules. This is a two-sided test at the 5% level.

Sample 1	Sample 2
10.1	9.8
13.6	9.6
12.5	11.4
11.4	9.1
	10.1

We have $\sum x_{1i} = 46.6$, $\sum x_{1i}^2 = 573.18$, $\bar{x}_1 = 11.90$, $n_1 = 4$, and $\sum x_{2i} = 50.0$, $\sum x_{2i}^2 = 502.98$, $\bar{x}_2 = 10.00$, $n_2 = 5$,

$$s^2 = \frac{573.18 - 47.6^2/4 + 502.98 - 50.0^2/5}{4 + 5 - 2}$$

$$= \frac{573.18 - 566.44 + 502.98 - 500.00}{7} = 1.3886$$

$$s = 1.18$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$$= \frac{11.90 - 10.00}{1.18} \sqrt{\frac{4 \cdot 5}{4 + 5}} = 2.40$$

The degrees of freedom involved in the pooled standard deviation are 7, $DF = (n_1 - 1) + (n_2 - 1)$. In the t table (see Table 10-10), for $p = 0.05$ and $DF = 7$ (two tails), the value of t given is 2.365. The value of t calculated is greater than this. Therefore, since the probability of these two samples being drawn from the same population is less than 0.05, we conclude that they were drawn from different populations. (This conclusion may be wrong 5 times in 100.) It can be stated that there is a statistically significant difference between the two samples.

among patients' responses to a drug, for example). In the paired test, the variability results from differences within experimental units. The within-individual variability should be less than the between-individual variability. (Theoretically, the measured between-individual variation includes the within-variation; therefore, the between-variation is larger than the within-variation.) Therefore, the paired-sample design has the advantage of reduced variability.

The paired-sample test also needs less experimental material. In a two independent sample design, comparing the response to two drugs, one might use 24 patients in each of two groups. In a paired design, each patient receives both drugs, on two different occasions if necessary. Thus, there is the need to recruit 24 patients rather than 48. For example, when testing a skin preparation, the products could be applied randomly to each arm of the 24 patients.

The paired design can be used only when there is a natural or easy way of pairing the experimental units. When comparing the dissolution of two different formulations, there seems to be

Example 11.

A drug is formulated to be dissolved more rapidly by substituting lactose for part of the lipoidal lubricant in the regular-release product. The formulator is convinced that this formulation change only could increase the rate of drug dissolution. A one-sided test at the 5% level is proposed when comparing the drug dissolution from the two products. The time to 50% dissolution for six tablets of each product in minutes is:

Original product:	25	22	29	30	26	24
Modified product:	18	23	24	22	19	16

For this test, H_0 and H_a are defined as:

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 > \mu_2$$

where μ_1 is the 50% dissolution time for the original product and μ_2 the respective dissolution time for the modified product.

If the test indicates rejection of the null hypothesis, it must be concluded that the new formulation has a faster dissolution time. If the test shows a nonsignificant difference, it is concluded that the data is insufficient to show that the new formulation reduces the dissolution time. Note that if the data show a longer dissolution time for the new formulation, a test would not be performed, but it would be concluded that the new formulation did not decrease the dissolution time. The average results and standard deviation for the two sets of data are:

$$\bar{x}_1 = 26 \quad \bar{x}_2 = 20.33$$

$$SD_1 = 3.033 \quad SD_2 = 3.141$$

$$t = \frac{26 - 19.17}{3.087/\sqrt{3}} = 3.18$$

Note that the pooled standard deviation is equal to:

$$\sqrt{\frac{SD_1^2 + SD_2^2}{2}}$$

when the sample sizes are equal in the two groups.

For a one-sided test, refer to one tail of the t distribution. For 10 degrees of freedom ($6 + 6 - 2$), the t value, leaving 5% of the area in the upper tail, is 1.812 (see Table 10-10). Therefore, it is concluded that the new formulation causes faster dissolution of the drug. Note that it is easier to get significance with a one-sided test. Had the test been two-sided, the t would have had to exceed 2.228 at the 5% level for significance, according to Table 10-10.

no obvious way of pairing the tablets from the two different formulations, as is the case of applying two treatments to the same individual. In animal experiments, litter mates may be paired. Pairing implies that the paired units are more alike than are two different units. In clinical studies, test units may be paired or matched on the basis of certain characteristics such as sex, age, or severity of disease. Then each subject in the pair is assigned to one of the experimental treatments.

A disadvantage of the paired design is that if treatments cannot be applied concurrently, as may be the case where two drugs administered orally are to be compared, the time to complete the experiment can be extended. In the case of clinical studies, this may be an important detriment because time usually is of the essence. Also, as these studies are prolonged, the chances of patient dropouts increase, and time can influence the progress of the disease.

In the paired design, a missing value means that the single unpaired datum is of no value. In this design, each experimental unit (e.g., each patient) essentially acts as its own control. That is, the comparison is made within each experimental unit. If one of the two paired values is missing, a comparison cannot be made.

Another potential disadvantage is that a *carryover effect* may be present. This means that effects from one treatment may affect the results of the other. For example, in a bioavailability study, if the first drug administered is not eliminated completely before the second drug is given, blood levels of the second drug will be contaminated. Or, in a clinical study, the first drug administered may modify the disease condition so that the effect of the second drug is not comparable directly with that of the first drug.

In any event, there are many situations where the advantages of the paired-sample design strongly suggest its use. For computational purposes, the formula is:

$$t = \frac{\bar{d}}{s/\sqrt{n}}$$

where $\bar{d}$ is the mean of the differences $x_1 - x_2$ of the n pairs of observations,

$$s^2 = \frac{\sum d_i^2 - (\sum d_i)^2 / n}{n - 1}$$

where $\sum d_i^2$ is the sum-of-squares of the n differences, $\sum d_i$ is the sum of the n differences, and n is the number of differences or pairs of observations.

TESTS FOR PROPORTIONS

The t test is applicable for continuous data that is distributed normally. Much of the data that is seen in pharmaceutical experiments is dichotomous. That is, answers to a questionnaire regarding filling a prescription for a specified drug may be yes or no, or a bottle of tablets may be acceptable or not acceptable, or a patient may be cured or not cured. Tests similar to the t test may be constructed for binomial data. The principle is to compute proportions that are probable, based on the sample proportion. If the probable proportions do not include the hypothetical proportion, the null hypothesis is rejected.

For large sample sizes (n is large), such computations can be tedious and difficult. Therefore, the normal approximation to the binomial is used whenever possible. Fortunately, in most practical cases, the normal approximation is applicable. When comparing proportions from two independent samples when the normal approximation is clearly not applicable, the Fisher Exact test can be used.⁸ (Statistical software programs can compute exact probabilities.) In general, use the rule that np and $n(1-p)$ should be equal to or greater than 5 in order to use the normal approximation. In practice, this rule may be relaxed somewhat. When in doubt, a professional statistician should be consulted.

Simple statistical tests for proportions are analogous to the t tests. For a one-sample test, where the hypothetical value defined by H_0 is p_0 , the ratio:

$$z = \frac{p - p_0}{\sqrt{p_0(1 - p_0)/n}}$$

Example 12.

The duration of loss of the righting reflex in minutes was measured in 16 mice, following treatment with a barbiturate. The drug was administered in the morning and the afternoon on two different occasions; the order of giving the morning or the afternoon dose was randomized in each mouse. It was desired to test the null hypothesis that the duration of loss of the righting reflex is the same in the morning and the afternoon (Table 10-11).

$$s^2 = \frac{354 - 40^2/16}{16 - 1} = 16.93$$

$$s = 4.11$$

$$t = \frac{\bar{d}}{s/\sqrt{n}} = 2.43$$

$$DF = n - 1 = 15$$

In the t table (see Table 10-10), for $p = 0.05$ and $DF = 15$ (two tails), the value of t is 2.131. The value of t calculated is greater than this. Therefore, as the probability of the morning and afternoon values being the same is less than 0.05, we conclude that they are different. Apparently, the duration of loss of the righting reflex in mice tested on the barbiturate in the morning was longer than when tested in the afternoon.

Note the similarity of the one-sample t test and the paired t test. The test is identical after differences between pairs have been calculated in the paired test. The null hypothesis in the paired test almost always is of the form $H_0: \delta = 0$, where δ is the hypothesized difference of the true means. It is hypothesized that the mean results of the two treatments are identical.

Table 10-11. Loss of Righting Reflex on 16 Mice

Mouse No	am X_1	pm x_2	Difference $D = x_1 - x_2$
1	75	73	2
2	86	89	-3
3	93	89	4
4	87	79	8
5	91	95	-4
6	87	81	6
7	76	77	-1
8	83	89	-6
9	87	82	5
10	95	91	4
11	91	87	4
12	86	86	0
13	83	78	5
14	76	69	7
15	82	78	4
16	93	88	5
			$\sum d_i = 40$ $\sum d_i^2 = 354$ $\bar{d} = 2.5$ $n = 16$

Example 13.

A one-sample test for proportions. A questionnaire was sent to pharmacists, asking which of two cold medications the pharmacist would recommend to customers. A statistical test was proposed to decide which product was most recommended. The null hypothesis was that the two products, A and B, were recommended equally: $H_0: p_a = p_b = 0.50$. The test is two-sided at the 5% level. Two hundred and fifty (250) pharmacists responded; Product A was recommended 145 times, and Product B was recommended 105 times. The scientist conducting the experiment had sent out 400 questionnaires and was rightfully concerned about the nonresponders. However, she decided that there was no reason to suspect a bias because of the lack of 100% response and proceeded to analyze the data. The observed proportion of successes (A is a success) is $145/250 = 0.58$ (or the observed proportion could be 0.42 as well). The absolute value of the numerator of the z statistic will be the same for $p = 0.42$ or $p = 0.58$.

$$z = \frac{|0.58 - 0.50|}{\sqrt{0.5 \cdot 0.5/250}} = 2.53$$

Since 2.53 exceeds the tabled value for α of 5% (1.96), it is concluded that Product A is the more recommended product. The normal approximation is improved if $1/(2n)$ is subtracted from the absolute value of the numerator (*Yates correction*), although the effect of the continuity correction is more evident for small sample sizes. In the present example, the corrected value of z is 2.47. A 95% confidence interval for the proportion recommending Product A also was reported.

$$0.58 \pm 1.96\sqrt{0.58 \cdot 0.42/250} = 0.519 \text{ to } 0.641$$

Exercise 4.

Compute z , with and without the continuity correction, if 141 of 250 pharmacists recommended Product A. Determine whether the result is significant by using a two-sided test at the 5% level.

Answers: $z = 2.02$ without continuity correction (significant) and $z = 1.96$ with continuity (not significant).

may be computed, where p is the observed proportion and n is the number of binomial trials, the sample size.

The calculated value of z is compared to the standard normal distribution rather than to the t distribution. Refer to Table 10-9 or to the last line in the t table, Table 10-10. For a two-sided test at the 5% level, the test statistic, z , must exceed 1.96 for the difference to be considered significant.

CHI-SQUARE TEST

To test for differences of two proportions from two independent samples, the chi-square test is used. Chi-square (χ^2) is a probability distribution derived from the sum of squares of normal variables. The chi-square distribution is not symmetrical and can have only positive values. Table 10-12³ is a short table of chi-square probabilities. This table is used in the same way as the normal and t tables: first compute a chi-square statistic, and if the value exceeds the tabled value, a significant effect is declared.

TABLE 10-12. The Chi-Square Table Probability

DF	P = 0.20	P = 0.10	P = 0.05	P = 0.01
1	1.64	2.71	3.84	6.64
2	3.22	4.61	5.99	9.21
3	4.64	6.25	7.82	11.34
4	5.99	7.78	9.49	13.28
5	7.29	9.24	11.07	15.09
6	8.56	10.64	12.59	16.81
7	9.80	12.02	14.07	18.48
8	11.03	13.36	15.51	20.09
9	12.24	14.68	16.92	21.67
10	13.44	15.99	18.31	23.21
20	25.04	28.41	31.41	37.57
30	36.25	40.26	43.77	50.89

(Data from Fisher RA, Yates F. *Statistical Tables for Biological, Agriculture and Medical Research*. New York: Hafner, 1963.)

Example 14.

In tossing a coin, 50% tails and 50% heads are expected. Suppose a coin is tossed 40 times, and 25 heads and 15 tails are obtained, whereas 20 heads and 20 tails are expected. Is the coin biased or weighted in some way?

$$\chi^2 = \frac{(25 - 20)^2}{20} + \frac{(15 - 20)^2}{20} = 2.5$$

The degrees of freedom (DF) associated with χ^2 are one less than the number of categories. Here $\chi^2 = 2.5$ with 1 DF. The greater the disagreement between expected and observed, the larger will be χ^2 . See Table 10-12 for probabilities of getting this value or larger. For 1 DF the probability of getting a value larger than 2.5 is somewhere between $\alpha = 0.20$ and $\alpha = 0.10$. To say that there is a statistically significant departure from the expected values, χ^2 would have to be larger than 3.84, which is the value for $\alpha = 0.05$ at 1 DF. A value of χ^2 larger than 6.64 for 1 DF would indicate a statistically highly significant ($\alpha < 0.01$) departure of the observed from the expected values.

Example 15.

Table 10-13 gives the survival rates for drug-treated and control pigs with swine dysentery. The survival rates for the drug-treated and control pigs are $pD = 25/39 = 64\%$ and $pC = 21/43 = 49\%$, respectively. To test the null hypothesis that there is no difference in the survival rates of drug-treated and control pigs, χ^2 is calculated.

The expected values in each of the four cells can be obtained by multiplying the column total by the row total and dividing this result by the grand total. The expected value for Cell a is $46.39/82 = 21.9$. The expected frequencies for Cells b , c , and d are 17.1, 24.1, and 18.9, respectively. Note that the sum of the expected frequencies across any row or column equal the totals for the row or column. For example, the expected frequencies for b and d are 17.1 and 18.9, which sums to 36, the total number who died. The calculation of χ^2 is:

$$\frac{(25-21.9)^2}{21.9} + \frac{(14-17.1)^2}{17.1} + \frac{(21-24.1)^2}{24.1} + \frac{(22-18.9)^2}{18.9} = 1.91$$

The DF associated with an $R \times C$ contingency table is $(R-1)(C-1)$, so that for a 2×2 contingency table, we have 1 DF. Table 10.12 shows that for 1 DF the probability of getting a value of χ^2 larger than the calculated value 1.91 is greater than $\alpha=0.10$. Since α is not equal to or less than 0.05, we conclude that there is insufficient evidence to indicate that the survival rates for the drug-treated and control pigs are different.

Table 10-13. Survival Rates in Swine Dysentery

Treatment	Survived	Died	Total
Drug	$a=25$	$b=14$	$a+b=39$
Controls	$c=21$	$d=22$	$c+d=43$
Totals	$a+c=46$	$b+d=36$	$N=82$

Example 16.

Two different types of penicillin (Type I and Type II) were given to each of 22 patients in random order, on successive occasions, and the presence (+) or absence (–) of a detectable blood level was determined (Table 10-14). The percentage of patients with detectable blood levels for the two forms of penicillin are $pl=16/22=73\%$ and $pII=8/22=36\%$. To test the null hypothesis that there is no difference in the percentage of patients with detectable blood levels for the two forms of penicillin, we calculate:

$$\chi^2 = \frac{(b-c-1)^2}{b+c} = \frac{(|10-2|-1)^2}{10+2} = 4.08$$

In Table 10-13 for $\alpha=0.05$ and $DF=1$, the value of χ^2 given is 3.84. The value of χ^2 calculated is greater than this. Therefore, as the probability of the percentages for type I and type II penicillin being the same is less than 0.05, we conclude that they are different.

Note that this test compares the number of patients who are positive on one test and negative on the other.

Table 10-14. Data for Example 16

		Type II		Totals
		+	–	
Type I	+	$a=6$	$b=10$	16
	–	$c=2$	$d=4$	6
Totals		8	14	22

Chi-square is calculated as:

$$\chi^2 = \frac{(\text{observed frequency} - \text{expected frequency})^2}{\text{expected frequency}}$$

The chi-square test commonly is used for comparing two percentages in a 2×2 or fourfold contingency table (Table 10-13).

The chi-square test for comparing two correlated percentages for paired data takes a somewhat different form.

The chi-square distribution is an approximation of the discrete distribution represented by the fourfold table. The

approximation can be improved by applying a *correction factor* for the Observed-Expected values. If the absolute difference is an exact integer (e.g., 4.0), subtract 0.5 from the absolute difference; 4.0 would become 3.5. If the absolute difference has a decimal between 0.5 and 0.99 (e.g., 3.8), change the decimal to

Exercise 5.

Calculate the corrected chi-square for Example 15.

Answer: 1.79.

0.5; 3.8 would become 3.5. If the decimal is between 0 and 0.5, reduce the absolute difference to its integer value; 4.1 would become 4. In *Example 15*, the absolute difference of Observed-Expected values would be reduced to 3.0.

THE F DISTRIBUTION AND TESTS OF SIGNIFICANCE

The F distribution is suitable for a statistical test comparing two means. The F distribution is used to compare two variances, F being defined by the ratio of the variances, with n_1-1 DF in the numerator and n_2-1 DF in the denominator of the ratio:

$$F_{n_1-1, n_2-1} = \frac{s_1^2}{s_2^2}$$

Similar to the chi-square distribution, the F distribution consists of only positive values and is a skewed distribution. The ratio of two variances is compared to values in the F table (Table 10-15¹⁰) with the appropriate DF in the numerator and denominator to test for statistical significance. If the calculated ratio exceeds the value in the table at a given α level, the variances differ at the α level of significance. The following two examples describe the use of the F test for comparing variances for independent and dependent samples.

Example 17 shows a test to compare the variances of two independent samples using the F test. Example 18 shows the test for comparing variances in related or paired samples, which uses the t statistic to determine significance. To compare the variances of samples from two independent populations, the calculation is:

$$F = \frac{s_1^2}{s_2^2} \quad \text{with } s_1^2 > s_2^2$$

where,

$$s_1^2 = \frac{\sum x_{1i}^2 - (\sum x_{1i})^2 / n_1}{n_1 - 1}$$

$$s_2^2 = \frac{\sum x_{2i}^2 - (\sum x_{2i})^2 / n_2}{n_2 - 1}$$

Example 17.

Two treatments showed the results in Table 10-16. Entering Table 10-15 with $f_1=6$ and $f_2=5$ DF, we find that the tabulated values of F are 4.95 and 6.98 for $\alpha=2 \cdot 0.05=0.10$ and $\alpha=2 \cdot 0.05=0.05$, respectively. Thus, the probability of getting a value of F larger than the calculated value 5.75 is between $\alpha=0.05$ and $\alpha=0.10$. Since P is not equal to or less than 0.05, we conclude that there is insufficient evidence to indicate that the two variances are different.

$$f_1 = n_1 - 1 = 7 - 1 = 6$$

$$f_2 = n_2 - 1 = 6 - 1 = 5$$

$$F = s_1^2 / s_2^2 = 13.81 / 2.40 = 5.75$$

Table 10-15. The F Table 10%, 5%, 2.5%, and 1% Points for the Distribution of F

f1 Degrees of Freedom (For Greater Mean Square)																	
F2	P	1	2	3	4	5	6	7	8	9	10	20	30	40	60	120	∞
5	0.10	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.31	3.17	3.16	3.14	3.12	3.10
	0.05	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.56	4.50	4.46	4.43	4.40	4.36
	0.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.33	6.23	6.18	6.12	6.07	6.02
10	0.01	16.26	13.27	12.06	11.39	10.97	10.67	10.45	10.27	10.15	10.05	9.55	9.38	9.29	9.20	9.11	9.02
	0.10	3.28	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.20	2.16	2.13	2.11	2.08	2.06
	0.05	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.77	2.70	2.66	2.62	2.58	2.54
15	0.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.42	3.31	3.26	3.20	3.14	3.08
	0.01	10.04	7.56	6.55	5.99	5.64	5.39	5.21	5.06	4.95	4.85	4.41	4.25	4.17	4.08	4.00	3.91
	0.10	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	1.92	1.87	1.85	1.82	1.79	1.76
20	0.05	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.33	2.25	2.20	2.16	2.11	2.07
	0.025	6.20	4.76	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.76	2.64	2.58	2.52	2.46	2.40
	0.01	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.36	3.20	3.12	3.05	2.96	2.87
25	0.10	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.79	1.74	1.71	1.68	1.64	1.61
	0.05	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.12	2.04	1.99	1.95	1.90	1.84
	0.025	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.46	2.35	2.29	2.22	2.16	2.09
30	0.01	8.10	5.85	4.94	4.43	4.10	3.87	3.71	3.56	3.45	3.37	2.94	2.77	2.69	2.61	2.52	2.42
	0.10	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.72	1.66	1.63	1.59	1.56	1.52
	0.05	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.01	1.92	1.87	1.82	1.77	1.71
40	0.025	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.30	2.18	2.12	2.05	1.98	1.91
	0.01	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.21	3.13	2.70	2.54	2.45	2.36	2.27	2.17
	0.10	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.67	1.61	1.57	1.54	1.50	1.46
60	0.05	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	1.93	1.84	1.79	1.74	1.68	1.62
	0.025	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.20	2.07	2.01	1.94	1.87	1.79
	0.01	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.06	2.98	2.55	2.38	2.29	2.21	2.11	2.01
120	0.10	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.61	1.54	1.51	1.47	1.42	1.38
	0.05	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.84	1.74	1.69	1.64	1.58	1.51
	0.025	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.07	1.94	1.88	1.80	1.72	1.64
∞	0.01	7.31	5.18	4.31	3.83	3.51	3.39	3.12	2.99	2.88	2.80	2.37	2.20	2.11	2.02	1.92	1.81
	0.10	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.54	1.48	1.44	1.40	1.35	1.29
	0.05	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.75	1.65	1.59	1.53	1.47	1.39
	0.025	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	1.94	1.82	1.74	1.67	1.58	1.48
	0.01	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.20	2.03	1.93	1.84	1.73	1.60
	0.10	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.48	1.41	1.37	1.32	1.26	1.19
	0.05	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.66	1.55	1.50	1.43	1.35	1.25
	0.025	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	1.82	1.69	1.61	1.53	1.43	1.31
	0.01	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.03	1.86	1.76	1.66	1.53	1.38
	0.10	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.42	1.34	1.30	1.24	1.17	1.00
	0.05	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.57	1.46	1.39	1.37	1.27	1.00
	0.025	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.71	1.57	1.48	1.39	1.27	1.00
	0.01	6.64	4.60	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	1.57	1.69	1.59	1.47	1.32	1.00

(Data from Snedecor GW, Cochran WG. Statistical Methods. 7th edn, Ames: Iowa State University Press, 1980.)

Table 10-16. Data for Example 17

	A	B
	6	15
	4	4
	3	10
	7	10
	6	5
	4	11
		9
ΣX_1	30	64
Σx_1^2	162	668
n_1	6	7
S_1^2	2.40	13.81
f_1	5	6

Example 18.

A characteristic was measured before and after aging for each of 10 items (Table 10-17). Has the variability changed with aging?

$$\begin{aligned}
 \Sigma x_B^2 &= 2393.81 \\
 \Sigma x_A^2 &= 2252.72 \\
 \Sigma x_B x_A &= 2298.92 \\
 &= 2393.81 - 148.5^2/10 = 188.59 \\
 &= 2252.72 - 147.2^2/10 = 85.94 \\
 &= 2298.92 - 148.5 \cdot 147.2/10 = 113.00 \\
 t &= \frac{([x_B^2] - [x_A^2])\sqrt{n-2}}{2\sqrt{[x_B][x_A] - [x_B x_A]^2}} = \frac{(188.59 - 85.94)\sqrt{8}}{2\sqrt{188.59 \cdot 85.94 - 113.00^2}} = 2.476 \\
 DF &= n - 2 = 8
 \end{aligned}$$

In the t table (see Table 10-10), for $\alpha=0.05$ and $DF=8$ (two tails), the value of t given is 2.306. The value of t calculated is greater than this. Therefore, because the probability of the variance before and after aging being the same is less than 0.05, it is concluded that they are different. Apparently, the variability decreased after aging.

Table 10-17. Measurement Before and After Aging

Item No	Before Aging	After Aging
1	8.3	9.3
2	8.4	10.9
2	14.9	13.2
4	12.2	12.8
5	12.5	16.0
6	15.0	15.2
7	17.1	16.8
8	19.2	16.2
9	22.0	17.9
10	18.9	18.9
	$\Sigma x_B = 148.5$	$\Sigma x_A = 147.2$

To test for significance, the F ratio is referred to in the F table (see Table 10-15) with $f_1=n_1-1$ and $f_2=n_2-1$ DF. If the calculated value is greater than the value of the table, the null hypothesis that the two variances are the same is rejected (at the $2P$ level of significance).

If it is desired to compare the variances from paired data, the F test described above would be inappropriate. Instead, proceed as exemplified below.

The F distribution is used most often for the comparison of more than two means through the analysis of variance and is equivalent to the t test if used to compare two means.

ANALYSIS OF VARIANCE (ANOVA) AND EXPERIMENTAL DESIGN

There is almost always more than one way to conduct an experiment to achieve a given objective. It was seen that when comparing the means of two treatment groups, two independent groups or a paired design could be used. For example, in a clinical study two treatments could be applied to two separate and independent groups of patients, or each patient could take both treatments.

In the paired experiment, a further refinement could be added with regard to treatment order. For example, if drugs cannot be administered concurrently, the order of drug administration can be balanced. Half of the patients receive Drug A at the first administration, and the other half receive A on the second administration. The patients who receive A on the first occasion will receive B on the second occasion and vice versa. This is known as a *crossover design*.

An alternative design is to assign each patient an order of administration randomly. In the latter case, it is likely that a balanced allocation would not be achieved, as in the crossover design. In fact, there is always a possibility, albeit a small one, that all patients will receive one and the same treatment first and the other treatment second, a situation that intentionally is avoided in the crossover design.

When more than two treatments are to be used in an experiment, a variety of designs are possible. In these cases, one design usually will be optimal, depending on the nature of the experiment, the treatments, and the experimental units or subjects. A common feature of most good designs is *symmetry*. That is not to say that all good designs are symmetrical. In some special cases, an asymmetrical design may be optimal, but this is not the usual circumstance.

The most simple analysis of variance design is known as a *one-way analysis of variance* (one-way ANOVA) or completely randomized design. This is the ANOVA analogy of the two independent sample t test. In the ANOVA design, there is interest in comparing the means of two or more treatment groups. As has been noted before, in the jargon of clinical trials, this design often is one of a class known as parallel-groups designs.

In the following description, an example from clinical trials will be used. However, one should understand that tablets, bottles, or consumers could be substituted for patients and the process is the same: n patients are available for the experiment with t treatments. For example, 150 patients are to be assigned to three treatment groups, one placebo and two active. The n patients are assigned randomly to the three groups (see the discussion on random assignment). The optimal assignment in the examples discussed in this chapter will result in equal numbers in each group, n/t units per group. Note that n is chosen to be divisible by t . If there are three treatments and $n=150$, we randomly would assign 50 units per treatment. A loss of observations will not invalidate the analysis, as is also true in the two independent groups t test. Observations are made, and the null hypothesis that all t -means are equal is tested by an ANOVA procedure.

The ANOVA separates the total sum-of-squares, $\Sigma(x_i - \bar{x})^2$ into parts determined by the experimental structure. For the one-way ANOVA, the sum-of-squares consists of the between (among) and within sum-of-squares. The between sum-of-squares (BSS) represents differences among treatments, large values indicating large treatment differences (e.g., if the

Example 19.

Groups of three subjects each were given one of ten food regimens and showed the weight gains (lb) in Table 10-19. These are unpaired data, and this type of study is referred to as a completely randomized experiment. There are only two sources of variation; the variation between regimens and the variation within regimens, as indicated in Table 10-18.

The sums-of-squares are obtained as:

$$\text{Total SS} = \sum x^2 - \frac{(\sum x)^2}{n} = 934 - \frac{148^2}{30} = 203.87$$

$$\begin{aligned} \text{Between regimens SS} &= \frac{(\sum x_1)^2}{n_1} + \frac{(\sum x_2)^2}{n_2} + \dots + \frac{(\sum x_{10})^2}{n_{10}} - \frac{(\sum x)^2}{n} \\ &= \frac{7^2}{3} + \frac{3^2}{3} + \dots + \frac{16^2}{3} - \frac{148^2}{30} \\ &= 160.54 \end{aligned}$$

$$\text{Within regimens SS} = 203.87 - 160.54 = 43.33$$

The mean squares are obtained by dividing the sums-of-squares by their corresponding DF. The mean square within regimens, s^2 , is the pooled variance for the ten samples. Since this is the only variance that can be identified as random sampling error (the mean square between regimens has in addition a component due to the variability among regimens), it becomes the denominator in the F ratio, so that:

$$F = \frac{\text{mean square between regimens}}{\text{mean square within regimens}} = \frac{17.81}{2.17} = 8.22$$

To test for significance, the F ratio is referred to in the F table (see Table 10-15) with $f_1 = t - 1 = 9$ and $f_2 = \sum (n_i - 1) = 20 = 2 \text{ DF}$.

We find that the calculated value 8.22 is larger than the tabulated value 3.45 for $P=0.01$. Therefore, as the probability of these ten samples being drawn from the same population is less than 0.05 (actually, it is less than 0.01), it is concluded that they are not all the same (i.e., not all the means are equal).

treatment means are identical, the BSS will be 0 on the average). The within sum-of-squares (WSS) represents differences within treatments, or error; i.e., the differences among objects (subjects in clinical studies) within a treatment is a measure of the variability of the observations.

An ANOVA table is prepared consisting of source of variation, degrees of freedom, sums-of-squares, and mean square. In the one-way ANOVA, the sources consist of the between, within, and total terms. The sum-of-squares divided by the DF is known as the mean square, *between mean square* (BMS), and *within mean square* (WMS) in the one-way ANOVA (Table 10-18).

For a one-way ANOVA, the DF for treatments is $t-1$. The DF for error (within treatments) is $n-t$, where n is the total number of observations. The total sum-of-squares (SS) is exactly the sum of the between and within sums of squares. The error mean square (WMS) corresponds to the variance for the test, and in the case of two treatments, corresponds to the pooled variance in the t test.

The ratio BMS/WMS has an F distribution under the null hypothesis, with $(t-1)$ DF in the numerator and $(n-t)$ DF in the denominator. If the ratio exceeds the appropriate F value found in the table, then at least two of the treatments tested are significantly different. The computations consist of simple

Table 10-18. ANOVA for Example 19

Analysis of Variance				
Source of Variation	Degrees of Freedom	Sums-of-Square	Mean Squares	F Ratio
Between regimens	$t-1 = 9$	160.54	17.81	8.22
Within regimens	$\sum (n_i - 1) = 20$	43.33	$S^2 = 2.17$	
Total	$N - 1 = 29$	203.87		

* $\sum (nt - 1 = 9) - N - t$.

Table 10-19. Weight Gains in Ten Food Regimens

Food Regimen											
	A	B	C	D	E	F	G	H	I	J	(t = to REGIMENS)
	2	1	2	4	9	3	6	7	4	4	
	3	2	4	8	8	3	5	6	4	6	
	2	0	1	7	11	6	6	5	7	6	
											Sums
Σx ₁	7	3	7	19	28	17	17	19	15	16	Σx = 148
	17	5	21	129	266	109	97	121	81	88	Σx² = 934
n ₁	3	3	3	3	3	3	3	3	3	3	N = 30
n _i -1	2	2	2	2	2	2	2	2	2	2	Σ(n _i - 1)= 20
X ₁	2.3	1.0	2.3	6.3	9.3	5.7	5.7	6.3	5.0	5.3	

arithmetic, summing individual values, and their squares. The following numerical example illustrates the computations and should clarify these concepts. Although it always is useful to practice some calculations, computer programs are available that should be used for most practical situations.

MULTIPLE COMPARISONS IN ANOVA

If the F test is significant and more than two treatments are included in the experiment ($t > 2$), it may not be obvious immediately which treatments are different. Some or all of the treatments may be different. Various multiple-comparison procedures have been proposed to solve this problem. It is not always apparent when a particular procedure is best, given the variety of procedures available:

- Least Significant Difference Procedure
- Studentized Range Procedure
- Duncan's New Multiple Range Procedure
- Dunnett's Procedure

Several of these tests are described here, with discussion of their application. The general procedure is to list the ranked means from lowest to highest and underline the means that are not statistically significantly different from each other. Sometimes brackets or parentheses are used instead of an underline. The procedure is carried out by calculating a 5% allowance, which is defined as the critical difference between means which allows one to reject the null hypothesis ($\mu_i = \mu_j$) and accept the alternative hypothesis ($\mu_i \neq \mu_j$) for any two sample means $\bar{x}_i$ and $\bar{x}_j$ at $\alpha = 0.05$. To calculate the 5% allowance, the following data is required:

s^2 :	pooled variance from the analysis of variance
DF :	degrees of freedom for the pooled variance from the analysis of variance
n_i, n_j :	the number of observations from which the means $\bar{x}_i$ and $\bar{x}_j$ were determined, respectively
t	a critical value at $\alpha = 0.05$, which depends upon the DF and the degree of conservatism desired, as exemplified by the multiple comparison procedures described below

- *Least Significant Difference Procedure*—For this procedure:

$$5\% \text{ allowance} = t \sqrt{s^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

where t is the value of t from Table 10-10 (two tails). This is the least conservative procedure and assures that the probability that any one comparison is judged to be significant by chance alone is 5%. However, the probability of one or more comparisons being judged significant would be greater than 5%. Applied to the results of Example 19,

$$s^2 = 2.17$$

$$n_i = n_j = 3$$

$$DF = 20$$

and $t = 2.086$ from Table 10-10 for 20 DF and $\alpha = 0.05$ (two tails).

$$5\% \text{ allowance} = t \sqrt{s^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} = 2.086 \sqrt{2.17 \cdot (1/3 + 1/3)} = 2.51$$

Thus, any two means differing by 2.51 or more are judged to be different.

Ranked Means						
B	A,C	I	J	F,G	D,H	E
1.0	2.3	5.0	5.3	5.7	6.3	9.3

or, $(BAC)(IJFGDH)(E)$.

Any two means underscored by the same line (or included in the same parentheses) do not differ statistically at $\alpha = 0.05$.

Any two means not underscored by the same line (or not included in the same parentheses) are statistically significantly different at $\alpha \leq 0.05$.

- *Studentized Range Procedure* For this method:

$$5\% \text{ allowance} = \frac{Q}{\sqrt{2}} \sqrt{s^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

where Q is the studentized range value for k treatments from Table 10-20.¹¹ This is one of the more conservative procedures, and it ensures that the probability of one or more comparisons being judged significant by chance alone is 5%. Applied to the results of Example 19,

$$Q = 5.01 \text{ from table 10-20}$$

for $k = 10$ treatments, 20 DF and $\alpha = 0.05$.

$$5\% \text{ allowance} = \frac{2}{\sqrt{2}} \sqrt{s^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} = \frac{5.01}{\sqrt{2}} \sqrt{2.17 \cdot (1/3 + 1/3)} = 4.26$$

Thus, any two means differing by 4.26 or more are judged to be different.

Ranked Means						
B	A,C	I	J	F,G	D,H	E
1.0	2.3	5.0	5.3	5.7	6.3	9.3

or, $(BACI)(ACIJFGDH)(JFGDHE)$.

- *Duncan's New Multiple Range Procedure* For this method:

$$5\% \text{ allowance} = \frac{t_k}{\sqrt{2}} \sqrt{s^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

where t_k are values for 2, 3, ..., k treatments obtained from Table 10-21.¹² The critical values will be $A_2, A_3, \dots, A_k$, depending upon how many means are included in the range of ranked means being compared. This is next to the least conservative procedure. Applied to the results of Example 19,

$$5\% \text{ allowance} = \frac{t_k}{\sqrt{2}} \sqrt{s^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} = \frac{t_k}{\sqrt{2}} \sqrt{2.17 \cdot (1/3 + 1/3)}$$

Values of t_k from Table 10-21 for $k = 2$ to 10 treatments, 20 DF , and $\alpha = 0.05$ give the allowances in Table 10-22. Thus, the critical difference between E and B is 2.89 because the range includes 10 means, the critical difference between E and H is 2.64 because the range includes three means, and so on.

Ranked Means						
B	A,C	I	J	F,G	D,H	E
1.0	2.3	5.0	5.3	5.7	6.3	9.3

or, $(BAC)(IJFGDH)(E)$.

Table 10-20. The Q Table Upper 5% Points, Q, in the Studentized Range

k (Number of Treatments)																				
DF	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
10	3.15	3.88	4.33	4.66	4.91	5.12	5.30	5.46	5.60	5.72	5.83	5.93	6.03	6.12	6.20	6.27	6.34	6.41	6.47	
11	3.11	3.82	4.26	4.58	4.82	5.03	5.20	5.35	5.49	5.61	5.71	5.81	5.90	5.98	6.06	6.14	6.20	6.27	6.33	
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.40	5.51	5.61	5.71	5.80	5.88	5.95	6.02	6.09	6.15	6.21	
13	3.06	3.73	4.15	4.46	4.69	4.88	5.05	5.19	5.32	5.43	5.53	5.63	5.71	5.79	5.86	5.93	6.00	6.06	6.11	
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36	5.46	5.56	5.64	5.72	5.79	5.86	5.92	5.98	6.03	
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.31	5.40	5.49	5.57	5.65	5.72	5.79	5.85	5.91	5.96	
16	3.00	3.65	4.05	4.34	4.56	4.74	4.90	5.03	5.15	5.26	5.35	5.44	5.52	5.59	5.66	5.73	5.79	5.84	5.90	
17	2.98	3.62	4.02	4.31	4.52	4.70	4.86	4.99	5.11	5.21	5.31	5.39	5.47	5.55	5.61	5.68	5.74	5.79	5.84	
18	2.97	3.61	4.00	4.28	4.49	4.67	4.83	4.96	5.07	5.17	5.27	5.35	5.43	5.50	5.57	5.63	5.69	5.74	5.79	
19	2.96	3.59	3.98	4.26	4.47	4.64	4.79	4.92	5.04	5.14	5.23	5.32	5.39	5.46	5.53	5.59	5.65	5.70	5.75	
20	2.95	3.58	3.96	4.24	4.45	4.62	4.77	4.90	5.01	5.11	5.20	5.28	5.36	5.43	5.50	5.56	5.61	5.66	5.71	
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.01	5.10	5.18	5.25	5.32	5.38	5.44	5.50	5.55	5.59	
30	2.89	3.48	3.84	4.11	4.30	4.46	4.60	4.72	4.83	4.92	5.00	5.08	5.15	5.21	5.27	5.33	5.38	5.43	5.48	
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.74	4.82	4.90	4.98	5.05	5.11	5.17	5.22	5.27	5.32	5.36	
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	4.73	4.81	4.88	4.94	5.00	5.06	5.11	5.15	5.20	5.24	
120	2.80	3.36	3.69	3.92	4.10	4.24	4.36	4.47	4.56	4.64	4.71	4.78	4.84	4.90	4.95	5.00	5.04	5.09	5.13	
∞	2.77	3.32	3.63	3.86	4.03	4.17	4.29	4.39	4.47	4.55	4.62	4.68	4.74	4.80	4.84	4.89	4.93	4.97	5.01	

(Data from Snedecor GW, Cochran WG. *Statistical Methods*. 7th edn. Ames: Iowa State University Press, 1980.)

Table 10-21. The Multiple Range Table. Duncan's New Multiple Range Test at the 5% Level of Significance

K (Number of Treatments)									
DF	2	3	4	5	6	8	10	14	20
10	3.15	3.30	3.37	3.43	3.46	3.47	3.47	3.47	3.48
12	3.08	3.23	3.33	3.36	3.40	3.44	3.46	3.46	3.48
14	3.03	3.18	3.27	3.33	3.37	3.41	3.44	3.46	3.47
16	3.00	3.15	3.23	3.30	3.34	3.39	3.43	3.45	3.47
18	2.97	3.12	3.21	3.27	3.32	3.37	3.41	3.45	3.47
20	2.95	3.10	3.18	3.25	3.30	3.36	3.40	3.44	3.47
24	2.92	3.07	3.15	3.22	3.28	3.34	3.38	3.44	3.47
30	2.89	3.04	3.12	3.20	3.25	3.32	3.37	3.43	3.47
60	2.83	2.98	3.08	3.14	3.20	3.28	3.33	3.40	3.47
100	2.80	2.95	3.05	3.12	3.18	3.26	3.32	3.40	3.47
∞	2.77	2.92	3.02	3.09	3.15	3.23	3.29	3.38	3.47

(Data from Duncan DB. Multiple range and multiple F tests. *Biometrics* 1955; 11: 1–42.)

Table 10-22. Critical Values using Duncan's Test for Example 19

<i>k</i>	<i>Tk</i>	<i>Ak</i>	<i>k</i>	<i>Tk</i>	<i>Ak</i>
2	2.95	2.51	7	3.34	2.84
3	3.10	2.64	8	3.36	2.86
4	3.18	2.70	9	3.38	2.87
5	3.25	2.76	10	3.40	2.89
6	3.30	2.81			

- **Dunnett's Procedure** The three procedures previously described are appropriate when it is desired to compare all possible pairs of means. Dunnett¹³ considered the problem when the objective of the study is to compare several treatments with a standard or control. In his method

$$5 \text{ allowance} = t_d \sqrt{s^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

where t_d is Dunnett's tD value for k treatments (excluding the standard or control), obtained from Table 10-23.

Like the studentized range procedure, this is one of the most conservative procedures, and it ensures that the probability of one or more comparisons between treatments and a standard or control being judged significant by chance alone is 5%. The one-tail values (listed in tables for $\alpha=0.10$) are used when the objective of the study is to select only those treatments that have higher (or lower) means than the standard or control. The two-tail values (listed in the table for $\alpha=0.05$) are used when the objective of the study is to select those treatments that are either higher or lower than the standard or control. Of course, the decision to carry out a one-tailed or a two-tailed test must be made before the study begins.

In Example 19, suppose J is a standard regimen, and it is desired to determine which regimens show different weight gains from J . Here, $tD = 2.95$ from Table 10-23 for $k = 9$ treatments, 20 DF, and $\alpha = 0.05$ (two-tails).

$$5\% \text{ allowance} = t_d \sqrt{s^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} = 2.95 \sqrt{2.17 \cdot (1/3 + 1/3)} = 3.55.$$

Thus, any regimen mean that differs from the mean for regimen J by 3.55 or more is judged to be different from J .

Ranked Means

B	A,C	I	J	F,G	D,H	E
1.0	2.3	5.0	5.3	5.7	6.3	9.3

or, $(B)(ACIJFGDH)(E)$.

It would be concluded that B showed a statistically significant smaller weight gain than J , E showed a statistically significantly larger weight gain than J , and there was insufficient evidence to indicate that the other regimens were different from J .

In the same example, if regimen A is a control group and we knew beforehand that all of the other regimens had to be at least as good as the control or better, it may be desired to select those regimens that are statistically, significantly better. We would proceed as follows:

$$t_d = 2.60 \quad \text{from table 12.23 for } k = 9 \text{ treatments } 20 \text{ DF,}$$

$$\alpha = 0.10 \quad (\text{this corresponds to a one-tail } P = 0.05)$$

$$5\% \text{ allowance} = t_d \sqrt{s^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} = 2.60 \sqrt{2.17 \cdot (1/3 + 1/3)} = 3.12$$

Thus, any regimen mean that is larger than the mean for regimen A by 3.12 or more is judged to be better than A .

It can be concluded that F , G , D , H , and E showed a statistically significantly better weight gain than A , and that there is insufficient evidence to indicate that B , C , I , and J were any better than A .

OTHER ANOVA DESIGNS COMMON TO PHARMACEUTICAL PROBLEMS

A somewhat more complex design is the *two-way* ANOVA. This design is analogous to the paired t test but consists of more than two treatments; i.e., more than one treatment is applied to the same experimental unit (e.g., patient) or related units (e.g., litter mates, males between 50 and 60 years, etc.). This design has the same advantages and disadvantages as the paired t test described earlier in this chapter. The ANOVA table is similar to the one-way table but includes some new terms. The between-treatments term has the same interpretation as that in the one-way analysis, representing differences between treatments. A new term, between rows, represents the variability of the units to which the treatments have been applied (e.g., patients). Finally, the table contains an error term, sometimes referred to as row $\times$ treatment interaction (patient $\times$ drug in a clinical trial).

Table 10-23. The t_p Table Values of t_p for Dunnett's Procedure for Comparing Several Treatments With a Control at the 5% Level of Significance (Use $P = 0.10$ Values for a One-Tailed Test and $P = 0.05$ Values for a Two-Tailed Test.)

k (Number Treatments, Excluding the Control)									
DF	p	2	3	4	5	6	7	8	9
10	0.10	2.15	2.34	2.47	2.56	2.64	2.70	2.76	2.81
	0.05	2.57	2.76	2.89	2.99	3.07	3.14	3.19	3.24
11	0.10	2.13	2.31	2.44	2.53	2.60	2.67	2.72	2.77
	0.05	2.53	2.72	2.84	2.94	3.02	3.08	3.14	3.19
12	0.10	2.11	2.29	2.41	2.50	2.58	2.64	2.69	2.74
	0.05	2.50	2.68	2.81	2.90	2.98	3.04	3.09	3.14
13	0.10	2.09	2.27	2.39	2.48	2.55	2.61	2.66	2.71
	0.05	2.48	2.65	2.78	2.87	2.94	3.00	3.06	3.10
14	0.10	2.08	2.25	2.37	2.46	2.53	2.59	2.64	2.69
	0.05	2.46	2.63	2.75	2.84	2.91	2.97	3.02	3.07
15	0.10	2.07	2.24	2.36	2.44	2.51	2.57	2.62	2.67
	0.05	2.44	2.61	2.73	2.82	2.89	2.95	3.00	3.04
16	0.10	2.06	2.23	2.34	2.43	2.50	2.56	2.61	2.65
	0.05	2.42	2.59	2.71	2.80	2.87	2.92	2.97	3.02
17	0.10	2.05	2.22	2.33	2.42	2.49	2.54	2.59	2.64
	0.05	2.41	2.58	2.69	2.78	2.85	2.90	2.95	3.00
18	0.10	2.04	2.21	2.32	2.41	2.48	2.53	2.58	2.62
	0.05	2.40	2.56	2.68	2.76	2.83	2.89	2.94	2.98
19	0.10	2.03	2.20	2.31	2.40	2.47	2.52	2.57	2.61
	0.05	2.39	2.55	2.66	2.75	2.81	2.87	2.92	2.96
20	0.10	2.03	2.19	2.30	2.39	2.46	2.51	2.56	2.60
	0.05	2.38	2.54	2.65	2.73	2.80	2.86	2.90	2.95
24	0.10	2.01	2.17	2.28	2.36	2.43	2.48	2.52	2.57
	0.05	2.35	2.51	2.61	2.70	2.76	2.81	2.86	2.90
30	0.10	1.99	2.15	2.25	2.33	2.40	2.45	2.50	2.54
	0.05	2.32	2.47	2.58	2.66	2.72	2.77	2.82	2.86
40	0.10	1.97	2.13	2.23	2.31	2.37	2.42	2.47	2.51
	0.05	2.29	2.44	2.54	2.62	2.68	2.73	2.77	2.81
60	0.10	1.95	2.10	2.21	2.28	2.35	2.39	2.44	2.48
	0.05	2.27	2.41	2.51	2.58	2.64	2.69	2.73	2.77
120	0.10	1.93	2.08	2.18	2.26	2.32	2.37	2.41	2.45
	0.05	2.24	2.38	2.47	2.55	2.60	2.65	2.69	2.73
∞	0.10	1.92	2.06	2.16	2.23	2.29	2.34	2.38	2.42
	0.05	2.21	2.35	2.44	2.51	2.57	2.61	2.65	2.69

(Data from Dunnett CW. A multiple comparisons procedure for comparing several treatments with a control. *Am Stat Assoc J* 1955; 50: 1096–1121.)

The treatment mean square is divided by the error mean square (EMS) to form an F ratio, for purposes of performing a statistical test. Some complications can exist in the interpretation of this table and the F ratios. The examples here consider treatments as including all treatments of interest, and rows as a random selection of experimental units taken from a large population of such units.

For example, to compare a placebo, a generic drug, and a standard drug (three treatments), use a random selection of patients as the experimental units, with each patient to take each of the three treatments. Another example is the comparison of five analytical methods (five treatments), in which ten analysts, selected at random, each perform assays with each method.

CROSSOVER DESIGN

A design that is popular in experimental research is the crossover design. This is in the class of paired-sample or two-way designs in that all treatments are applied to each experimental unit. For example, in practically all human bioequivalence studies, each subject takes all of the treatments. That is, if a

control marketed drug is to be compared to two new formulations, each subject takes all three products.

The difference between the crossover and the two-way design (also known as a randomized block design) is that in the two-way design, the order or placement of treatments are assigned randomly to each patient. In the crossover design, an additional constraint, *order* or *balance*, is imposed on the experiment. For example, in a bioequivalence study of three products, these are taken sequentially during three periods. In the crossover design, each product appears an equal number of times in each period.

Table 10-26 shows how three products, A , B , and C , may be assigned to nine subjects in a bioavailability study. Note that treatments A , B , and C appear exactly three times in each period and that each subject takes all three products. The balancing of order of administration compensates for period effects. If any extraneous variables affect the outcome differently in one period compared to another, all treatments may be affected equally. This would result in a fair comparison of the different treatments. In a purely random assignment of treatments, it would be unlikely that treatments would be assigned in such a balanced

Example 20.

Three variations of an acne preparation and a control are to be tested for skin irritation. The four products, **A**, **B**, **C**, and the control, each are applied to sites on the backs of eight patients. The assignment of the four products to the four sites on the patient is random; i.e., a random assignment of treatments to the four sites on each patient is done for each patient, using a random-number table. The products are applied, and after 24 hours, the degree of irritation is determined by assessing irritation subjectively on a scale of one to ten. A value of one means no irritation, and a value of ten means extreme irritation. The results are shown in Table 10-24.

Table 10-24. Skin Irritation Test

Patients	Treatment				Σx	Σx^2
	A	B	C	Control		
1	7	5	5	4	21	115
2	4	3	5	2	14	54
3	8	9	7	6	30	230
4	8	6	4	5	23	141
5	7	7	4	2	20	118
6	6	7	5	4	22	126
7	5	6	4	5	20	102
8	4	7	5	4	20	106
Totals	49	50	39	32	170	992

The computations are similar to those for the one-way ANOVA. The sum-of-squares for treatments is obtained as before. The sum-of-squares for patients is determined exactly as for treatments except the operation is across rows. This is the same as rotating the table by 90 and treating the rows as columns in the table matrix. The EMS (expected sum of squares) is obtained by subtracting the row and column sum-of-squares from the total sum-of-squares.

$$\text{Total SS} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}, \text{ where } (\sum x_i)^2 / n = CT$$

$$= 992 - 170^2/32 = 88.875$$

$$\text{Between treatments SS} = \frac{49^2 + 50^2 + 39^2 + 32^2}{8} - \frac{170^2}{32}$$

$$= 27.625$$

$$\text{Between patients SS} = \frac{21^2 + 14^2 + \dots + 20^2}{4} - CT \frac{3705}{4} - CT$$

$$= 23.125$$

$$\text{Error} = \text{Total SS} - \text{Between treatments SS} - \text{Between subjects SS}$$

$$= 88.875 - 27.625 - 23.125 = 38.125$$

Table 10-25 shows the ANOVA. Since the F ratio for treatments (5.07) exceeds the tabled F value with 3 and 21 DF at the 5% level, it can be concluded that at least two of the treatments differ. Although one may apply one of the a posteriori tests discussed under one-way ANOVA, inspection of the results suggests that results for treatments **A** and **B** are similar and both are greater in magnitude than treatments **C** and the control.

Table 10-25. ANOVA for Data of Table 10-24

Analysis of Variance				
Source of Variation	DF	Sums-of-Squares	Mean Square	F Ratio
Between Treatments	3	27.625	9.203	5.07
Between patients	7	23.125	3.304	
Error	21	38.125	1.815	
Total	31	88.875		

Table 10-26. Example of Crossover Design

Subject	Period 1	Period 2	Period 3
1	B	C	A
2	A	C	B
3	B	A	C
4	C	B	A
5	A	B	C
6	C	A	B
7	B	A	C
8	C	B	A
9	A	C	B

Example 21.

Three drug formulations were administered to nine subjects in a bioavailability study, according to the crossover design illustrated in Table 10-26. The area under the blood level curves were computed for each dosing, and the results are shown in Table 10-27.

Table 10-27. Results of Bioavailability Study

Subject	Period 1	Period 2	Period 3	Sum
1	B = 107	C = 102	A = 99	308
2	A = 100	C = 106	B = 89	295
3	B = 98	A = 90	C = 128	316
4	C = 71	B = 54	A = 63	188
5	A = 92	B = 111	C = 107	310
6	C = 113	A = 115	B = 91	319
7	B = 169	A = 187	C = 195	551
8	C = 88	B = 95	A = 77	260
9	A = 122	C = 168	B = 155	445
Period sum	I: 960	II: 1028	III: 1004	2992
Treatment sum	A: 945	B: 969	C: 1078	
Treatment average	105	107.7	119.8	

The ANOVA (Table 10-28) separates the total variance into four parts: subjects, period (order of administration), treatments, and error.

Table 10-28. ANOVA for Bioavailability Study

Analysis of Variance				
Source of Variation	DF	Sums-of-Squares	Mean Square	F Ratio
Between subjects	8	29834.1	3729.3	
Between treatments	2	1116.5	558.3	3.15
Order	2	264.3	132.1	0.75
Error	14	1947.2	177.0	
Total	26	33162.1		

$$\sum x_i = 2992 \quad \sum x_i^2 = 364,720$$

$$\text{Total SS} = \sum x_i^2 - \frac{(\sum x_i)^2}{n} = 33,162.1$$

$$\text{Subject SS} = \frac{\sum \text{row}^2}{3} - \frac{(\sum x_i)^2}{n} = 29,834.1$$

$$\text{treatment SS} = \frac{\sum \text{treat.sum}^2}{9} - \frac{(\sum x_i)^2}{n} = 1116.5$$

$$\text{Order SS} = \frac{\sum I^2 + \sum II^2 + \sum III^2}{9} - \frac{(\sum x_i)^2}{n} = 264.3$$

$$\text{Error SS} = \text{Total SS} - \text{Subject SS} - \text{treatment SS} - \text{Order SS} = 1947.2$$

Neither treatments nor order are significant (see Table 10-15). For 2 and 14 DF, an *F* value of 3.70 is needed for significance. Treatment C has a higher average result but fails to reach significance in this study. In the early days of bioequivalence testing, bioequivalence studies were designed to have a power of 0.8 to detect a difference of 20% between treatments. This means that a sufficient number of subjects should be included in the study so that if a true difference of 20% or more exists between two treatments, there will be at least an 80% chance of finding a significant difference. This method of evaluating equivalence has been replaced by a more meaningful confidence interval approach.⁸

order. In an unbalanced design, differences due to periods would not affect treatments equally, resulting in a potential bias and a larger experimental error – the experimental error would include the usual causes of variability and also variability due to period effects. Thus, the crossover design can be considered an improvement over the two-way design in that the error has been reduced and the experiment made more efficient.

Many such designs are available, but care should be exercised to apply the correct design to each experimental situation. The crossover design is related to the Latin square design. Several very good references are available on principles of experimental design. In particular, the book by Cox⁶ is recommended because it is not overly technical and can be understood without resorting to too much mathematics.

Example 22.

A pilot study to compare the effects of an antihypertensive drug versus placebo was designed with four patients on drug and four on placebo. Blood pressure changes from baseline were measured for six weeks at biweekly intervals. The results are shown in Table 10-29.

Table 10-29. Reduction in Diastolic Blood Pressure from Baseline

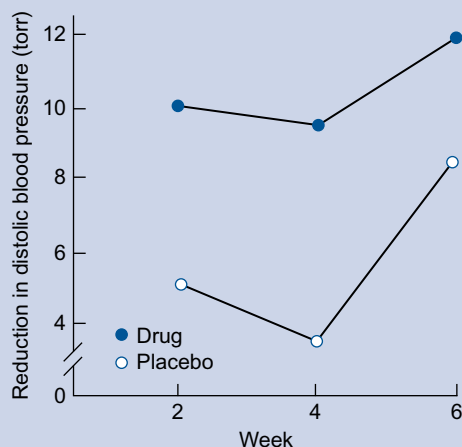
Drug Week				Placebo Week			
Patient	2	4	6	Patient	2	4	6
1	10	8	12	2	10	8	12
3	8	6	14	5	6	2	10
4	12	14	8	6	4	0	2
7	10	10	14	8	0	4	10
Average	10.0	9.5	12.0		5.0	3.3	8.5

The ANOVA is shown in Table 10-30. The terms of interest are *treatments* and *treatment × times*. The former term measures differences of the overall average results of the two treatments. The error term for treatments is the mean square for patients. The treatment × times term compares the time trends for the two treatments. The error term for the treatment × times is patient × times (treatments). If the trends are parallel, this term will not be significant. Significance for this term indicates a lack of parallelism, suggesting that differences between treatments depend on the time of observation.

Table 10-30. ANOVA for Example 22

Source	DF	SS	MS	F
Patients	6	109	18.2	
Treatments	1	140.2	140.2	7.7
Times	2	60.3	30.2	3.5
Treatment $\times$ times	2	6.3	3.2	0.4
Patient $\times$ times (treatments)	12	104	8.7	
Total	23	419.8		

As with most experimental data, a graphic display is recommended. Figure 10-8 is a plot of the average results versus time. The significant difference between treatments ($\alpha=0.05$) is apparent from the plot and the ANOVA. The time trends of both treatments are similar and can be explained by the experimental variability (treatment $\times$ times is not significant).

**Figure 10-8.** Plot of average results for Example 22.

If the crossover design becomes unbalanced, due to drop-outs or other conditions, a computer analysis can be used (e.g., SAS).

Another experimental design common in clinical trials is the repeated-measures design, often called a *split-plot design*. For example, two treatments are compared by making observations in two independent groups of patients over time. Although an equal number of patients in each group is desirable, it is not necessary for the data analysis. The observations are made at the same time periods in both groups. The example shows the basic design and ANOVA table. The details of the calculations are not shown. Usually, a software program is used to analyze and summarize the data. The details of the analysis are given in Bolton⁸ and Winer¹⁴.

NONPARAMETRIC TESTS OF SIGNIFICANCE

The validity of the t test for comparing two means depends to some extent (especially for small samples) on the assumptions that the two populations sampled are distributed approximately normally and have essentially equal variances. A procedure for testing the equality of variances has been discussed previously. Statistical procedures that do not depend on the assumption of normality or any other distribution are called nonparametric tests. Three commonly used procedures are the rank sum test for unpaired data (Rank Sum Test of Significance), the *Signed-Rank Sum Test of Significance*, and the *Sign Test* for paired data.

• **Rank Sum Test of Significance** The rank sum test of significance is the nonparametric analog of the two-independent sample t test. The n_1 and n_2 observations are taken from two independent groups. After the n_1 and n_2 observations are arranged in order of size, the combined values are ranked from 1, for the lowest, to (n_1+n_2) for the highest, and the sum of the ranks T of the n_1

Example 23.

Data were available on the duration of loss of the righting reflex (minutes) for 10 mice given a standard barbiturate and for 11 mice given a test barbiturate (Table 10-32). Adding the respective ranks for standard drug and test drug we find $T = 150$ and $T' = n_1(n_1 + n_2 + 1) - T = 10(10 + 11 + 1) - 150.5 = 69.5$. Entering Table 10-31 with $n_1=10$ and $n_2=11$, we find that the calculated T' value 69.5 is less than the tabulated value 73 for $\alpha=0.01$. Therefore, because the probability of the standard drug and test drug values being the same is less than 0.05 (actually, it is less than 0.01), it is concluded that they are different. This test compares the medians of the two-populations sampled. The median of an ordered set of observations is defined as the middlemost value for an odd number of observations, and as the average of the two middlemost values for an even number of observations. Thus, the median for the standard drug is, $(130+148)/2=139$ and the median for the test drug is 103.

Example 24.

The procedure is illustrated for data given in Example 12 (Table 10-34). Entering Table 10-33 with $n = 15$ and $T = 22.5$, we find that the calculated T value 22.5 is less than the tabulated value 25 for $\alpha=0.05$. Therefore, because the probability of the morning and afternoon values being the same is less than 0.05, it is concluded that they are different.

observations in the smaller sample is computed. Values that are tied are given average ranks. Also calculate $T' = n_1(n_1 + n_2 + 1) - T$, and enter Table 10-31¹⁵ with n_1 , n_2 , and T or T' , whichever is smaller. If the calculated T (or T') is equal to or less than the tabulated value, the null hypothesis is rejected at the significance level P .

• **Signed-Rank Sum Test of Significance** The signed-rank sum test of significance is the nonparametric analog of the paired t test. The differences between the n paired values are ranked in order of absolute size from 1, for the lowest, to n , for the highest, ignoring zero differences. Tied values are assigned

Example 25.

The procedure is illustrated for the data given in Examples 12 and 24.

b = number of positive differences = 11

c = number of negative differences = 4

$$\chi^2 = \frac{(b - c - 1)^2}{b + c} = \frac{(11 - 4 - 1)^2}{11 + 4} = \frac{36}{15} = 2.40$$

Table 10-12 shows that for 1 DF the probability of getting a value of χ^2 larger than the calculated value 2.40 is between $\alpha=0.10$ and $\alpha=0.20$. Since P is not equal to or less than 0.05, it is concluded that there is insufficient evidence to indicate that the morning and afternoon values are different. This conclusion is not in agreement with that of the t test and the signed-rank test. The reason for this is that the statistical sign test considers only the sign of the difference and not the magnitude and, thus, is a less-sensitive test in borderline situations such as this one.

Table 10-31. The Rank Sum Table Values of T or T', Whichever is Smaller, Significant at the 10%, 5%, and 1% Levels

n1 (Smaller Sample)																			
n2	P	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
8	0.10	15	23	31	41	51													
9	0.10	16	24	33	43	54	66												
	0.05	14	22	31	40	51	62												
	0.01	11	18	26	35	45	56												
10	0.10	17	26	35	45	56	69	82											
	0.05	15	23	32	42	53	65	78											
	0.01	12	19	27	37	47	58	71											
11	0.10	18	27	37	47	59	72	86	100										
	0.05	16	24	34	44	55	68	81	96										
	0.01	12	20	28	38	49	61	73	87										
12	0.10	19	28	38	49	62	75	89	104	120									
	0.05	17	26	35	46	58	71	84	99	115									
	0.01	13	21	30	40	51	63	76	90	105									
13	0.10	20	30	40	52	64	78	92	108	125	142								
	0.05	18	27	37	48	60	73	88	103	119	136								
	0.01	14	22	31	41	53	55	79	93	109	125								
14	0.10	21	31	42	54	67	81	96	112	129	147	166							
	0.05	19	23	38	50	62	76	91	106	123	141	160							
	0.01	14	22	32	43	54	67	81	96	112	129	147							
15	0.10	22	33	44	56	69	84	99	116	133	152	171	192						
	0.05	20	29	40	52	65	79	94	110	127	145	164	184						
	0.01	15	23	33	44	56	69	84	99	115	133	151	171						
16	0.10	24	34	46	58	72	87	103	120	138	156	176	197	219					
	0.05	21	30	42	54	67	82	97	113	131	150	169	190	211					
	0.01	15	24	34	46	58	72	86	102	119	136	155	175	196					
17	0.10	25	35	47	61	75	90	106	123	142	161	182	203	225	249				
	0.05	21	32	43	56	70	84	100	117	135	154	174	195	217	240				
	0.01	16	25	36	47	60	74	89	105	122	140	159	180	201	223				
18	0.10	26	37	49	63	77	93	110	127	146	166	187	208	231	255	280			
	0.05	22	33	45	58	72	87	103	121	139	158	179	200	222	246	270			
	0.01	16	26	37	49	62	76	92	108	125	144	163	184	206	228	252			
19	0.10	27	38	51	65	80	96	113	131	150	171	192	214	237	262	287	313		
	0.05	23	34	46	60	74	90	107	124	143	163	182	205	228	252	277	303		
	0.01	17	27	38	50	64	78	94	111	129	147	168	189	210	234	258	283		
20	0.10	28	40	53	67	83	99	117	135	155	175	197	220	243	268	294	320	348	
	0.05	24	35	48	62	77	93	110	128	147	167	188	210	234	258	283	309	337	
	0.01	18	28	39	52	66	81	97	114	132	151	172	193	215	239	263	289	315	

(Adapted from Tate MW, Clelland RC. Nonparametric and Shortcut Statistics. Danville IL: Interstate Print. 1957: 137.)

Table 10-32. Data for Example 23

Standard Drug	Rank	Test Drug	Rank
96	4.5	0	1
109	8	91	2
126	13	92	3
130	15	96	4.5
130	15	99	6
148	17	103	7
153	18	117	9
158	19	118	10
169	20	119	11
Died	21	120	12
	$T = 150.5$	130	15
	$n_1 = 10$		$n_2 = 11$

$$T' = n_1(n_1 + n_2 + 1) - T = 10(10 + 11 + 1) - 150.5 = 69.5$$

Table 10-33. The Signed-Rank Sum Table Values of T for Signed-Rank Test, Significant at the 10%, 5%, and 1% Levels

p				p			
n	0.10	0.05	0.01	n	0.10	0.05	0.01
5	0			18	47	40	27
6	2	0		19	53	46	32
7	3	2		20	60	52	37
8	5	3	0	21	67	58	43
9	8	5	1	22	75	65	49
10	10	8	3	23	83	73	55
11	14	10	5	24	91	81	61
12	17	13	7	25	100	89	68
13	21	17	9	26	110	97	75
14	25	21	12	27	120	106	83
15	30	25	16	28	130	116	91
16	35	29	19	29	141	126	100
17	41	34	23	30	152	136	109

an average rank. After the differences are ranked, the signs of the differences are attached to the ranks, and the sum of the positive ranks and of the negative ranks are obtained. Enter Table 10-33, with n being the number of non-zero differences and the sum T of positive or negative ranks, whichever is smaller. When the calculated T is equal to or less than the tabulated T , the null hypothesis is rejected at the significance level P .

• **Sign Test** The sign test is also used for paired data, but it is not as powerful as the signed-rank test; it is more difficult to find significant differences when they exist with the sign test. Count the number of positive differences (b) and the number of negative differences (c), ignoring zero differences, and calculate:

$$\chi^2 = \frac{(b - c - 1)^2}{b + c}$$

This is referred to the chi-square table (see Table 10-12) with $DF = 1$, the test being essentially the same as the chi-square test illustrated in Example 16.

EXACT TESTS

Statistical distribution functions enable us to calculate the probability of certain events under the null hypothesis -- but only if their prerequisites hold. And even then they are sometimes only approximations to the real distribution of the test

Table 10-34. Signed Ranks from Example 24

Differences	Signed-Ranks
2	2
-3	-3
4	6
8	15
-4	-6
6	12.5
-1	-1
-6	-12.5
5	10
4	6
4	6
0	Ignore
5	10
7	14
4	6
5	10
Sum of positive ranks = 97.5	
Sum of negative ranks = 22.5 = T	
$n = 15$	

statistic. For example, in a Mann-Whitney U test, the tabulated critical values cannot be used if the number of samples is too small or there are too many ties (i.e., a group of subjects for which the experiment results in the same value). In certain models of diseases, however, (e.g., experimental autoimmune encephalomyelitis, see below) only a few different outcomes are defined, and most of the animals will show results of moderate severity. So in this kind of experiment, many ties will occur, and the prerequisites for the Mann-Whitney are not met. Figure shows the theoretical distribution of the test statistic U of the Mann-Whitney test under the null hypothesis (which is listed in statistic books or calculated by statistic software), compared with the real distribution if ties are present.

However, the computing power of modern computers allows us even to find the exact distributions of events under the null hypothesis. The kind of statistical tests that are based on these exact distribution data are called *permutation tests*. Their working principle is as follows:

Let us suppose that the null hypothesis holds -- so for each subject the result of an experiment is not due to any treatment effects. In other words: Any subject from the control group would have shown the same result as if it had been assigned to the treatment group and vice versa. Thus, the exact null distribution can be found by consideration of all possible assignments of the subjects to the respective treatment groups.

As an example, we take the data from table. It shows severity scores of experimental autoimmune encephalomyelitis in mice. The animals in group 2 were treated with a drug candidate, while group 1 is a control group without medication. We see that the median in group 2 is lower than in group 1, but is this difference statistically significant?

It is common practice in pharmaceutical development to look for positive or negative impacts of the treatment. Therefore, we are going to perform a two-tailed test. So we state the null hypothesis H_0 :

The treatment has no influence on the severity of the EAE symptoms. (Differences in the severity scores are just caused by chance.)

The alternative hypothesis H_0 for the two-tailed test is:

The treatment leads to either an amelioration or aggravation of the symptoms.

We will reject the null hypothesis, if the chance to find results that are at least as extreme as the actual result of the experiment

Table 10-35.

Score	Group 1	Group 2	Total
0			2
0.5			0
1			9
1.5			0
2			9
2.5			0
3			4

is less than $\alpha = 5\%$ (significance level). If the chances that the difference in the two groups is merely due to chance is greater than α , we do not reject to the null hypothesis. However, we must bear in mind that the null hypothesis may actually be wrong and that the experiment (and the statistical evaluation) is just too weak to show the significance of the differences (type II error).

If we add up the ranks for the observations in group 1 to the sum R_1 and the ranks for group 2 to R_2 we find the test characteristic U of the Mann-Whitney test, with n_1 and n_2 denoting the number of observations in both groups as the smaller of the two values:

$$U_1 = R_1 - \frac{n_1(n_1 + 1)}{2} \quad \text{and} \quad U_2 = R_2 - \frac{n_2(n_2 + 1)}{2}.$$

In the absence of ties, precalculated tables can be used to obtain critical values of the U distribution for a given level of significance. For large samples, U is approximately normally distributed with mean $mn/2$ and variance $mn(m+n+1)/12$.

However, if we look at the counts of the scores of this example we find a lot of ties (Table 10-35).

In fact, we observed only four different ranks:

The rank sums for each treatment group are $R_1 = 191$ and $R_2 = 109$, so $U = 109 - 78 = 31$.

Since the tabulated U distribution is not applicable in the presence of so many ties, we need to compute our p -value from the *exact distribution* of the U values.

The general idea is that with the null hypothesis being valid, it does not matter which of the 24 animals were assigned to the control group. If there is no effect from therapy, every animal from control and treatment group would have shown the same severity score regardless of which group it was assigned to. Thus every possible combination of 12 animals out of the 24 animals that participate in the test leads to an U -value that would have occurred if only the random assignment to the control and treatment group had lead to a different partitioning. These U -values form the null distribution from which we can judge the significance of our result.¹⁶

In total there are $\binom{24}{12} = 2704156$ different possibilities (combinations) to arrange 24 animals in two groups of twelve. Though this number is fairly high, modern computers calculate the U -values for each of them in an instant. (However, computing time rises exponentially with the number of subjects, so there are still limits for the use of permutations tests.)

The exact null distribution is shown in figure together with the conventional U distribution. We find that the ranks in group 2 are significantly lower than those in group 1 ($p = 0.016$).

REJECTION OF ABERRANT OBSERVATIONS

It is common practice among chemists and others working in the physical sciences to make observations in duplicate or triplicate. This is usually done for the purpose both of obtaining a more accurate result and detecting mistakes in dilution, weighing, and so on. It is quite a common practice to reject the most extreme of the three results if it appears to disagree with the others.

Youden,^{17,18} a chemist as well as a statistician, made a study of the problem of rejection of observations in an attempt to answer three questions:

1. If the extreme observation of triplicates is always rejected when only normal variation is present, how accurate is the result?
2. Is the average of the two closest observations as good an estimate as the average of all three?
3. By how much should the outlying observation of triplicates differ from the other two in order to be reasonably assured that this difference is due to a blunder rather than normal variation?

He found that rejection of the outlying observation resulted not only in the variation being greatly underestimated but in the mean being biased.

If one wished to follow a simple rule of rejection^{17,18} of observations in samples of three, so as to reject not more than 5% of the extreme observations arising from normal variation, a rejection ratio of D/d greater than 20 would be required.

$D/d > 20$ where D is the difference between the most extreme observations and its closest neighbor and d is the difference between two closest observations

In the USP there is an excellent chapter on the design and analysis of biological assays, in which are included some tests for rejection of outlying observations. These and other tests can be applied to chemical as well as biological assays.¹⁹ Two criteria are presented here, one for rejecting single suspect observations in one group and the other for rejecting a whole group of observations.

To use the first criterion, arrange the observations in the group in order of their magnitude, and number them from 1 to n , beginning with the supposedly erratic or outlying observation, thus:

$$y_1, y_2, y_3, \dots, y_n$$

where y_1 is the suspect observation. If there are 3 to 7 observations in the group, calculate:

$$G_1 = \frac{y_2 - y_1}{y_n - y_1}$$

If there are 8 to 10 observations in the group, and the smallest value seems suspect, again arrange them in order from lowest to highest, and calculate:

$$G_2 = \frac{y_2 - y_1}{y_{n-1} - y_1}$$

If there are 11 to 13 observations, follow the same procedure, but use the statistic:

Table 10-36.

rank position	1	2	3	4	5	6	7	8	9	10	11	12	13	14		18	19	20	21	22	23	24
score	0	0	1	1	1	1	1	1	1	1	1	2	2	2	...	2	2	2	2	3	3	3
rank	1.5		7								16								22.5			

Table 10-37. Criteria for Testing Extreme Value

Statistic	n_i Number of Observations	Critical Values
$G_1 = \frac{y_2 - y_1}{y_n - y_1}$	3	.988
	4	.889
	5	.780
	6	.698
	7	.637
$G_2 = \frac{y_2 - y_1}{y_{n-1} - y_1}$	8	.683
	9	.635
	10	.597
$G_3 = \frac{y_3 - y_1}{y_{n-2} - y_1}$	11	.679
	12	.642
	13	.615
$G_4 = \frac{y_3 - y_1}{y_{n-2} - y_1}$	14	.641
	15	.616
	16	.595
	17	.577
	18	.561
	19	.547
	20	.535
	21	.524
	22	.514
	23	.505
	24	.497
	25	.489

Example 26.

Suppose among the gains in weight of six rats after a feeding experiment, one weight was found to be much less than the other five. Can that observation be discarded? The six gains in weight are 36, 40, 38, 42, 20, and 39.

Rearrange these in order from smallest to largest and label $y_1, \dots, y_6$, where $n=6$.

$$y_1 = 20, y_2 = 36, y_3 = 38, y_4 = 39, y_5 = 40, y_6 = 42$$

$$G_1 = \frac{y_2 - y_1}{y_6 - y_1} = \frac{36 - 20}{42 - 20} = \frac{16}{22} = 0.727$$

Referring to the value of G_1 for $n=6$ in the table, $G_1 = 0.698$ for $\alpha=0.01$. Since the calculated value of G_1 is larger than this value, reject the value of 20 and work with the remaining five values.

$$G_3 = \frac{y_3 - y_1}{y_{n-1} - y_1}$$

If there are 14-25 observations, follow the same procedure, but use the statistic:

$$G_4 = \frac{y_3 - y_1}{y_{n-2} - y_1}$$

If the largest value is open to suspicion as possibly being aberrant, arrange the observations in order from highest to lowest and number them, always labeling the suspect observation y_1 .

If the calculated value of G_1, G_2, G_3 , or G_4 is larger than the tabled value (which gives the probability of a value being so extreme as that observed), it can be assumed that the observation truly does not belong to the group, and the observation is rejected. The values of G for a probability $\alpha=0.01$, that an outlier could occur at either end are shown in Table 10-35. This same criterion could be used for testing whether the largest or smallest average in a group of averages differs significantly from the remainder of the averages (Table 10-37).

The second criterion for an aberrant observation, as given in the USP, compares the variation or range between various groups. It is a test for the homogeneity of the ranges (the range is again the highest value in a group minus the lowest value) and is for the purpose of locating outliers within one group of values. This method and its accompanying table are presented in considerable detail in the USP. The rejection of outliers, using only statistical criteria, is controversial. A knowledge of the characteristics or properties of the chemical or biological systems being studied should be used when making decisions to reject outlying data.

QUALITY CONTROL METHODS

A very short explanation is given here regarding the quality control methods that were developed primarily by Dr Walter Shewhart of the Bell Telephone Laboratories. A more complete explanation can be found in two short publications of the American Standards Association^{20,21} and many texts, including Dixon and Massey.²²

The quality control method for variables involves plotting the data as dots on a graph, with the variable measured on the vertical axis, and time (hours, days, etc.) on the horizontal axis. The control is maintained by inserting, on the chart, the grand average and control limits that have been calculated from accumulated experience and drawn on the chart as parallel horizontal lines, as shown in Figure 10-9. When all the dots fall within the limits, the results are said to be in a state of statistical control. When a dot falls outside the limits, a potential problem is indicated.

In a control chart, usually each dot is an average for a sample consisting of, say, four observations. The standard error of the average then is calculated for each group of four observations, and an average value for the standard error of the average is obtained. This is designated by $\bar{\sigma}_x$. The grand average of all the averages plotted also is calculated and is labeled $\bar{x}$. The 3-sigma control limits used on the control chart can be obtained from:

$$\text{Upper limit} = \bar{\bar{x}} + 3\bar{\sigma}_x$$

$$\text{Upper limit} = \bar{\bar{x}} + 3\bar{\sigma}_x$$

Thus, it can be seen that the control-chart technique is a graphic means of investigating whether the variation exhibited over a very short period of time is the same as the variation that

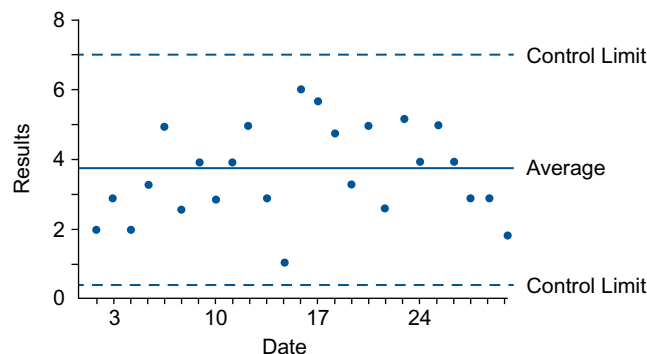
**Figure 10-9.** A typical quality control chart.

Table 10-38. Calculation of Standard Deviation from Range

Size of Sample (n)	Average Number of Standard Deviations in T-1E Average Range (d)	Size of Sample (n)	Average Number of Standard Deviations in the Average Range (d)
2	1.128	7	2.704
3	1.693	8	2.847
4	2.059	9	2.970
5	2.326	10	3.078
6	2.534		

Table 10-39. Factors for 3-Sigma Limits^a

Size of Sample (n)	Factor for $\bar{R}$ Chart		Factor for $\bar{X}$ Chart A_2
	D_2	D_4	
2	0	3.27	1.880
3	0	2.57	1.023
4	0	2.28	0.729
5	0	2.11	0.577
6	0	2.00	0.483
7	0.08	1.92	0.419
8	0.14	1.86	0.373
9	0.18	1.82	0.337
10	0.22	1.78	0.308

^a This table contains data from the tables in Appendix 1 of Z1.3—1958.²⁰

occurs over a long period of time. If the two variations are identical and all of the plotted dots fall within the control limits, the experiments or processes that produced the data are said to be in a state of *statistical control*.

For many pharmaceutical processes, particularly heterogeneous processes, typical Shewhart control charts do not describe the process adequately. The process seems to not be in control. In these cases, alternative methods should be considered.²³ Control charts are a valuable tool for process validation.

It is possible to calculate the control limits by using the range in each group of four instead of calculating the standard deviation. This is because, on the average, for samples of less than ten, the range and the standard deviation are related very closely. Given the number of observations in the sample, the standard deviation can be calculated by dividing the range by the appropriate figure given in Table 10-38 for the size of the sample, n . The factors for calculating 3-sigma limits from the range are given as column A_2 in Table 10-39.

Example 27.

A drug manufacturer keeps a record of the uniformity of the machine that is filling a given weight of a drug into ampuls. Samples of the finished product are taken at definite time intervals. The data are accumulated and arranged into groups of four ampuls according to the order in which they were taken from a filling machine. The average and the range are computed for each group of four, as given in Table 10-40, according to the time the samples are taken. The resulting quality-control charts are shown in Figure 10-10.

Table 10-40. Calculations for a Quality-Control Chart On Averages and Ranges for Samples of 4 from a Filling Machine

Time	Average (G)	Range (G)	Time	Average (G)	Range (G)
Jan 6			Jan 7		
8 AM	38.1	1.5	8 AM	37.6	2.1
9 AM	37.6	2.1	9 AM	39.1	1.4
10 AM	38.3	1.1	10 AM	38.5	1.1
11 AM	36.5	2.4	11 AM	37.7	1.9
12 M	38.9	3.1	12 M	38.1	2.3
1 PM	37.8	2.8	1 PM	38.5	2.4
2 PM	38.5	1.7	2 PM	37.6	1.6
3 PM	39.4	1.6	3 PM	37.9	1.8
4 PM	36.4	2.5	4 PM	38.6	1.0

Grand average = $\bar{X} = 38.1$

Average range = $\bar{R} = 1.9$

Control limits^a for average = $\bar{X} \pm A_2\bar{R} = 38.1 \pm 0.729(1.9)$

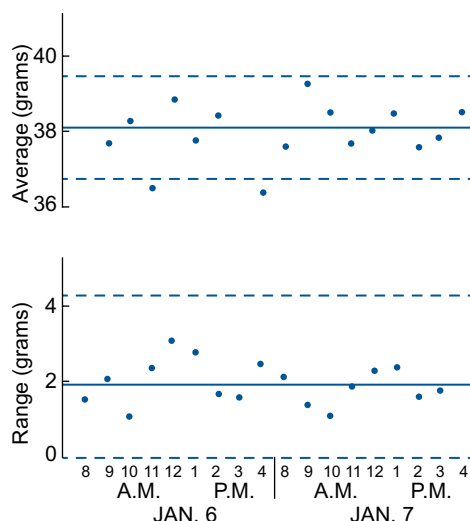
Upper limit = 39.49

Lower limit = 36.71

Control limits^b for range, are $D_3\bar{R}$ or 0(1.9) and $2.28(1.9)$ which equal 0 and 4.33, respectively.

^a A_2 is the factor for using the range to calculate 3-sigma limits for the average (i.e., 3 times the standard error of the average). See Table 10-37 for $N=4$.

^b D_3 and D_4 are the factors for using the range to calculate 3-sigma limits for the range (i.e., 3 times the standard error of the range). These values are taken from Table 10-39. In two instances the point plottings fell below the lower control limit, indicating a lower average fill than one might expect, i.e., there is a lack of statistical control.

**Figure 10-10.** Quality control charts for data from Table 10-40.

Control charts using 3-sigma limits can be obtained by the use of figures given in Table 10-37. With $\bar{R}$ being the average range, the formulas are:

$$\text{Upper limit for average} = \bar{\bar{x}} + A_2 \bar{R}$$

$$\text{Lower limit for average} = \bar{\bar{x}} - A_2 \bar{R}$$

$$\text{Upper limit for ranges} = D_4 \bar{R}$$

Example 28.

A department head in the capsule department of a pharmaceutical company keeps a record of the number of defective capsules found in sections of large lots of capsules (Table 10-41). Each section consists of approximately 19000 capsules. In Table 10-39 and Figure 10-11, where points fall above the upper control limit, a greater number of defects are present than may be expected – there is a lack of statistical control. These sections are reinspected carefully, and action is taken at the machine to correct the causes of bad quality.

$$\bar{p} = \frac{\text{Total number of defectives}}{\text{Total number inspected}} = \frac{131}{31 \times 300} = \frac{131}{9300} = 0.01408$$

$$\begin{aligned} \text{Control limits for } \bar{p} &= \bar{p} \pm 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \\ &= 0.01408 \pm 3\sqrt{\frac{0.01408(1-0.01408)}{300}} \end{aligned}$$

$$\text{Upper limit} = 0.0349$$

$$\text{Lower limit} = 0$$

The sample size, n , from each section is 300 capsules, and typical data are shown in Table 10-41, plotted in Figure 10-11. Note that Sections 21 and 29 appear to be out of control. These sections were subjected to 100% reinspection. Approximately 4.5% of the capsules were defective and were removed.

Table 10-41. Data Collected from the Process in Example 28

Section Number	Number Defectives	Fraction Defective	Section Number	Number Defectives	Fraction Defectives
1	3	0.01	17	2	0.0067
2	2	0.0067	18	3	0.01
3	3	0.01	19	4	0.0133
4	5	0.0167	20	5	0.0167
5	4	0.0133	21	12	0.04
6	3	0.01	22	5	0.0167
7	3	0.01	23	7	0.0233
8	5	0.0167	24	7	0.0233
9	1	0.0033	25	2	0.0067
10	2	0.0067	26	7	0.0233
11	1	0.0033	27	3	0.01
12	3	0.01	28	4	0.0133
13	2	0.0067	29	13	0.0433
14	2	0.0067	30	9	0.03
15	2	0.0067	31	4	0.0133
16	3	0.01			

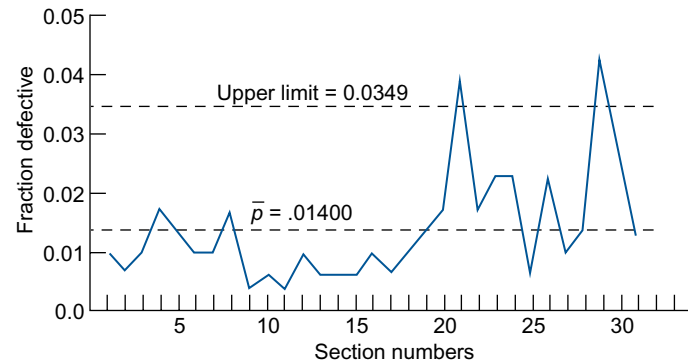


Figure 10-11. Control chart for fraction defective.

$$\text{Lower limit for ranges} = D_3 \bar{R}$$

These calculated limits are drawn on the charts as described above.

CONTROL CHART FOR FRACTION DEFECTIVE

The control chart for fraction defective may be applied to results of an inspection that accepts or rejects individual items of a product. It is designed with the same objectives in mind as the $\bar{x}$ and s , $\bar{x}$ and s_2 charts. Its most effective use is in the improvement of quality, although it also discloses the presence of assignable causes of variation. It provides management with an effective quality history. Fraction defective, P , may be defined as the ratio of the number of defective articles found in any inspection, or series of inspections, to the total number of articles actually inspected. This is expressed nearly always as a decimal fraction (Fig 10-11). The formula for the control limits on a fraction defective chart is:

$$\bar{p} \pm 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

ACCEPTANCE SAMPLING

Acceptance sampling has become one of the major fields of statistical quality control. It is used in many phases of manufacturing, such as inspection of incoming materials, process inspection at various points in the manufacturing operations, and final inspection of the finished product. Sampling inspection usually is used in lieu of 100% inspection for several reasons:

1. The cost of 100% inspection is prohibitive.
2. 100% inspection is fatiguing and may result in the inspectors making errors.
3. The inspection operation may involve destructive testing.

In sampling, one must consider the laws of probability. The risk of rejecting good-quality material and the risk of accepting bad merchandise should be appraised. Sampling plans can be designed and applied in such a manner as to reduce these risks to a minimum and, over a period of time, give assurance of quality products.

The graph illustrating the performance of a sampling plan (i.e., ability to discriminate between acceptable and unacceptable lots) is called an *operating characteristic curve* (OC curve). For any given quality of submitted material, it is possible to determine the probability of acceptance.

Figure 10-12 is an example of an OC curve for the sampling plan described in Example 29. The government publication MIL-STD-105E²⁴ gives many different sampling plans with their corresponding OC curves. A plan that is appropriate for a product is chosen depending on lot size and seriousness of the defect.

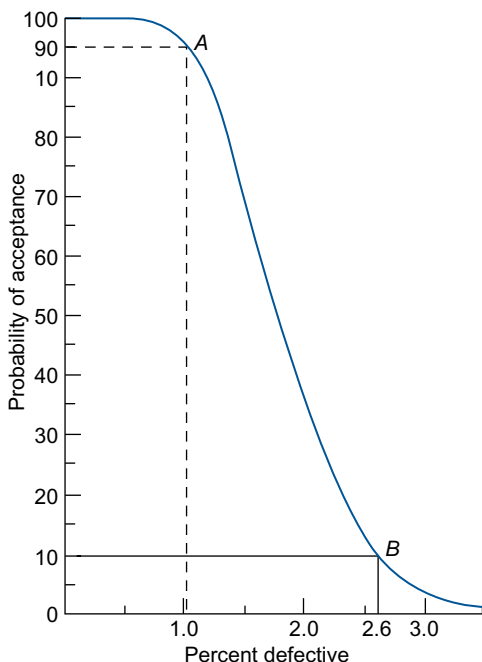


Figure 10-12. Operating characteristic curve. (From Military Sampling Procedures and Tables for Inspection by Attributes. MIL-STD-105E. Washington DC: USGPO, 1989.)

Example 29.

Example of a **statistical sampling plan**. A pharmaceutical manufacturer receives empty bottles of a particular size from a supplier in lots of 20 000 bottles each. The drug firm would like the producer to submit material that is not more than 1.0% defective most of the time, or specifically 95% of the time. See point **A**, Figure 10-12. However, the pharmaceutical firm has agreed to take one chance in 10 of accepting a lot that is 2.6% defective. See point **B**, Figure 10-12.

The acceptance sampling plan that complies with these specifications is as follows. Take a random sample of 540 bottles. Inspect the bottles for defectives. If zero to nine bottles are found defective, accept the lot; if ten or more defectives are found, reject the lot. The operating characteristic curve for this plan is illustrated in Figure 10-12.²⁴

One also can see that, using this sampling plan, submitted lots having 0.5% defective will be accepted about 99 times in 100 (probability of acceptance = 0.99) and thus rejected about one time in 100. Submitted lots having 1.75% defective will be accepted 50 times in 100 (probability of acceptance = 0.50) and rejected half the time.

STATISTICS OF THE STRAIGHT LINE

The use of straight lines to illustrate and define relationships or to help interpret data is common in research investigations. In pharmaceutical research, straight lines may be used for predictive purposes in stability studies or to estimate future events, such as sales figures in market research studies. Straight lines are found in many theoretical relationships in physical and biological chemistry. First-order and zero-order kinetics can be expressed in a linear fashion. Michaelis-Menten kinetics and the Arrhenius relationship can be transformed to a linear

form. Dose-response curves often are linearized if the response is plotted versus log dose. In fact, it is almost always desirable to express a relationship in the form of a straight line, if at all possible.

Reasons for the desirability of straight-line relationships include the ease of interpolation – and in well justified cases, extrapolation – as well as the simplification of the determination of the parameters of the line, the slope, and the intercept. The straight line is defined by these two parameters, and these often have biological and/or physical significance. Consider the example of a first-order kinetic relationship:

$$C = C_0 \cdot e^{-kt}$$

where C is the concentration at time t , C_0 is the concentration at time 0, and k is the first-order rate constant.

This equation is not linear—a plot of C versus t will not result in a straight line. If the experimental data are gathered for C as a function of time, one usually is interested in defining the first-order relationship, in particular to evaluate the parameters k and C_0 . This is done most easily by linearizing the equation, using a logarithmic (log) relationship. Using log to the natural base e , the following linear relationship is obtained.

$$\log C = \log C_0 - k \cdot t$$

This has the form of a straight line. The general equation of a straight line can be expressed as:

$$y = a + bx$$

with the dependent variable y , the independent variable x , the slope of the line b , and the intercept at the y -axis a .

Figure 10-13 shows this linear relationship and calculation of the parameters.

The linearized first-order kinetic equation will show a straight line when $\log C$ is plotted versus time, with intercept $\log C_0$ and slope $-k$. The linearized form makes it easy to obtain the values of C_0 and k . C_0 is the antilog of the intercept a , and k is the negative of the slope b :

$$C_0 = e^a, \quad k = -b.$$

One of the problems in estimating these values from real data is the variability; a plot does not clearly define a straight line. If variability is large, it may be very difficult to decide how to draw the line. Figure 10-14 shows real data from a pharmacokinetic study, in which plasma drug concentrations are measured following an intravenous bolus injection of drug (a one-compartment model).

When confronted with a relationship that should be linear from a theoretical viewpoint but where the x, y values do not lie exactly on a single line, the lack of an exact fit will be considered to be due to variability (error) in y (the dependent variable). In most cases that are encountered, the x variable (the independent variable) tends to have little error relative to the

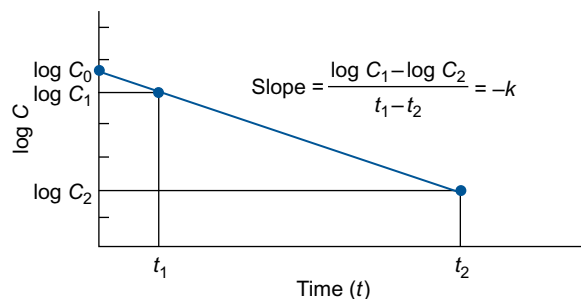


Figure 10-13. Plot of $\log C$ versus time.

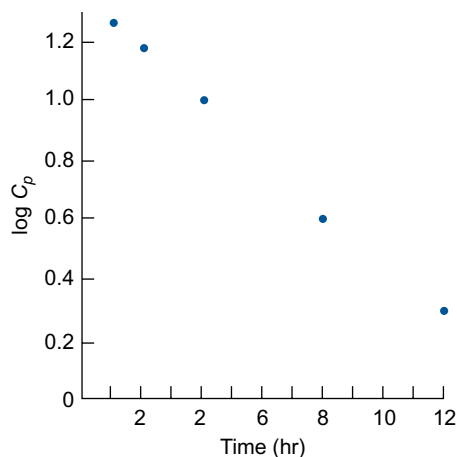


Figure 10-14. Drug plasma levels following an intravenous bolus dose of a drug.

y variable. For example, in a dose-response relationship, the drug is carefully prepared so that an almost unerring dose is administered. However, the response is unpredictable due to the biological variability of the natural material (e.g., animals or bacteria). In a kinetic study, the x variable, time, can be measured with great accuracy. The dependent variable, concentration, is variable due to analytical error, for example. The best line for such variable data is called the least-squares (LS) line. This line is such that the sum of the squared deviations of each point from the line is minimized. That is, if the vertical distance from each point to the LS line is calculated, and the squares of these distances are summed, the LS line would minimize the sum-of-squares. Deriving the sum of squares with respect to a and b leads to the equations:

$$a \cdot \sum x_i + b \cdot \sum x_i^2 = \sum x_i y_i, \quad a \cdot n + b \cdot \sum x_i = \sum y_i$$

and thus

$$a = \frac{\sum x_i^2 \cdot \sum y_i - \sum x_i \sum x_i y_i}{n \cdot \sum x_i^2 - (\sum x_i)^2}, \quad b = \frac{n \cdot \sum x_i y_i - \sum x_i \sum y_i}{n \cdot \sum x_i^2 - (\sum x_i)^2}$$

Example 30.

Consider the data from Figure 10-14. See Table 10-42. The equation of the LS line is $\log C = 1.345 - 0.0886 \cdot t$, thus:

$$b = \frac{16.036 - 27(4.334)/5}{229 - 27^2/5} = -0.0886$$

$$a = 0.8669 + 0.0886(5.4) = 1.34534$$

$$C \approx 22.14 \cdot e^{-0.0886 \cdot t}$$

Table 10-42. Concentration vs Time for Example 30

Time (x)	1 hr	2 hr	4 hr	8hr	12 hr
Concentration ($\mu\text{g/mL}$)	18	15	10	4	2
Log concentration	1.255	1.176	1.000	0.602	0.301

Example 31.

In analytical procedures for drugs, a calibration curve often is constructed using known concentrations of the material to be analyzed. The relationship of drug concentration and the analytical measurement usually is linear. In spectrometric methods, absorption usually is proportional to concentration. The data in Table 10-43 were obtained for the construction of such a calibration curve.

These data and the LS line are plotted in Figure 10-15. The LS intercept, a , and slope, b , are:

$$a = \frac{3000 \cdot 2.4210 - 100 \cdot 72.67}{4 \cdot 3000 - 100^2} \approx -0.0020,$$

$$b = \frac{4 \cdot 2.67 - 100 \cdot 2.4210}{4 \cdot 3000 - 100^2} \approx -0.0243$$

The estimate of the variance of y , $S_{y,x}^2$ is

$$s_{y,x} = \frac{\sum y_i^2 - (\sum y_i)^2 / n - b^2 (\sum x_i^2 - (\sum x_i)^2 / n)}{n - 2}$$

$$= \frac{1.7607 - 2.421^2 / 4 - 0.0249^2 (3000 - 100^2 / 4)}{2}$$

$$= 0.000208$$

The value of the numerator is the sum-of-squares of the difference between the actual values of y and the value of y on the LS line, for each y . The divisor, $n - 2$, is the number of data pairs minus 2, the DF. Thus, the estimate of the variance of y in this example has 2 DF. The reason that 2 is subtracted from the number of data points to obtain the DF is that two parameters are being estimated in the case of a straight line. In previous examples, such as the t test, only the mean is estimated for a treatment group, and $DF = n - 1$.

Table 10-43. Absorbance vs Concentration

Concentration	10 mg/L	20 mg/L	30 mg/L	40 mg/L
Absorbance	0.241	0.492	0.710	0.978

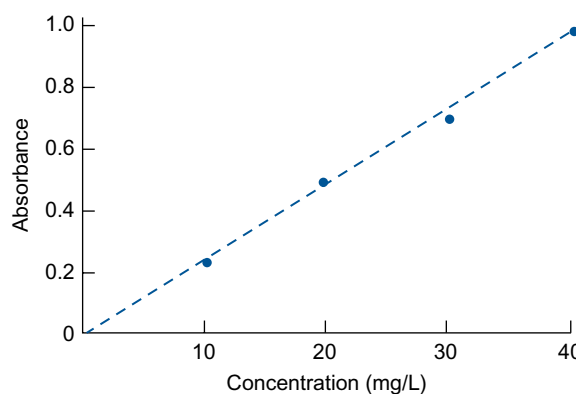


Figure 10-15. Beer's law plot.

Please note that the denominators in both expressions are equal and need only to be computed once.

It must be noted that, using the formulas for linear regression on transformed data, as in this example, the result does not minimize the sum of squares between measured and estimated concentrations but rather the sum of squares between the logarithms. In order to find the real least squares parameters C_0 and k , we would need to apply *nonlinear regression techniques*.

This procedure can be used to fit a line for any two variables. If statistical inference procedures are to be applied to the line, certain assumptions about the data are necessary.

1. The x variable is measured without error. In practical situations, the error in x should be small, compared to the error in the y variable.
2. The y variable is distributed normally with a true mean equal to $A+Bx$ (A and B are the true values of the intercept and slope) and with the same variance, σ^2 , at all values of x .

With these assumptions, the confidence intervals for the line can be computed and statistical tests performed on the parameter estimates, a and b .

With an estimate of the variance, statistical procedures can be applied to these data if the assumptions stated above hold. Concentration is measured with little error, whereas the spectrometric readings, y , have error due to instrumental variability, sample processing, and handling (diluting, pipetting, etc.), among other sources of variability. If it is assumed that the variance is the same at each concentration value and that the concentration values are distributed normally, the following statistical procedures can be used.

CONFIDENCE LIMITS AND TEST OF THE SLOPE

As in the statistical-hypothesis testing procedures described previously in this chapter, a test of the slope versus a hypothetical value can be performed. Also, confidence limits can be placed on the slope.

CONFIDENCE LIMITS AND TEST OF THE INTERCEPT

Tests for the intercept and confidence limits are analogous to those presented immediately above for the slope. The variance estimate of the intercept is:

$$S_a^2 = S_{y,x}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2} \right)$$

In Example 32, the calibration curve, a reasonable test would be to compare the intercept to zero. That is, one discovers whether zero concentration could correspond to a reading of zero. This would be a reasonable assumption if no interfering substances are present and if the optical density versus concentration relationship is a straight line from 0 to the highest concentration tested.

$$t = \frac{|-0.002 - 0|}{\sqrt{0.00028 \cdot (1/4 + 625/500)}} = 0.1132$$

Since 0.1132 is less than the tabled value at the 5% level (see Table 10-10) with 2 DF, it is concluded that the intercept is not significantly different from zero.

A 95% confidence interval for the intercept is:

$$-0.002 \pm 4.3 \sqrt{0.00028 \cdot (1/4 + 625/500)} = -0.002 \pm 0.082$$

These ideas as applied to analytical data are discussed in some detail by Youden.²⁵

Example 32.

Suppose that a value of 0.025 for the slope of the line is reported in an authoritative publication on this assay procedure. It is desired to determine if the slope in the experiment of Example 31 is different from 0.025 (ie, $H_0: B = 0.025$). The estimate of the variance of a slope is:

$$s_b^2 = \frac{s_{y,x}^2}{\sum (x_i - \bar{x})^2}$$

The test is a two-sided t test with $n - 2$ DF of the following form:

$$t = \frac{|b - B|}{\sqrt{s_b^2}} = \frac{|0.02429 - 0.025|}{\sqrt{0.000208/500}} = 1.10$$

Since t is less than the t value in the tables with 2 DF at the 5% level (see Table 10-10), it is concluded that the observed slope is not significantly different from 0.025. Note that a relatively large difference from 0.025 would be necessary to obtain significance because of the few DF in this test (the test is not very powerful). To increase the DF, more observations would be needed.

A confidence interval for the slope can be constructed in a manner similar to that described for means. A 95% confidence interval is:

$$\begin{aligned} b \pm t \sqrt{s_b^2} &= 0.02429 \pm 4.30 \sqrt{0.000208/500} \\ &= 0.00243 \pm 0.0028 = [0.0215, 0.0271] \end{aligned}$$

FITTING A LINE WITH AN INTERCEPT OF ZERO

In some situations, it is desirable to force the LS line to have a y intercept equal to zero.

CONFIDENCE INTERVAL FOR Y AND X

Many situations arise where a confidence interval for y at some specified x is of interest.

Figure 10-17 shows 95% confidence intervals (confidence band) for the line calculated from the data of Figure 10-16. Note the hyperbolic shape, the interval being smallest at $\bar{x}$ and wider as x deviates more from its mean value. Using the lower line of the confidence interval to compute the time to 90%

Example 33.

In the Beer's Law line in Example 32, if it is known that there are no interfering substances and that the relationship is linear throughout the region of concentration being tested, the assumption that the line must pass through the origin is valid. The slope of this line is calculated as:

$$b = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{72.67}{3000} = 0.02422$$

The slope of the line with 0 intercept is very close to that obtained above where the intercept was computed with no constraints on the value of the intercept.

Example 34.

The data from Figure 10-16 show the results of a kinetic stability study, in which drug content in tablets is measured as a function of time. The labeled content is 100 mg. The LS line was calculated as $p = 103.3 \text{ mg} - 0.483 \text{ mg/month } x$, where p is tablet potency and x is time. Note that the intercept is greater than 100 mg because a slight overage is built into the manufacturing process.

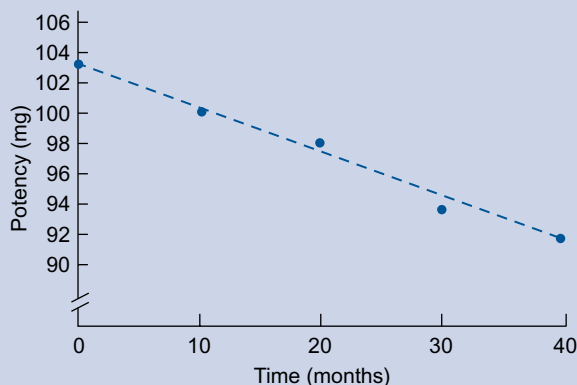


Figure 10-16. Tablet stability study.

The variance estimate, $S^2_{y,x}$, with 3 DF is equal to 0.367. In such stability studies, it often is of interest to predict the time for drug potency to reach 90% of the labeled amount in order to estimate shelf life or an expiration period. Substituting 90 mg for p (potency) and solving for time x yields,

$$x = \frac{103.3 \text{ mg} - 90 \text{ mg}}{0.483 \text{ mg/month}} = 27.54 \text{ month}$$

Therefore, the best estimate of the time to 90% potency is 27.54 months.

When establishing an expiration date, a conservative approach would take into account the error in the estimated values. A two-sided confidence interval can be constructed for the true value of y at a given x , using:

$$y \pm t \sqrt{S^2_{y,x} \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right)}$$

where y is a point on the LS line. The value of t (3 DF) for a two-sided 95% interval is 3.182. The width of the confidence interval depends on the value of x , being minimal when $x = \bar{x}$. The value of y when $x = \bar{x}$ is

$$y = 103.3 \text{ mg} - 0.483 \text{ mg/month} \cdot 12 \text{ month} = 97.5 \text{ mg}$$

The 95% confidence interval when $x = \bar{x} = 12$ is:

$$97.5 \pm 3.18 \sqrt{0.367 \left(\frac{1}{5} + \frac{0}{360} \right)} = [96.64, 98.36]$$

Exercise 6.

Calculate the 95% confidence interval for potency when $x = 24$ months.

Answer: 90.21 to 93.19.

potency yields a conservative estimate. In the example in Figure 10-17, a reasonable estimate of the expiration date would be 24 months. A one-sided interval (below the line) has been proposed as being more appropriate for stability data, for one usually is concerned with the loss of potency. For a 95% one-sided confidence interval, for 3 DF, the value of t from Table 10-10 is 2.353. Using this value of t to calculate the one-sided confidence interval, the one-sided confidence band shown in Figure 10-18 is obtained. For example, when $x = 12$ (months), the lower limit has the value 96.86 mg.

With this approach, the expiration date would be set at 25.1 months (see Fig 10-18).

A confidence interval can be computed for x at a given value of y . This is sometimes known as inverse estimation. In the stability example, interest would be in computing a confidence interval for the time at which 90% of the potency remains. This time was estimated as 27.5 months. The formula for the confidence interval for x is more complex than that for y :

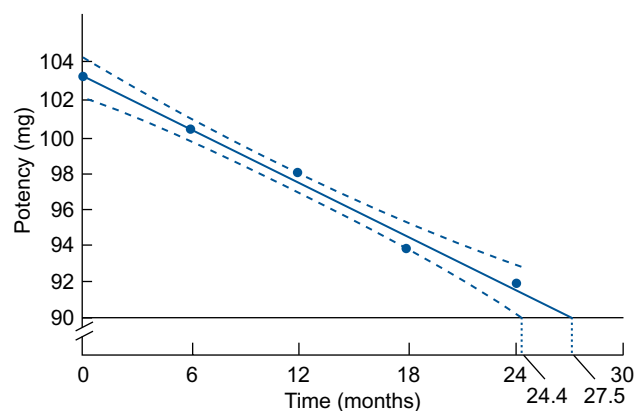


Figure 10-17. Two-sided confidence interval for stability data.

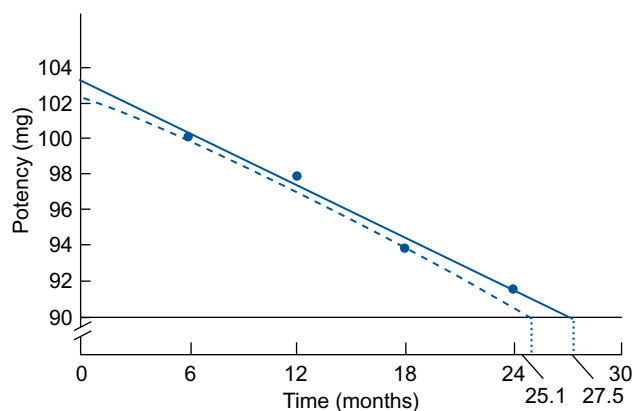


Figure 10-18. One-sided confidence interval for stability study.

Exercise 7.

Use the above formula to show that a one-sided lower interval for x (time) at a 90% potency is 25.1.

Answer: This answer corresponds to the value of time taken from Figure 10-18.

Example 35.

The line for the stability data depicted in Figure 10-16 has a slope of -0.483 , with a variance estimate of 0.367 with 3 DF.

The value of $\sum (x_i - \bar{x})^2$ is 360 . Another formulation was prepared and tested for stability. Ten sampling times were used for the stability study, and the slope was determined to be -0.533 . The variance estimate (8 DF) was 0.289 , and $\sum (x_i - \bar{x})^2$ was equal to 2565 . The test for equality of the slopes (rate of decomposition) is:

$$t = \frac{|0.533 - 0.483|}{\sqrt{0.310(1/2565 + 1/500)}} = 1.84$$

Since 1.84 is less than the tabled t value 2.20 for the 5% significance level with 11 DF, 2.20 (3 from one line and 8 from the other), it can be concluded that the slopes of the two lines are not significantly different.

$$\frac{(x - c^2 \bar{x}) \pm t [s_{y,x}/b] \sqrt{(1 - c^2)/n + (x - \bar{x})^2 / \sum (x_i - \bar{x})^2}}{1 - c^2}$$

where

$$c^2 = \frac{(t \cdot s)^2}{b^2 \sum (x_i - \bar{x})^2}$$

COMPARISON OF THE SLOPES OF TWO LINES

A statistical test may be performed to compare the slopes of two lines, using a t test. The null hypothesis is:

$$H_0: B_1 = B_2 \quad \text{or} \quad B_1 - B_2 = 0$$

The t test compares the difference of the two slopes to the standard error of the difference. The variances of y for the two lines are assumed to be equal, and the estimates are pooled as in the two-sample t test:

$$s_{\text{pooled}}^2 = \frac{s_{y,x}^2(n_1 - 2) + s_{y,x}^2(n_2 - 2)}{n_1 + n_2 - 4}$$

A two-sided t test with $(n_1 + n_2 - 4)$ DF results in:

$$t = \frac{|b_2 - b_1|}{\sqrt{s_{\text{pooled}}^2 \left(\frac{1}{x_1^2} + \frac{1}{x_2^2} \right)}}$$

where, x_1^2 and x_2^2 are $\sum (x_i - \bar{x})_1^2$ and $\sum (x_i - \bar{x})_2^2$, respectively.

In a biological assay, a common procedure is to determine the relative potency of two or more substances using the *parallel line assay*. In this procedure, the lines from a plot of response versus log dose are forced to be parallel, and the distance between the lines is a measure of the relative potency. Before performing this procedure, a test is made to ensure that the lines are parallel. Nonparallel lines will cross, suggesting that at low doses one product gives a greater response, whereas at higher doses the other product gives the greater response. Figure 10-19²³ illustrates the principle of this assay. The computations are tedious, and the book by Finney²⁶ should be consulted for those who wish more detail on the statistical treatment of this and other biological assay methods.

CORRELATION

Correlation is related to, but should not be confused with, linear regression. It is a measure of the linear relationship between two variables but does not prove linearity. In fact, the

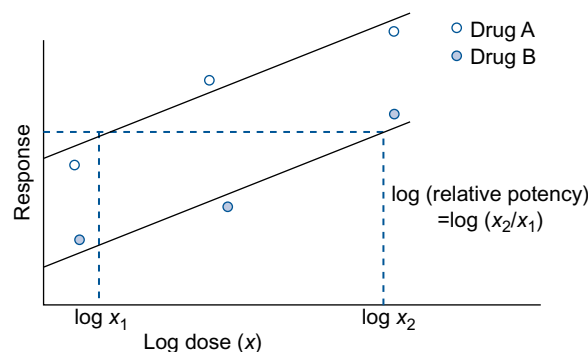


Figure 10-19. Relative potency estimate using a parallel line assay. Doses shown for $\log x_1$ and $\log x_2$ give the same response for Products A and B, respectively. (From Bolton S. *Pharmaceutical Statistics*. New York: Marcel Dekker, 1984; 416, 463.)

Exercise 8.

Compute the correlation between x and y for $x = -2, -1, 0, +1, +2$, for the relationship $y = x^2$.

Answer: $r = 0$.

usual formulas for determining the significance of the correlation assume that the variables already are related linearly. The question that is usually posed indirectly when testing the correlation is, can the value of one of the variables be used to predict the value of the second variable? This amounts to testing the slope of the line relating the variables versus 0 . If the slope is significantly different from 0 , then the variables have a *significant correlation*. Correlation is used when both variables are subject to error. If one variable is not subject to error (fixed), the linear regression approach to establish the relationship of the variables is more appropriate.

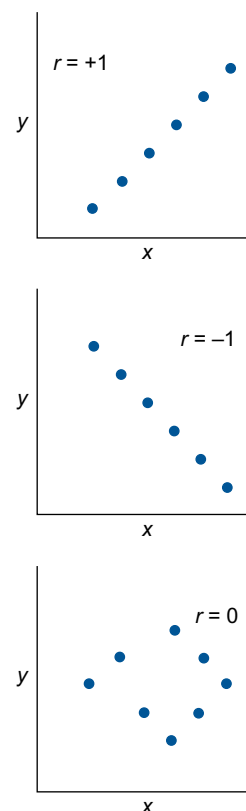


Figure 10-20. Correlation diagrams (scatter plots).

Example 36.

An experiment was performed to examine the relationship of tablet hardness to tablet dissolution. Dissolution was measured as the time (minutes) for 50% of the drug to dissolve in the USP Dissolution Test. Hardness was measured in kilograms. The following results were obtained for 12 tablets:

Hardness/ kg	6.8	5.3	5.8	7.2	6.9	6.0	6.8	8.1	7.5	6.3	
Dissolution/ min	18	17	21	26	28	20	25	29	31	18	

These data are plotted in Figure 10-21, known as a *scatter plot*. This plot suggests a trend toward slower dissolution as hardness increases.

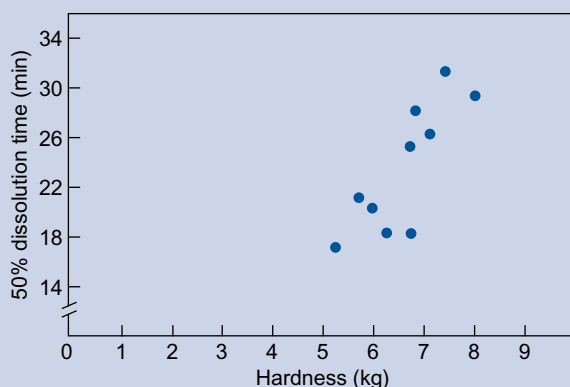


Figure 10-21. Scatter plot of hardness versus dissolution data (Example 36).

In this example, r is equal to:

$$\frac{1585.5 - 233 \cdot 66.7/10}{\sqrt{236.1 \cdot 6.321}} = 0.81$$

The test of the significance of the correlation coefficient shows $t=3.94$ with 8 DF. It is concluded that r is significantly different from 0, and hardness and dissolution are correlated ($\alpha < 0.05$; see Table 10-10; $t=2.228$ for significance at $\alpha=0.05$).

The measure of association is the correlation coefficient, r .

$$r = \frac{\sum x_i y_i - \sum x_i \sum y_i / n}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}}$$

The correlation coefficient can vary between +1 and -1. A correlation coefficient of +1 would result if all points fall exactly on a single line with positive slope; this is a perfect positive correlation. Similarly, if all points fall on a line with negative slope, $r = -1$, a perfect negative correlation is observed. If $r=0$, the variables are not correlated. These three cases are shown in Figure 10-20.

In real situations, these extreme results are seen rarely, but rather some intermediate value of r is observed. The statistical question of interest usually is concerned with the significance of the correlation---a test of r versus 0. One should appreciate, however, that the meaning of the correlation should be considered carefully. For example, if n , the number of data pairs, is large, correlation coefficients that are very small (practically insignificant) will be deemed statistically significant. Also, data that is not linear but clearly related may show small correlation coefficients.

The test of the correlation coefficient versus 0 is:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}, \quad (\text{DF} = n-2)$$

PRINCIPAL COMPONENT ANALYSIS (PCA)**Introduction**

PCA is not a statistical method to infer parameters or to test hypotheses. Instead it provides a means to “reduce a complex data set to a lower dimension to reveal the sometimes hidden, simplified structure that often underlie it.”²⁷ Imagine a data set that forms a cloud in a very high dimensional system e.g., in spectroscopy: each wavelength channel may be regarded as an axis in a coordinate system of many dimensions (about 600). Every spectrum can now be represented as a point in this coordinate system. Spectra of different batches of a given substance or formulation form a cluster in this system. The question we ask to PCA is now whether there is a way to modify the coordinate system (in rotation and stretching) so that as much information as possible can be found along one of the new axes. This new axis is the principal components of the data. The axis orthogonal to the principal component, along which the second most information is located, forms the second axis, and so on.

But what is the direction with most information? In PCA it is the direction along which most of the variance is spread.

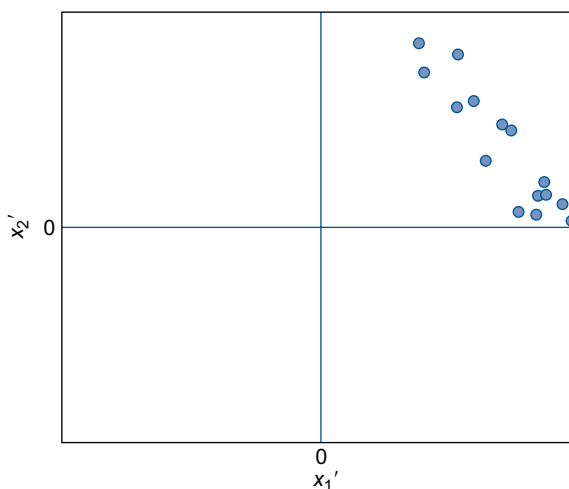


Figure 10-22. The original data set for the principal component analysis.

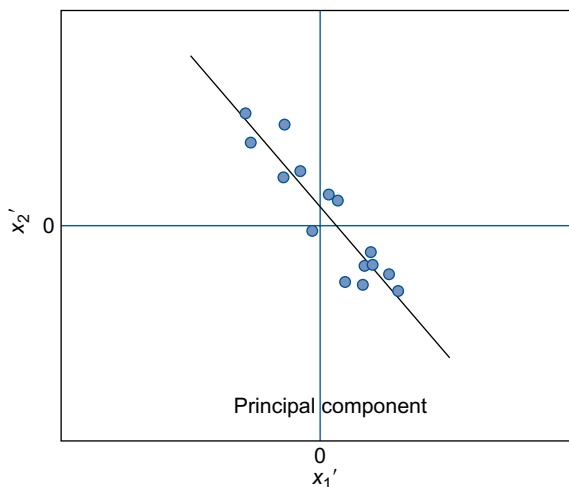


Figure 10-23. The principal component is found in the centered data.

As an example we will consider the following simple data set, consisting of 15 measurements in a two dimensional space (see Figures 10-22, 10-23, 10-24):

(1.03, -1.12)	(0.99, -1.65)	(0.2, 0.87)	(-0.18, -0.14)	(1.17, -0.74)
(-0.85, 1.35)	(-0.46, 1.52)	(-0.82, 2.82)	(1.81, -1.82)	(1.22, -1.09)
(0.58, -1.57)	(0.41, 0.7)	(-1.73, 3.13)	(-1.61, 2.32)	(1.6, -1.35)

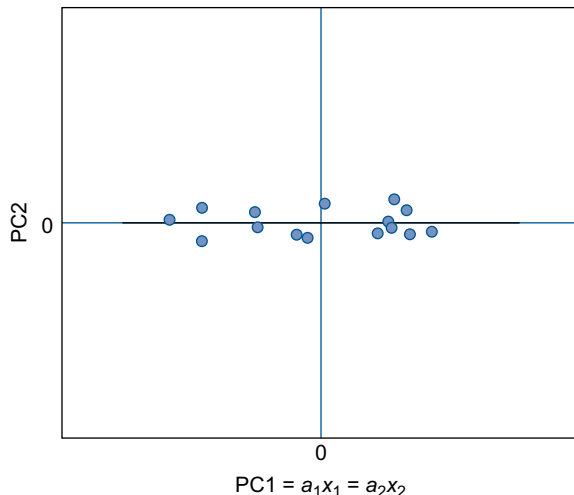


Figure 10-24. The coordinate system is rotated so that the principal-components lie on the axes.

Conduct of a PCA

Suppose you had a set of n measurements (15 in our example), each of which consists of m dimensions (2 in our example). The data is represented in a $m \times n$ matrix A : each column of the matrix consists of the m elements of one measurement. Our sample data looks like this in the matrix representation:

$$A' = \begin{pmatrix} 1.03 & 0.99 & 0.2 & -0.18 & 1.17 & -0.85 & -0.46 & -0.82 \\ -1.12 & -1.65 & 0.87 & -0.14 & -0.74 & 1.35 & 1.52 & 2.82 \\ 1.81 & 1.22 & 0.58 & 0.41 & -1.73 & -1.61 & 1.6 & \\ -1.82 & -1.09 & -1.57 & 0.7 & 3.13 & 2.32 & -1.35 \end{pmatrix}$$

The problem is now to rotate the matrix A in such a way that the covariances between the rotated dimensions vanish. This can be done by means of centralization and normalization.

Centralization and Normalization

The first step of performing a PCA is centralization of the data: For each dimension of the original data set, the mean in this dimension is subtracted from all data points. We achieve a new data set in which the mean in all dimensions is zero. It is generally a good idea, though by no means mandatory, to normalize the data, too: for each measurement each of its dimensions is divided by the standard deviation of that dimension. This makes the results independent of the units in which the data were measured; after normalization a data set measured in inches will result in the same normalized data as measurements that were taken in millimeters.

In the sample data the mean of the first dimension x_1 in row 1 has a mean of $\bar{x}_1 = 0.224$, the mean of the second row is $\bar{x}_2 = 0.215$. Subtracting these values from our data and dividing by the respective standard deviation results in:

$$A = \begin{pmatrix} 0.72 & 0.68 & -0.02 & -0.36 & 0.84 & -0.95 & -0.61 & -0.93 \\ -0.78 & -1.09 & 0.38 & -0.21 & -0.56 & 0.66 & 0.76 & 1.52 \end{pmatrix}$$

$$\begin{pmatrix} 1.41 & 0.88 & 0.32 & 0.17 & -1.74 & -1.63 & 1.22 \\ -1.19 & -0.76 & -1.04 & 0.28 & 1.70 & 1.23 & -0.91 \end{pmatrix}$$

Now the sum of each row is zero, and the respective standard deviations are equal to 1.

Calculation of the Covariance Matrix

In the next step we have to calculate the so called covariance matrix of the data. For all the m dimensions, we calculate all possible covariances between any two dimensions $d_i, d_j, i \in \{1, \dots, m\}, j \in \{1, \dots, m\}$ (c.f. page). The covariances are arranged in a square $m \times m$ matrix. Since $\text{Cov}(x, y) = \text{Cov}(y, x)$ the covariance matrix is symmetrical. Since $\text{Cov}(x, x) = \text{Var}(x)$ the values along the main diagonal are the variances of the data along the respective dimensions.

The covariance matrix of the data in our example is:

$$C = \begin{pmatrix} 1.00 & -0.92 \\ -0.92 & 1.00 \end{pmatrix}$$

As expected, matrix C is symmetric. The variances on the diagonal equal 1 because we have normalized our data. The negative sign of the covariance tells us that the values of x_2 tend to decrease with increasing x_1 .

Our goal is now to find a transformation of the covariance matrix that makes all the elements off the diagonal equal to zero. Fortunately diagonalization is a common problem in linear algebra, and there exist algorithms to perform this task.

Calculation of the Eigenvectors and Eigenvalues of the Covariance Matrix

One method to diagonalize a covariance matrix C involves computing its eigenvectors. An eigenvector of a matrix is a vector different from the null vector that does not change its direction if transformed by the matrix. (The transformation of a vector $v \rightarrow$ with a Matrix A is the product $A \cdot v \rightarrow$.) Since the direction of the vector remains the same, the transformation is just the same as a stretching of the vector by some value λ . This value is the eigenvalue that corresponds to the eigenvector: $A \cdot v \rightarrow = \lambda \cdot v \rightarrow$.

The covariance matrix C of our sample data has the eigenvectors:

$$\vec{v}_1 = \begin{pmatrix} -0.707 \\ -0.707 \end{pmatrix} \quad \text{and} \quad \vec{v}_2 = \begin{pmatrix} 0.707 \\ 0.707 \end{pmatrix};$$

the corresponding eigenvalues are:

$$\lambda_1 = 1.92 \quad \text{and} \quad \lambda_2 = 0.08.$$

Please notice that the length of both of the eigenvectors is $\sqrt{2} \cdot 0.707^2 = 1$

Any $m \times m$ covariance matrix has exactly m eigenvectors. (Well, since any multiple of an eigenvector is again an eigenvector, there are actually infinitely many of them, but only m are of unit length, and we count only them.) These m eigenvectors point in different directions; in fact, they are orthogonal to each other. The m eigenvectors can be glued together to a matrix E , and with the covariance matrix C , we have $E^{-1}CE = D$ where D is a diagonal matrix — exactly what we were looking for. It is now fortunate that the eigenvectors are orthogonal, because that makes E an orthogonal matrix, which can be easily inverted: We just need to transpose it, i.e., change rows to columns.

In our example, we get E from $v \rightarrow 1$ and $v \rightarrow 2$:

$$E^{-1} = E^T = \begin{pmatrix} -0.707 & 0.707 \\ -0.707 & -0.707 \end{pmatrix}^T = \begin{pmatrix} -0.707 & -0.707 \\ 0.707 & -0.707 \end{pmatrix}$$

Indeed, we find:

$$E \cdot E^T = \begin{pmatrix} -0.707 & 0.707 \\ -0.707 & -0.707 \end{pmatrix} \cdot \begin{pmatrix} -0.707 & -0.707 \\ 0.707 & -0.707 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

It turns out that the elements of matrix D are the eigenvalues of C :

$$D = E^{-1}CE = E^TCE = \begin{pmatrix} 0.08 & 0.00 \\ 0.00 & 1.92 \end{pmatrix}$$

The eigenvector with the highest corresponding eigenvalue points in the direction in which most of the variance of the measured data is found, so most of the information in the original data set lies along this direction, and this eigenvector is the principle component of the data. The eigenvalue tells us how much of the total variance in the raw data falls along the principal component; in our example, the principal component accounts for a fraction of 0.96 of the total variance because $\lambda_1/\text{Vartot}=0.96$. (Because of normalization the variance is 1 in each dimension, so it adds up to m — or 2 in our example.)

The eigenvector with the second largest eigenvalue points in the direction that is orthogonal to the first eigenvector and contains the most of the remaining variance of the data. This, again, is the direction along which most of the remaining information is located. Because $\lambda_2/\text{Vartot}=0.04$, the information along the second axis may be considered as neglectable.

Reduction of the Data

The goal of principal component analysis is to describe as much information of the original raw data with as few dimensions as possible. If we omit the second dimension of the transformed data in our sample, we keep as much as 96% of the original information.

But how do we access the direction of the principal component? Once we have calculated the eigenvectors, this task is easy: all the base vectors of the transformed system are linear combinations of the original dimensions, and the coefficients are just the elements of the corresponding eigenvectors. Thus, in our example, the principal component is $P_1 = -0.707x_1 + 0.707x_2$, and the data set reduced to one dimension becomes (see Figure 10-25).

Consider the information in an IR spectrum with about 1000 channels. If we are able to express say 80% of the original information in only two or three new dimensions (the principal components), the data become manageable and even visually perceivable.

DATA TRANSFORMATIONS

Probabilities calculated from statistical analyses are based on assumptions underlying the nature of the data. The typical analyses presented in this chapter often assume normality of data and variance homogeneity. When dealing with means of

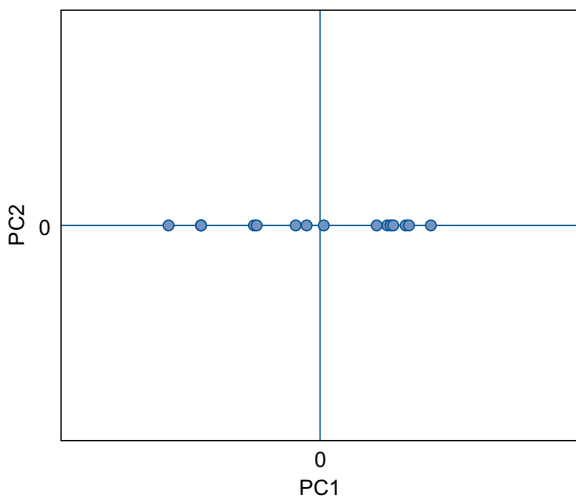


Figure 10-25. The data set is reduced by omitting the information along PC2.

a sufficiently large sample size, the assumption of normality is not critical. However, small sample sizes and a large deviation from normality can result in a significant violation of the normality assumption. When comparing samples from two or more groups, lack of homogeneity of variance (heteroscedasticity) is an important problem that can result in an unreliable analysis. One way of overcoming these problems is the use of *transformations*. Each data point is transformed, resulting in data that more closely fits the normality and variance homogeneity assumptions.¹⁹

Example 37.

The means of two treatment groups are to be compared where it is known that large values are associated with proportionally larger standard deviations. The measurements are 50% dissolution time in minutes (Table 10-44).

Table 10-44. 50% Dissolution Time for Two Formulations

	Formulation A	Formulation B
	27	65
	55	60
	33	98
	69	47
	36	57
	39	43
Mean	43.17	61.67
SD	15.75	19.59

A two independent sample t test (two-sided) comparing the means shows:

$$t = \frac{|43.17 - 61.67|}{19.11\sqrt{1/3}} = 1.75$$

A log transformation results in the data in Table 10-45.

The logarithmic, square root, and arcsine transformations will be presented here as examples of the more popular data transformations.

LOGARITHMIC (LOG) TRANSFORMATION

This transformation (log to the base 10 or log to the base e , $\ln$, may be used) is most applicable for skewed data of the form illustrated in Figure 10-26. These data typically show a relatively constant coefficient of variation (CV). That is, the larger the value, the larger is the SD; the SD is proportional to the mean

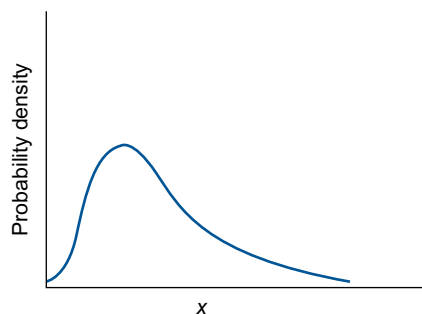


Figure 10-26. Example of a skewed distribution.

Table 10-45. Log Transformation of Data of Table 10-45

	Formulation A	Formulation B
	1.431	1.813
	1.740	1.778
	1.519	1.991
	1.839	1.672
	1.556	1.756
	1.591	1.633
Mean	1.613	1.774
SD	0.150	0.126

Neither test is significant at the 5% level, but the log-transformed values in this example result in a test with a lower probability level.

Exercise 9.

A bioequivalence study comparing two dosage forms, **A** and **B**, with six subjects in a paired design, resulted in the following ratios of AUC_a/AUC_b :

1.27, 1.06, 0.90, 1.30, 1.15, 0.96

Calculate the mean, standard deviations and a 95% confidence interval for the data, using the untransformed data and a log transformation. For the log transformation, calculate the antilogs for the lower and upper limit of the confidence interval. Repeat the calculations for the mean and standard deviation of the ratio of AUC_a/AUC_b . (Note that this is the reciprocal of the data presented above.) What can be said about the confidence intervals for the two kinds of ratios, A/B and B/A ?

Answer:

transformation	Mean	SD	CI
none	1.107	0.1627	1.107±1.171
log	0.000	0.0646	0.04±0.0678
reciprocal:			
none	0.92	0.1376	[0.938, 1.282]
log	-0.0400	0.0646	

(SD/x is constant). This transformation is applicable to data that meet the above conditions and also are greater than 0; the log of 0 or a negative number is undefined.

This probably is the most common transformation for data in the pharmaceutical sciences. Many physical and biological measurements show larger variability as the size of the measurement increases. This is logical for many types of data. For example, the measurement of a large value, such as the assay of a concentrated solution, may be expected to show considerable variability about its mean (e.g., 1000 ± 50 , a 5% variability). The assay of a dilute solution, 10, cannot show very large variation, particularly on the low side where zero (0) is the lower limit. If the CV were 5% (10 ± 0.5), a log transformation would be suitable. If the data are skewed (see Fig 10-22) and the CV is constant, a log transformation will tend to normalize the data distribution and equalize the variances.

When data are presented as ratios, a log transformation often is appropriate. Unless the data are extremely variable, the conclusions using the original or transformed data should be

Exercise 10.

Compute the mean and standard deviation of the following data before and after applying the square-root transformation ($\sqrt{x}$). Draw a histogram of the original and transformed values. Note the greater symmetry of the transformed data.

	Mean	SD
Original data:	5.625	4.272
Transformed data:	2.00	1.206

similar. The conclusions using the transformed data, however, will be more reliable if the transformation is appropriate.

Care should be taken that the log transformation does not help to improve one assumption while making another less valid. For data that are skewed but have constant variance, the normality assumption may improve while causing problems with the variance homogeneity assumption. Fortunately, the log transformation, when indicated, does not seem to cause such difficult and perplexing problems.

Example 38.

Use a normal test to compare the proportion of rats who developed tumors in placebo-control and active-drug groups. In the placebo group, 15 of 100 animals developed tumors, whereas in the drug group, 22 of 100 developed tumors. The arcsines of 0.15 and 0.22 are 22.786 and 27.972, respectively. The normal test statistic Z is:

$$Z = \frac{|27.972 - 22.786|}{\sqrt{821/100 + 821/100}} = 1.28$$

The proportions are not significantly different.

Exercise 11.

Calculate the value of chi-square for the test of these two proportions.

Answer: 1.625. Note that $X^2 = Z^2$ in this example.

For bioequivalence data, a log transformation of AUC and C_{max} is currently recommended.

SQUARE-ROOT TRANSFORMATION

A square-root transformation is useful for data where the sample means are proportional or equal to the variances. The transformation will cause the data to have approximately homogeneous variance. This transformation may be used to replace the log transformation when the data consist of small numbers. If the numbers are less than 10 and zeros are present, $x+1$ may be an appropriate transformation.²⁸ This transformation, like the log transformation, will tend to normalize distributions skewed to the right (distributions with a relative small number of very large values).

ARCSINE (INVERSE SINE) TRANSFORMATION

The arcsine (inverse sine) transformation is used for binomial data or data expressed as percentages or proportions. The

transformation is arcsin p , where p , the proportion or probability, is expressed as a decimal. The variance of a binomial proportion is pq/n , where p is the proportion of successes and q the proportion of failures in n binomial observations. If p varies in different treatment groups, the variance will vary. The arcsine transformation applied to the proportions tends to equalize the variances and normalize the data.¹⁹ The variance of the transformed proportion is $821/n$ when the transformed data are in degrees. This transformation assumes that all proportions transformed have the same value of n . If n is approximately equal for the different groups, the transformation may still be used.

REFERENCES

- Ott L, Longnecker M. *An Introduction to Statistical Methods and Data Analysis*. Belmont: Brooks/Cole, 2010.
- Kish L. *Survey Sampling*. New York: Wiley, 1995.
- Fisher RA, Yates F. *Statistical Tables for Biological, Agriculture and Medical Research*. New York: Hafner, 1963: 134 (table 38).
- Fisher RA. *The Design of Experiments*, 5th edn. Edinburgh: Oliver & Boyd, 1986.
- Cochran WG, Cox GM. *Experimental Design*, 2nd edn. New York: Wiley, 1957.
- Cox DR. *Planning of Experiments*. New York: Wiley, 1958.
- United States Department of the Army. *Tables of the Binomial Probability Distribution*. Applied Mathematics Series No. 6 Washington, DC: USGPO, 1952.
- Bolton S. *Pharmaceutical Statistics*, 3rd edn. New York: Marcel Dekker, 1997.
- Snedecor GW, Cochran WG. *Statistical Methods*, 8th edn. Ames: Iowa State University Press, 1989: 97.
- Snedecor GW, Cochran WG. *Statistical Methods*, 7th edn. Ames: Iowa State University Press, 1980: 476 (table A14).
- Snedecor GW, Cochran WG. *Statistical Methods*, 7th edn. Ames: Iowa State University Press, 1980: 480 (table A15).
- Duncan DB. *Biometrics II* 1948 1.
- Dunnett CW. *Am Stat Assoc J* 1955; 50: 1096.
- Winer BJ. *Statistical Principles in Experimental Design*, 2nd edn. New York: McGraw-Hill, 1971.
- Tate MW, Clelland RC. *Nonparametric and Shortcut Statistics*. Danville IL: Interstate Print, 1957: 137 (table L).
- Good PI. *Permutation, Parametric and Bootstrap Tests of Hypotheses*. New York: Springer, 2005.
- Youden WJ. *Sci Monthly* 1953; 77: 143.
- Youden WJ. *Natl Bur Std(US)Tech News Bull* 1949: 33(July).
- Dixon WJ, Massey Jr FJ. *Introduction to Statistical Analysis*, 3rd edn. New York: McGraw-Hill, 1969: 328.
- Control Chart Method of Controlling Quality During Production (Std Z1. 3)*. New York: American Standards Association, 1958.
- Guide for Quality Control (Std Z1. 3)*. New York: American Standards Association, 1958.
- Dixon WJ, Massey Jr FJ. *Introduction to Statistical Analysis*, 3rd edn. New York: McGraw-Hill, 1969: 142.
- Bolton S. *Pharmaceutical Statistics*. New York: Marcel Dekker, 1984 :416, 463.
- Military Sampling Procedures and Tables for Inspection by Attributes*. MIL-STD-105E. Washington DC: USGPO, 1989.
- Youden WJ. *Statistical Methods for Chemists*. New York: Wiley, 1951.
- Finney DJ. *Statistical Method in Biological Assay*, 4th edn. New York: Hafner, 1980.
- Shlens J. *A tutorial on Principal Component Analysis*. La Jolla: Salk Institute for Biological Studies, 2009.
- Steel RGD, Torrie JH. *Principles and Procedures of Statistics*. New York: McGraw-Hill, 1960.

BIBLIOGRAPHY

EXPERIMENTAL DESIGN

- Cox DR. *Planning of Experiments*. New York: Wiley, 1992.
- Fisher RA. *Statistical Methods for Research Workers*, 13th edn. Edinburgh: Oliver & Boyd, 1970.
- Montgomery DC. *Design and Analysis of Experiments*, 4th edn. New York: Wiley, 1996.
- Chow S-C, Liu J-P. *Statistical Design and Analysis in Pharmaceutical Science*. New York: Dekker, 1995.

STATISTICAL QUALITY CONTROL

- Grant EL. *Statistical Quality Control*, 5th edn. New York: McGraw-Hill, 1980.
- Mandel J. *The Statistical Analysis of Experimental Data*. New York: Dover, 1984.
- Montgomery DC. *Introduction to Statistical Quality Control*, 3rd edn. New York: Wiley, 1996.
- Weber RT. *An Easy Approach to Acceptance Sampling: How to Use MIL-STD-105E*. Milwaukee, WI: American Society for Quality Control, 1991.

SAMPLING

- Cochran WG. *Sampling Techniques*, 3rd edn. New York: Wiley, 1997.
- Deming WE. *Some Theory of Sampling*. New York: Wiley, 1984.
- Yates F. *Sampling Methods for Censuses and Surveys*. New York: Hafner, 1981.

BIOLOGICAL ASSAY

- Bliss CI. *The Statistics of Bioassay with Special Reference to the Vitamins*. New York: Academic Press, 1952.
- Bliss CI. *Am Sci* 1957; 45: 449.
- Finney DJ. *Statistical Method in Biological Assay*, 3rd edn. New York: Hafner, 1978.
- Finney DJ. *Probit Analysis*, 4th edn. London, Cambridge University Press, 1980.

GENERAL

- Bennett CA, Franklin NL. *Statistical Analysis in Chemistry and the Chemical Industry*. New York: Wiley, 1954.
- Bolton S. *Pharmaceutical Statistics*. 3rd edn. New York: Marcel Dekker, 1997.
- Brownlee KA. *Statistical Theory and Methods in Science and Engineering*. New York: Wiley, 1960.
- Buncher CR, Tsay J. *Statistics in the Pharmaceutical Industry*. New York: Marcel Dekker, 1981.
- Chow SC, edn., *Encyclopedia of Biopharmaceutical Statistics*. New York: Marcel Dekker, 2000.
- Davies OL. *The Design and Analysis of Industrial Experiments*. New York: Hafner, 1954.
- Peace KE. *Biopharmaceutical Statistics for Drug Development*. New York: Marcel Dekker, 1988.
- Snedecor GW, Cochran WG. *Statistical Methods*, 8th edn. Ames, Iowa State University Press, 1989.

STATISTICAL SOFTWARE PACKAGES (EXAMPLES)

- R, The R Foundation for Statistical Computing, c/o Institute for Statistics and Mathematics, Vienna, Austria. Available at <http://www.r-project.org>.
- BMDP, Biomedical Computer Programs, University of California, Los Angeles, CA.
- NCSS NCSS Statistical Software, 329 North 1000 East, Kaysville, Utah 84037 (<http://www.ncss.com>; phone: 800-898-6109).
- SAS, SAS Institute Inc, Cary, NC (<http://www.sas.com/>; e-mail: software@sas.com).

EXACT HYPOTHESIS TESTS

- Good PI. *Resampling Methods*. Boston: Birkhäuser, 2001.

PRINCIPAL COMPONENT ANALYSIS

- Shlens J. *A Tutorial on Principal Component Analysis*. La Jolla, Ca: System Neurobiology Laboratory, Salk Institute for Biological Studies, 2005.

Chapter 11

Molecular Structure, Properties, and States of Matter

Thomas Rades, PhD; Keith C. Gordon, PhD and Kirsten Graeser, PhD

MOLECULAR STRUCTURE AND PROPERTIES 177

STATES OF MATTER 189

MOLECULAR STRUCTURE AND PROPERTIES

INTRODUCTION

Pharmaceuticals are made up of molecules, the properties of which are determined by their shape and electronic structure (i.e., the nature of their bonds). These, in turn, are determined by the constituent atoms and how those atoms are held together (bonded). A striking example of this is the element carbon. This can exist in a variety of forms – including nanotubes and buckyballs – but the two forms that illustrate the difference in bonding most strongly are graphite and diamond. Both are entirely composed of carbon atoms, but one is a hard crystalline material, colorless in its pure form (diamond), the other a black greasy material used in pencils and as a lubricant.

Atoms are built from neutrons, protons, and electrons; the former are much more massive than the electrons. The properties of these subatomic units are given in Table 11-1. The fundamental ideas for atomic structure and bonding were developed in the early part of the twentieth century.

HISTORICAL PERSPECTIVE

From the early 1800s it had been known that elements had certain properties in relation to their interaction with other elements. This was expounded in Dalton's atomic theory, which stated that

1. Elements are composed of atoms.
2. The atoms for a specific element are unique to that element.
3. Atoms of one element can combine with atoms of other elements to form compounds. Each compound has a specific ratio of the elements of which it is composed, and these have simple integral ratios.
4. Atoms are not created or destroyed in making compounds – they simply associate in differing ways in differing compounds.

By the late nineteenth century, a number of observations had been made that suggested atoms had an internal structure that had some sort of periodicity based on their mass. The weights of the elements had been measured, and it was found that if they were arranged in order, then there were similarities between each eighth element. For example, neon, helium, and krypton were all inert, and lithium, sodium, and potassium reacted in a ratio of 1:1 with chlorine, whereas magnesium and calcium (also eight different) reacted with chlorine in a 1:2 ratio. The English chemist Newland noted that this behavior was similar to the eighth note in a musical octave, and

these findings thus became called the law of octaves. Around 1869 Mendeleev and Meyer developed what we now call the periodic table, in which all of the elements known at that time were listed. Mendeleev left gaps for gallium, scandium, and germanium – he even went so far as to predict the properties of these elements. These elements were discovered and purified subsequent to Mendeleev; for example, gallium was discovered in 1875, scandium was purified in 1937, and germanium in 1885. These elements had many of the properties Mendeleev had predicted.

This became known as the periodic law, and although it is not the weights of the elements but rather the atomic number that is periodic, Mendeleev's periodic table led to the idea of some sort of repeating internal structure within the atoms.

The key issues in examining the periodic table are:

1. Each element has an atomic number (Z), which corresponds to the number of protons in the nucleus and is a multiple of the electronic charge.
2. Atomic weight is the average mass for a particular element versus the current standard, which is based on the mass of the carbon-12 isotope. This was given the atomic weight of 12.
3. Isotopes are forms of the same element that differ in the number of neutrons in the nucleus.

At the beginning of the twentieth century, J. J. Thomson proposed a model of the atom based on the data available at the time. In his model (euphemistically called the plum pudding model) the atom was a mass that contained protons and neutrons with the electrons embedded, rather like raisins in a plum pudding. The key issue of this model was that all of the particles occupied the same volume of space. This model was shattered by a series of experiments conducted by Marsden and Geiger in 1909 and interpreted by Rutherford in 1911. The experiment, which became known as the Rutherford backscattering experiment, took high energy particles and fired them at a thin gold foil. The particles in question were alpha particles. These are comprised of two neutrons and two protons, thus carrying a +2 charge. In the plum pudding model, the alpha particles would be unimpeded in penetrating the gold foil since each gold atom was a small mass with no charge because the electrons and protons were in close proximity. However, Marsden and Geiger observed that the high energy alpha particles scattered back towards them. This was a striking result, which Rutherford described as "...quite the most incredible event that has ever happened to me in my life. It was almost as incredible as if you fired a 15-inch shell at a piece of tissue paper and it came back and hit you."¹

Table 11-1. Masses and Charges for the Neutron, Proton and Electron

Subatomic particle	Mass/kg	Relative mass	Charge/C	Relative charge
Proton	1.673×10^{-27}	1837	1.602×10^{-19}	1
Neutron	1.675×10^{-27}	1839	0	0
Electron	9.109×10^{-31}	1	-1.602×10^{-19}	-1

Rutherford explained these findings by hypothesizing a new model of the atoms, in which the nucleus contained only protons and neutrons and the electrons were buzzing around this central positively charged core. This model certainly was consistent with the results of the gold foil experiments, but it created some difficulties since an electron orbiting a positive proton would be attracted into the centre. Thus, all atoms would be unstable.

The Bohr model

In 1913 Bohr developed the first model of modern atomic theory. Bohr was trying to develop a model that would explain the nature of emission spectra from atoms such as hydrogen. In the 1880s Rydberg had recognized that the emission spectrum of hydrogen followed a mathematical relationship, such that the wavelength of light (λ) observed was related to integers, n_1 and $n_2, \dots$ and a constant called the Rydberg constant (R_∞)

$$\frac{1}{\lambda} = R_\infty \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right) \quad (1)$$

For example, in a series of emission lines observed at about 400 nm (the Balmer series), the wavelength of each line corresponded to $n_1 = 2$ and $n_2 = 3, 4, 5, 6, 7$, etc.

In Bohr's model (Figure 11-1) he stated that

1. The electrons were restricted to set orbits at fixed distances from the nucleus. These fixed orbits were labeled by a quantum number n .
2. An electron in one of these orbits has an energy associated with being in that orbit (also called the energy level).
3. The jumping from one orbit to another corresponds to the electron moving from one orbit to another. The Balmer series occurs because the electron in a series of high orbits falls to the orbit of $n = 2$.

Bohr's model received further experimental support in a series of experiments conducted by Franck and Hertz on mercury vapor. They found, as they bombarded mercury with electrons of varying energy, that at discrete energies the electrons lost all of their kinetic energy due to absorption of that energy by the mercury atoms.

The Bohr model had a number of shortcomings, not least the unanswered question of why these orbits exist. An elegant explanation of this came in 1925 when de Broglie postulated that a particle such as an electron in an orbit would have a wave associated with it, such that

$$\lambda = \frac{h}{mu} \quad (2)$$

Where: λ is the wavelength; h is the Planck constant (6.63×10^{-34} Js); m is the mass of the particle; and u is the velocity of the particle.

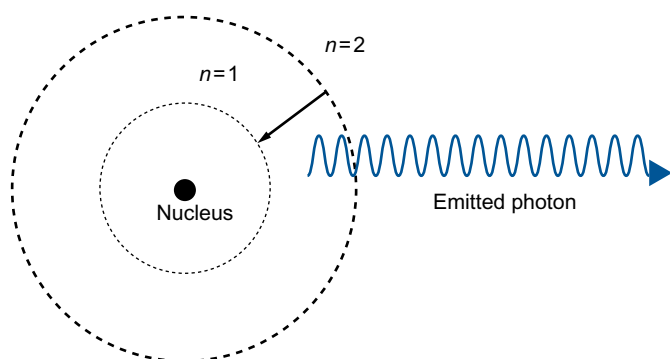


Figure 11-1. The electronic transition from $n = 2$ to $n = 1$ emits a photon matching the energy difference between the orbits.

In the circular orbits proposed by Bohr, standing waves will exist if

$$n\lambda = 2\pi r \quad (3)$$

Where: n is an integer; r is the radius of the orbit.

In the Bohr model, the centripetal force of the electron spinning about the nucleus must be balanced by the electron's angular momentum (Figure 11-2) so that:

$$\frac{Ze^2}{4\pi\epsilon_0 r^2} = \frac{mu^2}{r} \quad (4)$$

Where: Z is the charge on the nucleus (Table 11-1) or atomic number; e the charge of the electron; r the radius of the orbit; m the mass of the electron; u the electron velocity; and ϵ_0 the permittivity of free space.

The total energy of the electron in the orbit is a sum of its kinetic energy and potential energy. This is given by:

$$\text{Kinetic energy} = \frac{1}{2}mu^2 = \frac{Ze^2}{8\pi\epsilon_0 r} \quad (5)$$

$$\text{Potential energy} = -\frac{Ze^2}{4\pi\epsilon_0 r} \quad (6)$$

$$\text{Total energy} = \text{Potential} + \text{Kinetic} = -\frac{Ze^2}{8\pi\epsilon_0 r} \quad (7)$$

With these equations the radius of the orbit (r) may be solved:

$$r = \frac{n^2 h^2 \epsilon_0}{Z \pi m e^2} = \frac{n^2 a_0}{Z} \quad (8)$$

Bohr calculated that $a_0 = 0.0529$ nm (this is called the Bohr radius).

The total energy for each orbit is related to the inverse of the radius and thus the inverse of n^2 .

The energy of an orbit is thus: $E_n \propto \frac{1}{n^2}$ and when the electron jumps from one orbital to another:

$$\Delta E = E_{n_1} - E_{n_2} \propto \left(\frac{1}{n_2^2} - \frac{1}{n_1^2} \right) \quad (9)$$

The wavelength of the transition is related to the energy difference by:

$$\Delta E = \frac{hc}{\lambda} \quad (10)$$

These equations (9 and 10) show that the inverse of the wavelength is related to the inverse of the square of the energy level number, in this case n_1 and n_2 (equation 1). For example,

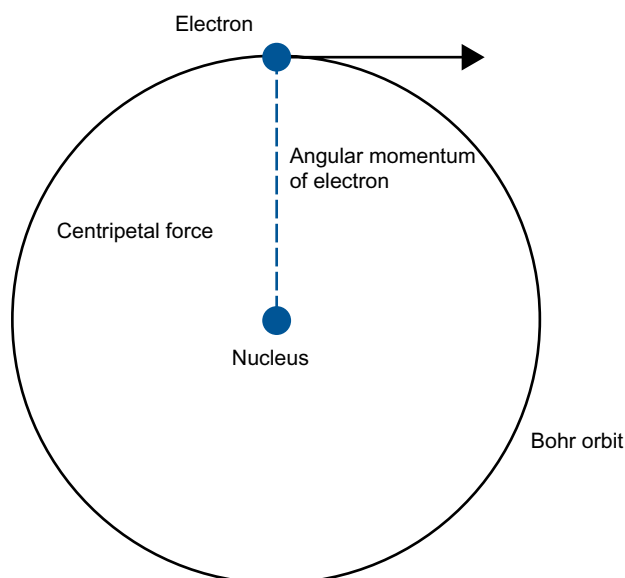


Figure 11-2. Forces acting on the electron in a Bohr model.

the Balmer emission series observed for the hydrogen atom has lines at 656.3, 486.1 and 434.1 nm; these correspond to emission transitions ($n_2 \rightarrow n_1$) of $3 \rightarrow 2$, $4 \rightarrow 2$ and $5 \rightarrow 2$, respectively.

The Bohr model, despite its shortcomings, was effectively the first modern model of atomic theory. The key issue was that the energy was quantized; this was related to the extent of motion of the electron in space (the orbits) and the fact that electrons could jump from one discrete level to another.

ELECTRONIC CONFIGURATIONS

Now that the existence of internal structure within the atom has been determined, we have to ask the following questions: (1) how are the levels labeled and (2) how do electrons fill the available quantized energy levels. First, these levels are defined by a series of quantum numbers. The principal quantum number n is referred to as the electron shell. Each shell can contain n^2 orbitals, which have orbital angular momentum quantum numbers l , where l can have integer values of 0 to $n-1$; in turn, each value of l has $2l+1$ degenerate orbitals to accommodate electrons. Within each degenerate set of orbitals, a further labeling occurs, based on the magnetic quantum number m_l , which takes integer values from $-l$ to $+l$. For example, for $n = 1$ $l = 0$ there is only one orbital called the s , or $1s$, orbital. For $n = 2$ $l = 0$ and 1, there is a $2s$ orbital ($n = 2$, $l = 0$) and three $2p$ orbitals ($n = 2$, $l = 1$, with $m_l = -1, 0, +1$); in total, the $n = 2$ level has four orbitals (Table 11-2). The orbitals are labeled by letters running s , p , d and f for $l = 0, 1, 2$, and 3, respectively. These labels originate from the nature of the spectroscopic transitions associated with the orbitals which are sharp, principal, diffuse, and fundamental. The principal quantum number levels may also be labeled with letters as K , L , M shells for $n = 1, 2$, and 3. There is one final quantum number that relates to the electrons filling these orbitals: it is spin, denoted by m_s and taking values of $+1/2$ or $-1/2$.

The labeling system tells us how the electrons are to be organized in an atom. The second issue is how these levels are filled. The building up or Aufbau principle provides a method of determining how electrons will fill the available orbitals for the ground state (lowest energy) configuration. The filling of orbitals is depicted in Figure 11-3, and a series of electron configurations for the first 25 elements is presented in Table 11-3.

Two additional issues need to be considered in this picture; first, the *Pauli Exclusion Principle* states that no two electrons may have the same four quantum numbers (n, l, m_l, m_s) – this limits the occupancy of an orbital to two electrons ($+1/2$ and $-1/2$) – and secondly, *Hund's rule* states that if degenerate orbitals are filled, the electrons spread themselves out as much as possible. Thus, if three electrons occupy the $2p$ orbital set, then the configuration is $2p_x^1 2p_y^1 2p_z^1$ (Figure 11-4).

As atoms interact with each other, the picture becomes somewhat more complex, and indeed, as electrons fill orbitals, they affect the energy of the orbitals and can have a bearing on the configuration. For example, in the case of Cr, in which one might expect a $4s^2 3d^4$ configuration, this is not observed. Because the half-filling of the $3d$ subshell imbues additional stabilization, the lowest configuration is $4s^1 3d^5$.

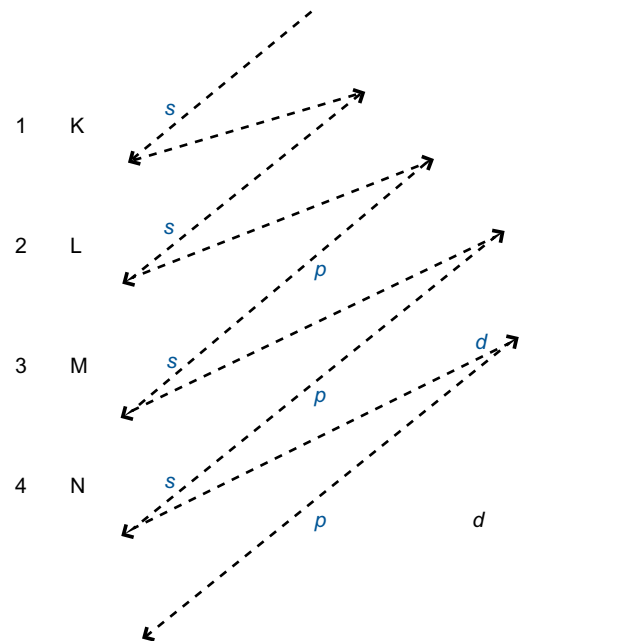


Figure 11-3. Filling diagram of shells and subshells. Levels are filled by following the dashed path.

The historic observation that reactivity has a periodic nature can be attributed to the electronic structure. For example, if halogens (fluorine, chlorine, bromine, and iodine) are reduced they become anions with a filled shell; that is, they have the electronic configuration of the noble gases. Thus, Cl^- and other halogen anions are stable. In a similar fashion, the alkali metals, lithium, sodium, and the like, readily lose an electron to adopt the noble gas closed shell electronic configuration. Thus, Li^+ and Na^+ salts may be readily formed, but Na^{2+} salts are not observed.

Although the labeling of the orbitals and shells may appear a book-keeping exercise, the quantum numbers have important consequences on the chemical and physical properties of the atoms. Furthermore, the quantum numbers affect the shape and energy of the orbitals; for example, s orbitals are spherical, but p and d orbitals have distinct shapes; p orbitals appear as dumb-bell shaped (Figure 11-4) and are directed along x , y or z axes, depending on the value of m_l . They also possess sign character; that is, the wavefunction that defines the p orbital has both a negative and positive part. This is normally depicted by color, as shown in Figure 11-4, using black and white.

FORMATION OF BONDS

The differing electronic configurations of the atoms mean that they can interact with each other to form larger structures, such as molecules or salts. Depending on the nature of the atoms, there are different interactions possible, at which we will take a closer look in the following section.

Table 11-2. Properties and Labels Used in Describing Orbitals

Orbital	Allowed quantum numbers		Number of orbitals
	l	m_l	
s	0	0	1
p	1	$-1, 0, +1$	3
d	2	$-2, -1, 0, +1, +2$	5
f	3	$-3, -2, -1, 0, +1, +2, +3$	7

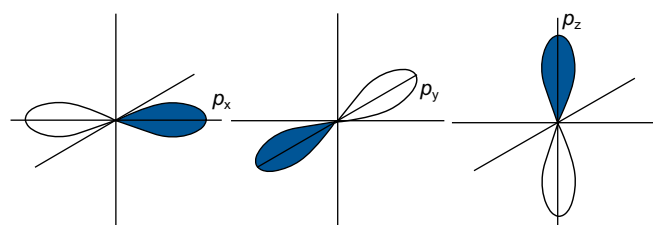


Figure 11-4. Depiction of $2p$ orbitals.

Table 11-3. Electronic Configurations for First 25 Elements in Their Ground State

Atomic number (Z)	Element	n , shell	1	2	3		4	3	
			K	L	M		N	M	
		l , subshell	0	0	1	0	1	0	2
			1s	2s	2p	3s	3p	4s	3d
1	H		1						
2	He		2						
3	Li		2	1					
4	Be		2	2					
5	B		2	2	1				
6	C		2	2	2				
7	N		2	2	3				
8	O		2	2	4				
9	F		2	2	5				
10	Ne		2	2	6				
11	Na		2	2	6	1			
12	Mg		2	2	6	2			
13	Al		2	2	6	2	1		
14	Si		2	2	6	2	2		
15	P		2	2	6	2	3		
16	S		2	2	6	2	4		
17	Cl		2	2	6	2	5		
18	Ar		2	2	6	2	6		
19	K		2	2	6	2	6	1	
20	Ca		2	2	6	2	6	2	
21	Sc		2	2	6	2	6	2	1
22	Ti		2	2	6	2	6	2	2
23	V		2	2	6	2	6	2	3
24	Cr		2	2	6	2	6	1	5
25	Mn		2	2	6	2	6	2	5

Ionic bonds coulomb interactions

One way in which atoms can form larger structures is through ionic bonding. This occurs when two elements combine, whereby one adds and the other one loses electrons to become ionic. For example, Na readily loses an electron to become the cation Na^+ (note that the electronic configuration of the Na^+ is a close shell and thus stable), and the chlorine atom can gain an electron to become an anion, Cl^- (this also now has a close shell configuration). These two types of ions attract each other because of their opposite charge, and in this way salts are formed. The bonding that holds these ions together is coulombic in nature, and the structure of the salt is a balance of these coulombic forces. In the case of NaCl, these forces balance to create a crystal of many thousands of atoms, in which the ions are held in a face-centered cubic arrangement (Figure 11-5, also see the section on crystalline solids).

Covalent bonds

Covalent bonds occur when electrons are shared between atoms. If we consider two H atoms interacting as they approach, there is an interplay between the coulomb interactions of the two nuclei and two electrons (which repel) and nuclei and electrons (which attract). At very short distances the coulombic forces of repulsion dominate, and the energy increases. At long distances the atoms cannot sense each other, and there is no interaction; however, at distances of 120 to 200 pm (1 pm = 10^{-12}m), the electrons may be shared between the atoms, and this creates a bond that requires energy to break. The electrons

that started in 1s atomic orbitals now occupy a new orbital, which is called a molecular orbital (MO), which is a result of the interaction between the atomic orbitals. The new molecular orbital forms from the spatial overlap of the 1s orbitals. In general, the strongest overlap occurs where atomic orbitals are similar in energy and occupy the same space as the nuclei approach (Figures 11.6 and 11.7).

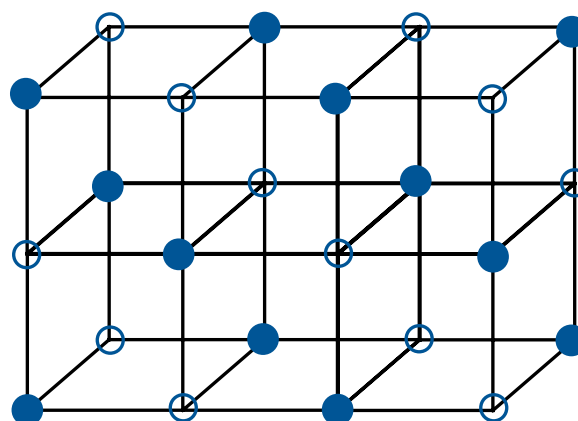


Figure 11-5. Depiction of a section of the NaCl crystal lattice with the Na^+ ions as open circles and Cl^- as closed circles.

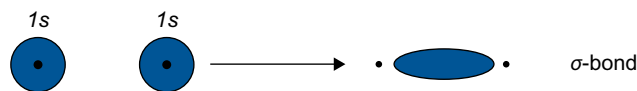


Figure 11-6. Overlap of two 1s orbitals to give a bonding σ -orbital.

Let us take a closer look at these MOs. As the 1s atomic orbitals overlap, two new orbitals are formed. The lower energy orbital is bonding in character and termed a σ MO, but there is also a corresponding antibonding MO, termed σ^* (Figure 11-8). In the case of H_2 , because there are only two electrons, only the bonding σ MO is populated, and a bond is formed. However, if the atom 1s atomic orbitals are filled, as is the case of He ($1s^2$), He_2 cannot form because the bonding interaction of the σ MO is counteracted by the population of the antibonding σ^* MO.

In the case of H_2 , only one type of bond can form because each atom has a $1s^1$ electron configuration, but for atoms with many electrons, a number of types of molecular orbitals may be generated. If we consider N atoms, the electron configuration is $1s^2 2s^2 2p^3$, with an electron in each of the p orbitals (Hund's rule). A number of molecular orbitals may be formed by the overlap of the atomic orbitals. If we consider the p -orbitals, there are two types of bonds that can be formed. For the $2p_z$ orbitals of each atom, these can overlap directly; the resulting molecular orbital holds electron density between the two N atoms – the σ -bond. But the $2p_x$ and $2p_y$ orbitals can also overlap in a face to face fashion. Each of these is a π -bond, in which electron density is not directly between the atoms but above and below the σ -bond (Figure 11-9). In the case of N_2 , there are two π -bonds that can be formed (Figure 11-10). A σ -bond typically has a bond energy of $100 \text{ kcal mol}^{-1}$; triple bonds (one σ - and two π -bonds) can have bond energies of $240 \text{ kcal mol}^{-1}$.

H_2 and N_2 are examples of bonding in which both atoms involved are the same, and thus the electrons are shared equally between the nuclei. However, for heteronuclear compounds (in which the atoms differ), the sharing of electrons will not be equal, and the electron density will lie to the more electronegative nucleus. This effect will occur in both σ - and π -bonds, and this results in a polarity in the bonding and dipole moment in the molecule.

The existence of polar bonding has its origin from the electronic configurations of the elements; for example, for halogen gases which are one electron short of a close shell configuration (and the additional stability that is brought by such a configuration), the sharing of electron density in a covalent bond between two differing atoms is not symmetric. The atom with a stronger ability to attract the electron (which can be described

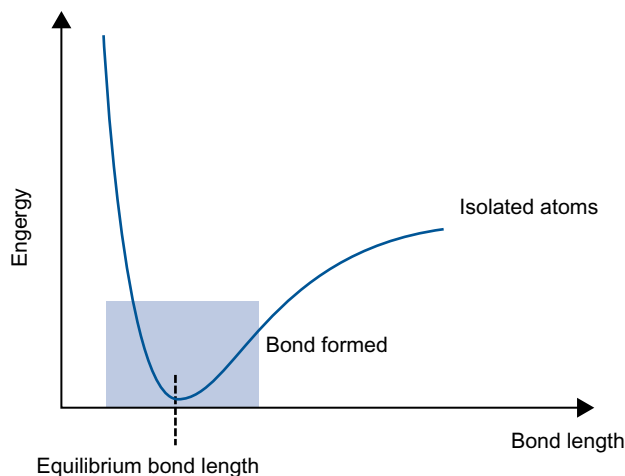


Figure 11-7. Bond length energy diagram, showing the region in which a bond may be said to be formed (gray shading).

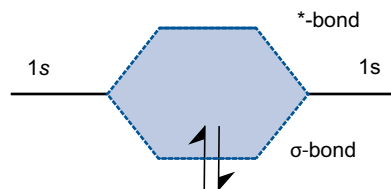


Figure 11-8. Depiction of the changes in energy when two 1s orbitals combine to form a σ and σ^* MO.

by the electronegativity of the element) will polarize the bond so that electron density lies closer to a more electronegative atom. This has the important consequence of creating a polar bond which has associated with it a dipole moment (a disparity of charge distribution). Polar bonds and dipole moments are important properties of molecules that have an effect on the bulk properties of the material. The electronegativities for some elements and dipole moments for some simple molecules are shown in Tables 11.4 and 11.5.

More than one bond from a central atom

Because a large number of pharmaceutically important compounds are carbon-based, we will take a closer look at the carbon atom. Carbon, with its $2s^2 2p^2$ valence shell, has some interesting properties. Most notable is that carbon readily forms up to four covalent bonds with other carbon atoms and thus, from carbon, many large molecular structures are obtained. It is worth considering how one might draw such structures as a variety of methods are used. The simplest is to draw the carbon atom with four bonds from it arranged at 90° ; this is easy to draw, but it is important to realize (as we will see shortly) that the molecules have a three dimensional shape. This is depicted in Figure 11-11, which shows a model of CH_4 , as well as two ways in which it can be drawn on paper.

A cursory consideration of the electronic structure of the carbon atom might suggest that only two covalent bonds can be formed as the 2s orbitals are filled. However, when atoms approach each other, the atomic orbitals may become reorganized

Table 11-4. Electronegativities of Selected Elements

Element	Electronegativity
F	3.98
O	3.44
N	3.04
Cl	3.16
Br	2.96
S	2.58
C	2.55
I	2.66
P	2.19
H	2.20
B	2.04
Si	1.90
Al	1.61
Mg	1.31
Li	1.00
Ca	1.00
Na	0.93
K	0.82

(Data from Pauling, L. (1932). "The Nature of the Chemical Bond. IV. The Energy of Single Bonds and the Relative Electronegativity of Atoms". *Journal of the American Chemical Society* 54 (9): 3570–3582.)

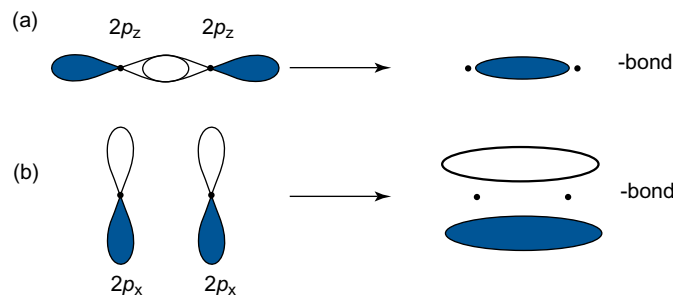


Figure 11-9. (a) Overlap of $2p_x$ orbitals to create a σ -bond, (b) Overlap of $2p_z$ orbitals to create a σ -bond.

and form hybrid orbitals. These hybrid orbitals may be considered as a combination or average of the atomic orbitals from which they are composed.

One way to think about this is that although the isolated carbon atoms may have a $2s^2 2p^2$ ground state electronic configuration, as carbon atoms approach each other, this $2s^2 2p^2$ configuration may not be the energetically favored state. In the case of diamond, which is a network of tetrahedrally bonding carbon atoms (each having four equivalent bonds), the valence shell of the atoms in this bonded state is $2s^1 2p^3$ and, more importantly, the shapes of the atomic orbitals are not maintained, and new molecular orbitals are generated. These specific orbitals are called sp^3 orbitals, and more generally this type of orbital is called a hybrid orbital because they come from a hybridization of the original atomic orbitals.

The sp^3 orbitals covalently link the carbon atom to other atoms through four σ -bonds. As each of the orbitals is equivalent, they repel each other and adopt a tetrahedral geometry. This sort of bonding is prevalent in alkane structures, that is, hydrocarbons with no double bond character. However, this is not the only way in which the valence electrons of carbon atoms may hybridize. It is possible that only three of the orbitals hybridize (the $2s$, $2p_x$ and $2p_y$ for example) to form a sp^2 centre. The three sp^2 orbitals lie in a plane (the xy plane in this case) with an angle between each bond of 120° . Once again, the hybrid

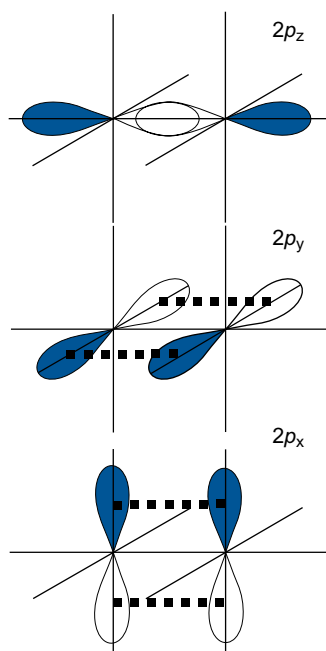


Figure 11-10. The bonding of valence orbitals in N_2 , showing one σ - and two π -bonds.

Table 11-5. Dipole Moments for a Selection of Heteronuclear Diatomic Molecules

Molecule	Dipole / D ¹
CO	0.11
HCN	2.98
HCl	1.05
HBr	0.82
HI	0.38

¹1D (one debye) corresponds to $1 \text{ D} = 3.344 \times 10^{-30} \text{ cm}$

orbitals are as far apart as is possible. Of course, in such an arrangement there still remains a p orbital ($2p_z$) at right angles to the sp^2 plane. The result of this hybridization is that the carbon centre may bind through three σ -bonds, but the additional perpendicular p orbital can form a π -bond. In this way, one of the linkages between two sp^2 carbon atoms will have both a σ - and π -bond to give double bond character. The electrons that occupy the π -bond are often referred to as π -electrons, and they possess a number of interesting electronic properties.

We know that sp hybridization occurs when only one of the carbon $2p$ orbitals interacts with the $2s$ orbital; this creates two sp hybrid orbitals that form two σ -bonds linear to each other. However, there remain two unused p orbitals, each of which contains an electron, so that two π -bonds may be formed between adjacent centers. This leads to carbon-carbon triple bonded species, the alkynes.

Hybridization shows how the original atomic orbitals may change shape considerably in forming molecular structures, as depicted in Figure 11-12.

There are a number of important differences between σ - and π -bonds. The most striking of these is that π -bonds are rigid; they cannot rotate because the interaction between adjacent $2p_z$ orbitals only occurs if these orbitals are in the same plane.

Resonance and bond delocalization

Benzene and other extended sp^2 bonded structures exhibit some interesting behavior with respect to the π -electrons. The six carbon $2p_z$ atomic orbitals that create the π -bonds do not have to overlap with only one adjacent $2p_z$ but can share electron density with both adjacent atomic orbitals. This can happen with each of the $2p_z$ atomic orbitals in the ring. This delocalization of electron density or extended conjugation imbues the benzene molecule with chemical stability because of a new series of π -molecular orbitals being created. This delocalization of electrons may be depicted by the π -molecular orbitals or by drawing single and double bonds and adding the possibility of resonance structures (Figure 11-13).

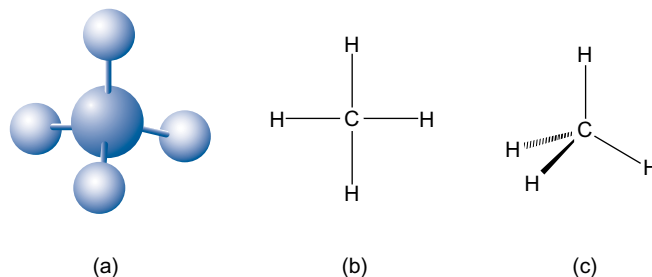


Figure 11-11. (a) Model of CH_4 and other drawing formats, (b), and (c). Note that (c) implies a three-dimensional structure with the thick wedge sticking out of the page and the dashed wedge sticking into it.

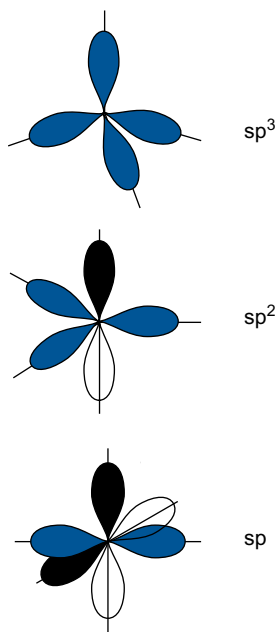


Figure 11-12. Depiction of hybrid orbitals (blue) and residual porbitals (black and white).

If one draws the structure of benzene with single and double bonds (the so-called valence bond structure), then the double bonds may be placed in different places. This is not a mere nicety: it actually implies that the structure is more complex than either one of the diagrams. This “resonance” between these two valence bond structures was first suggested by Kekulé in 1872.

Extended sp^2 systems demonstrate this delocalization in their electronic spectra, i.e., their color. The electronic spectrum can show the energy difference between the highest occupied molecular orbital (HOMO, in this case a π -orbital) and the lowest unoccupied MO (LUMO). For ethene, which has a single π -bond, the observed transition from HOMO to LUMO is at 171 nm, which corresponds to a difference in energy between HOMO and LUMO of 166 kcal mol⁻¹. For butadiene (which has two π -bonds conjugated) the transition is at 217 nm, corresponding to 131 kcal mol⁻¹. For β -carotene, which has 11 conjugated π -bonds, the transition is at 470 nm, corresponding to 60 kcal mol⁻¹. These examples demonstrate how the extending conjugation changes the electronic structure because of the delocalization of π -electrons.

Using hybridization and ideas of delocalization, it is easy to rationalize the properties of two of the allotropes of carbon, namely diamond and graphite. In the case of diamond, each carbon atom is sp^3 hybridized and diamond is an extended structure, in which each carbon has four bonds in a tetrahedral arrangement to four other carbon atoms. The rigidity of these σ -bonds creates the hardness diamond has (diamond is the hardest natural material, graded 10 on the scale of mineral hardness, the so-called Mohs scale). In the case of graphite, each carbon atom is sp^2 hybridized, which creates flat structures with delocalized electrons. Thus, the material has very low energy levels between its HOMO and LUMO – hence, it is black in color. Furthermore, graphite readily conducts electricity because of the delocalized π -bonds. The extended structure of graphite is a consequence of the $2p_z$ orbitals, which bond weakly between sheets of the sp^2 structures; this imbues graphite with its lubricating property.

Extended molecular systems that involve non-carbon atoms can also be bound through σ -bonds, as in C–Cl or π -bonds C=O, and heterocyclic systems, such as pyrrole. Because the electronegativity of atoms such as O, S, Cl, and other non-carbon

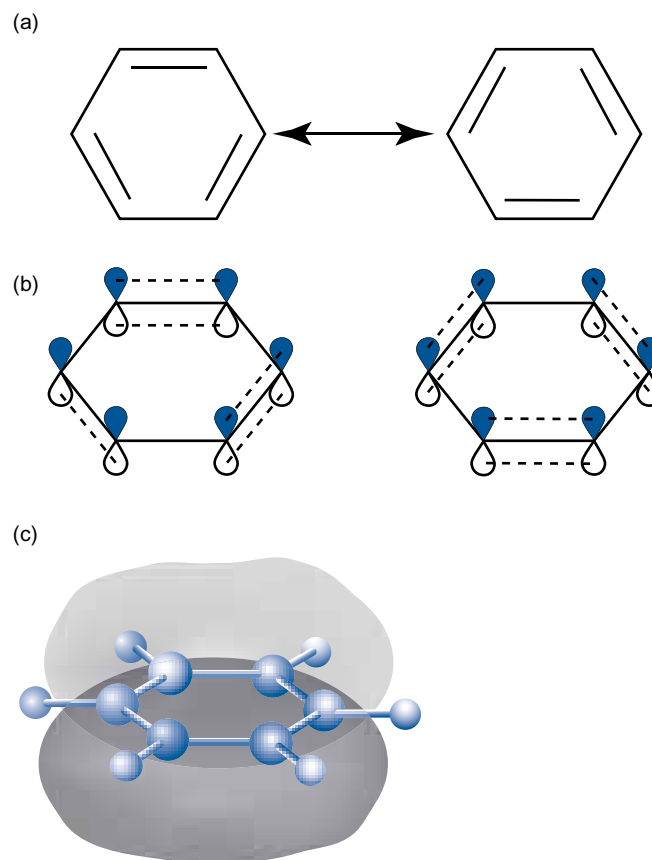


Figure 11-13. (a) Valence bond resonance structures for benzene, (b) overlap of $2p$ orbitals that results in delocalized bonds, and (c) resulting molecular orbital delocalized over the entire ring.

atoms differs from that of C, this results in polar bonds in the molecular structure containing these systems.

Additional types of bonding

A number of other bonding types are recognized that flow from the ideas of covalent and ionic bonding.

The *coordinate covalent bond* arises when overlap between atomic orbitals creates a molecular orbital, but only one of the atoms involved has electrons. This type of bonding is commonly found in amines because the nitrogen adopts an sp^3 configuration, but since it has an extra electron (N is $2s^2 2p^3$) compared to carbon, only three σ -bonds are formed, leaving a lone pair (Figure 11-14). This lone pair can create a bonding interaction if it overlaps with an empty orbital of another species. Boron, which has one less electron than carbon, can bind to an amine (Figure 11-15). As the amine donates electrons, it is termed a Lewis base, and the boron, which accepts electrons, is called a Lewis acid.

Coordinate bonding may also occur in metal complexes in which the binding species, the ligand, donates electrons to form the σ -bonding, and the d orbitals on the metal play a secondary role in bond formation. *Cis* platin (Figure 11-16) is probably the best example of a drug that has this bonding characteristic.

Metalloenes

In 1973 Fischer and Wilkinson shared the Nobel prize in chemistry for the discovery of ferrocene.^{2,3} Ferrocene is composed of two cyclopentadienyl rings with an iron in between (Figure 11-17). What is remarkable about ferrocene is that it is very stable; it can be boiled at 249°C. The compound was first reported in 1951 by Kealy and Pauson.⁴ Based on the knowledge of bonding that

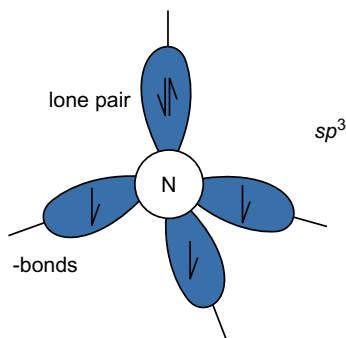


Figure 11-14. The sp^3 hybridized N atom centre.

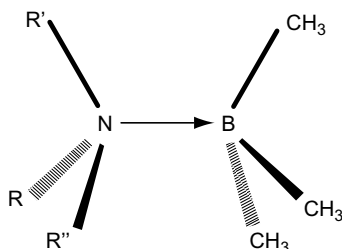


Figure 11-15. Formation of a coordinate bond between a tertiary amine and trimethylboron.

existed at that time, they proposed a σ -bonded structure. However, they did recognize that such a structure would not have great stability, and thus, that what they had made was unusual. In fact, ferrocene is bonded not by σ -bonds but rather by π -overlap.⁵ The cyclopentadienyl ring in ferrocene can be thought of as having five sp^2 hybridized carbon atoms (the ligand formally carries a negative charge). The residual electrons lie in unhybridized p orbitals. These are able to overlap and bond with the $3d$ orbitals on the iron (which is formally Fe^{2+}), a so-called metallocene or haptic bond. This makes the compound chemically stable, and ferrocene behaves like an aromatic system, such as benzene.

ISOMERS

One of the critical issues in dealing with molecules is that they can bond to each other in different ways yet have the same formula. For example, the formula C_2H_6O corresponds to ethanol or dimethyl ether. The former is a colorless liquid, which can be tolerated and metabolized by the human body (having an LD_{50} in rats of about 7000 mg/kg). The latter is a colorless gas used in aerosol cans and is much more flammable than ethanol and more toxic (LD_{50} in rats of about 1000 mg/kg). These are examples of structural isomers in which the atoms in the molecule are the same, but they are attached in different ways.

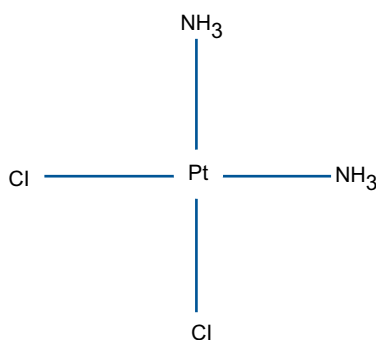


Figure 11-16. *cis* platin. The NH_3 and Cl ligands donate electrons for the four σ -bonds in this structure. The platinum actually adopts a planar structure.

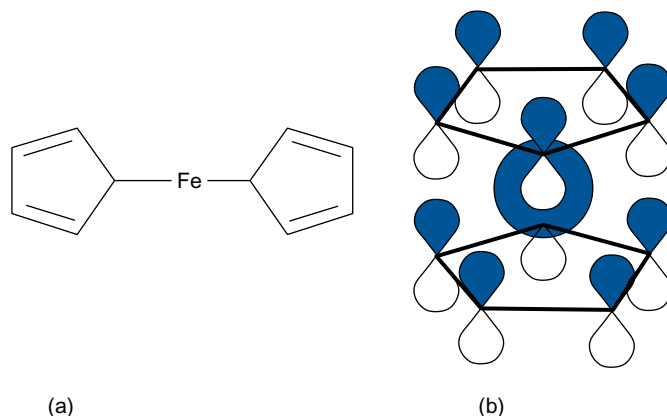


Figure 11-17. (a) The incorrect σ -bonded structure of ferrocene, and (b) the metallocene structure.

However, isomers in molecules can be more subtle than those discussed above. Geometric isomers exist where the bonding attachment between two compounds is the same, but because of the rigidity of the molecule, the atoms are not in identical positions. This is common in alkene systems because the double bond cannot rotate.

In the example shown in Figure 11-18, the compounds have identical attachments, in that each sp^2 carbon is bonded to one hydrogen, one chlorine, and another sp^2 carbon. But in one case the chlorine atoms lie on the same side of the π -bond, and in the other on opposite sides. The same-side isomer is called the *cis* isomer; the opposite-side is the *trans* isomer.

These geometric isomers are a subset of stereoisomers: compounds with the same structure but for which the atoms do not occupy the same arrangement in space. *Cis* and *trans* isomers can also exist if hydrogen bonding is present (see below the example of captopril) or in rings that restrict rotation of the bonds (Figure 11-19).

Conformations

Open chain compounds can create different conformations because of the ability to rotate the σ -bonds. If we consider C_2H_6 , ethane, the rotation of the central carbon carbon bond may be such that the hydrogen atoms are *staggered* with respect to each other, or *eclipsed*. These are not isomers but different conformers, and they have different energies. They may be depicted using a Newman projection, in which the molecule is viewed down the axis of the bond that is undergoing rotation (Figure 11-20).

The situation is slightly more complicated if there are different substituents attached through the σ -bond of interest. In the case of butane, the central carbon linkage has each carbon atom bonded to two hydrogens and one methyl group (Figure 11-21). The eclipsed and staggered conformations must also consider the interaction of the methyl groups. In this case, when the methyl groups are opposite, the conformation is fully eclipsed and has the highest energy as the repulsive interactions of the methyl groups are maximized. If the methyl groups are opposite hydrogen atoms, the conformation is eclipsed. In the staggered conformations when the methyl groups are a maximum distance apart, the conformation is called *anti*, and when they are staggered but closer, the conformation is called *gauche*.

Enantiomers

These are molecules that exhibit optical activity, in that they interact with light of different handedness in differing ways. They normally possess the same atoms around a tetrahedral central, such as carbon sp^3 , but they are non-superimposable, and they are called chiral centers (Figure 11-22). These enantiomers are chemically identical but interact with plane polarized light in different ways.

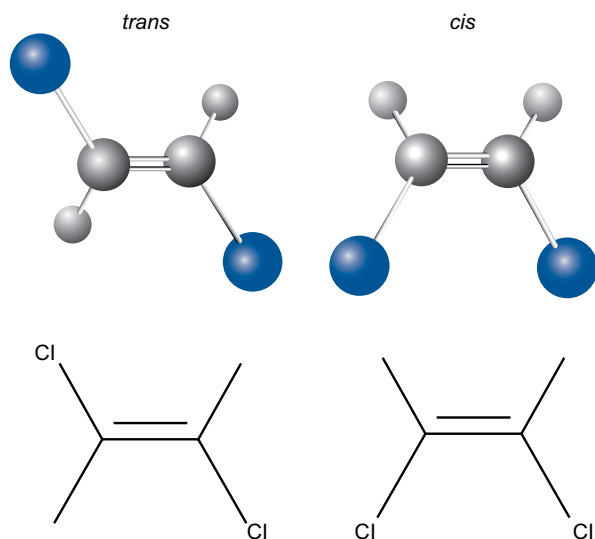


Figure 11-18. Depiction of *trans*- and *cis*-1,2-dichloroethene as a molecular model picture and as a chemical structure.

The absolute configuration of a chiral centre (termed *R* and *S* configurations) may be determined by using the following methodology.

1. Allocate substituent priorities.
2. The substituent with the atom bonded to the chiral centre with lightest mass is allocated as priority four, the next heaviest is three, and so on. In the event of equal mass bonded atoms, count down the chain until the substituents differ, and that with the lower mass is allocated the priority number up to four. In the example shown in Figure 11-22, the Cl is priority one; O is priority two; carbon, priority three; and hydrogen, priority four.
3. Rotate the sp^3 centre, such that the lowest priority substituent is pointing away from the viewer, i.e., into the page.
4. Determine if the priority labels are arranged in a clockwise or counter-clockwise fashion with regard to highest priority one to three. A clockwise orientation has an *R* absolute configuration, and a counter clockwise orientation has an *S* configuration (Figure 11-23).

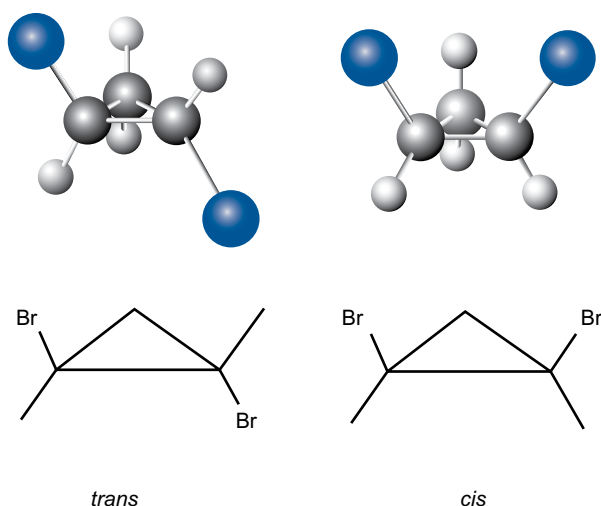


Figure 11-19. *Trans* and *cis* isomers of dibromocyclopropane; the strained ring system does not allow for bond rotation.

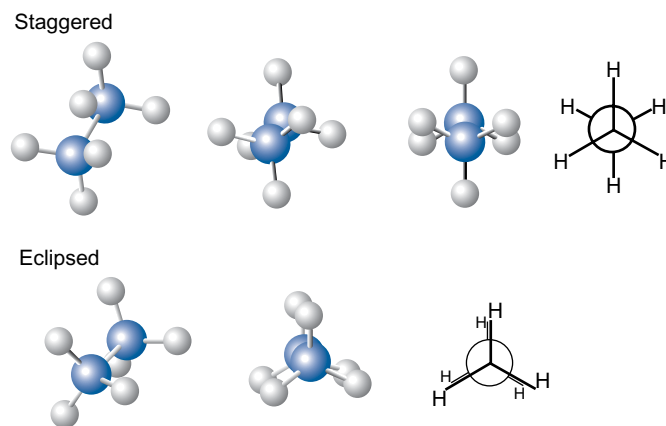


Figure 11-20. Staggered and eclipsed. Depiction of three dimensional models of the staggered and eclipsed conformers of ethane – rotated around into the orientation used in the Newman projection, shown on the right for each conformer.

Measuring optical activity

The most striking difference between two enantiomers is their interaction with light. In most other respects, such as physical properties like melting and boiling point, they are identical, but in their interaction with light, they may be differentiated.⁶ In 1815 Jean-Baptiste Biot noted that some materials could rotate plane polarized light. Subsequent to this, Louis Pasteur discovered that tartrate crystals could be separated by physically picking them out under microscopic examination and

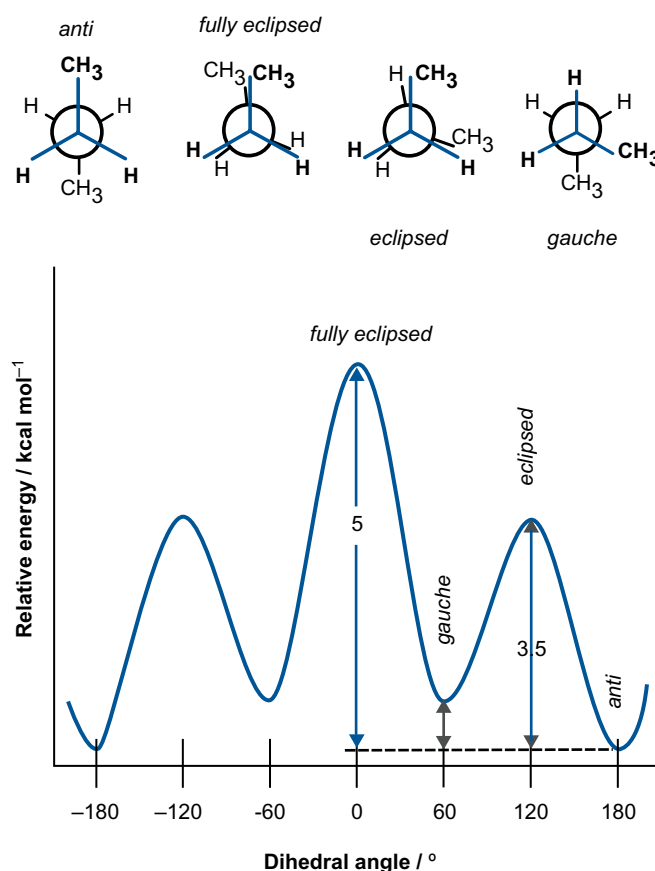


Figure 11-21. Depiction of different conformers for butane and the changes in relative energy that occur as the dihedral angle between the methyl groups is altered from 0° (fully eclipsed) to 180° (anti).

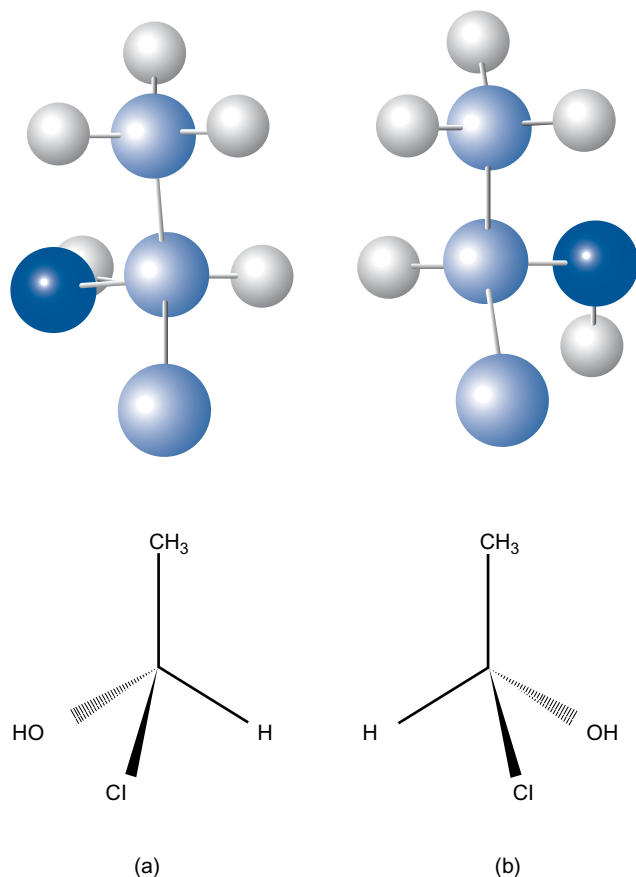


Figure 11-22. Models and chemical structures of two enantiomers.

that these two separate types of crystals rotated plane polarized light to the right or to the left in equal amounts. Yet a mixture of the two types of crystals had no effect on the polarization of light. The rotation of plane polarized light is a fairly subtle effect. To understand it, we need to consider the nature of light. We can think of light as an oscillating electromagnetic wave. The wave moves in the z -direction with the electric field oscillating in the xz plane and that of the magnetic field in yz . We need only concern ourselves with the electric field. The question arises, how can a chiral centre affect a plane wave? To answer this, we need to consider the plane wave as being made up of two components; each of these is a vector, but they rotate in a clockwise (E_R) and counter-clockwise (E_L) direction. When they are aligned (Figure 11-24), the electric field is at a maximum in the x axis; as they rotate, the strength of the field in the x -axis diminishes until they oppose each other, and after that the direction of the field changes until both vectors point in the $-x$ direction. Note that at no time during the precession of the two vectors is there any electric field in the y axis because they always cancel. What is remarkable about chiral centers is that they interact differently with the E_R and E_L vectors. This can be described in a number of ways, but generally it is said that the refractive index (n) for each direction differs; that is, n_R does not equal n_L . This literally means that the speed of light differs for each of the vectors; in other words, the speed with which E_R rotates differs from that of E_L . This is depicted in Figure 11.25, in which the E_R vectors rotates more slowly than E_L . The effect of this is that the resultant vector (E) has a component in the x -axis and thus starts to rotate. The longer the path length of the optically active medium, the more it will rotate, and furthermore, as soon as the light exits the medium, the two rotating vectors continue rotating at equal speed, and the light becomes plane polarized again just with a rotation imbued from the interaction with the optical material.

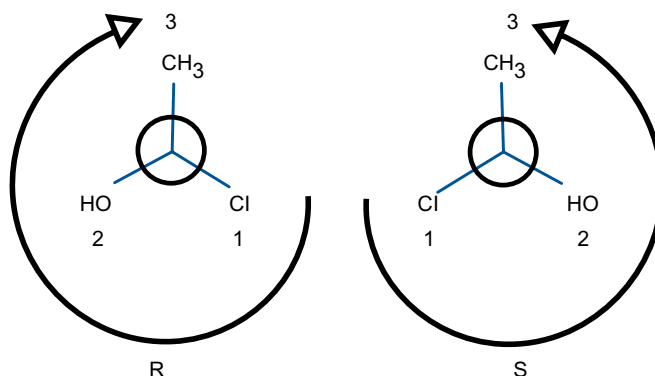


Figure 11-23. The rotated structures from Figure 11-22, showing absolute configuration.

The angle of rotation of light (α) is given by:

$$\alpha = \frac{1800}{\lambda} (n_L - n_R) \quad (11)$$

Where: α is called the angle of rotation [degrees per cm of medium]; λ is the wavelength of the polarized light [m]; and n_L and n_R the refractive indices of the left and right circularly polarized components of the plane polarized light.

It is very common to use the sodium D line at 589.3 nm for single wavelength determinations of angle of rotation. A more commonly used term is the specific rotation ($[\alpha]_\lambda^T$) which is given by:

$$[\alpha]_\lambda^T = \frac{\alpha}{lc} \quad (12)$$

Where: l is the pathlength of the solution of optically active material [dm]; and c the concentration [g mL^{-1}].

Both temperature (T) and wavelength (λ) need to be specified because they have a significant effect on the degree of rotation (see above). This effect is very small. For example, the angle of rotation of the optical material may be 2.00° ; this corresponds to a difference in refractive indices ($n_L - n_R$) of about 7×10^{-8} .

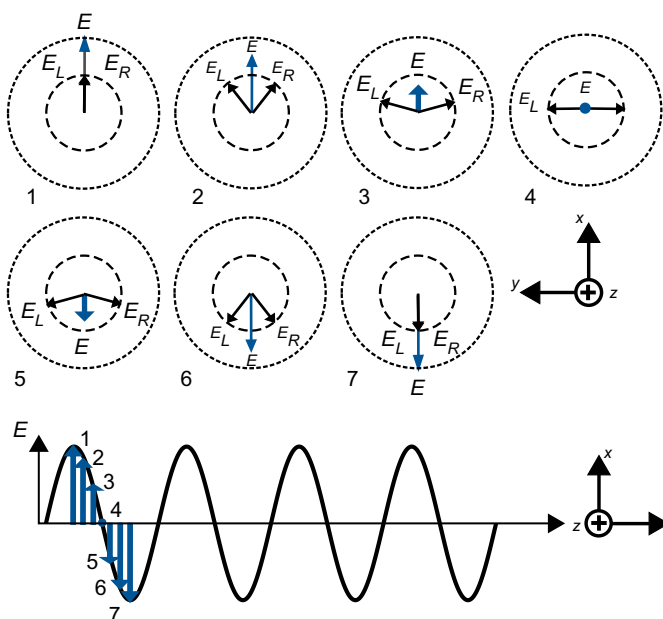


Figure 11-24. Depiction of the plane polarized light wave, made up of left-hand and right-hand rotating circular polarized vectors (E_L and E_R). As the vectors rotate, the amplitude of the resultant electric field goes from a maximum value at 1 through a node at position 4 to a minimum at position 7.

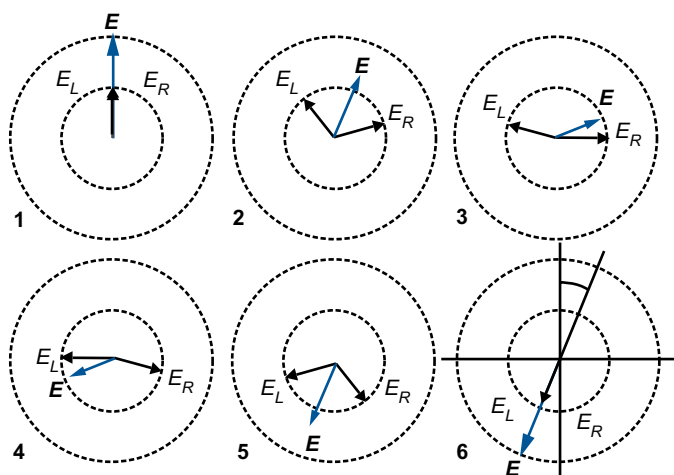


Figure 11-25. Depiction of the transmission of plane polarized light through an optically active medium (2-5). As the light passes through the medium, the left-handed circularly polarized light travels more slowly, and so the left and right parts get out of synch. Once the light exits, the medium both parts rotate at the same speed (6), and this results in a restoration of the plane polarized light field, but now it is rotated by an angle (α).

In addition to the ability of chiral centers to differentiate the circular components of plane polarized light, molecules with these centers can also absorb circularly polarized light to different extents. That is, the extinction coefficient (ϵ) is such that $\epsilon_R \neq \epsilon_L$. This phenomenon is referred to as circular dichroism (CD), and spectra of materials with this effect may be measured. Figure 11-26 shows the absorption, CD, and optical rotatory dispersion (ORD) spectra for a chiral molecule. The sign of the ORD response changes as the wavelength is scanned across the absorption spectrum, and the CD spectrum has a sign, depending on whether $\epsilon_R > \epsilon_L$ or vice versa. The CD plot is given by the molar ellipticity ($[\theta]$) versus wavelength, where:

$$[\theta] = 3300(\epsilon_L - \epsilon_R) \quad (13)$$

Where: $[\theta]$ is measured in degrees cm^{-2} per decimole and typically has values of approximately 0.08 degrees cm^{-2} per decimole.

If the light is rotated to the right, it has a positive value and is said to be dextrorotatory; if it is to the left, it has a negative value and is termed levorotatory. These terms are phenomenological

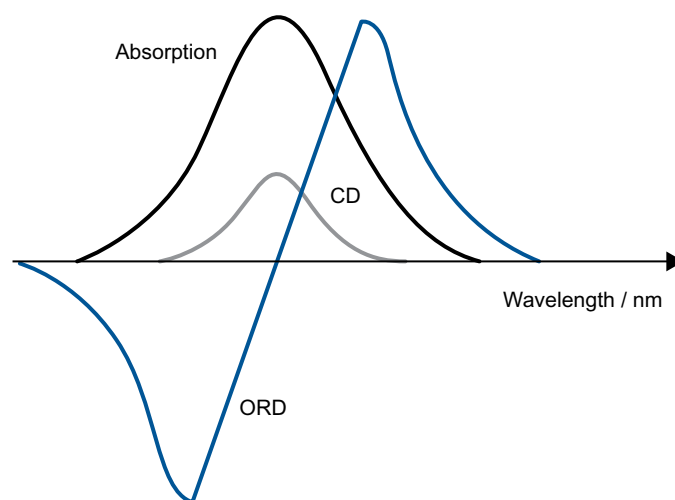


Figure 11-26. The spectral response of an optical absorption and the corresponding ORD and CD curves – note that the direction of the ORD and CD may be reversed depending of which handedness of light interacts with the optical material.

only, and the reasons for the direction of rotation cannot be readily associated with the absolute configuration (R or S) of the chiral centre.

MOLECULES WITH MANY CHIRAL CENTERS

It is not uncommon that molecules with more than one chiral centre exist. These molecules are termed diastereoisomers, and they show different physical properties in addition to rotation of plane polarized light. The naming of the molecules will reflect the atom number and the chirality associated with each centre. For example, in Figure 11-27a the molecule, 2,3,4-trihydroxybutanal is shown. The particular configuration depicted is 2S, 3S, based on the fact that carbon 2 has an S configuration and carbon 3 an S configuration. A convenient way to depict such structures is to use a method invented by Fischer in 1891. This involves rotating the three dimensional structure so that the carbon-carbon bond containing the two chiral centers is vertical in the plane of the page. The residual carbon backbone is also depicted vertically, although these now point backwards into the page, and the other substituents are coming out of the page. This depiction is then “flattened” to give the so-called Fischer projection. These Fischer projections allow one to conveniently depict multiple chiral centers. The four optical isomers of 2,3,4-trihydroxybutanal may be readily drawn, using Fischer projections (Figure 11-27b). Note that not all are enantiomers of each other, and indeed, if the substituents on the carbon atoms are the same, then *meso* configurations are possible, in which two of the depicted configurations become symmetry equivalent; this occurs in tartaric acid.

Intermolecular binding forces

Up till now we have focused on the interaction and binding of atoms to molecules. A critical aspect of the behavior of

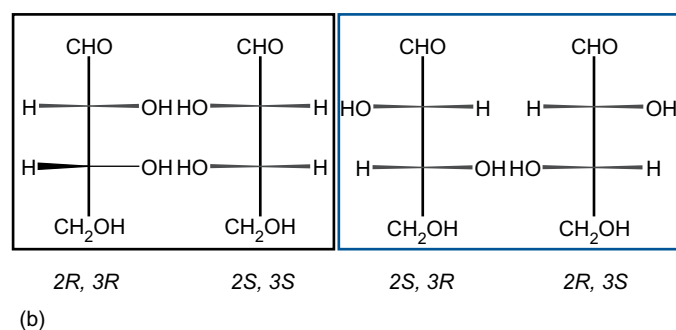
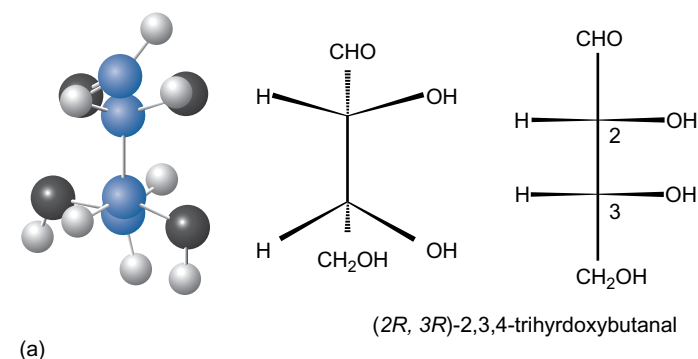


Figure 11-27. (a) Depictions of 2,3,4-trihydroxybutanal from a model structure (left) to the Fischer projection (right); the intermediate structure shows how the molecule is rotated so that the carbon backbone is in the plane of the page or behind it. (b) Fischer projections of 2,3,4-trihydroxybutanal. Absolute configurations are given, the 2R, 3R and 2S, 3S are enantiomers; the 2S, 3R and 2R, 3S are enantiomers. However, 2R, 3R and 2S, 3R are not enantiomers but diastereoisomers.

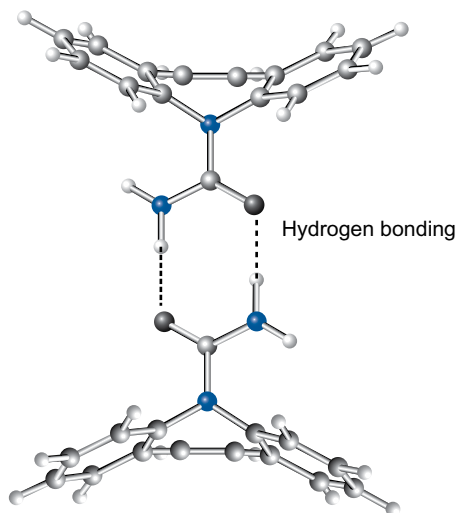


Figure 11-28. Structure of carbamazepine dimer with hydrogen bonding.

pharmaceuticals is the nature of the interactions between molecules, the intermolecular forces. The intermolecular forces may involve specific intermolecular bonds – such as hydrogen bonds – or they may involve weak attractive forces due to the electron distribution in the molecules.

Hydrogen bonding is a very important interaction, in which a pendant hydrogen interacts with the electronegative atom of a neighboring molecule. A good example of this is the anticonvulsant drug carbamazepine (CBZ), which shows distinctive hydrogen bonding between the H atom on the NH_2 group and the neighboring oxygen of the $\text{C}=\text{O}$ linkage (Figure 11-28). The energies of these interactions are rather weak, being less than 10 kcal mol^{-1} .

Dipole–dipole interactions occur where molecules have an asymmetric distribution of electrons, i.e., they contain polar bonds. These dipoles can align to lower the energy of the assembled molecules (Figure 11-29). Typically the energy of such interactions is weak, being less than 2 kcal mol^{-1} .

Dispersion forces occur even in molecules with no discernible dipole. The electrons buzz around the nuclei, and at a particular moment there may be an asymmetric electron distribution

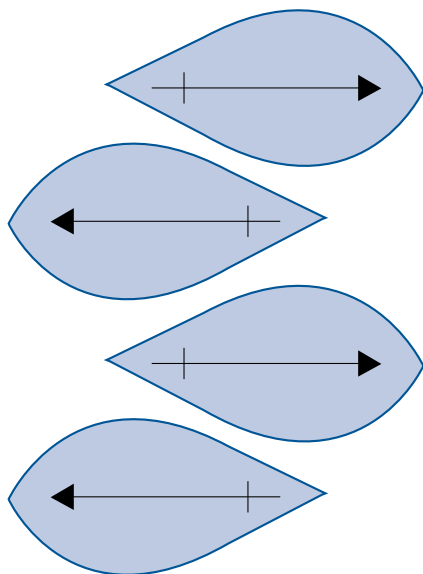


Figure 11-29. Molecules with dipole moments (depicted by arrows) align to lower energy through a dipole–dipole interaction.

and a brief temporary dipole. Despite the fleeting nature of these dipoles, an interaction does occur and may be referred to as either dispersion or London forces (after Fritz London who first described them). These are weak, being less than 1 kcal mol^{-1} , but they are present in all molecules.

Visualizing forces

It is now possible and relatively easy to calculate the structures of molecules using computational chemistry^{7,8} so the forces that hold the molecules together may be evaluated. Furthermore, it is possible to measure the forces that hold molecules together using spectroscopy. The strength of bonding is associated with the potential energy curve, depicted in Figure 11.7. In the case of a covalent bond, the bond can vibrate at discrete frequencies, and this may be detected by shining light of the appropriate energy at the molecule. For covalent bonds these resonances occur in the infrared region of the electromagnetic spectrum (40 to 4000 cm^{-1}), and most materials may be characterized by their infrared spectra; as the depth and steepness of the bonding interaction decreases, so too does the energy at which the resonances or absorptions occur. In fact, the absorption frequencies may be related to the force constants (k , N m^{-1}) that hold the molecule together and the masses (the reduced mass μ , kg) involved in the various vibrations. The upshot of this is that the stronger the bonding, the higher the frequency. If lightweight atoms, e.g., hydrogen, are involved in the vibration, then the frequency of the absorption will be high. The frequency of the absorption ($\tilde{\nu}$, cm^{-1}) is given by

$$\tilde{\nu} = \frac{1}{2\pi c} \sqrt{\frac{k}{\mu}} \quad (14)$$

Where: c is the speed of light [cm s^{-1}].

Some representative data on bond vibrations are given in Table 11.6.

The use of spectroscopy in understanding structure and bonding is well illustrated by the study of *cis* and *trans* captropil. Captropil, 1-[2(S)-3-mercapto-2-methyl-1-oxopropyl]-L-proline, is a drug for treatment of hypertension and heart failure. It exists in an equilibrium between *cis* and *trans* forms, with the *trans* form being the active form. The ratio of *trans*:*cis* is approximately 6:1 at room temperature in solution. The *trans* and *cis* isomers are created by the two conformers that exist about rotation around a C–N bond, with the *trans* form being locked by a hydrogen bond (Figure 11-30).

The captropil structure contains two carbonyl bonds ($\text{C}=\text{O}$), and the frequencies of vibrations for these is distinct, depending on the isomer formed. This is because the *trans* isomer is hydrogen bonded, which shifts the $\text{C}=\text{O}$ vibration from 1637 cm^{-1} down to 1591 cm^{-1} . Wang *et al.*,⁹ studied this drug in the solid

Table 11-6. Vibrational Spectroscopy and Bonding Parameters for Selected Bond Types

Bond type	$\tilde{\nu}/\text{cm}^{-1}$	k/Nm^{-1}	μ/kg	Bond strength/ kcal mol^{-1}
C–C	1130	450	9.96×10^{-27}	90 ^a
C=C	1600	900	9.96×10^{-27}	170
C≡C	1990	1400	9.96×10^{-27}	234
C–H	2880	450	1.53×10^{-27}	

^aEstimated from the bond enthalpy for C_2H_6 , C_2H_4 and C_2H_2 respectively.

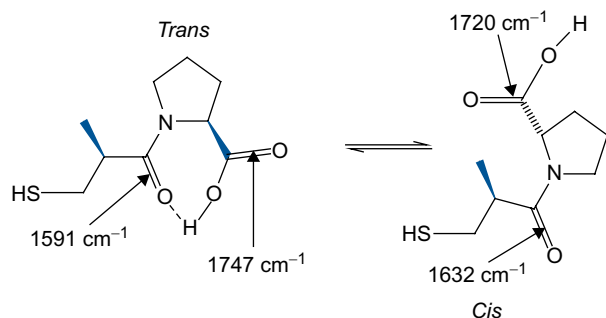


Figure 11-30. The structures of *trans* and *cis* forms of captopril and the frequencies observed for the various C=O bonds.

state and were able to determine the energetics between the two isomers.

For intramolecular interactions, the interactions are 10 to 20 times weaker than covalent bonds; thus, the interaction has very low energy infrared absorptions. These occur in the terahertz region of the spectrum (where $\tilde{\nu} < 100 \text{ cm}^{-1}$). Different polymorphs of a drug (see the section on polymorphism) will have very similar infrared spectra because the bonding interactions of the molecule that make up the pharmaceutical are virtually unchanged (unless there is hydrogen bonding), but the terahertz (THz) region of the electromagnetic spectrum is distinctive for each polymorph because the intermolecular forces differ between polymorphs. Carbamazepine shows a number of polymorphs, and these have been characterized using THz spectroscopy.¹⁰ The spectra of the form III polymorph shows distinctive bands at 31, 44 and 52 cm⁻¹ (Figure 11-31), whereas polymorph I has a band at 41 cm⁻¹.

STATES OF MATTER

INTRODUCTION

In everyday life we usually consider three states of matter: the solid state (solids), the liquid state (liquids), and the gaseous state (gases). For example, water can be differentiated into the three states: ice, liquid water, and water vapor, and at a given pressure (usually normal or atmospheric pressure, 1.013 bar) it is the temperature that decides in which state water is present; at normal pressure ice melts at 0°C and boils at 100°C. What we are observing at these temperatures are phase transitions from the solid state to the liquid state (melting) and from the liquid

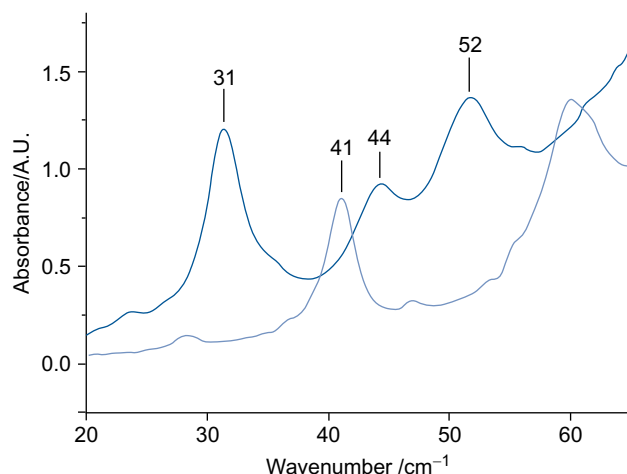


Figure 11-31. THz spectra of carbamazepine: Upper trace polymorph III, lower trace polymorph I.

state to the gaseous state (boiling). If we lower the temperature of water vapor below 100°C, we observe condensation (gas to liquid transition) and if we further lower the temperature below 0°C, we observe freezing of the liquid water to ice.

In this section we will describe the different states of matter (including a few “exotic states”, which nevertheless do play a role in some pharmaceutical applications) and highlight their most important characteristics. We will also discuss how we can describe the different states of matter in phase diagrams and identify a few experimental techniques that can be used to determine the characteristics of a given state of matter. Because most drugs and dosage forms used in the pharmaceutical field are solid at room temperature, we will discuss the solid state in a little more detail than the other states of matter.

SOLIDS, LIQUIDS AND GASES – A QUESTION OF INTERMOLECULAR FORCES

As outlined above, when we think about states of matter, we usually think of them in terms of solids, liquids, and gases. These distinctions are based on the interactions between the molecules (or in some cases atoms or ions) of the matter in question. Since in the pharmaceutical field we are most often dealing with molecules, the discussions below will mainly consider molecules as the components forming the various states of matter, and the reader should keep in mind that in many cases the word “molecule” could be exchanged for the words “atom” or “ion”.

In the solid state of a drug, the drug molecules are held together by comparatively strong interactions, which do not allow the molecules to change their position but only to vibrate around a fixed position. These intermolecular interactions result in some important properties of solids. A given mass of a solid at a given temperature and pressure has a defined volume and keeps its shape, if no external forces (such as in milling, agglomeration or tableting operations) are acting on the solid material. Similarly, a given mass of a liquid has a defined volume at a given temperature and pressure (like solids, liquids are practically incompressible), but in contrast to a solid, its shape can vary and will take on the shape of the container it is placed in. The reasons for this are again the interactions between the molecules in the liquid state. These are still considerably strong, but the kinetic energy of the molecules is now sufficiently high to allow the molecules to change their positions (not only to vibrate around a fixed position) and to move relative to each other by diffusion or convection. This is also the reason why the density of a liquid is usually lower than that of a solid (with the most important liquid in the pharmaceutical field being the exception: water has its highest density at normal pressure at 4°C, i.e., in the liquid state). A gas, on the other hand, does not have a defined shape or volume (in fact, gas molecules will fully occupy each volume available to them) and thus has a very low density, depending on the volume it occupies. This is due to the fact that intermolecular interactions in real gases are very weak and, in fact, are considered zero for an ideal gas. Since interactions between the molecules (or between atoms or ions) are very strong or strong in solids and liquids, these two states of matter are also summarized as condensed matter. In contrast, since the molecules in the liquid and gas state are able to change their positions, these two states of matter are summarized as fluid matter.

CHANGES BETWEEN THE STATES OF MATTER

Upon the heating (at constant pressure), for example, of a solid drug material (usually in a crystalline form, see below), the solid transforms into a liquid (melt) and further into a gas (if it is not chemically degraded in the heating process). Heat (Q , a form of energy, Unit: Joules, J) is transferred into the material during the heating process. This leads to an increase in the enthalpy ΔH [J] of the material.

$$Q = \Delta H \quad (15)$$

It follows that a gas has a higher enthalpy than a liquid and a liquid has a higher enthalpy than a solid for a given material. At the same time, the number of ways in which the molecules can be arranged with respect to each other also increases. This is known as the entropy S [J K^{-1}] of the system and is often described as the “disorder” of the system. It follows that upon a state change due to heating, both the enthalpy and the entropy of the material increase. Since every system has a tendency to exist in the lowest enthalpy and the highest entropy form, it depends on the temperature T of a given system (here considered solely as the material in question) in which form it is present. This can be summarized in form of the equation:

$$\Delta G = \Delta H - T\Delta S \quad (16)$$

Where: ΔG [J] is the free energy of the system.

If a liquid is heated, the temperature of the liquid will increase. How much the temperature increases as a function of the added heat energy depends on the heat capacity of the liquid. The heat capacity, C , can be defined as the amount of heat energy, Q , required to change the temperature, T , of a given material by a given temperature interval ΔT (measured in Kelvin, K):

$$C = Q / \Delta T \quad (17)$$

In this form C is an extensive variable, i.e., its value will depend on the mass of the material. To convert the heat capacity into an intensive variable, i.e., a variable whose value is independent of the mass of the material (measured in kg) or its amount (measured in mol), the heat capacity is usually expressed as either specific heat capacity, C_{sp} [$\text{J K}^{-1} \text{kg}^{-1}$], or molar heat capacity, C_{mol} [$\text{J K}^{-1} \text{mol}^{-1}$].

In any case, the result of heating the liquid for a given period of time is that the temperature of the liquid increases. However, when the liquid starts to boil, i.e., the state of the material changes from the liquid to the gaseous (vapor) state, we observe that the temperature of the material stays constant during this state transition. This temperature is the boiling point (T_b in Figure 11-32). Similarly, when a solid melts, the temperature of the material stays constant during this state transition (melting point, T_m in Figure 11-32). This is because when a substance boils, the liquid and vapor state are in equilibrium, and when a substance melts, the solid and liquid state are in equilibrium. Any heat that is applied to the system will be used to bring about the state transition until the liquid (or solid) has completely gone into the vapor state (or liquid state). Only then does the temperature rise further. The energy used for boiling the liquid is known as latent heat of vaporization (or boiling) and in the case of condensation, the energy released is known as latent heat of condensation. The amount of energy used in the vaporization process is identical to the heat of condensation released in the condensation process since these processes are reversible.

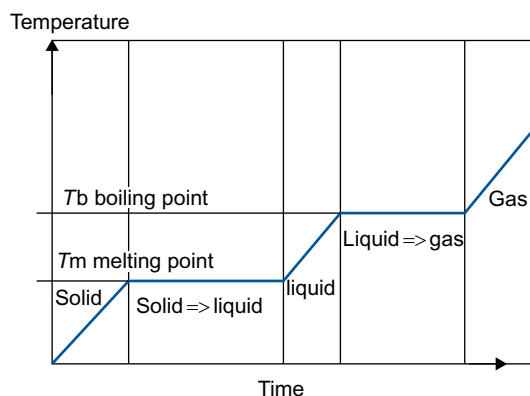


Figure 11-32. Heating curve of a material undergoing a state transition from solid to liquid and from liquid to gas. Note that the temperature stays constant during the state transitions.

During a phase transition the heat capacity of the material is infinitely high (all heat energy is used to bring about the state change), and such transitions are termed first-order phase transitions. The same holds true for the state transition from a solid to a liquid (latent heat of melting) or from a liquid to a solid (latent heat of freezing). In melting and boiling processes, the change is endothermic (the system absorbs energy), whereas in freezing and condensation processes, the change is exothermic (the system releases energy). At the melting and boiling point, the change in Gibbs free energy (ΔG) is zero. However, as discussed above, the enthalpy and the entropy of the material are increasing (i.e., ΔH and ΔS have positive values). Taking into account the relationship between the free energy and the enthalpy and entropy of the systems (see equation 16, $\Delta G = \Delta H - T\Delta S$, note the minus sign between ΔH and ΔS), we can conclude that a phase transition occurs when the free energy of the gas (or liquid) is lower than that of the liquid (or solid) at a given temperature for the material in question. At a given pressure, the temperature that denotes the boiling or melting point is the temperature at which ΔG for this process is zero, i.e., at exactly the boiling or melting point (for a given pressure), the two states can coexist.

So far we have considered the temperature as being decisive for a given material in what state of matter it is present. However, to fully describe a given state of matter, we require a second intensive variable, pressure, p . If we plot temperature on the x -axis and pressure on the y -axis of a coordinate system, and determine the state of a given material at any temperature–pressure combination, we obtain the phase diagram of the material in question. Let us again consider water.

Figure 11-33 shows the (partial) phase diagram of water as a function of pressure and temperature. We can differentiate three areas, those of solid water (ice), liquid water (water), and water vapor. To fully describe the state of water in each of these areas, we need to specify the temperature and pressure at which the material is present. For example, water exists in the liquid state at both temperature–pressure combinations indicated in the diagram by the symbol $\blacktriangle$ and, in fact, at an infinite possibility of other temperature–pressure combinations in the region of liquid water. The same holds true for the areas of solid water and water vapor. The different areas of the phase diagram are separated by boundary lines, known as the phase boundaries. At combinations of pressure and temperature which lie on these phase boundaries, two states of water exist in equilibrium. The curves a, b, and c in Figure 11-33 are known as the sublimation (or re-sublimation) curve (solid and gas in equilibrium), the melting (or freezing) curve (solid and liquid in equilibrium), and the boiling (or condensation) curve (liquid and gas in equilibrium). Note the orientation of the melting curve b in Figure 11-33: its slope is negative; that is, at constant temperature ice can melt if the pressure is increased (this is the principle of ice-skating). Most other compounds behave oppositely; they can solidify under pressure, so their melting curves will have a positive slope. This is termed the anomaly of water and also explains the observation that ice has a lower density than liquid water.

To determine, for instance, the position of the boiling curve in the phase diagram of water, we have to determine the vapor pressure of water at various temperatures. If we place a liquid in a vacuum at a temperature T , some of the molecules of the liquid will leave the liquid phase and will go into the free space above the liquid until an equilibrium is reached, in which the same number of molecules from the free space above the liquid will go into the liquid as are leaving the liquid (this is a dynamic equilibrium). As long as there is still liquid water left (note: it is not important how much water is left), the resulting pressure is known as the vapor pressure of the liquid, and will constitute a point on the boiling curve at the temperature T . Figure 11-34 shows the temperature dependence of the vapor pressure of water. In a closed system the vapor pressure increases with increasing temperature. This is true not only for water but for

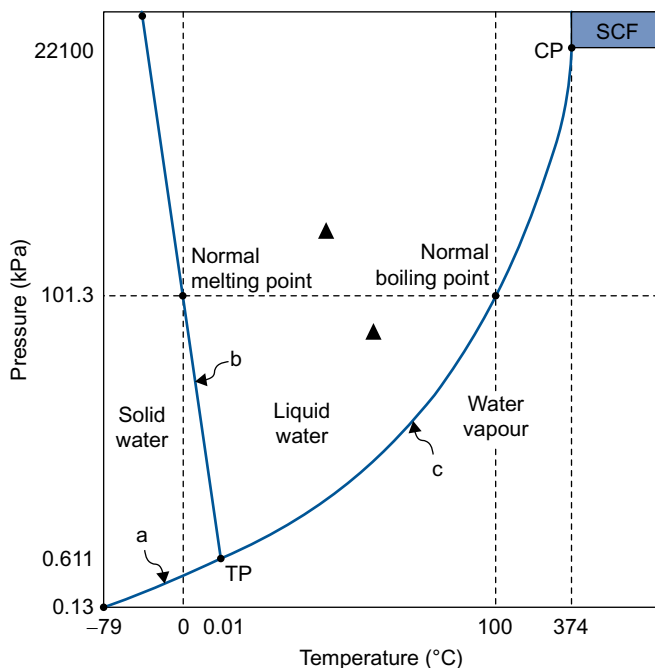


Figure 11-33. Partial phase diagram of water. The curves a, b, and c are the sublimation (or re-sublimation) curve (solid and gas in equilibrium), the melting (or freezing) curve (solid and liquid in equilibrium), and the boiling (or condensation) curve (liquid and gas in equilibrium), respectively. TP: triple point, CP: Critical point, SCF: region in which water exists as supercritical fluid. ▲ shows two different pressure and temperature combinations at which water exists in the fluid state. Note that the temperature and pressure axes are not linear.

all substances, since an increase in temperature increases the likelihood for molecules to leave the condensed phase, as the increased kinetic energy of the molecules counteracts the attractive forces holding the molecules together in the condensed phase. Different liquids have different vapor pressures at given temperatures. A liquid will boil at a temperature at which its vapor pressure equals the pressure exerted onto the liquid by the surrounding pressure of the environment. Taking again the example of water, at 100°C the vapor pressure of water is 101.3 kPa, so under normal (sea level) condition of atmospheric pressure, water will boil. At 20°C, however, the vapor pressure of water is only 2.3 kPa, so it will not boil under these conditions. An interesting consequence of this is that water, for example, on a high mountain (where the atmospheric pressure is lower than at sea level) will boil at a lower temperature than 100°C.

At this point it let us make a quick stop and consider the question why drying (for example, of wet granules, to take a pharmaceutical example) occurs even if the temperature is far below 100°C at normal atmospheric conditions (i.e., below the boiling point of water). To understand this, we have to differentiate between the processes of boiling and evaporation. We have seen that, in a closed system, molecules from the liquid will leave the liquid phase until the equilibrium vapor pressure for the given temperature is reached. However, in an open system (for example, wet granules in a laboratory) some molecules from the surface of the liquid will have a sufficiently high kinetic energy to leave the liquid and are then transported away. This process takes place at any temperature also below the boiling point of the liquid, and its rate depends on factors such as

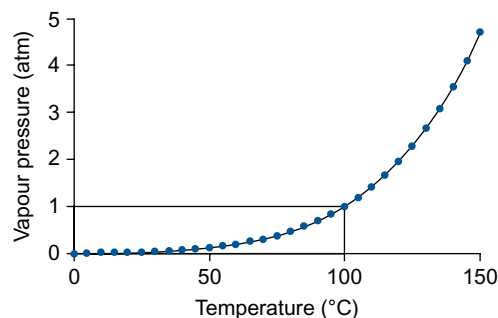


Figure 11-34. Temperature dependence of the vapor pressure of water.

- the surface area between the liquid and the air (evaporation is a surface phenomenon)
- the temperature of the liquid (the number of molecules with a sufficiently high kinetic energy to leave the liquid phase increases with increasing temperature)
- the air above the liquid (if there is an air stream, transport of the molecules away from the liquid is facilitated)
- the humidity of the air above the liquid (evaporation of water molecules from a liquid will be faster if the air is dryer), as well as other factors.

It should be noted that evaporation takes place from the surface of a liquid, whereas boiling takes place also in the liquid (that is the reason why water vapor bubbles are formed in boiling water).

We have seen how we can determine the boiling curve for a liquid to construct the phase diagram of a substance. Similarly, we can determine the sublimation curve, although experimentally this is more difficult due to the much lower vapor pressure of solids. The melting curve can be determined by thermo-analytical methods. This is usually done by determining the cooling curve of a material at various pressures as a function of time.

As we have already discussed for the boiling of a liquid, during the phase transition from the liquid to the solid, the temperature does not decrease. If we are heating a liquid in a closed container, the vapor pressure will also increase. However, since the space above the liquid is limited, further heating will increase the number of molecules in the space above the liquid, and thus the density of the gas phase will increase. At a certain, material-dependent temperature, the density of the gas phase will be equal to the density of the liquid, so that there will no longer be a phase boundary between the liquid and gas phase, and both are now indistinguishable (this is called a supercritical fluid). The temperature (critical temperature) and pressure (critical pressure) combination at which a supercritical fluid is formed is known as the critical point (CP in Figure 11-33). At temperatures above the critical temperature a liquid can no longer exist or be formed by increasing the pressure, and thus the critical point marks the end of the boiling curve. The region in the phase diagram at temperatures and pressures higher than that of the critical temperature is known as the supercritical fluid region (SCF). For water, the critical point is at approximately 374°C and 22.1 MPa, and for CO₂ the critical point is at approx. 31.1°C and 7.38 kPa. Supercritical fluids have gained pharmaceutical importance since this state of matter combines properties of gases, such as low viscosity and high diffusivity, with properties of liquids, such as a high capacity to dissolve other substances.

From Figure 11-33 we can see another critical point in the phase diagram, known as the triple point (TP in Figure 11-33). At this combination of temperature and pressure, all three states of water coexist in a dynamic equilibrium. For water the triple point is at 0.01°C and 0.61 kPa. The triple point also marks the lowest temperature at which a liquid can exist, so the temperature range for liquids to exist is between the triple

point and the critical point of the material. Similarly, sublimation and resublimation can only occur at temperatures below the triple point (see sublimation curve in Figure 11-33). Since sublimation (in the same way as boiling) occurs when the vapor pressure of the solid material equals that of the gas phase, sublimation should only occur under atmospheric conditions if the pressure at the triple point of the system is above 101.3 kPa (1 atm). For example, solid CO_2 (dry ice) sublimates at atmospheric pressure at -78.5°C since its triple point is at -56.4°C and 517 kPa. For the same reason, in the pharmaceutically relevant process of freeze drying (which at least in the primary drying phase is based on the sublimation of water), an aqueous solution or dispersion is initially frozen (below the temperature of the triple point of water) and is then subjected to a reduced pressure (below the pressure of the triple point of water). In the earlier mentioned real-life example, however, we also observe that wet clothes on a washing line (i.e., at atmospheric pressure) at or below freezing temperature (so the water in the clothes is frozen to ice) eventually will be dry. We thus observe sublimation also for substances for which the triple point has a pressure below 1 atm. The reason for this is that in an open system like that on a washing line, the partial pressure (see below) of the solid does not attain the vapor pressure of the material at the triple point. The situation is similar to the one discussed above for liquids in an open and closed system (evaporation versus boiling).

PHASES

In the examples discussed above, we have talked about states of matter, and we have considered the solid, liquid, and gaseous state. In the phase diagrams we have seen that these states of matter constitute different phases of a material. It is, however, possible and, in fact, very likely that a solid material can in its own right exist in different phases. We can define a phase as a volume element of a system, separated from other volume elements of the system by a phase boundary. The physical properties within the phase do not show abrupt changes, which means that the phase is physically homogeneous. If two phases coexist, such as at the phase boundaries in the phase diagram, the physical properties between the phases, however, are different. For example, at normal pressure and 0°C , the density of ice is approximately 920 kg/m^3 ; and that of liquid water is approximately 999.8 kg/m^3 . Note that water makes an exception from most other materials, in that its higher density is as a liquid rather than as a solid at the equilibrium of liquid water and ice.

Let us again consider the phase diagram of water, but now have a look at a larger temperature and pressure range (Figure 11-35). We can see that different forms of ice can exist (termed ice I to ice VII). These all constitute a solid state of water but are all different in the molecular arrangement of the water molecules. They therefore have different physical properties, and these change abruptly at the phase boundary. These different solid forms of water thus constitute different water phases. While these different ice phases may not be relevant in pharmaceutical applications, they nevertheless show that molecules in the solid state can be arranged in different ways and thus give rise to the existence of different solid phases. At any given temperature and pressure, only one such solid form can be thermodynamically stable, but other forms may exist as metastable forms for considerable times at conditions where another form is the stable form. This will become important when we are considering crystalline solids in more detail later in this chapter.

The phase diagram of water in Figure 11-35 also shows that for a given chemical compound, more than one triple point can exist. In fact, in Figure 11-35 we can see six other temperature–pressure combinations at which three phases coexist in equilibrium. However, unlike the triple point between ice I, water, and water vapor, these are conditions in which either two solid phases and liquid water or three solid phases coexist.

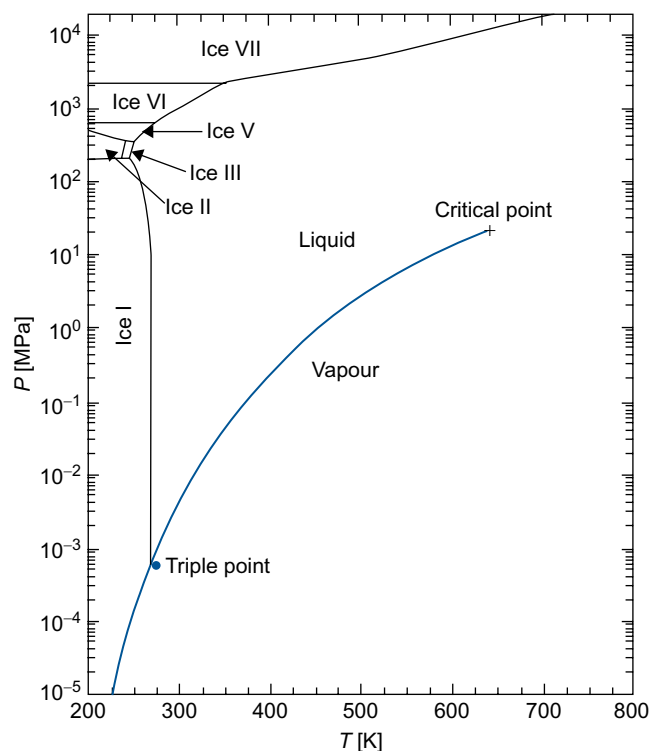


Figure 11-35. Phase diagram of water, showing several solid forms of ice (Ice I – VII). Note that the temperature is given in Kelvin [K], and the pressure axis is shown in logarithmic form.

Earlier, we have stated that to fully describe a system in one of the areas of the phase diagram, we have to specify both its pressure and temperature. Pressure and temperature are known as the degrees of freedom of the system. The number of degrees of freedom can be defined as the least number of intensive variables (i.e., of variables that do not depend on the amount of material or its mass) of the system that are independent of other variables of the system and that can be changed without changing the number of phases that are coexisting. For any sample within an area of the phase diagram (e.g., in the area of liquid water or in the area of ice III, etc.), we can freely change pressure and temperature (until we reach a phase boundary) without changing the number of phases present (one phase).

If two phases coexist (such as for systems on the phase boundaries), the situation is different. If we change the pressure of the system, we cannot freely change the temperature but have to change the temperature to a specific value, if we don't want to change the number of phases. In the case of two phases coexisting, the system thus only has one degree of freedom. This is important, for example, in moist heat sterilization, such as sterilization in an autoclave. To have both liquid water and water vapor present in the sterilization process (this is most effective in killing microorganisms) at 106 kPa, the temperature has to be 121°C (if no air is present in the autoclave). If we raise the temperature to 134°C , we have to increase the pressure to 312 kPa to maintain the coexistence of liquid water and water vapor.

Finally, if we are at a triple point in the phase diagram, i.e., three phases are coexisting, we cannot change either pressure or temperature without getting a change in the number of phases, which means that here the system has zero degrees of freedom.

From these considerations, Gibbs' phase rule can be formulated as follows:

$$F = 3 - P \quad (18)$$

Where: F is the number of degrees of freedom; and P is the number of phases coexisting (one, two, or three).

So far, we have only considered phase diagrams of a single chemical component (e.g., H_2O). It is, however, also possible to have a single phase system which consists of more than one component. Let us now consider, for example, a salt solution. Here we have a single phase system (the salt molecules are dissolved as sodium and chloride ions in water, and the system is homogeneous). However, in this case, we can vary not only pressure and temperature without getting a phase change but also the salt concentration until we get a saturated solution of the salt in water. In this case, we have three independent variables that can be changed without changing the number of phases coexisting. Of course, as with pressure and temperature changes, this can only be done within limits, i.e., until we reach a boundary line in the phase diagram. If we, for example, increase the salt concentration above the saturation solubility of NaCl at the given pressure and temperature, the excess salt will form a second, solid phase, coexisting with the salt solution phase.

In the example of a salt solution, we have to consider two chemical entities, H_2O and the salt (e.g., NaCl). These chemical entities must be independent of each other. So even though NaCl is dissociating into Na^+ and Cl^- ions, in the current discussion these do not constitute two chemical entities because the Na^+ concentration is not independent of the Cl^- concentration. The same holds true, for example, for H_3O^+ and OH^- ions from the auto-protolysis of water. The independent chemical entities of a system are called components and are defined as the least number of chemically independent entities required to describe the composition of the system. A little tongue-in-cheek, we might say that the phase is physically defined, whereas the component is chemically defined (however, note that chemically a Na^+ ion is very different to a Cl^- ion). For multicomponent systems we can formulate Gibbs' phase law as follows:

$$F = C - P + 2 \quad (19)$$

Where: C represents the number of components.

For the examples of a salt solution, $F = 2 - 1 + 2 = 3$. We can see pressure, temperature, and salt concentration here as the degrees of freedom of the system. In most cases, phase diagrams of two component systems are given at a constant pressure, usually normal atmospheric pressure (i.e., one degree of freedom is used up). In this case we can express Gibbs' phase law for a two component system as $F^* = 3 - P$, with F^* being the remaining degrees of freedom of the system. An example of such a phase diagram will be discussed later (see the section on eutectics).

CRYSTALLINE SOLIDS

The solid state of matter is perhaps the most important state in pharmaceuticals since the starting material for most formulations and the majority of orally administered dosage forms are in a solid form (most often as tablets or as powders, pellets, or granules in hard gelatine capsules). We will thus consider this state of matter in some detail, whereas the properties of other states of matter will be discussed more briefly since important properties of these are explored elsewhere in this book.

Crystals are defined by a high degree of order, with a three-dimensional periodicity in position for atoms or ions and, additionally, with configurational periodicity for molecules that form the crystal. Most solid forms of drugs are used in the crystalline state, and most pharmaceutically relevant crystals are formed by organic molecules. Not all solids, however, are crystalline (this will be discussed in more detail in the section on amorphous solids).

The reason for the high degree of order of crystals lies in the strong interactions between the atoms, ions, or molecules that form the crystal. As we have seen earlier, in the case of diamond, carbon atoms form a periodic three dimensional tetrahedral network of covalently bonded sp^3 hybridized carbon

atoms. Diamond is thus an example of a crystal formed by atoms through covalent bonds. Most inorganic materials form ionic crystals. A pharmaceutically relevant example is NaCl. Here the crystal is formed by sodium and chloride ions, held together by ionic interactions. In the case of NaCl, each sodium ion is surrounded by six chloride ions, and each chloride ion by six sodium ions. The number of ions (or atoms or molecules) that each ion has as its nearest neighbors is called the coordination number. The coordination number for NaCl is, therefore, six. For ionic crystals, the coordination number depends on the relative size of the cation and the anion. If the size difference is very large, the most often found coordination number is four, and if the size differences are small, a coordination number of eight is found. Higher coordination numbers are found in metal crystals (eight or twelve), and the coordination number of diamond is four.

If atoms or ions are held together by covalent or ionic interactions, the resulting crystals have very high melting points and are often hard and brittle. In metals, the crystalline structure is formed by cations, where the electrons are delocalized. This is the reason for both the high coordination numbers of metal crystals as well as the ductile behavior of many metals.

Most drugs are organic molecules. The crystals formed by these substances are held together by dipole-dipole interactions, hydrogen bonds, or London forces. Since these are at least an order of magnitude weaker than ionic or covalent interactions, the melting points of molecular crystals are lower than those of the aforementioned crystalline structures. For

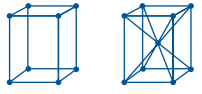
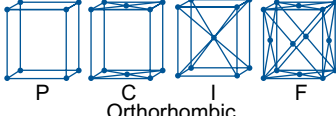
 Cubic P Cubic I Cubic F	Three axes at right angles; all equal: $a = b = c$; $\alpha = \beta = \gamma = 90^\circ$
 Tetragonal P Tetragonal I	Three axes at right angles; two equal: $a = b \neq c$; $\alpha = \beta = \gamma = 90^\circ$
 P C I F Orthorhombic	Three axes at right angles; all unequal: $a \neq b \neq c$; $\alpha = \beta = \gamma = 90^\circ$
 Monoclinic P Monoclinic C	Three axes, one pair not at right angles, all unequal: $a \neq b \neq c$; $\alpha = \gamma = 90^\circ \neq \beta$
 Triclinic P	Three axes not at right angles; all unequal: $a \neq b \neq c$; $\alpha \neq \beta \neq \gamma \neq 90^\circ$
 Rhombohedral	Three axes equally inclined, not at right angles; all equal: $a = b = c$; $\alpha = \beta = \gamma \neq 90^\circ$
 Hexagonal	Three axes coplanar at 120°, fourth axis at right angles to these: $a_1 = a_2 = a_3 \neq c$; $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$

Figure 11-36. The 14 Bravais lattices.

example, the melting points of paracetamol (molecular crystal), NaCl (ionic crystal), copper (metal crystal), and diamond (covalent crystal) are 169°C, 801°C, 1084°C, and 3550°C, respectively.

We have already seen that different types of crystals can have different coordination numbers. In fact, since crystals are periodically organized, we can define a specific unit cell characteristic for each crystal. A unit cell is defined as the smallest unit of a crystal, which, if repeated, could generate the whole crystal. For this we imagine the atoms, ions, or in the case of drug crystals, molecules, as points in a lattice, disregarding their actual shape. Within the crystal structure we now look for the smallest unit which, when endlessly repeated, forms the crystal structure. Since there may be different ways to do this, the unit cell of a given crystal is the one with the shortest dimensions, and that includes angles that deviate the least from a right angle. The unit cell is a geometric construction and does not take into account the size and shape of the crystal forming atoms, ions, or molecules, nor the type and strength of interactions between them. For this reason, it is possible to summarize the types of unit cells into only seven crystal systems, which can be understood as unique unit cell shapes and which can be differentiated with respect to the relative lengths on their axes a , b , and c and the angles α , β , and γ between the axes. These seven crystal systems are called cubic, tetragonal, orthorhombic, monoclinic, triclinic, rhombohedral (or trigonal), and hexagonal (Figure 11-36).

While there are only seven crystal systems, there are 14 Bravais lattices. This is because additional points can be present in a unit cell, other than the points at the corners of the unit cell, without changing the essential symmetry elements of the crystal system. If there is an additional point in the centre of the unit cell, these lattices are called body (or volume) centered (for example, body centered (cubic I in Figure 11-36)). If there are additional points in the centre of each side (face) of the unit cell, they are called face-centered (for example, face centered (cubic F in Figure 11-36), and if an additional point is present at the centre of two opposite faces of the unit cell, they are called base-centered (for example, base-centered (monoclinic C in Figure 11-36). Bravais lattices without extra points are called simple or primitive for example cubic P in Figure 11-36. The 14 Bravais lattices are shown in Figure 11-36.

Every crystalline material belongs to one of these Bravais lattice types but differs in their lattice parameters, and for each unit cell we can define three axes, with a length characteristic for the individual crystal, and three angles, also characteristic for this crystal. For example, NaCl crystals belong to the crystal system: cubic, Bravais lattice type: face-centered. The only lattice parameter we need is the length of one axis, $a = 5.6402 \text{ \AA}$, since for a cubic crystal the axes $a = b = c$, and the angles $\alpha = \beta = \gamma = 90^\circ$. Paracetamol (polymorphic form I) belongs to the crystal system: monoclinic, Bravais lattice type: primitive, $a = 11.74 \text{ \AA}$, $b = 9.396 \text{ \AA}$, $c = 7.117 \text{ \AA}$ and $\beta = 97.47^\circ$ ($\alpha = \gamma = 90^\circ$, for a monoclinic unit cell).

In the above discussion, we have considered the crystalline structure as infinitely periodic. In reality most crystals, however, will show defect structures, i.e., structures in which there are packing mistakes of some kind, so the crystalline structure of a real crystal will not be perfect. Depending on the number of these defects in a crystalline material, properties of the crystalline substance may vary, for example, its behavior in a milling process. Point defects occur at a single lattice point (Figure 11-37). For example, a single lattice point may be vacant (vacancy) or occupied by a different atom, ion, or molecule from the ones that form the crystal (substitutional point defect; this would constitute a chemical impurity). It is also possible that there may be additional lattice points (interstitial point defects), either from the same or a different atom, ion, or molecule.

The dislocation is an example of a line defect, and the grain boundary is an example of a planar defect (Figure 11-38).

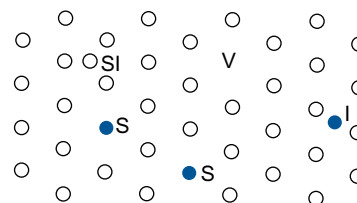


Figure 11-37. Point defects in a crystal lattice. V: vacancy, S: substitutional point defect (different atom, ion, or molecule from the ones that form the crystal), I: interstitial point defect (different atom, ion, or molecule from the ones that form the crystal), SI: self-interstitial point defect (same atom, ion, or molecule as the ones that form the crystal).

A grain boundary marks an interface between different crystalline regions, called crystallites. Such a material is termed polycrystalline and is the reality for most crystalline materials, including drug substances.

Polymorphism

An important property of many chemical entities, including most drug molecules, is that they are able to crystallize in several different crystalline forms, termed polymorphic forms. These different crystalline forms consist of the same molecules, but these are arranged in different ways. If the crystal former is an element (as we have seen in the case of carbon), different polymorphic forms are termed allotropes. For molecular crystals, the crystal is held together by secondary forces, as outlined above. Since these interactions are fairly weak (compared to ionic, metal, or covalent crystals), and because the molecules may adopt different molecular configurations in the crystalline structures, it is not surprising that most pharmaceutical drugs show polymorphism, and are often able to crystallize in more than two, and in some case even more than ten, different crystalline forms. These different polymorphs are also termed polymorphic modifications of the substance in question. It is important to note that different modifications of a drug may not necessarily form crystals with a different type of Bravais lattice. Although this is, of course, possible, it is also possible that two polymorphic forms of a drug both have, for example, a monoclinic unit cell. In this case, however, the lattice parameters of the two polymorphic forms will be different. Also, for some polymorphic forms, for example, of drug substances, the molecules present in the lattices of the different modifications will adopt different molecular conformations (termed conformational polymorphism), while other polymorphic pairs will contain the drug molecules with the same conformation (termed packing polymorphism).

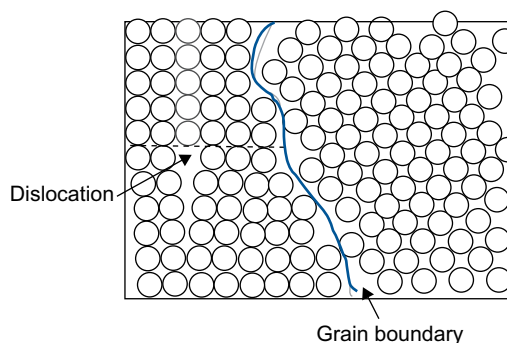


Figure 11-38. Dislocation (a line defect) and grain boundary (a planar defect).

Polymorphism is important in the pharmaceutical field, since the physical properties of different polymorphic forms are often considerably different, which can have consequences for important quality-relevant attributes of a drug substance or dosage form. Differences may be encountered in, for example, dissolution rate, solubility, flow properties, behavior under mechanical stress (for example during milling, tableting, agglomeration), hygroscopicity, etc. These differences may be sufficiently large so that different polymorphic forms may have different bioavailabilities. It is thus possible, or even common, that for a given drug regulatory approval is only given to a specific polymorphic form of the drug.

While often different polymorphic forms of a drug form macroscopically different crystals, such as needle-like versus plate-like crystals (the macroscopic appearance of a crystal is also known as the habit of the crystal), it is important to note that different polymorphic forms may not necessarily look different macroscopically, and thus it may be difficult to differentiate these by visual inspection or in the light or electron microscope alone. It is also possible that if one and the same polymorphic form of a crystal is crystallized, for example, from different solvents or from either a solution or the melt, it may adopt different crystal habits. So while the habit is important, most notably for further processing of the drug powder, it is also important to have techniques available to unambiguously identify the crystalline structure of a drug and its polymorphic form. This will be discussed in the section on methods to characterize solids, where some techniques to identify and characterize polymorphs (or indeed solid state forms in general) will be mentioned.

If a drug can crystallize in different polymorphic forms, it should be noted that at any given condition of pressure and temperature, only one polymorphic form can be the thermodynamically stable form. All other polymorphic forms that may form under these conditions will be only metastable. Metastable here means that these polymorphs can form but eventually will convert to the stable form over time if the environmental conditions are not changed. The metastable forms, on the other hand, are not unstable (in contrast to amorphous forms which will be discussed in the next section), since (sometimes considerable) activation energy is required to bring about a phase transition from the metastable to the stable polymorphic form. For some metastable polymorphs, therefore, the polymorphic transition may be very slow and may take much longer than the shelf life of the drug product containing a metastable polymorphic form. In general, if not precluded by a too low solubility of the stable form, most drugs will be manufactured in solid dosage forms in the stable polymorphic form. However the drug ranitidine-HCl, which can exist in two polymorphic forms, is manufactured using both polymorphs, i.e., the metastable polymorph is “practically” stable enough to allow safe manufacturing.¹¹

It is often found that not the most stable, but a less stable (often the least stable) polymorphic form of a substance crystallizes first, for example, upon cooling of a melt. This is known as Ostwald's rule. While not universally applicable (not being a law, but rather a rule), it nevertheless highlights the importance in drug development to carefully check which polymorphic form of a drug one is indeed dealing with.

Let us consider a given set of two polymorphs of a substance. Because the solid state forms are different, their phase diagrams will also be different, and the overall phase diagram of this polymorphic substance will be the superimposed phase diagram of the two forms. This is illustrated in Figure 11-39. We can differentiate two possible scenarios, shown in Figures 11.39a and b. For the monotropic system (Figure 11-39a), the sublimation curve of form I is below the sublimation curve of form II. Since the stable polymorph is the one with the lower vapor pressure, in the case of a monotropic pair of polymorphs, one form (here form I) is always the stable form until either of the forms melts. Note that after melting there is no difference between the melts of the substance, irrespective of the polymorphic solid form.

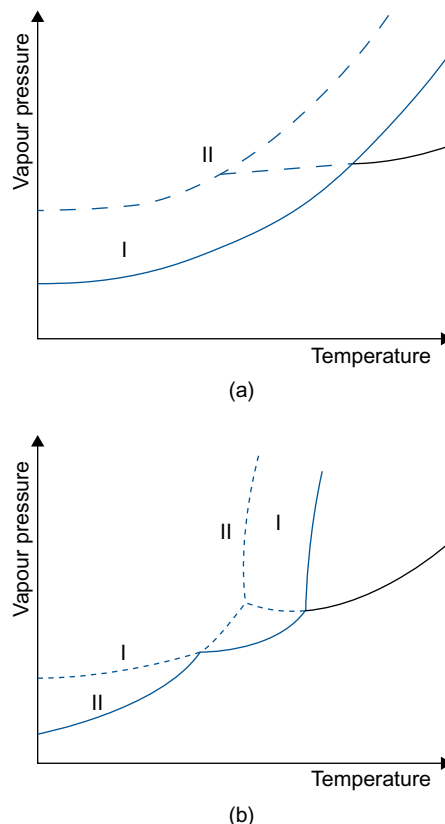


Figure 11-39. Phase diagrams of (a) a monotropic and (b) an enantiotropic polymorphic system. The solid lines represent stable states.

Should initially form II have been generated by cooling from a melt (according to Ostwald's rule this is not unlikely to happen in a practical scenario), then if conversion occurs to the stable form I, this conversion will be irreversible. The situation is different for enantiotropic systems (Figure 11-39b). Here form II has a lower vapor pressure than form I at low temperature (so it is the stable polymorph at these temperature conditions), but at higher temperatures there is a crossover in the sublimation curves of the two polymorphs before either of these forms melts. After the crossover, form I now has the lower vapor pressure and is the stable form. This means if we have form II present at low temperature, this is the stable form, but upon heating at a certain temperature, polymorphic conversion will take place and form I is the stable form. If we now cool down form I again, conversion back to form II will occur. In other words, the process is reversible.

AMORPHOUS SOLIDS

In the previous section we have seen that a crystalline solid is characterized by its long-range positional order. Amorphous solids lack the long-range order seen in crystals although short-range order over several molecular dimensions may exist. The molecular arrangement is thought to represent that of a frozen liquid with the rheological properties of a solid.

Heating of a crystalline material leads to a gradual increase in the thermodynamic properties such as enthalpy and entropy until the melting point, T_m , is reached (Figure 11.40). As we have seen, a significant increase in thermodynamic properties marks the first order transition (melting) from the solid to the liquid state. Above the melting temperature, the material exists as a molten liquid. This is a reversible process, so upon cooling, crystallization of the melt occurs whereby the molecules rearrange themselves in their crystal lattice. However, if the liquid is cooled sufficiently fast below its melting temperature, crystallization may be prevented so that a super-cooled liquid is formed, and the slope of the

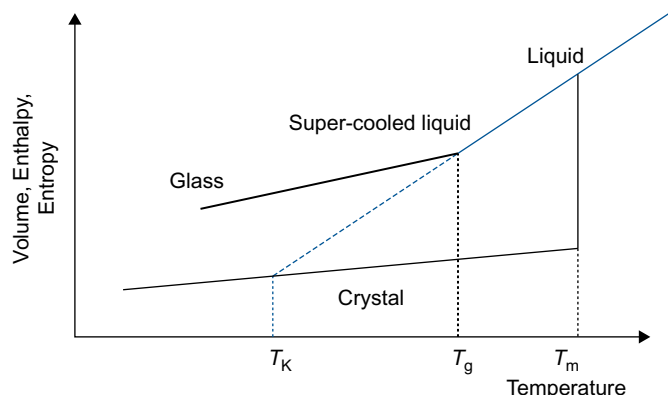


Figure 11-40. Thermodynamic relationship of crystalline and amorphous solids as a function of temperature. Shown are the melting temperature (T_m), the glass transition temperature (T_g), and the Kauzmann temperature (T_K).

equilibrium liquid line (Figure 11-40) may be followed, resulting in a gradual decrease of the thermodynamic properties below T_m .

At temperatures above the glass transition temperature (T_g , Figure 11-40) the material is said to be in the super-cooled liquid state (sometimes also called the rubbery state), the viscosity of which is typically between 10^{-3} and 10^{12} Pa s. In the super-cooled liquid state, the molecules are able to follow any further decrease of temperature to attain equilibrium conditions. However, cooling of the super-cooled liquid increases the viscosity of the system, and upon further cooling the molecules are not able to reach equilibrium any longer. At this temperature, the T_g , the molecules are “kinetically frozen in” and the super-cooled liquid solidifies into a glass. (Note that in a glass the molecules still show some mobility.) The T_g represents the temperature at which the system falls out of equilibrium. Amorphous solids are therefore non-equilibrium solids and the glass transition is a second order phase transition. Unlike melting or boiling, the heat capacity is not infinitely high at the phase transition, but jumps from one value to another. The non-equilibrium state of a glassy amorphous system has special implications: the fundamentals of thermodynamics only apply to systems in equilibrium, and therefore, the behavior of the glassy state cannot be predicted from regular thermodynamics.

Due to the high viscosity of the resulting glass (viscosity is usually greater than 10^{12} Pas), it appears as a solid. The T_g represents a characteristic thermal event for an amorphous system, but it should be noted that the T_g is a kinetic property of the material, and its exact position depends, for example, on the cooling rate of the super-cooled liquid, i.e., on its thermal history.

In principal, the amorphous state can be produced in two ways:

1. from a liquid state, either by rapidly cooling the molten liquid or by fast evaporation of a solvent from a solution, hence “freezing in” the molecular arrangement of the liquid (vitrification), or
2. by gradually inducing and increasing defects in the crystal lattice until the amorphous state is generated (amorphization).

The vitrification process (1) is also termed the thermodynamic pathway to form amorphous materials, and the amorphization process (2) has been named the kinetic pathway. A schematic representation of the conversion from crystalline to amorphous is depicted in Figure 11-41.

If the amorphous state is produced via the thermodynamic pathway, the crystalline material has to be transferred to a liquid

state, either by melting or by dissolution in an appropriate solvent. If the kinetic pathway is followed, the crystalline lattice is continuously disrupted by mechanical processing like milling, inducing crystal defects. As long as the crystal defects remain small in number, the material still exhibits crystalline properties. However, if a critical number of defects is introduced, the crystal cannot retain its structure and converts to the amorphous state. The main difference between these two approaches is that while the thermodynamic pathway abruptly destroys the long-range order of the crystal, the kinetic pathway gradually increases the number of crystal defects, thereby creating non-equilibrium crystalline states. Such materials, still crystalline but with increasing number of defects, are termed defective crystals.

While the thermodynamic pathway is commonly employed in order to deliberately generate the amorphous state, the kinetic pathway is usually associated with the unintentional amorphization of a crystalline material, for example when the actual objective of a milling process is to reduce the particle size of the crystalline material, not to change its solid state form.

It should be noted that not every preparation technique to convert a crystalline drug into its amorphous counterpart is suitable for every drug, e. g. thermolabile drugs cannot be converted to the amorphous form by using heat based methods; drugs that are insoluble in a range of organic solvents cannot easily be spray-dried to amorphize the substance, and some methods (milling) may not completely convert certain drugs into the amorphous state. We have now seen a variety of methods that can be used to generate the amorphous state of a drug. The question arises whether the amorphous state generated is the same in all cases.

Let us take a closer look at the T_g , an important feature in the amorphous state. Above the T_g the system is still in an equilibrium state; below the T_g the system has fallen out of equilibrium. Let us imagine that the T_g did not exist: in Figure 11-40 we see that the decrease of the temperature would eventually lead to the super-cooled liquid line intersecting with the crystal line. This would cause the super-cooled liquid eventually to have a lower entropy than the crystal upon further cooling. This, however, is not considered possible. The theoretical temperature at which the super-cooled liquid would attain the same properties as the crystal is known as the Kauzmann temperature, T_K (see Figure 11-40). The kinetic nature of the glass transition is evident from its dependence on the heating and cooling rates, which means that the T_g does not exist at a defined temperature. Slow cooling rates lead to a prolonged super-cooled liquid state and result in a T_g with a lower temperature, as displayed by $T_{g, II}$ in Figure 11-42. As the super-cooled liquid is cooled at a slower rate, the system is able to adjust to the temperature dependent increase in viscosity, resulting in a lower temperature for the glass transition. Glasses created by slower cooling rates also show lower values of enthalpy and entropy than glasses that are created by faster cooling rates. The explanation for this is that due to the lower onset of the T_g , the super-cooled liquid state is maintained at lower temperatures, therefore decreasing the value of enthalpy and entropy further.

For practical purposes, this means that for a given material an amorphous system which is prepared by one method is not identical to the amorphous system prepared by another method. Even that, if the method is the same but parameters are varied, the resulting amorphous states will not be identical, either. So, unlike “the crystalline state,” which has defined properties, there is no “the amorphous state”. Depending on the preparation method, sample history, geometry and mass, and the experimental heating and cooling rates, the amorphous states can differ considerably in properties, especially in physical stability (recrystallization to the crystalline state).

From Figure 11-40 we can see that the amorphous state has higher thermodynamic properties, like enthalpy and entropy, compared to the crystal. What Figure 11-40 does not show, however, is that the properties of the amorphous state are time-dependent. As time passes, we observe that the amorphous

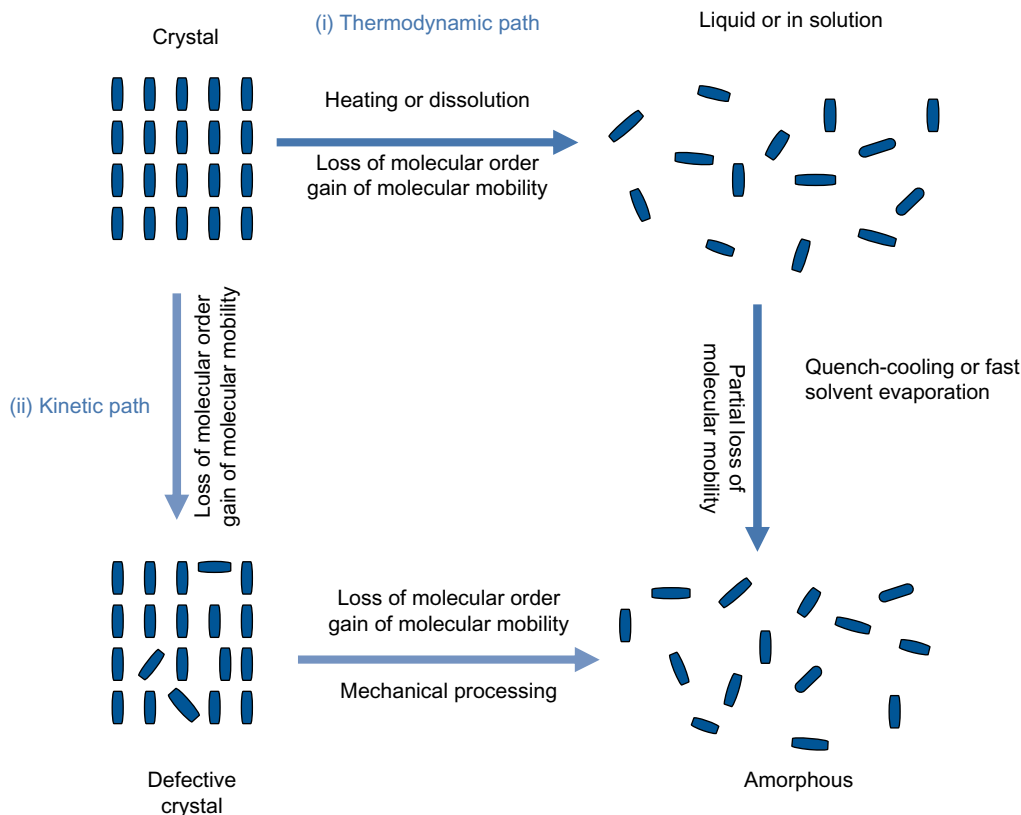


Figure 11-41. Schematic representation of the crystalline to amorphous conversion via the thermodynamic and kinetic pathways.

state decreases in properties such as enthalpy even though the temperature is kept constant. This phenomenon is called relaxation, and it has its origin in the instability of the amorphous state. The molecules may be frozen in, but they still have some molecular mobility left. This is sufficient for the molecules to relax over time and rearrange themselves in a more thermodynamically favorable way. By doing this, they decrease their enthalpy, and this is also considered an important factor in the stability of amorphous systems. This relaxation can be visualized in a DSC thermogram, in which the relaxation appears as an endothermic peak on top of the glass transition event.

In recent years the amorphous state of drugs has gained increasing interest in pharmaceuticals due to its favorable solubility properties, compared to its crystalline counterpart.¹² Many new drugs show insufficient solubility for oral dosage form development. Because of its higher energy, the solubility of amorphous forms is often orders of magnitude higher than for crystals and,

thus, this solid form is attractive in dosage form development. However, issues when dealing with the amorphous state, such as physical and chemical instabilities, remain. Also, to date, the prediction of physical and chemical stability of drugs in the amorphous state still proves challenging, since, for example, the application of stress tests is not straightforward for amorphous systems (they don't behave in an Arrhenius fashion).

SOLIDS WITH MORE THAN ONE COMPONENT

In the previous sections we have discussed crystalline and amorphous solids on the basis of single component systems, i.e., we were considering a single chemical entity (e.g., the drug molecules). In this section we will expand the discussion on solids by introducing a second component.

Solvates and co-crystals

Not only can most drugs crystallize in different polymorphic forms, but they can also incorporate other molecules to form crystalline structures. For example, during a crystallization process from a solvent, solvent molecules may be incorporated into the crystalline structure. This occurs if the solvent molecules can interact with the drug molecules through hydrogen bonding or other weak interactions. In this case a new crystalline structure of the drug and the solvent molecules may be formed. These crystalline structures are termed solvates. In the specific case of water as the additional molecule to be incorporated into the crystalline structure, the solvate is called a hydrate. Hydrates can also form from crystalline or amorphous non-hydrate drug substances by addition of water molecules from the air or during a dissolution process in water. Thus, hydrate formation of a drug (if the drug can indeed form hydrates, which is compound specific) may even occur during the oral administration of the drug. This is of pharmaceutical significance, for in most cases the resulting hydrate will have lower water solubility than the respective non-hydrate, which may even result in a lower bioavailability. For example, partial hydrate formation has been made responsible for erratic bioavailability

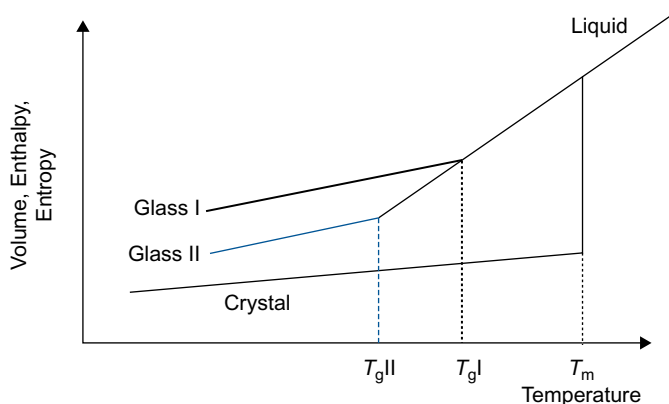


Figure 11-42. Thermodynamic relationship of glasses created by different cooling rates. Glass I has been cooled fast, and glass II has been cooled slowly.

of carbamazepine and theophylline after administration of the drug in a non-hydrate form.^{13,14}

Since the water molecules (or other solvent molecules) are incorporated into the crystalline lattice in a regular structure, the molar ratio of drug to solvent molecules will take on fixed ratios (stoichiometric hydrates). However, this molar ratio does not have to be 1:1 (monohydrates), and for example hemihydrates (the molar ratio of drug to water molecules in a hemihydrate is 2:1) or dihydrates (the molar ratio of drug to water molecules is 1:2) are not uncommon. Some drugs can form different hydrates (for example, mono- and dihydrates) and even within one group of hydrates, polymorphism may occur, further increasing the possibilities of a given material to crystallize in many different forms, with different physico-chemical properties.

Not all hydrates, however, have a fixed stoichiometric ratio between the drug and the water molecules. If the water molecules can be incorporated into channels that may be present in a non-hydrate crystalline structure, many different ratios of water to drug molecules are possible, depending, for example, on the relative humidity of the air surrounding the hydrate. These types of hydrates are called channel hydrates.

If a stoichiometric hydrate loses its water (for example, by heating), the water in most cases will leave the crystal at a defined temperature. If we are measuring the weight loss of a hydrate as a function of temperature (these types of measurements are called thermogravimetric measurements), typical weight loss curves, as shown in Figure 11-43, are obtained for stoichiometric hydrates. In case of a dihydrate, for example, water may be leaving the crystal initially to form a monohydrate

that is then losing the remaining water at a higher temperature, so two steps in the weight loss curve may be observed (Figure 11-43b). For a given mass of material, if the molecular weights of the drug and the solvate molecule (18 g/mol in case of water) are known, the stoichiometric ratio of drug to solvate molecules can be determined by these measurements. In some cases the water only leaves the crystal upon melting of the material. In contrast, for channel hydrates the water molecules are leaving the crystalline structure often before melting and over a broader temperature range. If the water in a stoichiometric hydrate leaves the crystal before melting this may result in a collapse of the crystalline structure altogether, leading to the formation of an amorphous material. In the case of channel hydrates, the result of water leaving the crystal may be a non-hydrate with almost the same crystalline structure as the hydrate (same crystal lattice type and similar, but not identical, lattice parameters), sometimes called a vacant hydrate, before further polymorphic conversion to the stable non-hydrate form occurs.

Of recently increasing pharmaceutical interest are co-crystals.¹⁵ Unlike salts, in which the drug (as an anion or cation, depending on the nature of the drug) is bound by ionic interaction to the counter ion, in a co-crystal the drug and the co-crystal former (see below) are interacting via weaker interactions, in most cases by hydrogen bonding, as is the case for solvates. However, in contrast to solvates, both components of the co-crystal are solid at room temperature. If the hydrogen bond between the two molecules is formed by similar molecular moieties (for example, two carboxylic acid groups), this is known as a supramolecular homosynthon, whereas a hydrogen bond between two different moieties of the two components (for example, a phenol group and an amide group) is known as a supramolecular heterosynthon. Generally, the formation of a heterosynthon leads to a higher likelihood of formation of a stable co-crystal. Typical co-crystal formers used in the pharmaceutical field include maleic acid, tartaric acid, nicotinamide, saccharin, and many others. Since co-crystals have other properties, including solubility, hygroscopicity, and pharmaceutical processability, they offer interesting possibilities in drug development. Co-crystals (for example, the carbamazepine-nicotinamide co-crystal and the carbamazepine-saccharin co-crystals) can also show polymorphism.¹⁶

Eutectics

Above we have considered solid state materials containing two components but forming a single phase. Eutectics are an important example of a solid mixture of crystals (in contrast to co-crystals, which are “mixed crystals”, rather than a “crystalline mixture”). Eutectics thus contain two phases. However, in eutectic mixtures the two phases (two materials) are intimately mixed and thus often show different properties from simple physical mixtures of the two components.

Let us consider two crystalline materials, A and B, with meeting points T_{m_A} and T_{m_B} , that are miscible in the molten state, i.e., they form a single liquid phase. This is shown in form of a phase diagram in Figure 11-44. If we have a mixture of the two materials at the mixing ratio R1, at a temperature where both materials are molten (so we are in the liquid area of the phase diagram), and we now reduce the temperature of this melt, at the intersection with the melting curve, we will observe crystallization of compound A. As A will crystallize, but not B, this means that the composition of the remaining melt will change and it will become more B-rich. We thus follow the melting curve upon further lowering of the temperature until we reach the lowest point of the melting curve of A. If we lower the temperature further, now B will also crystallize. Similarly, if we have a B-rich mixture, B will crystallize when the temperature is lowered to the melting curve of B. This, in turn, will lead to the remaining melt having relatively more A than the original melt, and we follow the melting curve of B until we reach its lowest point, which is the same as for an A-rich system. The ratio of A and B at this point is called the eutectic mixture for these

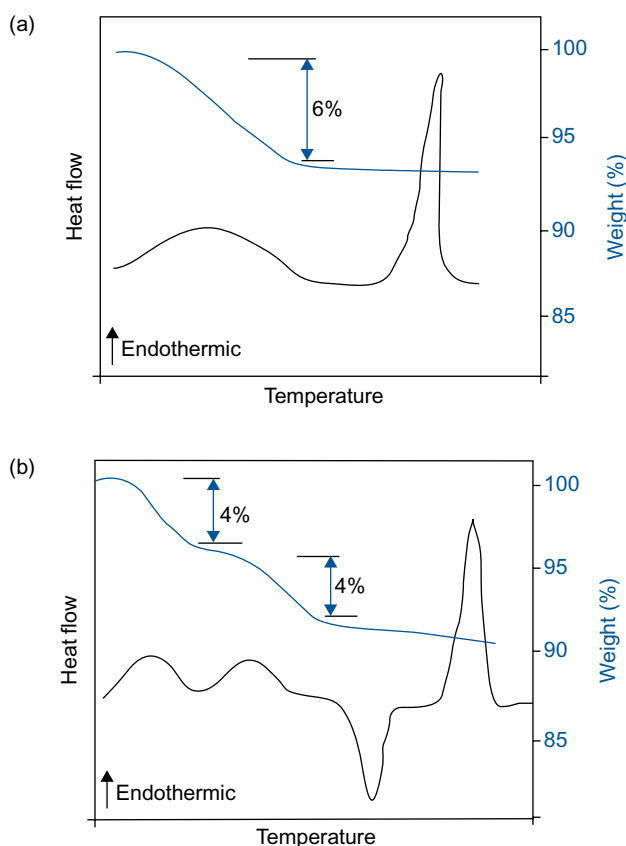


Figure 11-43. Thermogravimetric weight loss curves (—) and corresponding DSC curves (—) of (a) a hypothetical monohydrate and (b) a hypothetical dihydrate. Note that the water loss also appears as an endothermic event in the DSC thermogram. Other thermal events in the DSC curves (e.g., a polymorphic conversion, or melting) are not accompanied by a weight loss in the thermogravimetric curves. For more information on DSC, see the section Methods to characterize solids.

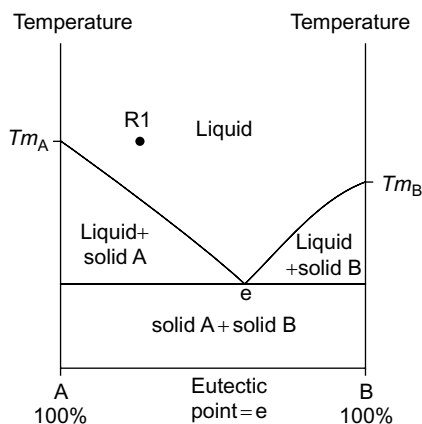


Figure 11-44. Phase diagram of two materials, A and B.

two compounds, and the temperature in the phase diagram is known as the eutectic temperature (e), the lowest temperature at which both compounds crystallize together. In other words, at the eutectic point both crystalline A and B and the melt are in equilibrium. For the eutectic mixture of sodium chloride and water, this occurs at a salt concentration of 22.4 percent and a eutectic temperature of -22°C .

For a given two compound system, the ratio of the compounds at the eutectic mixture can be estimated from the theoretical melting curves of the two compounds, using a simplified form of the Schröder–Van Laar equation:

$$\ln x = \frac{\Delta H_0}{R} \left(\frac{1}{T_0} - \frac{1}{T} \right) \quad (20)$$

Where: x is the mole fraction of one compound in the mixture; ΔH_0 and T_0 are the corresponding heat of fusion [J mol^{-1}] and melting temperature [K] of the pure compound, respectively; T is the melting point of the binary mixture at x ; and R is the gas constant [$8.314 \text{ J K}^{-1} \text{ mol}^{-1}$].

However, it should be noted that for this approach to be valid, a number of assumptions are made: there is no solid solution formation between the two compounds (see below), the melt has to be an ideal mixture, and the heat capacities of the pure compounds in the melt and as a solid should be fairly similar.

In the phase diagram shown above, we did not consider the possible formation of a solid solution between the two compounds. If this occurs, then there will be partial miscibility between the compounds in the solid state, and rather than pure compound A and pure compound B crystallizing together at the eutectic point, we will have the crystallization of an A-rich and a B-rich phase of the two corresponding solid solutions. This plays a major role in the formation of metal alloys.

Amorphous mixtures

We have discussed above that the amorphous form of a solid is thermodynamically unstable and tends to crystallize back to the stable crystalline form or a metastable polymorphic form. In order to make use of the advantages, including a higher dissolution rate and solubility that amorphous solids offer, especially for poorly water soluble drugs, it is therefore often necessary to stabilize the amorphous form. One way of achieving this is to use two components (for example, a drug and an excipient) in the formation of a solid amorphous material. In most cases this will be an amorphous, hydrophilic polymer. If an amorphous polymer (such as, for example, polyvinylpyrrolidone (PVP), PVP-vinylacetate copolymer (PVPVA), hydroxypropylmethylcellulose (HPMC), HPMC-acetatesuccinate (HPMCAS), and various polymethacrylates) is melted together with the drug to form a single phase melt, this can be cooled down rapidly to form a single phase amorphous material, known as a glass solution (do not confuse the amorphous glass

solution with the crystalline solid solution, mentioned above). In this solid form the drug and the polymer are molecularly mixed with each other. For this to be successful, it is required that the two compounds show mutual solubility, for example, have similar solubility parameters. It is often advantageous for the physical stability of glass solutions if some interactions between the drug and polymer occur (usually hydrogen bonding), since this will lower the molecular mobility (especially of the low molecular weight drug) and thus will slow down the processes of nucleation and crystal growth, which lead to the crystallization of the amorphous material. The polymers used for the formation of glass solutions often have a high glass transition temperature (T_g) so that the resulting glass transition temperature of the glass solution will be higher than the T_g of the pure amorphous drug, further stabilizing the drug against crystallization. If the glass transition temperatures of the two pure amorphous components are known, the resulting glass transition temperature of the glass solution in various mixing ratios of the two compounds can be estimated based on the Gordon Taylor equation:

$$T_{g12} = \frac{w_1 \cdot T_{g1} + K \cdot w_2 \cdot T_{g2}}{w_1 + K \cdot w_2} \quad (21)$$

Where: T_{g12} is the glass transition temperature [K] of the mixture; T_{g1} and T_{g2} are the glass transition temperatures [K] of the single amorphous compounds (e.g., the drug and the polymer); w_1 and w_2 are the weight fractions of each compound in the mixture; and K is a constant, depending on the glass transition temperatures and amorphous densities of the single compounds:

$$K = \frac{T_{g1} \cdot \rho_1}{T_{g2} \cdot \rho_2} \quad (22)$$

Where: ρ_1 and ρ_2 are the densities of the single amorphous components.

It should be noted, though, that this equation assumes volume additivity, i.e., the absence of specific molecular interactions and, as such, can often only be used as an estimate for the actual glass transition temperature of the glass solution.

In most cases the T_g of the glass solution is predicted by using the T_g of the drug and polymer. However, in real situations a third component can be present: moisture. It is known that moisture can accelerate crystallization from the amorphous state, and this effect can be explained by the plasticizing effect of water. The T_g of water is approx. 139 K, which is significantly lower than the T_g of commonly used pharmaceutical materials. In the same way that a high T_g polymer is added for stability, a low T_g compound can lower the T_g of the system. If the Gordon-Taylor equation is extended to a third component, it is obvious that even small amounts of a low T_g compound can have a significant effect on the T_g and, hence, on stability. Therefore, moisture should be excluded during manufacturing and storage of amorphous systems.

If the two components (for example, the drug and the polymer) are not miscible on a molecular level, the formation of a solid suspension may occur upon cooling of a melt of the two components. These systems will show two glass transition temperatures, one for the drug (or possibly a drug rich phase, if there is partial miscibility) and one for the polymer (or the polymer rich phase).

METHODS TO CHARACTERIZE SOLIDS

It is beyond the scope of this chapter to provide a detailed description of the various methods that have been developed to characterize the properties of solids. However, in a brief fashion we will have a look at several important techniques that allow the characterization of the many forms in which the solid state of matter, for example, of a given drug may manifest itself.

We have seen that a pharmaceutical solid can be present during development or in a final dosage form in various polymorphic forms. It can also be present in form of hydrates or

other solvates. Salt formation is possible, as is the formation of a co-crystal. Solvates, salts and co-crystals in their own right can show polymorphism. The drug may also be partly or fully amorphous, either on its own or in form of a glass solution or amorphous suspension. All of these solid state forms will differ in many of their physical properties, and this can be relevant for the drug and dosage form if, for example, solubility, dissolution rate, and chemical and physical stability are affected. It is thus obvious that we require a thorough identification and, in many cases quantification, of the various solid state forms of a drug that we may encounter in preformulation, formulation, during storage and upon administration of the dosage form. It is also obvious that this task is more complex for solids than, for example, if the drug is present in liquid form or in a solution, for in these cases all differences of the solid state forms disappear.

The list of analytical techniques that are useful to investigate solid state properties is almost endless, and it is convenient to differentiate the techniques available to the pharmaceutical scientist according to the different levels at which they probe the solid material. Most spectroscopic techniques, such as Raman and infrared spectroscopy, but also solid state NMR, near infrared spectroscopy and others are probing predominately molecular properties. If a drug, for example, can crystallize

in different polymorphic forms, these forms still contain the same molecules, and as such, the resulting spectra should be identical. Indeed, that is (almost) the case. However, the fingerprint regions of the resulting spectra will also be influenced by the neighboring molecules, for example, by the unit cell of the crystalline forms. Because these are different between different polymorphic forms, the resulting spectra, for example, for IR and Raman measurements, will also (subtly) be different. If the polymorphic pair shows configurational differences between the molecules, the differences in the resulting spectra will be larger than if the configuration of the drug molecules is the same in both polymorphs. It is often necessary to use multivariate data analytical tools, such as principal components analysis or partial least squares analysis, to identify the different solid state forms using IR or Raman spectroscopy. It is advisable, in any case, to use such techniques if a quantification of different solid state forms that are present at the same time is desired. Generally, differences in the spectra between a non-solvate and a solvate form are larger than between two polymorphic forms, and differences between the spectra of a crystalline and an amorphous form may be even larger, with the spectra in IR

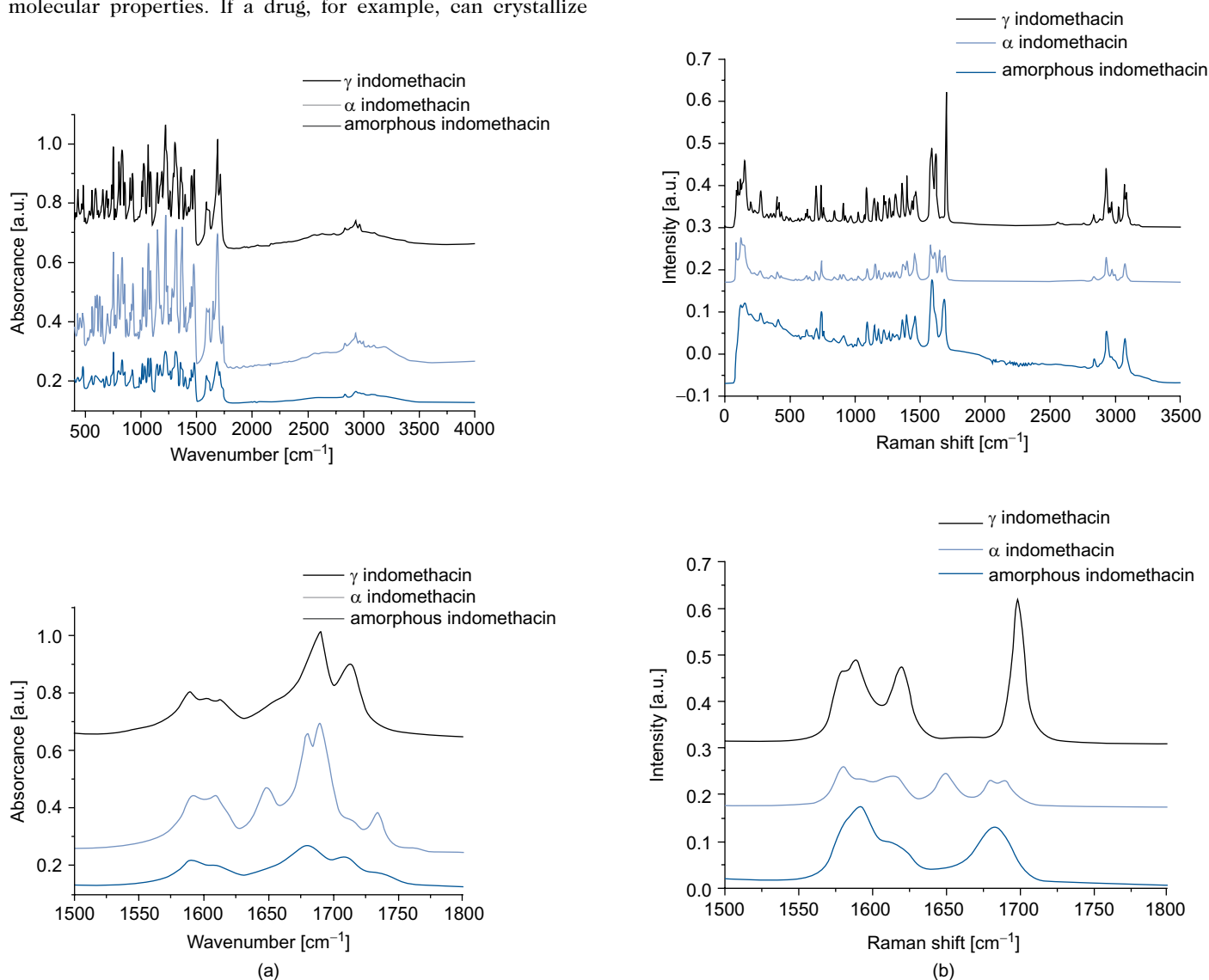


Figure 11-45. (a), Infrared spectra of amorphous indomethacin, and the α - and γ - polymorphic crystalline forms of indomethacin. The lower figure shows the wavenumber region of 1500–1800 cm⁻¹ (C=O stretch), where differences between the different solid state forms can be clearly visualised. Individual traces are offset for clarity. (b), Raman spectra of amorphous indomethacin, and the α - and γ - polymorphic crystalline forms of indomethacin. The lower figure shows the wavenumber region of 1500–1800 cm⁻¹ (C=O stretch), where differences between the different solid state forms can be clearly visualised. Individual traces are offset for clarity.

and Raman measurements being less well defined (peak broadening and peak merging) for amorphous forms, compared to their crystalline counterparts. Figure 11-45 shows example IR spectra (Figure 11-45a) and Raman spectra (Figure 11-45b) of different solid state forms of the drug indomethacin.

The next level of analytical techniques, and in many ways the “gold standard” techniques to characterize and quantify different solid state forms of a drug, are the techniques that probe an assembly of molecules. These techniques have been termed “particulate level techniques,” and include X-ray diffraction (mostly used in the forms of X-ray powder diffraction, XRPD), and various thermal analytical techniques (of which differential scanning calorimetry, DSC, plays the largest role in pharmaceutical analysis). These will be described briefly below. It should be noted, however, that also terahertz spectroscopy, probing the phonon modes of solids, is a particulate level technique, as are the various microscopic and electron microscopic techniques.

X-ray powder diffraction is one of the most important analytical techniques used in solid-state analysis. An X-ray powder pattern is the result of scattering of an incident monochromatic X-ray beam by the lattice of the crystalline material. The resulting diffractogram is based on constructive and destructive interference of the diffracted X-ray beams. This phenomenon is described by Bragg’s law:

$$n\lambda = 2d\sin\theta \quad (23)$$

Where: n is the order of reflection (an integer); λ is wavelength of the incident beam [Å]; d is distance between the planes in the crystal [Å]; and θ is angle of beam diffraction.

Constructive interference takes place when the scattered X-rays are in phase, i.e., the phase difference that the scattered X-rays reflected from two neighboring planes of the crystal lattice is an integer (Figure 11-46). Each crystalline material will exhibit a diffractogram, which is unique to the specific compound and its solid state form, and individual peaks may be attributed to the crystalline structure. As amorphous solids do not possess an ordered lattice, no diffraction occurs and no diffraction pattern is obtained. Rather, a diffuse “amorphous halo,” showing one or several broad maxima, is observed. XRPD is particularly useful to detect small amounts of crystallinity in an amorphous sample, for the residual crystallinity results in small diffraction peaks on an amorphous halo. However, the technique is of limited use for the detection of low levels of amorphous content since the halo will be obscured by the stronger crystalline diffraction peaks.

XRPD is one of the most commonly used techniques in the pharmaceutical industry for solid-state analysis because it may be used for the identification and quantification of polymorphs, solvates, hydrates and amorphous forms, based on their unique diffraction patterns. It does not require large sample sizes, and unlike thermal methods, it is a non-destructive technique. This is potentially advantageous in the early stages of pharmaceutical development, and despite its limitations in detecting low levels of amorphous material, XRPD has been used both qualitatively and quantitatively in the characterization of partially

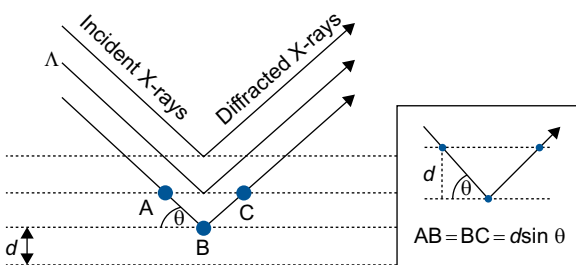


Figure 11-46. Diffraction of X-ray beams by planes of a crystal lattice. A, B, and C are points in the crystal planes. d is the distance between two planes.

crystalline systems. Quantification of the crystalline content can be achieved by measuring the area or the height of the main peaks in the diffractogram and of the amorphous content by measuring the amorphous scattering in the diffractogram. However, quantitative analysis using XRPD has to be carried out carefully, for physical characteristics such as sample height, surface character and shape, and the presence of microcrystallinity may influence the outcome of the measurements. Figure 11-47 shows example diffractograms of different solid state forms of the drug indomethacin.

DSC is used in the study of both crystalline and amorphous materials. Two types of DSC instruments are available: power-compensated DSC and heat-flux DSC. In power-compensated DSC a reference and a sample pans are placed on two individual furnaces, and the temperature difference between them is maintained at zero by varying the heat that is required to keep both pans at the same temperature. The parameter recorded is the difference in energy input which is required to maintain the samples at the same temperature, for example, during a melting process. In heat-flux DSC one heating element is used for sample and reference. Heat is applied to both via an electrically heated thermoelectric disc, and the temperature difference between sample and reference is monitored. The obtained signal is converted to heat flow [mW] via the following relationship:

$$\Delta Q = \frac{T_s - T_r}{R_\gamma} \quad (24)$$

Where: Q is heat energy [J]; R_γ is the thermal resistance [KJ⁻¹]; T_s is the sample temperature [K]; and T_r is the reference temperature [K].

In DSC, thermal events connected to the sample are presented in a thermogram, in which endothermic and exothermic energy deviations from the reference pan are recorded as a function of time or temperature. Endothermic events include processes such as evaporation, melting, and some solid–solid transitions. In contrast, the recrystallization of an amorphous form and other solid–solid transitions are exothermic. It is possible to use DSC not only to determine the temperature of these events but also to calculate the heat of the respective reaction.

As we have seen above, any change in the enthalpy of a system at constant pressure can be described by the change in heat. Hence, the area under a DSC peak is directly proportional to the heat absorbed or released by the thermal event if the sample mass is known.

The determination of the heat capacity, C_p , for the amorphous state is of particular interest, for the glassy and the super-cooled liquid state exhibit different values for their respective C_p s, and

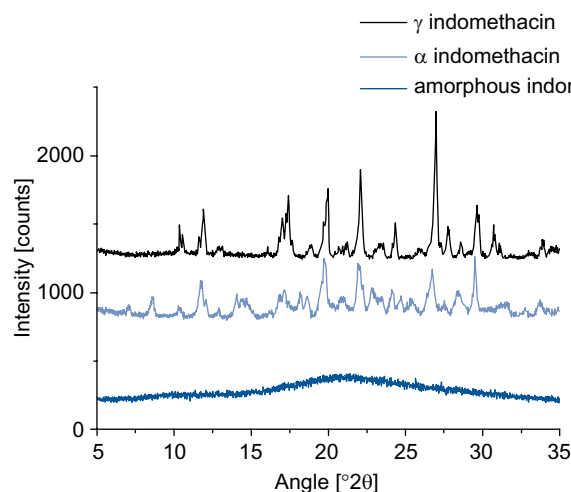


Figure 11-47. X-ray powder diffraction patterns of amorphous indomethacin, and the α - and γ - polymorphic crystalline forms of indomethacin. Individual traces are offset for clarity.

the heat capacity will change in a stepwise fashion at the glass transition temperature. The extent of the difference in heat capacity (ΔC_p) is an important characteristic for amorphous materials. The enthalpy of a system increases as its temperature is increased. The heat capacity can be obtained as the slope of a plot of enthalpy versus temperature at constant pressure:

$$C_p = \left(\frac{\delta h}{\delta T} \right)_p \quad (25)$$

The advantages of DSC for analytical purposes lie in the requirement of only small sample sizes (typically 2–6 mg), the broad temperature range, and in the speed of the experiments (heating rates of 5–20 K/min are typically employed). DSC can be used to analyze both crystalline and amorphous materials, and unlike XRPD, it is possible to directly detect the amorphous content through the presence of a T_g .

The thermal behavior of a solid sample may be analyzed, for example, in terms of

- melting point and melting enthalpy of polymorphic crystalline materials
- conversion temperature of two enantiotropic polymorphic forms
- enthalpy and temperature of the loss of solvate molecules from solvates
- heat capacity difference and temperature of the glass transition of amorphous material
- enthalpy and temperature of any crystallization events of amorphous material
- temperature of decomposition
- extent of interactions in mixtures (e.g., with excipients).

Quantification, for example, of amorphous content in a sample is also possible. However, despite the many advantages that make DSC one of the most widely used techniques, it has to be considered that the analyzed sample is subjected to thermal treatment, which destroys or at least changes the sample and may induce thermal artifacts. Therefore, care has to be taken in the interpretation of thermograms to avoid misleading conclusions. Moreover, DSC thermograms are often difficult to interpret if multiple overlapping thermal processes are involved. Every DSC thermogram is a compromise between sensitivity and resolution as both are dependent on the heating rate, but in opposite ways. To increase the sensitivity, the scan rate may be increased; however, this may lead to decreased resolution,

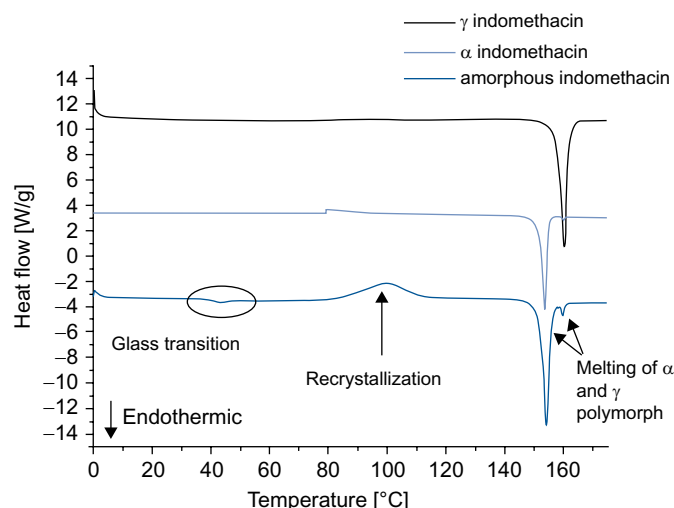


Figure 11-48. DSC thermograms of amorphous indomethacin, and the α - and γ - polymorphic crystalline forms of indomethacin. The thermogram of the amorphous form shows a glass transition and recrystallization event, in addition to the melting. Individual traces are offset for clarity.

for the scan rate may be too rapid to separate close thermal events. Figure 11-48 shows example DSC thermograms of different solid state forms of the drug indomethacin.

Modulated temperature DSC (MTDSC) is an alternative to conventional DSC.¹⁷ In MTDSC a sine wave modulation is applied to a linear temperature program. The obtained heat flow signal is analyzed by applying a discrete Fourier Transform algorithm, which results in the deconvolution of the measured heat flow signal into reversing and non-reversing components. The non-reversing signal comprises the kinetically controlled events that are dependent on the absolute temperature alone, and the reversing signal is a function of the samples' heat capacity and rate of temperature change. The heat flow signal of a sample will be a combination of these two components. The total heat flow signal may, therefore, be expressed as

$$\frac{dQ}{dt} = \frac{C_p}{dT} + f(t, T) \quad (26)$$

Where: dQ/dT is the heat change over time (heat flow) [$J s^{-1}$ or W]; C_p is the "thermodynamic" heat capacity [$J K^{-1}$]; T is the temperature [K]; and $f(t, T)$ is the kinetic response.

Conventional DSC records the reversing and non-reversing heat flow simultaneously and is not able to resolve them. MTDSC is able to separate the heat capacity (reversing) and kinetic (non-reversing) components, due to their different responses to the underlying and modulated heating rates. Processes that occur at defined temperatures, such as crystallization, evaporation, relaxation, and decomposition will appear in the non-reversing heat flow signal. As the reversing component is related to the samples' heat capacity, for example, the glass transition event is visible in the reversing signal. This enables the separation of the glass transition event from any endothermic relaxation.

LIQUID CRYSTALS

We have seen that crystalline solids are defined by long range order in three dimensions. In contrast, liquids lack long-range order (as do amorphous solids). When some solid materials are heated, they do not directly show a phase transition to the liquid state but instead enter into a different state of matter, between the solid and the liquid state. This state of matter is called the liquid crystalline state. Liquid crystals (also called mesophases) show structural properties which are intermediate between those of crystalline solids and liquids. It should be noted, however, that liquid crystals are not simply a mixture of solids and liquids but, indeed, a separate state of matter. Also, it is important to note that not every organic material is able to exist in a liquid crystalline state. We will discuss a few structural considerations of molecules that can exist in the liquid crystalline state below. These molecules are termed mesogens and usually are anisotropic organic molecules, mostly of rod-like shape.

In a thermogram we observe a first order phase transition, when we, for example, cool a melt that at a certain temperature transforms into a liquid crystal. Similarly the transition from a solid to a liquid crystal is a first order transition. Being situated between the solid and the liquid state of matter, liquid crystals also are a condensed state of matter.

There are two principal types of liquid crystals: thermotropic liquid crystals (TLC) and lyotropic liquid crystals (LLC). TLCs are formed by heating a crystalline solid. In contrast, LLCs are formed by the addition of a liquid (in most pharmaceutical cases, water), to a usually solid mesogen. If we are applying Gibbs' phase law, we can say that the minimum number of components in a TLC is one and in a LLC is two, if a single phase liquid crystal is formed. The degrees of freedom to describe a TLC is two (temperature and pressure) and for a LLC is three (temperature, pressure, and the concentration of the mesogen).

Lyotropic liquid crystals

LLCs can be understood as associations of differently shaped micelles. Micelles are colloidal assemblies of amphiphilic

molecules, e.g., surfactant molecules and phospholipids, i.e., molecules that have a polar head group and a lipophilic, non-polar tail group. We can differentiate the following types of LLC, shown in Figure 11-49:

Lamellar phase (Figure 11.49a): This is pharmaceutically perhaps the most important liquid crystalline phase and consists of a layered packing of indefinitely extended disc-like micelles, leading to a bilayer structure as repetition unit. As such, the lamellar phase has a one-dimensional, long-range, positional order (from layer to layer) but only long-range orientational order within the layers. It is also possible that a lamellar liquid crystalline phase may be dispersed in excess water (so that we get a two phase system). The resulting lamellar liquid crystalline particles form concentric bilayers and are termed liposomes. Phospholipids are pharmaceutically relevant examples of lamellar liquid crystal forming molecules.

Hexagonal phase (Figure 11.49b): This type of LLC consists of hexagonally packed rod-like micelles. As such, the hexagonal phase has a two-dimensional, long-range, positional order (normal to the symmetry axis of the rods) but only long-range, orientational order within the rods. Depending on the nature of the solvent (polar or non-polar), hexagonal phases can exist in a normal and in a reversed form. In the normal form the polar head groups of the mesogens are pointing outwards, and in the reversed form inwards into the rod-like micelles that are forming the hexagonal phase. As we have just discussed for the lamellar phase, also the hexagonal phase of some amphiphilic molecules may be dispersed into colloidal particles, known as hexosomes.

Cubic phase (Figure 11.49c): This type of LLC consists, in the simplest case, of a cubic packing of spherical micelles. As such, the cubic phase has a three-dimensional, long-range,

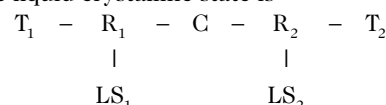
positional order, like a crystalline solid. In contrast to a solid crystal, however, the mesogens in the micelles forming the cubic phase show rotational-diffusional motion and not merely vibrations, as in the case of crystals. Micellar cubic phases can also exist in a normal and in a reverse form, depending on the nature of the solvent. Moreover, bicontinuous cubic phases exist, which cannot simply be interpreted as assemblies of spherical micelles, and some bicontinuous cubic phases may also be dispersed into particles called cubosomes.

Lyotropic nematic phase (Figure 11.49d): This LLC is composed of rod-like micelles and shows a long-range orientational order with respect to the symmetry axis of the micelle. There is no long-range positional order in this type of LLC.

As we could expect from the increasing degree of order, the viscosity of the liquid crystals increases as we go from lyotropic nematic to lamellar, hexagonal, and cubic liquid crystals. Many surfactant and phospholipid mesogens may also show liquid crystalline polymorphism, i.e., they are able to self-assemble into different LLCs, depending on the concentration of the mesogens and temperature.

Thermotropic liquid crystals

The typical structure of a rod-like mesogen able to exist in a thermotropic liquid crystalline state is



The rod-like shape of the mesogen is usually formed by two ring systems (R_1 and R_2 , which can be aromatic or aliphatic and may also consist of condensed ring systems). These ring systems are connected by a polar central group (C) and linked to two terminal groups (T_1 and T_2), of which at least one usually is an alkyl chain. However, not all these structural elements necessarily have to be present in a mesogen. Additionally, lateral substituent groups may be present in the mesogen molecular structure (LS_1 and LS_2).

As was the case for LLC, several types of TLC can be differentiated. It is beyond the scope of this chapter to give a detailed description of all TLC types that have been described in the literature (there are over 20 different types of smectic TLC alone). We will thus restrict the description of different TLCs here to three types: nematic phases, chiral nematic phases (cholesteric phases), and fluid smectic phases (including smectic A and smectic C). Schematic representations of these TLCs are shown in Figure 11-50.

Above, we have already described a lyotropic nematic phase, formed by elongated micelles. In case of a thermotropic nematic phase, individual rod-like molecules (not micelles) show an orientational long-range order, again in the absence of any positional long-range order. The orientation of the long axes of the rod-like molecules, however, is not the same for all molecules and varies around a preferred orientation, known as the director of the nematic phase (n). This can be expressed using the order parameter S :

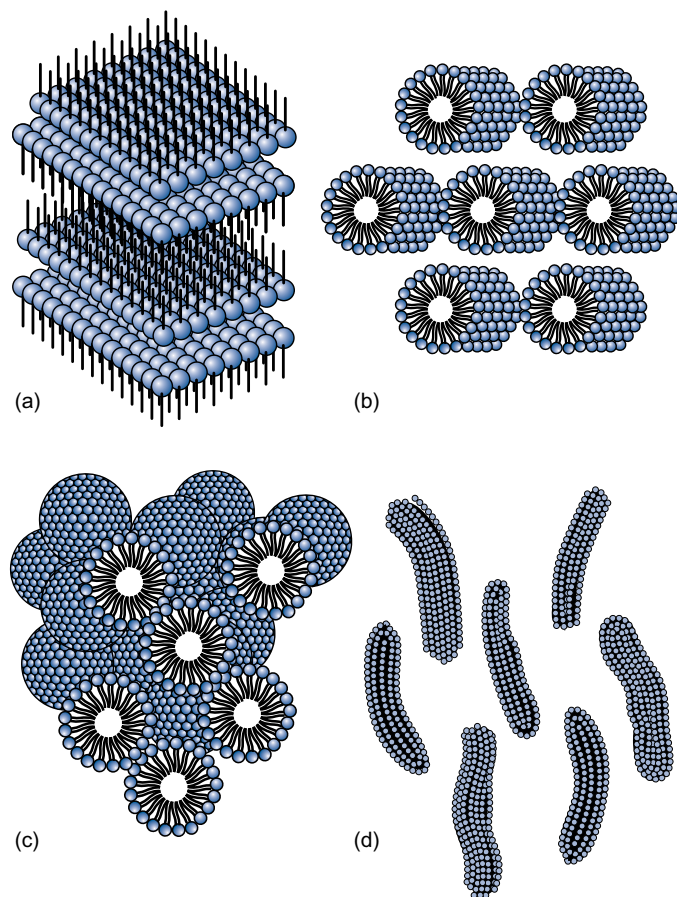


Figure 11-49. Schematic representation of several types of lyotropic liquid crystals. (a) lamellar phase, (b) hexagonal phase, (c) cubic phase, and (d) lyotropic nematic phase.

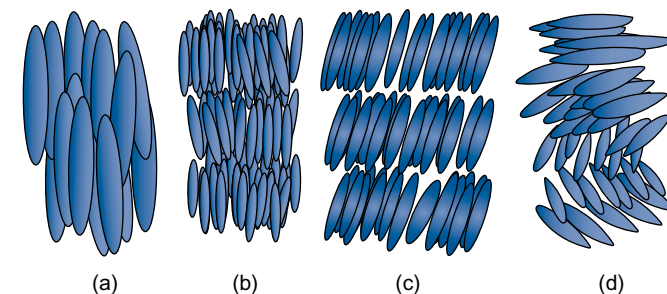


Figure 11-50. Schematic representation of several types of thermotropic liquid crystals: (a) nematic, (b) smectic A, (c) smectic C, and (d) chiral nematic (cholesteric) phase.

$$S = 0.5 < 3 \cos^2 \Theta - 1 > \quad (27)$$

Where: Θ is the angle between the director and the orientation of the molecular length axis of the molecule (see Figure 11.51).

This is, of course, an average value, depending, for example, on the temperature. In the ideal case of complete parallel arrangement of the molecules, S would have the value of 1.

The chiral nematic phase can be regarded as a special form of a nematic phase. If the mesogen is chiral (see above), the director field may adopt a helical superstructure, so that we have nematic layers whose directors are shifted by a constant angle from layer to layer. The distance between two nematic layers with the same director is known as the pitch of the chiral nematic phase. Since the pitch height of some cholesteric mesophases is in the order of the wave length of visible light, they may be colored, and since the pitch height is sensitive to temperature, such liquid crystals may be used as sensitive thermometers, with a specific color indicating a specific temperature.

Smectic phases have a layered structure, similar to the lamellar structure of LLCs. However, in contrast to the lamellar phase, the smectic phase usually does not form a bilayer structure. Depending on the angle the rod-like molecules have with respect to the layer, we can differentiate the smectic A phase (in which the molecules are oriented perpendicular to the layers) from the smectic C phase (in which the molecules are tilted with respect to the layers).

Several drugs have been identified to be able to exist as thermotropic liquid crystals at elevated temperatures,¹⁸ including fenoprofen sodium, fenoprofen calcium,¹⁹ and itraconazole.²⁰

Both LLC and TLC can be identified by characteristic appearances (textures) in the polarizing light microscope, as well as by electron microscopic investigation. Also, the liquid crystals show characteristic peak sequences when investigated by X-ray diffraction. Since the distances between the repetition units in liquid crystals are larger than in crystals, the diffraction angles are smaller and, therefore, small angle X-ray diffraction is usually used as an analytical technique, in contrast to wide angle X-ray diffraction, which is usually used in powder diffraction of crystalline solids.

LIQUIDS

The liquid state of matter is an intermediate state between the solid (and potentially the liquid crystalline state) and the gaseous state of matter. It is thus not surprising that the liquid state of matter shares some commonalities with both solids and gases. As for solids, liquids are practically incompressible under pressure (except if very high pressures are used) and have a high density. This is the reason that solids and liquids are summarized as condensed states of matter. Usually, however, the density of liquids is lower than that of solids for a given material, with water being an important exception, as we have discussed above. Unlike solids, however, the molecules in a liquid can move and change places by diffusing under the influence of heat energy (diffusion). This is a property that liquids have in common with gases and the reason that gases and liquids are summarized as fluid states of matter. While liquids show in general a much higher tendency to be miscible with each other than solids, not all liquids are miscible (unlike gases, which are always completely miscible with each other), and in many cases mixing two liquids will lead to the formation of a two phase system of the two immiscible liquids or

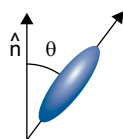


Figure 11-51. Schematic showing the angle Θ between the director ($\hat{n}$) and the orientation of the molecular length axis of a molecule in a nematic liquid crystal.

a two phase system of a liquid A-rich phase and a liquid B-rich phase if the liquids show partial miscibility. In these cases the two liquids will form an interface with a corresponding interfacial tension. Since the interfacial tension γ is the intensive variable of the interfacial free energy G_S (with the interfacial area A being the extensive variable; $G_S = \gamma A$), two phase liquid systems have a higher energy if the interfacial area between the liquids is increased. An example for this is an emulsion, in which one liquid phase is dispersed in the form of droplets in the other. The system will try to minimize its free energy, and thus the dispersed droplets will fuse together (coalescence) over time to reduce the interfacial energy by reducing the interfacial area. To reduce the surface free energy of emulsions and thereby to stabilize the dispersed nature of the emulsion, surfactants can be added to the emulsion to reduce the interfacial tension between the two liquid phases.

Generally, the principle “like dissolves like” can be applied, and liquids with similar polarity (for example, water and ethanol, or benzene and cyclohexane) will be miscible, whereas liquids with different polarities (for example, water and benzene) will be immiscible or show partial miscibility.

Another property liquids have in common with gases is that they are isotropic. This means their physical properties are uniform in all directions or orientations. This is not the case for crystalline solids and liquid crystals, which are anisotropic (with the exception of cubic crystals and liquid crystals). The anisotropic nature of most crystals (and liquid crystals) is also the reason for their optical birefringence, making them visible in the polarized light microscope. Conversely, liquids will be invisible in the polarized light microscope under cross polarization.

If we investigate a liquid with X-ray diffraction, we only measure the characteristic halo we have already discussed in the context of amorphous solid materials (amorphous materials are also isotropic and show no long-range order, which would give rise to peaks in a diffractogram). However, also like amorphous materials, liquids are not without near order, and certain distances between the molecules of a liquid may be preferred over others, due to intermolecular interactions. This can also be determined from X-ray diffraction measurements, by calculating the radial distribution function $g(r)$, which can be generated by a Fourier transformation of the diffraction pattern, and presents the average spatial distribution in a liquid (or in a amorphous solid).

An important property relevant for many pharmaceutical applications of liquids is their viscosity. Viscosity is the resistance to flow or, in other words, the resistance of the liquid to be deformed under a shear stress (due to the intermolecular interactions of molecules in a liquid). If a shear stress (a force F [N] applied parallel to the fluid's surface A [m²], F/A [N/m²]) is applied to a liquid, this will result in a velocity gradient in the liquid ($d\mathbf{v}/d\mathbf{x}$ [s⁻¹]). The ratio between the two is called viscosity η [Pa s]

$$\eta = (F / A) / (d\mathbf{v} / d\mathbf{x}) \quad (28)$$

This equation is usually rearranged to give Newton's law of viscosity:

$$F / A = \eta d\mathbf{v} / d\mathbf{x} \quad (29)$$

Liquids can be differentiated by their viscosities and the type of viscosity they have. A Newtonian liquid (ideal viscosity) is a liquid in which η remains constant if the shear stress or shear rate is increased. Water and ethanol are examples of Newtonian liquids. Not all liquids, however, show ideal flow behavior, and these liquids are called non-Newtonian liquids. Many pharmaceutically relevant liquids are non-Newtonian. If the viscosity decreases with increasing shear stress, the liquid shows pseudoplastic flow behavior (also called shear thinning), and in the opposite case, when the viscosity increases as shear stress is increased, the liquid shows dilatant flow behavior (also called shear thickening). Time, however, does not change the viscosity for a given shear stress for pseudoplastic and dilatant liquids (nor for Newtonian liquids). This is, however, not always the case. A pseudoplastic liquid shows a certain viscosity for a certain shear stress. If the stress is kept constant, for

some liquids a further decrease in viscosity can be observed. When the shear stress is removed, the system does not immediately return to its original state, but the viscosity gradually increases. This behavior is termed thixotropy. For pharmaceutical suspensions this is advantageous, for the suspension can be homogenized by shaking and can be poured easily. After some time, the viscosity increases again, stabilizing the suspension against sedimentation.

While many polymer solutions used in pharmaceutical applications are pseudoplastic liquids, concentrated pastes may show dilatant flow behavior, and many suspensions and “semi-solids”, such as creams and ointments, show thixotropy. Note that a semisolid is a pharmaceutical term used to describe creams, ointments, etc. These are solids at rest, but they start to flow (become liquid) at low shear stresses.

To describe pseudoplastic and dilatant liquids, we can extend Newton's viscosity equation to a power law, known as the Ostwald-de Waele equation:

$$F/A = K(dv/dx)^n \quad (30)$$

Where: K is the flow consistency index [Pa s^n]; and n is the flow behavior index [no unit]. $n = 1$ for a Newtonian liquid, $n < 1$ for a pseudoplastic liquid and $n > 1$ for a dilatant liquid.

The viscosity of liquids decreases with increasing temperature, due to the higher molecular mobility. Gases, in contrast, not only have much lower viscosities than liquids but also show an increase in viscosity as temperature is increased (due to higher collision rates of the gas molecules at elevated temperatures). For example, water has a viscosity of $1.002 \times 10^{-3} \text{ Pa s}$ at room temperature and $0.35 \times 10^{-3} \text{ Pa s}$ at 80°C . Air, on the other hand, has a viscosity $1.79 \times 10^{-5} \text{ Pa s}$ at room temperature. Amorphous solids in the glassy state have viscosities in the order of 10^{12} Pa s or higher.

GASES

Gases play an important role in the pharmaceutical field, for example, in the use of medical gases (e.g., oxygen, nitrous oxide, carbon dioxide, compressed air, etc.), as functional excipients in aerosols and inhalers, and in quality control of drugs and dosage forms. Further, oxygen from the air may contribute to oxidative degradation of drugs, and water molecules from air can initiate hydrolytic degradation. Water molecules in air (especially at high relative humidity) may also decrease the physical stability of amorphous drugs or glass solutions, as we have seen, since water is a potent plasticizer, lowering the glass transition temperature of amorphous materials and thus increasing the likelihood of crystallization.

The gaseous state of matter is characterized by very weak interactions between the molecules (for real gases), which in fact are assumed to be completely absent for ideal gases (see below). The consequence of this is that in gases the molecules constantly change their positions by diffusion or convection and have a very low viscosity and density. This makes gases easily compressible under pressure (in contrast to liquids and solids). Also, in contrast to solids and liquids, different gases are always completely miscible with each other. For example, air is a mixture of several gases: approximately 78.1 percent nitrogen, 20.9 percent oxygen, 0.9 percent argon, 0.03 percent carbon dioxide, and small concentrations of neon, helium, methane, and other gases.

If we are completely disregarding the weak interactions between gas molecules, we can fully describe the behavior of gas by the ideal gas law. Historically, the properties of gases have been described in a set of laws, which we will have a quick look at first:

Boyle's law states that the product of absolute pressure (p) and volume (V) is constant (K_1) at constant temperature:

$$pV = K_1 \quad (31)$$

Charles' law states that if a given mass of a gas is heated at constant pressure, the volume of the gas is directly proportional (K_2) to its absolute temperature (T):

$$V/T = K_2 \quad (32)$$

Avogadro's law states that equal volumes of pure gases at constant temperature and pressure contain the same number of particles (n):

$$V/n = K_3 \quad (33)$$

The volume that one mole of gas molecules occupies (the molar volume) is 22.4 L at 0°C and a pressure of 1 atm. The number of molecules in a mol of gas is: $6.022 \times 10^{23} \text{ mol}^{-1}$.

These three laws can be regarded as special cases of the ideal gas law:

$$pV = nRT \quad (34)$$

Where: R is the ideal gas constant [8.314 J/(mol K)]. Note that K_1 in Boyle's law is nRT , K_2 in Charles' law is nR/p , and K_3 in Avogadro's law is RT/p in the ideal gas law.

Finally, if we take into account the ideal gas law (which assumes no interactions between the molecules forming the gas), we arrive at Dalton's law, which states that the total pressure (p_t) of a mixture of gases equals the sum of their partial pressures (p_1 to p_n):

$$p_t = p_1 + p_2 + \dots + p_n \quad (35)$$

Related to Dalton's law is Henry's law, which states that at constant temperature the amount of a gas (n) that dissolves in a given volume V ($c = n/V$) of a given liquid is directly proportional (K_4) to the partial pressure of that gas in equilibrium with the liquid (p_1).

$$c/p_1 = K_4 \quad (36)$$

While all the above laws hold true for many gases over reasonable temperature and pressure ranges, and thus can be used fairly well to describe the behavior of gases, they do not take into account the fact that interactions between the gas molecules do take place and that the gas molecules occupy a part of the total volume of the gas (albeit usually a very small part). Deviations of the behavior of gases from the ideal situation become more important as, for example, pressure is increased or temperature is decreased. We can also understand this by the fact that gases can be liquefied if the temperature is lowered (this would not be possible if the gas molecules did not interact with each other). On the other hand, many gases behave very closely to ideal gases, and this can be seen, for example, if we determine the molar volumes of some gases. For the gases that make up more than 99 percent of air, we find that nitrogen, oxygen, and the noble gases indeed have molar volumes of 22.4 L. Other gases, though, already at 0°C and a pressure of 1 atm, have molar volumes that differ slightly from that of ideal gases. Methane and carbon dioxide, for example, have molar volumes of 22.3 L and ammonia of 22.1 L.

As stated above, for real gases we have to take into account that the gas molecules occupy a certain volume in the given volume of a gas and that interactions between the gas molecules can occur. These interactions, for example, could be dipole-dipole interactions if the gas molecules contain polar bonds (these interactions are termed Keesom forces). Even in the absence of permanent dipoles in the gas molecules, interaction can occur through transiently induced dipoles (London forces). Note that in the literature the expression “van der Waals force” is sometimes used to describe Keesom and London forces and sometimes specifically for London forces.

The van der Waals equation is one model that takes into account the non-zero volume of the gas molecules and the possible interactions between the gas molecules:

$$nRT = (p + an^2/V_m^2)(V_m - b) \quad (37)$$

For one mol of gas the equation simplifies to

$$RT = (p + a/V_m^2)(V_m - b) \quad (38)$$

Where: p is the pressure; T is the temperature; R is the ideal gas constant; V_m is the molar volume; and n is the number of molecules.

The parameters a and b are parameters that have to be empirically determined for each gas. The parameter a appears in the pressure term of the equation and relates to intermolecular interactions, whereas the parameter b appears in the volume term of the equation and relates to the fact that the gas molecules themselves occupy a certain volume of the gas.

It should be noted that the van der Waals equation is just one of many models that have been developed to better describe the behavior of gases, especially at extreme conditions, and that many other models exist. This is, perhaps, an opportune moment to remind us that a model is a mathematical description of a physical reality, and not the physical reality itself. It is up to the given situation (and possibly the required precision) as to which model is best used in a particular situation.

PLASMA

We will finally briefly discuss the plasma state of matter. Plasma is actually the most common state of matter in the universe (for example, stars are matter in the plasma state). Plasma is formed by a phase transition from gases if these are heated to very high temperatures. At a certain temperature the electrons of the gas molecules or atoms leave the atoms, leading to an ionization of the gas. Plasma thus contains charged molecules or atoms and free electrons and conducts electricity (unlike gases). Due to the very high temperature of plasma, it is not particularly useful in the pharmaceutical field. Not all atoms however, need to be ionized for plasma to gain its typical properties. So most plasma consist of ions, electrons, and neutral molecules, which nevertheless are in an activated state, e.g., in the form of radicals (molecules containing one or more unpaired electrons).

It is also possible to create the plasma state from gases by electric discharge under reduced pressure, leading to "cold plasma". Cold plasma created in this way has found medical applications²¹ and, for example, cold plasma from argon or oxygen has been used to irradiate polymer coats of tablets and granules.^{22,23} This plasma irradiation leads to the emission of intensive UV radiation, which can lead to changes in the polymer structure, influencing, for example, drug release kinetics from these tablets or creating gastro-retentive, floating drug delivery systems.²⁴

REFERENCES

- Laylin LK. *Nobel Laureates in Chemistry 1901–1992*. Darby: Diane Publishing Co., 1993.
- Page AJ, Wilkinson G. The polarographic chemistry of ferrocene, ruthenocene and the metal hydrocarbon ions. *J Am Chem Soc* 1952; 74: 6149–50.
- Rausch M, Fischer EO, Grubert H. The aromatic reactivity of ferrocene, ruthenocene, and osmocene. *J Am Chem Soc* 1960; 82: 76–82.
- Kealy TJ, Pauson PL. A new type of organo-iron compound. *Nature* 1951; 168: 1039.
- Beer PD, Hayes EJ. Transition metal and organometallic anion complexation agents. *Coord Chem Rev* 2003; 240: 167–89.
- Moss GP. Basic terminology of stereochemistry (IUPAC Recommendations 1996). *Pure Appl Chem* 1996; 68: 2193–222.
- Foresman JB, Frisch A. *Exploring Chemistry with Electronic Structure Methods*, 2nd edn. Pittsburgh: Gaussian Inc., 1993.
- Hehre WJ. *A Guide to Molecular Mechanics and Quantum Chemical Calculations*. Irvine: Wavefunction, Inc., 2003.
- Wang SL *et al.* Solid-state trans-cis isomerization of captopril determined by thermal Fourier transform infrared (FT-IR) microspectroscopy. *J Pharm Sci* 2001; 90: 1034–9.
- Zeitler JA *et al.* Terahertz pulsed spectroscopy and imaging in the pharmaceutical setting—a review. *J Pharm Pharmacol* 2007; 59: 209–23.
- Chieng N *et al.* Effect of milling conditions on the solid-state conversion of rantidine hydrochloride form 1. *Int J Pharm* 2006; 327: 36–44.
- Murdande SB *et al.* Aqueous solubility of crystalline and amorphous drugs: Challenges in measurement. *Pharm Dev Technol* 2011; 16: 187–200.
- Herman J, Visavarungroj N, Remon JP. Instability of drug release from anhydrous theophylline-microcrystalline cellulose formulations. *Int J Pharm* 1989; 55: 143–6.
- Kahela P *et al.* Pharmacokinetics and dissolution of two crystalline forms of carbamazepine. *Int J Pharm* 1983; 14: 103–112.
- Vishweshwar P *et al.* Pharmaceutical co-crystals. *J Pharm Sci* 2006; 95: 499–516.
- Porter W, W. Elie S, Matzger AJ. Polymorphism in carbamazepine cocrystals. *Cryst Growth Des* 2008; 8: 14–16.
- Coleman NJ, Craig DQM. Modulated temperature differential scanning calorimetry: A novel approach to pharmaceutical thermal analysis. *Int J Pharm* 1996; 135: 13–29.
- Bunjes H, Rades T. Thermotropic liquid crystalline drugs. *J Pharm Pharmacol* 2005; 57: 807–16.
- Rades T, Muller-Goymann CC. Melting behaviour and thermotropic mesomorphism of fenopropen salts. *Europ J Pharm Biopharm* 1994; 40: 277–82.
- Six K *et al.* Investigation of thermal properties of glassy itraconazole: identification of a monotropic mesophase. *Thermochim Acta* 2001; 376: 175–81.
- Fridman G *et al.* Applied Plasma Medicine. *Plasma Processes and Polymers* 2008; 5: 503–33.
- Ishikawa M *et al.* A new drug delivery system using plasma-irradiated pharmaceutical aids. V. Controlled release of theophylline from plasma-irradiated double-compressed tablet composed of a wall material containing polybenzylmethacrylate. *Chem Pharm Bull* 1995; 43: 2215–20.
- Ishikawa M *et al.* A new drug delivery system using plasma-irradiated pharmaceutical aids. VII. Control release of theophylline from plasma-irradiated polymer-coated granules. *Chem Pharm Bull* 1996; 44: 1232–7.
- Kondo S *et al.* Preparation of floating drug delivery system by pulsed-plasma techniques. *J Photopolym Sci Technol* 2004; 17: 149–52.

BIBLIOGRAPHY

- Alberty RA, Silbey RJ, Bawendi MG. *Physical Chemistry*, 4th edn. Indianapolis: John Wiley & Sons, Inc., 2004.
- Atkins P, de Paula J. *Physical Chemistry*, 8th edn. New York: W.H. Freeman, 2006.
- Djerassi C. *Optical Rotary Dispersion*. New York: McGraw-Hill, 1960.
- Engel T, Reid P. *Physical Chemistry*, 2nd edn. Upper Saddle River: Pearson Education, 2010.
- Fessenden RJ, Fessenden JS. *Organic Chemistry*, 4th edn. California: Brooks/Cole, 1990.
- Hilfiker R ed. *Polymorphism in the Pharmaceutical Industry*. Weinheim: Wiley-VCH Verlag, 2006.
- Hillert M. *Phase Equilibria, Phase Diagrams, and Phase Transformations: Their Thermodynamic basis*, 2nd edn. New York: Cambridge University Press, 2008.
- Levine H ed. *Amorphous Food and Pharmaceutical Systems*. Cambridge: The Royal Society of Chemistry, 2002.
- Levine IN. *Physical Chemistry*, 6th edn. New York: McGraw Hill, 2009.
- Mortimer RG. *Physical Chemistry*, 3rd edn. Burlington: Elsevier Academic Press, 2008.
- O'Connell JP, Haile JM. *Thermodynamics, Fundamentals for Application*. New York: Cambridge University Press, 2005.
- Pauling L. *The Nature of the Chemical Bond*, 3rd edn. Ithaca: Cornell University Press, 1960.
- Poling BE, Prausnitz JM, O'Connell JP. *The Properties of Gases and Liquids*. 5th edn. New York: McGraw-Hill, 2001.
- Turns SR. *Thermodynamics, Concepts and Applications*. New York: Cambridge University Press, 2006.
- Ymén I. Introduction to the solid state – Physicals properties and processes. In: Storey RA, Ymén I, eds. *Solid State Characterization of Pharmaceuticals*. West Sussex: Blackwell Publishing, 2011; pp. 1–31.
- Zallen R. *The Physics of Amorphous Solids*. Weinheim: Wiley-VCH Verlag, 2005.

Chapter 12

Thermodynamics

Timothy S. Wiedmann, PhD

INTRODUCTION 207

THE FIRST LAW 207

WORK AND HEAT 208

ENTROPY AND THE SECOND LAW 211

THE THIRD LAW 212

FREE ENERGY 213

STANDARD MOLAR GIBBS FREE ENERGY 213

NON-IDEALITY 213

EQUILIBRIA 214

SOLUBILITY 215

HEAT OF SOLUTION 215

TEMPERATURE DEPENDENCE OF EQUILIBRIUM
CONSTANTS 215

PARTITIONING BEHAVIOR 216

PROTEIN BINDING 216

INTRODUCTION

Thermodynamics rests upon three basic laws that took over 500 years to establish. Although quantum mechanics has defined the limits of its scope, the concepts embodied in the three laws have remained unchanged for over a century. Thus, thermodynamics provides unassailable, certain answers. As such, the great value, but also the main challenge, is using the simple concepts to understand complicated physical and biological phenomena. There are many texts available on the subject of thermodynamics ranging from introductory to complex, rigorous treatments. The approach in this chapter will involve the development of the concepts with pharmaceutically relevant examples. The purpose is to demonstrate how to make judicious assumptions to allow use of the rigorous definitions in thermodynamics to reveal the certainty in the complex field of pharmaceutics.

To begin, the definitions of system, surroundings, universe, and boundaries are introduced. These, as well as all other definitions, must be clearly understood, otherwise thermodynamics will not be properly used. A *system* is that part of the *universe* under consideration and, as such, is separated from the *surroundings* or, equivalently, the rest of the universe. The focus of the analysis will center on how the properties of a system are altered through an interaction with the surroundings. Properties are those qualitative characteristics (e.g., type of phase, which may be solid, liquid, or gas) and quantitative characteristics (e.g., temperature, pressure) that describe the system. The interaction between the system and surroundings is controlled by the boundary and is revealed in the change in properties.

When a sufficient number of properties of the system have been specified with fixed values, then the system is at equilibrium. Certain systems at equilibrium have a simple equation that provides a relationship among the values of the properties. For example, a system containing an ideal gas (one component) has the properties of pressure, P , volume, V , number of mols, n and temperature, T (K), related by:

$$PV = nRT \quad (1)$$

where R is the gas law constant. Such a relationship is referred to as an *equation of state* because it specifies the relationship among the quantitative properties of a system in a definite state. Furthermore, if the system is at equilibrium, only three properties need to be specified, as the fourth may be calculated from the equation of state. Most pharmaceutically relevant systems are extremely complicated, because there are typically many components; the properties of each must be specified for the system to be at equilibrium.

THE FIRST LAW

The first law of thermodynamics is a statement of the principle of conservation of energy; energy may neither be created nor be destroyed. It is mathematically written as:

$$dE = \delta q - \delta W \quad (2)$$

where dE is the differential change in the internal energy, δq is the differential change in the absorbed heat, and δW is the differential change in the expended work or work done.

The change in internal energy of a system in going from state A to state B is given by:

$$\Delta E = \int_A^B dE \quad (3)$$

The internal energy is a state function. A change in the value of a state function depends only on the initial and final state and does not depend on how the change in state was achieved. The symbol, d , is used and represents an exact differential. Another example of a state property is the temperature; that is, the change in temperature only depends on the initial and final temperatures and does not depend on how the temperature was changed (that is, the path followed). A cyclic change in a system describes a series of changes where the initial and final states have the same values for the state properties; one may say, the system “comes to a full circle.” For all cyclic processes, the change in all state properties is zero. Finally, the first law provides only a relation for the change in the internal energy and does not provide an absolute value. Thus like temperature, a zero point must be defined as well as a scale for a unit of energy, which were arbitrary until the scientific community reached a consensus.

Although it seems too obvious to point out that if the system has the same value of energy in the initial and final state that the change in energy would be zero, there are quantities that may not be zero even when a system undergoes a cyclic change. Two examples are heat flow and work done, which in the first law are described by inexact differentials, and the unique symbol, δ , is used. Neither the heat nor the work is a state function and thus, the integral of the differential depends on the path taken. Mathematically, they are written as:

$$q = \int_A^B \delta q \quad (4)$$

$$W = \int_A^B \delta W \quad (5)$$

The work and heat may only be determined if more information is provided concerning how the change in the state of the system was achieved.

WORK AND HEAT

The concept of *work* in thermodynamics is expressed as a product of an *intensity factor* and a *capacity factor*; for example, mechanical work is given as:

$$\delta W = F dl \quad (6)$$

where the differential quantity of work done, δW , is the product of the force, F (intensity factor), and a differential change in distance, dl (capacity factor). Other types of work encountered in introductory physics include *gravitational* (gravitational potential and mass), *electrical* (potential difference and quantity of charge), and *volume expansion* or *PV work* (pressure and volume) Table 12.1. In pharmaceuticals, *surface work* (intensity and capacity factors are surface tension and area) is fundamental to understanding the instability of dispersed dosage forms. Chemical reactions can also be cast into the formalism of thermodynamic work as in the production of energy (work) in the conversion of ATP to ADP where the intensity factor is the chemical potential and the capacity factor is the number of moles. The essence of thermodynamics lies in understanding the conversion of one type of energy into another. Thus, in the following practical examples, the formalism is developed to reveal how these conversions can be described.

Some peculiarities of the work are that it appears only at the boundary and thus may be thought of as flowing into or out of the system. Work may be generated only through a change in the state of the system, and it is an algebraic quantity that may be positive or negative with units of energy, Kgm^2/s^2 . The convention chosen here is that if the system does work on the surroundings, the work is a positive quantity; conversely, if the surroundings does work on the system, the work is a negative quantity.

The other quantity appearing in the first law is *heat*. It shares many properties with work. Specifically, it also appears at the boundary and only with a change in the state of the system. By definition, heat flow into the system is taken to be a positive quantity, which results in an increase in the internal energy of the system. For a system where there is no work done, $\delta W = 0$, the change in internal energy is given by the heat flow, so q often is referred to as a transfer of thermal energy.

Before providing the relationship among internal energy, heat flow, and temperature, the above definitions will be used in some examples of PV work associated with mechanical ventilation of the lung. It is important to note that the following represents one model constructs. Many others can be developed to describe breathing, where different assumptions are made. Because each construct involves assumptions, none is strictly correct. Nevertheless, the approach does provide insight and mechanistic understanding of the process of breathing.

In Figure 12-1a, a system (shown with dotted background) is shown as two connected chambers, each with a movable boundary (double lines) and a rigid boundary (single line). Everything else is in the surroundings. One side of the system has a spring, which represents the lung, and the other side represents air external to the lung but still is part of the system. In the initial state, the spring has no tension. In mechanical ventilation, a positive pressure is applied to inflate the lung, which

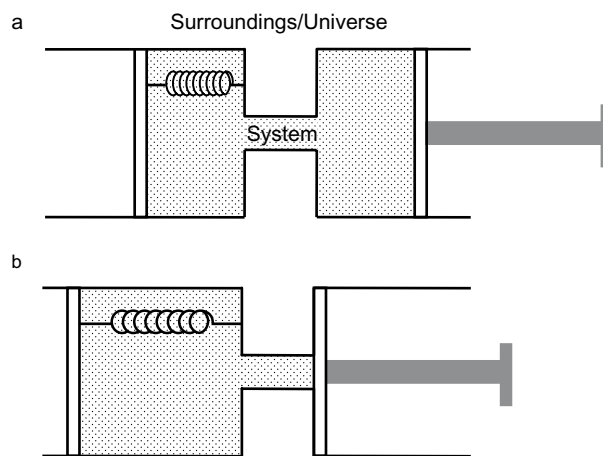


Figure 12-1. Schematic representation of the lung in the (a) initial state and (b) final state. The dotted region represents the system, which is separated from the surroundings by the boundary.

can be thought as a depression of the piston. As the piston is depressed, the pressure rises and gas will flow from the right to the left.

More information about the type of boundary must be provided. The pistons are allowed to move and thus that portion of the boundary is *movable*; other portions are *rigid*. For the first case, the boundary will be stipulated to be *diathermal*, which allows the passage of heat, and in the later case, it will be *adiabatic*, thereby preventing heat flow. Finally, the boundary is *closed* rather than *open*, so no mass transfer is allowed.

From the literature, the residual volume of an adult lung (i.e. volume without contraction of the diaphragm) is about 500 ml (0.5 L or 0.0005 m^3) at a temperature of 37°C (310 K) and 1 atmosphere pressure ($1 \text{ atm} = \text{ca } 1.0 \times 10^5 \text{ Newtons/m}^2$, N/m^2 or Pascal, Pa). The lung on the left will be inflated with the 0.0005 m^3 air from the right hand side, also assumed to be at 310 K. Thus in the system, the initial total volume will be 0.001 m^3 of air, which can be taken to be an ideal gas. This allows the use of the equation of state to calculate the total number of moles of gas. That is:

$$n = PV/RT = [(1.0 \times 10^5 \text{ N/m}^2)(0.001 \text{ m}^3)] / [(8.314 \text{ J/molK})(310 \text{ K})] \\ n = 0.0388 \text{ moles} \quad (7)$$

In Figure 12-1b, the system is shown in the final equilibrium state. In the final state, the movement of the gas from the right chamber into the ‘lung’ caused the spring to stretch. In fact, with mechanical ventilation, a positive pressure is applied to inflate the lung, which is typically about 35 cm of water. The latter is a peculiar unit for pressure, because it seems to have units of ‘length’. However, a column of water with a height of 35 cm exerts a pressure, which is calculated as:

$$\Delta P = \text{density} \times \text{acceleration constant} \times \text{height of liquid} \\ = (1000 \text{ kg/m}^3)(9.8 \text{ m/s}^2)(0.35 \text{ m}) \\ = 3430 \text{ N/m}^2 \quad (8)$$

Table 12-1. Types of Work and the Associated Intensity and Capacity Factors.

Work	Intensity factor	Capacity factor	Typical units	SI units
Mechanical	Force	Distance	Newton-meter	Kgm^2/s^2
Gravitation	Acceleration	Mass	$(\text{Meter}^2/\text{s}^2)$ -kilogram	Kgm^2/s^2
Electrical	Electrical potential	Charge	Volt-Coulomb	$\text{VC} = \text{Kgm}^2/\text{s}^2$
Pressure-volume	Pressure	Volume	$(\text{Newton}/\text{meter}^2)$ -meter ² Also Pa-m ³	Kgm^2/s^2
Surface work	Surface tension	Area	Newton/m-meter ²	Kgm^2/s^2
Chemical work	Chemical potential	Moles	$(\text{Joules}/\text{moles})$ -moles	

An initial state and final state have been given, which allows the change in state to be determined. It perhaps is worth emphasizing that thermodynamics provides only information on the change between the initial and final state. The initial pressure was $1.0 \times 10^5 \text{ N/m}^2$, and the final pressure is higher at $1.034 \times 10^5 \text{ N/m}^2$. Because the pressure is higher and the temperature is assumed constant (isothermal), there must have been a reduction in volume. That is:

$$W = \int PdV \quad (9)$$

This type of change in state is described as an isothermal compression of an ideal gas. The work can be obtained using the following equation:

$$W = \int PdV \quad (10)$$

However, to calculate the work, more information is needed as to how the lung was inflated; really how the gas was compressed. In essence, the path must be given. An unrealistic, but simple path would be to assume that the gas was instantaneously compressed to the final pressure of $1.034 \times 10^5 \text{ Pa}$, which was maintained constant. The work done by compressing the gas with a constant pressure can be calculated as follows, where the pressure is moved in front of the integral:

$$W = P_{\text{ext}} \int dV = P_{\text{ext}} (V_f - V_i) = (1.034 \times 10^5 \text{ N/m}^2)(-0.000034 \text{ m}^3) = -3.41 \text{ J} \quad (11)$$

The negative quantity reflects the fact that the surroundings did work on the system.

Assuming a different path, which represents perhaps a better assumption, the pressure is initially at $1.0 \times 10^5 \text{ N/m}^2$, but the applied pressure increases continuously as the compression occurs. To integrate the PdV term in the expression for the work, the pressure, which now changes, can be replaced using the equation of state, $P = nRT/V$. The expression becomes:

$$W = \int (nRT/V) dV \quad (12)$$

We have stipulated that the temperature is constant, and so n , R , and T can be moved in front of the integral:

$$\begin{aligned} W &= nRT \int (dV/V) = nRT \ln(V_f/V_i) \\ &= (0.0388 \text{ mol})(8.314 \text{ J/molK})(310 \text{ K}) \ln(0.000967/0.001) \\ &= -3.36 \text{ J} \end{aligned} \quad (13)$$

First, it should be noted that the changes in state properties, which are exact differentials, were used to calculate the work. Second, the changes in the state properties for the two processes are identical-i.e. independent of path. Third, the amount of work required to compress the system differs and therefore is seen to be dependent on how the compression was carried out.

By virtue of the definition of an ideal gas, the internal energy is only a function of temperature. Because our system was isothermal with no change in temperature, the internal energy of the system is constant, i.e., $\Delta E = 0$, and therefore from the first law:

$$q = W \quad (14)$$

Therefore, during the isothermal compression of an ideal gas with a constant pressure, the heat was -3.41 J , which as a negative quantity represents the amount lost to the surroundings. Compression with a continually rising pressure, the heat was -3.36 J , representing a smaller absolute value of heat loss to the surroundings. The latter case is actually unique in that it represents the minimum heat and work required to achieve the change in state. Moreover, we could return the system to the original state by following the same path in the opposite direction. That is, an isothermal expansion resulting in 3.36 J of work done by the system and 3.36 J of heat flow into the system. This unique type of change in the system is referred to as being "reversible". All other paths are denoted as irreversible. It should also be noted that all changes in the real world are "irreversible", whereas many of the calculations of thermodynamics involve the assumption of reversibility. As such, thermodynamics

often provides the theoretical maximum, which can be used to limit practical possibilities.

The spring was placed in our diagram to represent the stretching of lung tissue. From this construct, the elasticity of the tissue can be assessed. The work in stretching a spring is given by the product of an intensity factor (spring constant, k) and a capacity factor (length, l):

$$W_{\text{El}} = \int kdl \quad (15)$$

where the elastic work will equal the PV work, if this is the only biologically relevant aspect of lung expansion (it is not), then,

$$W_{\text{El}} = W_{\text{PV}} = -3.36 \text{ J} \quad (16)$$

Because alveoli double in volume with lung expansion and a typical size of alveoli is $300 \mu\text{m}$ or $3 \times 10^{-4} \text{ m}$ (relative length change, $\Delta l/l = 0.5$), Young's modulus, Y , of the lung can be estimated to be:

$$Y = W_{\text{El}}/(\Delta l/l) = |-3.36|/0.5 = 6.72 \text{ Nm} \quad (17)$$

This is equivalent to a spring that stretches 1 meter when adding a 1 kg weight, $(1 \text{ kg})(9.8 \text{ m/s}^2)(1 \text{ m}) = 9.8 \text{ Nm}$. It should be not lost to the reader that a quantitative analysis has just been provided for the conversion of one form of work or energy (PV) to another form of work (elastic), which is can be generalized to any process of energy conversion.

To demonstrate the influence of other types of barriers, consider the process of compression with an adiabatic barrier that does not allow the flow of heat, $q = 0$. In this case, the work done causes an increase in the internal energy. For compression against a constant pressure, the work is still -3.41 J , but the change in internal energy must now be calculated as:

$$\Delta E = \int dE = -W = 3.41 \text{ J} \quad (18)$$

Therefore, work was done on the system resulting in an increase in internal energy of 3.41 J . This increase in internal energy will result in a change in temperature and/or volume.

The relationship between internal energy and temperature is expressed by the heat capacity. The value of the heat capacity depends on the path, that is, whether the temperature change occurred under constant volume or constant pressure. For heat capacity at constant volume, C_v , the definition is:

$$C_v \equiv dq_v / dT \quad (19)$$

Rearranging:

$$\int dq_v = \int C_v dT \quad (20)$$

and integrating with the assumption that the heat capacity is constant over the range of ΔT :

$$q_v = C_v \Delta T \quad (21)$$

However, the heat flow at constant volume will equal the change in internal energy, because the work is zero when only PV work is allowed.

$$\Delta E = q_v + \int PdV = q_v = C_v \Delta T \quad (22)$$

In our example, the change in temperature may be calculated, since the value of C_v for a monatomic ideal gas is $(3/2)R$, where $1/2R$ is allotted to each mode of translation motion in the x , y , and z direction. Because air is composed of nitrogen and oxygen, which are diatomic molecules, a better estimate of C_v would be $(5/2)R$, which provides an additional $1/2R$ to the two possible modes of rotational motion. The actual value of the heat capacity can also be found on the internet, 20.76 J/Kmol , which is very close to $(5/2)R$. Thus, the temperature change is calculated by:

$$\Delta T = \Delta E/nC_v = (3.41 \text{ J})/[(0.0388 \text{ mol})(20.76 \text{ J/Kmol})] = 4.23 \text{ K} \quad (23)$$

With the number of moles and final pressure and temperature, the volume may be calculated, 0.00098 m^3 , which is larger than that obtained for the isothermal compression, but consistent with the higher temperature.

The variables in the latter two terms can be separated and then integrated as:

$$\int (C_v/T) dT = -\int (nR/V) dV \quad (36)$$

To yield:

$$C_v \cdot \ln(T_2/T_1) = -nR \cdot \ln(V_2/V_1) \quad (37)$$

Simplifying:

$$(T_2/T_1)^{(C_v/nR)} = V_1/V_2 \quad (38)$$

With $C_v = 5nR/2$ and $V_2 = 0.00096 \text{ m}^3$.

We note that $\Delta E = \int nC_v dT = nC_v \Delta T = (0.04)(5/2)R^*(305-300) = 4.16 \text{ J}$, which is also negative of the work. Finally, we note that ΔH must be zero, because there is no heat flow.

Step 2: Isobaric expansion to a final temperature of 310. With the pressure constant, then:

$$T_3/T_2 = V_3/V_2 \quad (39)$$

Solving for V_3 :

$$(310/305) \cdot 0.00096 = 0.000976 \text{ m}^3 \quad (40)$$

Work is straightforward with the pressure constant:

$$P_2 = P_3 = nRT/V = 1.057 \times 10^5 \text{ Pa}$$

$$W = P \int dV = P \Delta V = (1.057 \times 10^5) (0.000976 - 0.00096) = 1.69 \text{ J} \quad (41)$$

The change in internal energy, ΔE :

$$\Delta E = C_v \cdot \Delta T = 0.04 \cdot (5/2)R^*(310 - 305) = 4.16 \text{ J} \quad (42)$$

The heat flow may be calculated from the work and change in internal energy:

$$q = \Delta E + W = 4.16 + 1.69 = 5.85 \text{ J} \quad (43)$$

The change in enthalpy is equal to the heat flow at constant pressure, thus:

$$\Delta H = qp = 5.85 \text{ J}$$

$$\text{But } \Delta H = \Delta E + \Delta(PV) = \Delta E + \Delta(nRT) = 4.16 + (0.04 \cdot 8.314 \cdot 5) = 5.85 \text{ J} \quad (44)$$

Step 3: Adiabatic expansion to a final temperature of 305 K, which much like step 1.

$$\Delta E = \int C_v dT = (0.04)(5/2)R(305 - 310) = -4.16 \text{ J} \quad (45)$$

At zero heat flow ($\Delta H=0$):

$$W = -\Delta E = 4.16 \text{ J} \quad (46)$$

Step 4: Isobaric compression, which is much like step 2.

$$W = P \int dV = (1.0 \times 10^5) \cdot (0.001 - 0.001014) = -1.4 \text{ J} \quad (47)$$

All of the values are given in Table 12-2 and Table 12-3.

In the Table 12-3 final example, the thermodynamic approach that led to the discovery of lung surfactant is given. In 1972, Clements and co-workers reported the following experiment. An experimental animal was anesthetized and the expansion of the lung was measured as a function of applied pressure. The lungs were then rinsed with water to remove a frothy substance at the surface, and the measurement was repeated. It was noted that the work, which is associated with lung expansion:

$$W = \int PdV \quad (48)$$

Table 12-2. Values of the Pressure, Temperature, and Volume for the Cyclic Change Described in Text and Schematically Illustrated in Figure 12-3.

Point	A	B	C	D
P, Pa	1.0×10^5	1.057×10^5	1.057×10^5	1.0×10^5
V, m ³	0.001	0.00096	0.000975	0.001014
T, K	300	305	310	305

Table 12-3. Values for the Work and Heat and the Change in Energy and Enthalpy for Each Step for the Cyclic Change Described in Text and Illustrated in Figure 12-3.

Step	1	2	3	4	Total
W, J	-4.16	1.69	4.16	-1.4	0.29
Q, J	0	5.85	0	-5.85	3.04
ΔE , J	4.16	4.16	-4.16	-4.16	0
ΔH , J	0	5.85	0	-2.74	0

as revealed in the area under the curve of a plot of pressure as a function of volume, was significantly greater in the lung that was rinsed relative to the control. It was argued that rinsing the lung would not be expected to affect tissue elasticity, and thus, an additional contribution to the work of expansion must exist. Moreover, they noted that the frothy material extracted from the lung caused a reduction in the surface tension at the air/water interface. They proposed that the total work of lung expansion involves a contribution from elasticity and from surface work:

$$W_{PV} = W_{El} + W_{sur} \quad (49)$$

Where:

$$W_{sur} = \int \gamma dA \quad (50)$$

Additional work revealed that the material responsible for lowering the surface tension was a mixture of lipids and proteins, which was called, lung surfactant. The production of lung surfactant occurred late in the pregnancy, because it is not required until birth, when the infant must inflate the lungs to replace the amniotic fluid with air. The conclusion was that prematurely born infants had insufficient lung surfactant, which could be treated by providing a replacement. In the early 1990's, the first replacement therapy was introduced. Moreover, the improvement in premature infant survival that has since been realized can be attributed in part to this discovery and subsequent development and administration of these products.

ENTROPY AND THE SECOND LAW

Although the first law provides the framework for calculating the change in energy associated with chemical reactions or physical changes in state, there is insufficient information to allow prediction of the likelihood of whether the change will occur. Consider a system composed of two parts that are at different temperatures, T_1 and T_2 , separated by an impermeable, adiabatic partition. When the partition is removed, heat will flow from the part at a higher temperature to the part at a lower temperature. According to the first law, the energy of the whole system, the sum of parts one and two, has not changed.

Intuitively it is known that the above change will occur regardless of the fact that the first law does not provide a method of predicting the occurrence. Such changes are described as *spontaneous*, for the obvious reason that they occur without additional stimulation. It should be noted that this spontaneous change involved an increase in the disorder or, if you will, the randomness of the system. Thus, the system initially was separated into two parts at different temperatures, but after thermal contact, a uniform temperature was reached. The *entropy*, S , is the function that provides a quantitative description of the randomness or disorder of the system and is fundamental for predicting the spontaneity of chemical reactions and physical changes. The entropy is a state function that depends only on the initial and final state of the system.

The definition of the entropy change is given by the seemingly surprising form:

$$dS = \delta q_{\text{rev}} / T \quad (51)$$

where the subscript *rev* denotes that the heat flow occurs in a reversible manner. By carrying out the integration, the change in entropy for a reversible, isothermal change from state 1 to state 2 is given by:

$$\Delta S = \int \delta q_{\text{rev}} / T = q_{\text{rev}} / T \quad (52)$$

With the introduction of entropy, the second law may be stated as follows: For any spontaneous process in an isolated system, there is an increase in the value of entropy. Alternatively, the first and second laws may be combined with the classic thermodynamic statement, “the energy of the universe is constant; the entropy is increasing.”

Historically, Carnot analyzed the efficiency of such a theoretical cycle, which has come to be known as the Carnot cycle. This can be found in most texts on thermodynamics, although it typically involves isothermal and adiabatic processes in the context of what is known as a heat engine. Nevertheless, the premise is that by permitting the flow of heat into a system, work may be done by the system. The second law dictates that not all of the heat may be converted into work, even if all changes occur in a reversible manner. In fact, the maximum work, W_{max} , that may be obtained is specified by the heat flow into the system and the temperature difference over which the heat engine is operating; that is:

$$W_{\text{max}} = q_1(T_1 - T_2) / T_1 \quad (53)$$

where $T_1 > T_2$. The proof of this equation can be established by calculating the heat and work for each step and noting that not all of the heat energy may be converted into work. This can be seen from Table 12-3. The heat flow into the system occurs with the second step. The efficiency of an engine ϵ , is given by the amount of work extracted divided by the heat flow into the system:

$$\epsilon = W_{\text{tot}} / q_1 \quad (54)$$

where $W_{\text{tot}} = W_{\text{max}}$ for reversible changes. The efficiency is proportional to the difference in temperature between the heat reservoirs. In addition, there is no work done unless there is a difference in temperature between the heat reservoirs. Finally, 100% efficiency is obtained only when $T_2 \rightarrow 0$, which as will be noted from the third law, is impossible.

The above analysis elucidates the connection between entropy and heat flow. The second law provides the quantitative limit on the amount of work that can be done with a cyclic operation performed in a reversible manner.

Given the above background, the entropy changes associated with several reversible processes may be determined. Consider the vaporization of water in the context of the lung, which is known to carry out the important function of humidifying air. Let us assume that initially that the inhaled air contains no water vapor, but is at 310 K, and the final state is air saturated with water at 310 K. To calculate the amount of water that must be evaporated to saturate the air, the saturation vapor pressure of water in air can be found. The value from the internet is 47 mm Hg. Using the approach from above, the moles of water that must undergo the phase change from a liquid at 310 to a vapor at the same temperature is 37 mm Hg or (37 * 133 Pa/mm Hg =) 4921 Pa. The number of moles is found with the ideal gas law:

$$n/V = P/RT = (4921)/(8.314*310) = 1.91 \text{ moles/m}^3 \quad (55)$$

or for a typical inhalation volume of 0.001 m³, 1.91 × 10⁻³ moles.

The entropy of vaporization, ΔS_{vap} , is given by:

$$\int dS = \int \delta q / T_b \quad (56)$$

which may be determined from $\int \delta q = q_p = \Delta H$, but this is the value at the boiling point. We will return to this point below, but for now, consider the problem as occurring at the boiling point.

The molar heat of vaporization of water at the boiling point is 40.65 kJ/mol. Thus:

$$\begin{aligned} \int dS &= \Delta H_{\text{vap}} / T_m \\ \Delta S &= 40650/373 = 109 \text{ J/mol K} \end{aligned} \quad (57)$$

In our problem, the number of moles was 1.91 × 10⁻³ leading to the total entropy change of 0.208 J/K. The positive change in entropy confirms the intuition concerning such an event; gases are in a state of more disorder than liquids, thus the entropy also is greater in the vapor state in comparison to the liquid state.

The discussion, thus far, has been limited to reversible processes that are strictly impossible to achieve in the laboratory (even though such a process may be approximated very closely). The question arises as to what the entropy change is for an irreversible change. Here the entropy change is given by:

$$dS > \delta q_{\text{irr}} / T \quad (58)$$

Thus, all real processes may be written as:

$$dS \geq \delta q / T \quad (59)$$

This concept may be extended to determine the condition of spontaneity. Consider a system that is transformed irreversibly from state 1 to state 2 and then reversibly from state 2 back to state 1. The overall change is given by:

$$\int_{\text{State 1}}^{\text{State 2}} \delta q_{\text{irr}} / T + \int_{\text{State 2}}^{\text{State 1}} \delta q_{\text{rev}} / T < 0 \quad (60)$$

$$\int_{\text{State 1}}^{\text{State 2}} \delta q_{\text{irr}} / T + \int_{\text{State 1}}^{\text{State 2}} dS < 0 \quad (61)$$

which may be rearranged, with changing the limits of integration, to yield:

$$\int_{\text{State 1}}^{\text{State 2}} \delta q_{\text{irr}} / T < \int_{\text{State 1}}^{\text{State 2}} dS \quad (62)$$

or, equivalently, for infinitesimal changes:

$$\delta q_{\text{irr}} / T < dS \quad (63)$$

This is known as the *Clausius inequality*. For isolated systems, where boundaries do not permit the passage of energy or matter, $\delta q_{\text{irr}} = 0$, the result is given:

$$dS > 0 \quad (64)$$

That is, for every spontaneous change in an isolated system there is an increase in the entropy.

The second law may be generalized in another way. The total entropy for any process is given by the sum of the entropy of the system and the surroundings; that is:

$$dS_{\text{tot}} = dS_{\text{sys}} + dS_{\text{surr}} \quad (65)$$

For reversible processes, the entropy change in a system is the negative of the entropy change produced in the surroundings. The total entropy, therefore, is zero. For irreversible processes, the total entropy, system plus surroundings, increases. The mathematical statement of this relationship is:

$$\sum \Delta S_{\text{tot}} = 0 \text{ reversible process} \quad (66)$$

$$\sum \Delta S_{\text{tot}} > 0 \text{ irreversible process} \quad (67)$$

THE THIRD LAW

The third law of thermodynamics simply defines the zero point of the entropy scale. The entropy of a pure, perfectly crystalline substance is zero at absolute zero. Intuitively, a system that has perfect three-dimensional order should have no entropy at the lowest possible temperature. The defining of a zero for the entropy is unlike the other state functions introduced previously. Thus, the value of the entropy, S , of a system in any state, in principle, may be calculated.

What would be the entropy of a crystalline solid at 150 K, S_{150} ? This may be calculated as:

$$\Delta S = S_{150} - S_0 \quad (68)$$

$$\Delta S = \int (c_p / T) dT - 0 \quad (69)$$

or:

$$\Delta S = S_{150} \quad (70)$$

If the heat capacity over the range of 0 to 150 K is known, the value of the entropy may be calculated.

FREE ENERGY

The concept of *free energy* is probably the most useful aspect of thermodynamics. The criteria for determining the spontaneity of a chemical reaction or phase change were presented above; however, it involved carrying out the change in an isolated system. One can imagine how inconvenient and often impossible it would be to apply such a constraint to the laboratory setting. For this sake, additional state functions have been defined to allow prediction of the spontaneity of a change in state. The rationale for the development of other functions was to allow maximum flexibility in their application. The two functions introduced are *Helmholtz free energy*, A , and *Gibbs free energy*, G . The functions for predicting spontaneity are:

1. Isolated system: $dS > 0$
2. Isothermal and isochoric system: $dA < 0$
3. Isothermal and isobaric: $dG < 0$
4. Constant volume and entropy: $dE < 0$

Helmholtz free energy is defined as:

$$A \equiv E - TS \quad (71)$$

Helmholtz free energy is the energy available to do pressure-volume work for reversible isothermal processes; a decrease in the Helmholtz free energy is equal to the capacity of the system to do work. An alternative view is that, for systems at constant volume and temperature, a change in state is spontaneous if, and only if, there is a decrease in the Helmholtz free energy. Thus, with the introduction of ΔA , the spontaneity of changes occurring at constant volume and temperature may be predicted.

As most reactions carried out in the laboratory are under conditions of constant pressure and temperature, Gibbs free energy is the most useful function and is defined as:

$$G \equiv E + PV - TS \quad (72)$$

which can be converted into a more usable form by an analogous method used with the Helmholtz function. Taking the differential and applying the constraints of constant pressure and temperature yields:

$$dG = dE + PdV - TdS \quad (73)$$

but $dE = \delta q - \delta W = TdS - \delta W$; thus, upon substitution:

$$-dG = \delta W - PdV \quad (74)$$

A decrease in Gibbs free energy is equal to the non-PV work done by the system or, equivalently:

$$dG = -\delta W_{(\text{non-PV})} \quad (75)$$

which also provides the conditions of a spontaneous change under the constraints of constant temperature and pressure. A direct application of the relationship between Gibbs free energy and non-PV work is used in potentiometry.

These relationships for predicting spontaneity often are expressed in a differential form, which presents the state functions in a concise manner as well as facilitating their use to specific problems. The four differential equations are:

$$dE = TdS - PdV \quad (76)$$

$$dH = TdS + VdP \quad (77)$$

$$dA = -SdT - PdV \quad (78)$$

$$dG = -SdT + VdP \quad (79)$$

These expressions represent the four fundamental equations of thermodynamics, which in reality are four ways of looking at one fundamental equation describing the conditions of spontaneity.

Q—One mol of liquid water is vaporized reversibly at 100° and 1 atm pressure. The molar heat of vaporization is 40.65 kJ/mol; what are qp , ΔH , ΔE , ΔA , ΔG , and ΔS ?

A—The value of qp actually is given in the question, as the heat required to vaporize 1 mol of liquid is the definition of the molar heat of vaporization; thus, $qp_3 = 40.65$ kJ. Recognizing that the pressure is constant, $\Delta H = qp = 40.65$ kJ. To calculate ΔE , the work first must be determined. The work is given by:

$$W = \int PdV = P\Delta V \quad (80)$$

$$W = P(V_g - V_l)$$

However, the volume of the gas, V_g , is much larger than the volume of the liquid, V_l , which implies that the work is given by:

$$W \approx PV_g \quad (81)$$

Assuming the gas is ideal, the work is:

$$W = nRT = (1 \text{ mol})(8.314 \text{ J/mol K})(373 \text{ K}) = 3100 \text{ J} \quad (82)$$

From the above, ΔE may be calculated from:

$$\Delta E = q - w = 40650 - 3100 = 37550 \text{ J} \quad (83)$$

The change in entropy was determined above:

$$\Delta S = \Delta H / T_m = 40650 / 373 = 109 \text{ J/K} \quad (84)$$

Helmholtz free energy is given by:

$$\Delta A = \Delta E - T\Delta S = 40650 - (373)(109.0) = -3100 \text{ J} \quad (85)$$

which also may have been obtained by recognizing that:

$$\Delta A = -W_{\text{rev}} = -3100 \text{ cal} \quad (86)$$

Finally, the change in Gibbs free energy is determined from:

$$\Delta G = \Delta E + P\Delta V - T\Delta S = 37550 + 3100 - (373)(109) = 0 \text{ cal} \quad (87)$$

which, too, may have been obtained by recognizing the absence of non-PV work.

STANDARD MOLAR GIBBS FREE ENERGY

The fundamental equation for Gibbs free energy has been given as:

$$dG = -SdT + VdP \quad (88)$$

Consider the change in free energy with pressure at constant temperature. One may begin by defining a standard free energy, $G^\circ(T)$, which corresponds to the free energy of the ideal gas under a pressure of 1 atm. Because the temperature is constant, the change in free energy is given by:

$$\int dG = \int VdP \quad (89)$$

between the limits of 1 atm and the pressure, P . For an ideal gas, the volume is a strong function of pressure; thus, with substitution, and after integration, the result is:

$$G(T, P) - G^\circ(T) = \int (nRT / P) dP \quad (90)$$

Simplifying yields:

$$G = G^\circ + nRT \ln P \quad (91)$$

Dividing through by the number of mols gives:

$$G/n = G^\circ/n + RT \ln P \quad (92)$$

Molar free energy, G/n , is encountered so frequently it is given a special symbol, μ , and Equation 82 is written as:

$$= \mu^\circ + RT \ln P \quad (93)$$

The molar free energy also is referred to as the *chemical potential*.

NON-IDEALITY

Equation 93 describes the molar free energy of an ideal gas, but for real gases the molar free energy is not related directly to the pressure. Thus, a function, the *fugacity*, f , has been introduced, which provides the same functional form of equation for a real gas:

$$= \mu^\circ + RT \ln f \quad (94)$$

The fugacity is related to the pressure by the following equation, which is provided without derivation:

$$\ln f = \ln P + (1/nRT) \int (V - V_{id}) dP \quad (95)$$

where V_{id} represents the volume of an ideal gas. This equation may be justified by considering the assumptions of an ideal gas, which are that the molecules are point particles without volume, and no intermolecular attractive or repulsive forces. Both of these effects have a direct impact on the measured volume; thus, the fugacity may be conceived as a function that corrects for inaccuracies of these assumptions. Clearly, the fugacity approaches the pressure as the real volume approaches the ideal volume.

A similar approach is applied when dealing with mixtures. Consider the molar free energy of a mixture of gases. From Raoult's law, the partial pressure, P_i , of a gas is given by:

$$P_i = x_i P \quad (96)$$

where x_i is the mol fraction of the i th component and P is the total pressure. Molar free energy is:

$$= {}^\circ(T) + RT(\ln P + \ln x_i) \quad (97)$$

For the purposes of evaluating mixtures, it generally is more convenient to define a new standard state, $\mu^\circ(T, P)$, which consists simply of the pure gas at 1 atm, thereby yielding:

$$\mu_i = \mu_i(\text{pure})(T, P) + RT \ln x_i \quad (98)$$

In fact, this equation is applicable not only to the gas state but any ideal state of aggregation. This becomes more apparent by letting the mol fraction go to unity (that is, a pure substance), whence the logarithmic term goes to zero and the molar free energy is equal to the standard-state molar free energy.

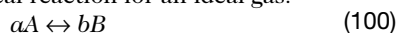
For solutions, there is a corresponding term that describes the departure for an ideal mixture, the *activity*, a . Equation 88 is applicable only for ideal mixtures. However, with the introduction of the activity, a , the above expression may be written as follows, which is general for all mixtures:

$$\mu_i = \mu_i^\circ(T, P) + RT \ln a_i \quad (99)$$

For solutions, $\mu_i^\circ(T, P)$ is the molar free energy of the liquid in the pure state.

EQUILIBRIA

Equilibrium is related intimately to spontaneity, thus the functions above used to predict the spontaneity also may be used for establishing conditions of equilibrium. In essence, if no spontaneous change is predicted, the system is at equilibrium. Consider the following chemical reaction for an ideal gas:



For this reaction, the equilibrium constant is written as:

$$K = P_B^b / P_A^a \quad (101)$$

Let the molar free energy of each component for the condition of equilibrium be defined as G_A and G_B , and as G_A' and G_B' for the nonequilibrium state. The changes in free energy at equilibrium and in a nonequilibrium state are given as:

$$\Delta G = bG_B - aG_A \quad (102)$$

$$\Delta G' = bG_B' - aG_A' \quad (103)$$

The change in free energy between these two states is given by the difference between the changes in free energy; that is:

$$\Delta G' - \Delta G = b(G_B' - G_B) - a(G_A' - G_A) \quad (104)$$

The difference between G_B and G_B' may be calculated for an ideal gas with the use of the fundamental equation given above:

$$dG = VdP - SdT \quad (105)$$

which is related, under the condition of constant temperature, as:

$$dG = VdP \quad (106)$$

$$\Delta G = \int VdP \quad (107)$$

$$\Delta G = \int (nRT/P) dP \quad (108)$$

After integration between limits of the two states, it yields:

$$\Delta G = nRT \ln(P_B'/P_B) \quad (109)$$

Substituting into the Equation 104, to find the overall change in ΔG :

$$\Delta G' - \Delta G = bRT \ln(P_B'/P_B) - aRT \ln(P_A'/P_A) \quad (110)$$

The quantity under the logarithm is given a special definition because it may be generalized to other cases not involving ideal gases; thus, the reaction quotient is defined as:

$$Q = [B']^b / [A']^a \quad (111)$$

where including the equilibrium constant yields:

$$\Delta G' - \Delta G = RT \ln Q - RT \ln K \quad (112)$$

Under conditions of both constant pressure and temperature, dG and $\Delta G = 0$; thus:

$$\Delta G' = RT \ln Q - RT \ln K \quad (113)$$

In a similar fashion to the standard enthalpies of formation of specific compounds, a standard Gibbs free energy, ΔG° , for the above reaction, may be defined as the free energy associated with the conversion of a mols of reactants to b mols of products when the pressure and temperature are held constant, or $\Delta G' = \Delta G^\circ$ and $\ln Q = 0$. The concentrations (or in this example, the pressures) are equal to unity, thus:

$$\Delta G^\circ = -RT \ln K \quad (114)$$

This equation is of great importance because it provides the energy per mole of any chemical reaction, provided the equilibrium constant is known under standard conditions. Alternatively, the equilibrium constant may be calculated if the free energy is known. The universal applicability of thermodynamics also is displayed. Although it was derived for an ideal gas, it is equally applicable to reactions conducted in solution or even in the solid state.

A special case of the van't Hoff equation is known as the *Claapeyron equation*. The interesting feature is that the free-energy change with temperature, in connection with phase changes, is approached in a different manner, with the same result. Consider a liquid in equilibrium with a vapor, such that the free energy associated with each phase, liquid and vapor, may be written from the fundamental equations as:

$$dG_l = -S_l dT + V_l dP \quad (115)$$

$$dG_v = -S_v dT + V_v dP \quad (116)$$

However, because the phases are in equilibrium, $dG_l = dG_v$, or equating the above relationships:

$$-S_l dT + V_l dP = -S_v dT + V_v dP \quad (117)$$

these can be rearranged to yield:

$$(S_v - S_l) dT = (V_v - V_l) dP \quad (118)$$

or with separation of the differential with the incremental changes to give:

$$dP/dT = (S_v - S_l) / (V_v - V_l) \quad (119)$$

or equivalent:

$$dP/dT = \Delta S / \Delta V \quad (120)$$

The interesting aspect of this equation is that the derivative, dP/dT , is related to the discontinuous changes that occur with a phase change. Although this relation was derived for a liquid-vapor equilibrium, it is general and may be applied to any phase change.

The relationship may be manipulated further by recalling that $\Delta S = \Delta H/T$, where T is the temperature at the point of equilibrium. Thus, by substitution, the following is obtained:

$$dP/dT = \Delta H / T \Delta V \quad (121)$$

An approximation may be made, as before, by noting that $V_v - V_l \approx V_v$, which for 1 mol of an ideal gas is $V_v = RT/P$ and affords:

$$dP/dT = P \Delta H / RT^2 \quad (122)$$

The expression may be rearranged, thereby providing a means for measuring the change in enthalpy and entropy:

$$dP/P = (\Delta H/RT^2)dT \quad (123)$$

This is known as the *Clausius–Clapeyron equation*. Assuming ΔH is constant over the small temperature range between T_2 and T_1 , this expression may be integrated yielding:

$$\ln(P_2/P_1) = (\Delta H/R)[(1/T_1) - (1/T_2)] \quad (124)$$

Q—If the equilibrium constant is 1.3×10^{-2} at 25° and 1.7×10^{-1} at 150° , what are ΔH° , ΔG° , and ΔS° ?

A—The temperature of 25° is taken as the standard temperature; Gibbs free energy is given by:

$$\Delta G^\circ = -RT \ln(K_{298}) = -(1.987)(298) \ln(1.2 \times 10^{-2}) = 2.62 \text{ kcal/mol.} \quad (125)$$

The standard enthalpy change may be calculated from the temperature dependence as follows:

$$\Delta H^\circ = R \ln[K(423)/K(298)] / [(1/T_1) - (1/T_2)] = 5.15 \text{ kcal/mol} \quad (126)$$

Finally, ΔS° may be calculated, knowing the free energy and enthalpy changes, from:

$$\Delta S^\circ = -(1/T)(\Delta G^\circ - \Delta H^\circ) = -(1/298)(2620 - 5150) = 8.48 \text{ eu (entropy units)} \quad (127)$$

Q—Justify the following expression of the temperature dependence of the vapor pressure of a liquefied gas:

$$\ln P = -\Delta H_{\text{vap}}/RT + C \quad (128)$$

where C is a constant.

A—Taking the indefinite integral of Equation 86:

$$\int dP/P = \int (\Delta H_{\text{vap}}/RT^2)dT \quad (129)$$

the result is obtained directly:

$$\ln P = (-\Delta H_{\text{vap}}/RT) + C \quad (130)$$

SOLUBILITY

Consider a system consisting of solid drug in equilibrium with a saturated solution. The molar free energy of the solute is the same everywhere, thereby permitting the following:

$$\mu_{2,\text{solution}}(T, P, a_2) = \mu_{2,\text{solid}}(T, P) \quad (131)$$

from which:

$$\mu_{2,\text{liquid}}^\circ(T, P) + RT \ln a = \mu_{2,\text{solid}}(T, P) \quad (132)$$

where $\mu_{2,\text{liquid}}^\circ$ is the chemical potential of the pure liquid solute. Solving for the activity of the pure liquid solute yields:

$$\ln a_2 = [\mu_{2,\text{solid}}(T, P) - \mu_{2,\text{liquid}}^\circ(T, P)]/RT \quad (133)$$

The relationship between activity and solubility is given as follows:

$$a_2 = \gamma_2 X_2 \quad (134)$$

where γ_2 is the activity coefficient of the solute in water. From this analysis, the mole fraction solubility is seen to depend only on the chemical potential of the solute. Thus, the ideal solubility of a drug is the same in every solvent. Furthermore, the difference in solubility observed between solvents is related to the nonideality of the solute, which is quantitatively determined by the activity coefficient.

The cause of nonideality arises from intermolecular interactions, which may be favorable leading to high solubility and a low activity coefficient. Alternatively, the interactions may be unfavorable, which lead to low solubility and high activity coefficient. For this reason, the activity coefficient quantitatively describes the escaping tendency of the solute from the solution.

HEAT OF SOLUTION

When a compound is dissolved in a solvent, the resulting enthalpy change of the system is referred to as the *heat of solution*.

The heat evolved or absorbed reflects the energy required to disrupt the cohesive forces of the solid and the energy generated from interaction of the solute molecules with solvent molecules. There are two ways of expressing the enthalpy change per mol of material dissolved: the integral and differential heats of solution. The two conventions arise from the dependency of the heat of solution on the amount of solvent used to dissolve the solute. Thus, the *integral heat of solution* describes the enthalpy change when 1 mol of solute is dissolved to yield a specified concentration, perhaps a 1 molar solution, whereas the *differential heat of solution* provides a value of the enthalpy change when the amount of solute dissolved is negligible.

One way of understanding the difference between these representations, and also a way of remembering, is as follows:

- The integral heat of solution gives the enthalpy change for a discrete or integral change in concentration of the solution.
- The differential heat of solution provides the enthalpy change for an infinitesimally or differential (dC) change in concentration.

Q—If the differential heat of solution of two polymorphic forms of a drug were measured in water at standard temperature and pressure (STP) and form A had a larger heat than form B, which is more stable at STP?

A—More energy is required to dissolve form A; therefore, it must be the more stable polymorph at STP.

TEMPERATURE DEPENDENCE OF EQUILIBRIUM CONSTANTS

A related aspect is the question of the temperature dependence of the equilibrium constant or, from a different perspective, the temperature dependence of the change in free energy. Using the fundamental equations, it can be shown that Gibbs free energy is related to the following state functions:

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ \quad (135)$$

The determination of the temperature dependence is obtained through the Gibbs–Helmholtz equation, which is derived as follows. First, both sides are divided by the temperature:

$$\Delta G^\circ/T = \Delta H^\circ/T - \Delta S^\circ \quad (136)$$

then the derivative with respect to temperature is taken:

$$\partial(\Delta G^\circ/T)/\partial T = \partial(\Delta H^\circ/T)/\partial T \quad (137)$$

$$\partial(\Delta G^\circ/T)/\partial T = -\Delta H^\circ/T^2 \quad (138)$$

This is referred to as the *Gibbs–Helmholtz equation*, which provides the relationship between the change in Gibbs free energy with temperature and the enthalpy change. However, the standard free-energy change also is related to the equilibrium constant, which, in general, also is temperature dependent, as:

$$\partial(\Delta G^\circ/T)/\partial T = -\partial(R \ln K)/\partial T \quad (139)$$

Combining Equations 137 and 138 yields:

$$-\partial(R \ln K)/\partial T = -\Delta H^\circ/T^2 \quad (140)$$

which may be rearranged and integrated over the limits of T_1 and T_2 , assuming that ΔH° is not a function of temperature, which is a reasonable approximation for a small temperature range:

$$\int \partial(\ln K) = -\int (\Delta H^\circ/RT^2) \partial T \quad (141)$$

Thus, the temperature dependence of the equilibrium constant is given as:

$$\ln [K_2/K_1] = (\Delta H^\circ/R)[(1/T_1) - (1/T_2)] \quad (142)$$

where K_1 and K_2 are the equilibrium constants at temperature T_1 and T_2 , respectively. This is known as the *van't Hoff equation*; this equation is extremely important because of its wide applicability, to not only equilibrium constants of chemical reactions, but also other phenomena such as solubility, complexation, dissociation, and vapor pressure.

Q—If the equilibrium constant is 13.6 at STP, at what temperature will it be 20 if the standard enthalpy for the reaction is 8.3 kcal/mol?

A—Using the following equation:

$$(-R/\Delta H^\circ) \ln(K_2/K_1) + 1/T_1 = 1/T_2 \quad (143)$$

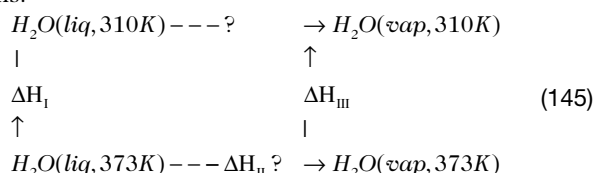
$$1/T_2 = (-1.987/8300) \ln(20/13.6) + 1/298 \quad (144)$$

$$T_2 = 3.26 \times 10^{-3} = 306 \text{ K} = 33^\circ$$

At this point the information presented is useful if every reaction was carried out under standard conditions of 298 K and 1 atm pressure; however, most reactions are not. Nevertheless, the heat of the reaction may be calculated at any other temperature by the temperature dependence of the heat capacity.

Consider the problem of the heat of reaction for the freezing of water at -10° (263 K). One should note that a simple phase change may be treated in a manner analogous to a chemical reaction for the purposes of calculating the enthalpy change.

The approach to the problem is to calculate the enthalpy change associated with the temperature change of the reactant (water) from 0° to -10° and the product (ice) from -10° to 0° , and then use the value for the enthalpy change at the melting point. The following schematic illustrates the approach and also provides insight into the concept of path independence of state functions.



Specifically, the molar enthalpy for the reaction at 310 K is given by:

$$\Delta H(310) = \Delta H_I + \Delta H_{II} + \Delta H_{III} \quad (146)$$

with:

$$\begin{aligned} \Delta H_I &= \int c_p(H_2O, 1) dT = c_p(1) \Delta T \\ \Delta H_I &= (74.5)(373 - 310) = 4693 \text{ J/mol} \\ \Delta H_{II} &= 40,650 \text{ J/mol (found in table 12-3)} \\ \Delta H_{III} &= \int c_p(H_2O, s) dT = c_p(s) \Delta T \end{aligned} \quad (147)$$

The value for the heat capacity of water vapor is about 1.9 kJ/kgK or 34.2 J/molK:

$$\Delta H_{III} = (34.2)(310 - 373) = -2052 \text{ J/mol} \quad (148)$$

Thus, the enthalpy for the conversion of water to ice at -10° is given by the sum or:

$$\Delta H(310 \text{ K}) = 4693 + 40650 - 2052 = 43291 \text{ cal/mol} \quad (149)$$

As shown, the direction around the circle dictates the limits of integration (final less initial), thus care must be taken not to confuse them. The heat capacity was assumed to be independent of temperature, which in this case is reasonable for such a small temperature change (10°). However, in general, heat capacity is a function of temperature, which will not present severe difficulties if the functional relationship is known. In this case the temperature dependence may be substituted into the equation and then integrated with the appropriate limits. For example, the molar heat capacity of oxygen over the range of 300 to 1500 K has been determined experimentally and is approximated closely by:

$$c_p = 6.0954 + 3.2533 \times 10^{-3} T - 10.171 \times 10^{-7} T^2 \quad (150)$$

The change in the molar heat of formation for oxygen from temperature T_1 to T_2 would then be given by:

$$\int dH = \int (6.0954 + 3.2533 \times 10^{-3} T - 10.171 \times 10^{-7} T^2) dT \quad (151)$$

$$\begin{aligned} \Delta H &= [6.0954(T_2 - T_1) + (1/2)(3.2533 \times 10^{-3})[T_2^2 - T_1^2] + \\ &\quad (1/3)(-10.171 \times 10^{-7})(T_2^3 - T_1^3)] \end{aligned} \quad (152)$$

PARTITIONING BEHAVIOR

In an analogous manner, the activity coefficients may be related to the partition coefficient. Consider a solute distributed between two immiscible solvents, A and B. At equilibrium, the chemical potential of the solute is the same in each phase, thus:

$$\mu_{2A} = \mu_{2B} \quad (153)$$

and:

$$\mu_2^\circ + RT \ln a_A = \mu_2^\circ + RT \ln a_B \quad (154)$$

The standard state for the solute in both solvents A and B, μ_2° , must be the same because it is based on the pure liquid solute, the activities must be equal.

$$a_A = a_B \quad (155)$$

The partition coefficient is defined in terms of the mole fractions of infinitely dilute solutions:

$$P_{o/w} = (X_o)^\infty / (X_w)^\infty = \gamma_w / \gamma_o \quad (156)$$

where the activities of the solute in each phase have been cancelled. Noteworthy, in using this system of standard states is that for ideal oil and water solutions of solutes, the partition coefficient is unity, because the activity coefficients, γ_o and γ_w , are equal to one.

These two concepts have been combined by Yalkowski and his coworkers to predict the aqueous solubility of drugs. They proposed the following:

$$\ln(X_{2w}) = \ln(X_{2oct}) - \ln(P_{o/w}) \quad (157)$$

where X_{2w} is the mole fraction in water, $P_{o/w}$ is the octanol/water partition coefficient, and X_{2oct} is the mole fraction solubility of the solute in octanol. From above, the P_o/w is given by the ratio of the activity coefficients, γ_w/γ_o . However, the drug in octanol generally forms an ideal solution, i.e., $\gamma_o = 1$. As such, the octanol/water partition coefficient is a measure of the activity coefficient of the drug in water. Finally, the mole fraction of the solute in an ideal solution of octanol may be determined from the heat of fusion and the melting point. The resulting predictive equation for the water solubility, which has yielded highly correlated data, is:

$$\ln X_2 = -(\Delta H_{fus}/R)(1/T - 1/T_m) - \ln P_{o/w}^\infty \quad (158)$$

The melting point and heat of fusion may be readily measured, and the partition coefficient can be estimated by group contribution methods. Thus, a conceptually elegant foundation has been developed into a practically useful scheme to estimate the water solubility of drugs from the thermal properties and the octanol/water partition coefficient of drugs.

PROTEIN BINDING

As a final example of equilibria, the protein-binding of drugs should be mentioned. Consider the case where a protein has a single binding site for a drug. A mass-balanced equation may be written as:

$$[P] + [D] = [PD] \quad (159)$$

where $[P]$ is the concentration of unbound protein, $[D]$ is the concentration of unbound drug, and $[PD]$ is the concentration of the drug-protein complex. The equilibrium constant may be written for this reaction as:

$$K_a = [PD]/[P][D] \quad (160)$$

where K_a is the association constant. Assuming ideality, the standard free energy for the above equilibrium may be immediately identified as:

$$\Delta G^\circ = -RT \ln(K_a) \quad (161)$$

This concept often is taken a step farther in order to characterize the nature of the binding site of the drug. This has application in structure-activity relationships used for predicting pharmacological activity.

Suppose the association constants of two structurally related drugs, K'_a and K''_a , were determined experimentally. The standard free energy of each association is given as $\Delta G^{\circ'}$ and $\Delta G^{\circ''}$. The effect of the change in the chemical structure on the energetics of the association then can be calculated from the association constants as:

$$\Delta\Delta G^\circ = \Delta G^{\circ''} - \Delta G^{\circ'} = RT \ln(K'_a / K''_a) \quad (162)$$

where $\Delta\Delta G^\circ$ is the standard free-energy change of protein-binding associated with the specific chemical modification.

Q—At room temperature and a protein concentration of 2 μM , the fraction of penicillin G and penicillin V bound was found to be 0.65 and 0.80, respectively. Calculate the change in the standard free energy of binding associated with the replacement of the benzyl group in penicillin G by the phenoxy moiety in penicillin V.

A—The fraction bound may be related to the equilibrium constant by assuming there is only one binding site on each protein molecule. The fraction of drug bound, F , is defined as:

$$F = [DP] / ([D] + [DP]) \quad (163)$$

and because the concentration of the drug-protein complex is given by:

$$[PD] = K_a [P][D] \quad (164)$$

This may be substituted into the above equation yielding

$$F = K_a [P][D] / ([D] + K_a [P][D]) \quad (165)$$

After canceling terms and solving for K_a , the desired expression is obtained:

$$K_a = F/[P](1 - F) \quad (166)$$

The association constants for each drug then are calculated:

$$\begin{aligned} K_a G &= (0.65) / [2 \mu\text{M}] (1 - 0.65) = 0.93 \mu\text{M}^{-1} \\ K_a V &= (0.80) / [2 \mu\text{M}] (1 - 0.80) = 2.0 \mu\text{M}^{-1} \end{aligned} \quad (167)$$

From the association constants, the change in standard free energy associated with replacing the benzyl group with a phenoxy moiety is:

$$\begin{aligned} \Delta\Delta G^\circ &= \Delta G^\circ(V) - \Delta G^\circ(G) = RT \ln(0.93/2.0) \\ \Delta\Delta G^\circ &= -453 \text{ cal/mol K} \end{aligned} \quad (168)$$

The change in the standard free energy is negative, in agreement with the concept that binding of the phenoxy group is more favorable than the benzyl group.

BIBLIOGRAPHY

INTRODUCTORY

- Levine IN. *Physical Chemistry*, 4th ed. New York: McGraw-Hill, 1995.
 Alberty RA, Silbey RJ. *Physical Chemistry A Basic Theory and Methods*, 2nd ed. New York: Wiley, 1996.
 Reiss H. *Methods of Thermodynamics*. New York: Dover, 1997.
 Connors KA. *Thermodynamics of Pharmaceutical Systems, An Introduction for Students of Pharmacy*, New York: Wiley Interscience, 2002.

COMPREHENSIVE

- Glasstone S. *Thermodynamics for Chemists*, New York: Van Nostrand, 1946.
 Lewis GN, Randall M. *Thermodynamics*. Revised by Pitzer KS, Brewer L. New York: McGraw-Hill, 1961.
 Kondepudi DK, Prigogine I. *Modern Thermodynamics: From Heat Engines to Dissipative Structures*. New York: Wiley, 1996.

Chapter 13

Solutions and Phase Equilibria

Pardeep K. Gupta, PhD

SOLUTIONS AND SOLUBILITY 219

PHARMACEUTICAL SOLVENTS 229

COLLIGATIVE PROPERTIES OF SOLUTIONS 232

METHODS TO INCREASE SOLUBILITY OF POORLY SOLUBLE DRUGS 237

SOLUTIONS AND SOLUBILITY

A solution is a chemically and physically homogeneous mixture of two or more substances. The term *solution* generally denotes a homogeneous mixture that is liquid, even though it is possible to have homogeneous mixtures that are solid or gaseous. Thus, it is possible to have solutions of solids in liquids, liquids in liquids, gases in liquids, gases in gases, and solids in solids. The first three of these are most important in pharmacy, and ensuing discussions will be concerned primarily with them.

In pharmacy different kinds of liquid dosage forms are used, and all consist of a dispersion of one or more substances in a liquid phase. Depending on the size of the dispersed particle, they are classified as *true solutions*, *colloidal solutions*, or *disperse systems*. If sugar is dissolved in water, it is supposed that the ultimate sugar particle is of molecular dimensions and that a true solution is formed. On the other hand, if very fine sand is mixed with water, a suspension of comparatively large particles, each consisting of many molecules, is obtained. Between these two extremes lie the colloidal solutions, the dispersed particles of which are larger than those of true solutions but smaller than the particles present in suspensions. Colloidal solutions, in general, are considered to be thermodynamically stable. In this chapter only true solutions will be discussed.

It is possible to classify broadly all solutions as one of two types. In the first type, although there may be lesser or greater interaction between the dispersed substance (the solute) and the dispersing medium (the solvent), the solution phase contains the same chemical entity as found in the solid phase; thus, upon removal of the solvent, the solute is recovered unchanged. One example would be sugar dissolved in water where, in the presence of sugar in excess of its solubility, there is an equilibrium between sugar molecules in the solid phase with sugar molecules in the solution phase. A second example would be dissolving silver chloride in water. Admittedly, the solubility of this salt in water is low, but it is finite. In this case the solvent contains silver and chloride ions, and the solid phase contains the same material. Removal of the solvent yields the initial solute in unchanged form.

In the second type the solvent contains a compound that is different from the one in the solid phase. The difference between the compound in the solid phase and solution is due generally to some chemical reaction that has occurred in the presence of solvent. An example would be dissolving aspirin in an aqueous solvent containing some basic material capable of reacting with the acid aspirin. Now the species in solution would not only be undissociated aspirin, but aspirin also as its anion, whereas the species in the solid phase is aspirin in only its undissociated acid form. In this situation, if the solvent were removed, part of the substance obtained (the salt of aspirin) would be different from what was present initially in the solid.

SOLUTIONS OF SOLIDS IN LIQUIDS

Reversible Solubility Without Chemical Reaction

From a pharmaceutical standpoint, solutions of solids in liquids, with or without accompanying chemical reaction in the

solvent, are of the greatest importance, and many quantitative data on the behavior and properties of such solutions are available. This discussion will be concerned with definitions of solubility, with the rate at which substances go into solution, and with temperature and other factors that control the rate and extent of solubility.

Solubility

When an excess of a solid is brought into contact with a liquid, molecules of the former are removed from its surface until equilibrium is established between the molecules leaving the solid and those returning to it. The resulting solution is said to be saturated at the temperature of the experiment, and the extent to which the solute dissolves is referred to as its *solubility*. The extent of solubility of different substances varies from almost imperceptible amounts to relatively large quantities, but for any given solute the solubility has a constant value at a given constant temperature.

Under certain conditions it is possible to prepare a solution containing a larger amount of solute than is necessary to form a saturated solution. This may occur when a solution is saturated at one temperature, the excess of solid solute is then removed, and the solution cooled. The solute present in solution, even though it may be less soluble at the lower temperature, does not always separate from the solution, and there is produced a supersaturated solution. Such solutions, formed by sodium thiosulfate or potassium acetate, for example, may be made to deposit their excess of solute by vigorous shaking, scratching the side of the vessel in contact with the solution, or introducing into the solution a small crystal of the solute. Supersaturated solutions are considered to be thermodynamically unstable systems and, therefore, usually return to a saturated solution by excluding the excess solute.

Methods of Expressing Solubility

When quantitative data are available, solubility may be expressed in several ways. For example, the solubility of sodium chloride in water at 25°C may be stated as

1. 1 g of sodium chloride dissolves in 2.786 mL of water. (An approximation of this method is used by the USP.)
2. 35.89 g of sodium chloride dissolves in 100 mL of water.
3. 100 mL of a saturated solution of sodium chloride in water contains 31.71 g of solute.
4. 100 g of a saturated solution of sodium chloride in water contains 26.47 g of solute.
5. 1 L of a saturated solution of sodium chloride in water contains 5.425 moles of solute. This also may be stated as a saturated solution of sodium chloride in water is 5.425 molar with respect to the solute.

In order to calculate item 3 above from items 1 or 2, it is necessary to know the density of the solution, which in this case is 1.198 g/mL.

To calculate item 5, the number of grams of solute in 1000 mL of solution (obtained by multiplying the data in item 3 by 10) is divided by the molecular weight of sodium chloride, namely 58.45.

Several other concentration expressions are also available. Molality is the number of moles of solute in 1000 g of solvent and could be calculated from the data in item 4 by subtracting grams of solute from grams of solution to obtain grams of solvent, relating this to 1000 g of solvent and dividing by molecular weight to obtain moles.

Mole fraction is the number of moles of a component divided by total number of moles in that solution. Mole percent may be obtained by multiplying mole fraction by 100. Normality refers to the number of gram equivalent weights of solute dissolved in 1000 mL of solution.

In pharmacy, use also is made of three other concentration expressions. Percent by weight (% w/w) is the number of grams of solute per 100 g of solution and is exemplified by item 4 above. Percent weight in volume (% w/v) is the number of grams of solute per 100 mL of solution and is exemplified by item 3 above. Percent by volume (% v/v) is the number of milliliters of solute in 100 mL of solution, referring to solutions of liquids in liquids. The USP indicates that the term percent, when unqualified, means percent weight in volume for solutions of solids in liquids and percent by volume for solutions of liquids in liquids.

In pharmacopeial texts, when it has not been possible, or in some instances not desirable, to indicate exact solubility, a descriptive term is used. Table 13-1 indicates the usual meaning of such terms.

Rate of Solution

It is possible to define quantitatively the rate at which a solute goes into solution. The simplest treatment is based on a model depicted in Figure 13-1. A solid particle dispersed in a solvent is surrounded by a thin layer of solvent having a finite thickness, l , in units of length e.g. centimeters. The layer is an integral part of the solid and, thus, is referred to characteristically as the *stagnant layer*. This means that, regardless of how fast the bulk solution is stirred, the stagnant layer remains a part of the surface of the solid, moving wherever the particle moves. The thickness of this layer may get smaller as the stirring of the bulk solution increases, but it is important to recognize that this layer will always have a finite thickness, however small it may get.

Physical model representing the dissolution process.

Using Fick's First Law of Diffusion, the rate of solution of the solid can be explained, in the simplest case, as the rate at which a dissolved solute particle diffuses through the stagnant layer to the bulk solution. The driving force behind the movement of the solute molecule through the stagnant layer is the difference in concentration that exists between the concentration of the solute, C_1 , in the stagnant layer at the surface of the solid and its concentration, C_2 , on the farthest side of the stagnant layer. The greater the difference in concentrations ($C_1 - C_2$), the faster the rate of dissolution, assuming that the thickness of the stagnant layer stays constant.

According to Fick's Law, the rate of solution is also directly proportional to surface area of the solid, A in cm^2 , exposed to solvent and inversely proportional to the length of the path through which the dissolved solute molecule must diffuse.

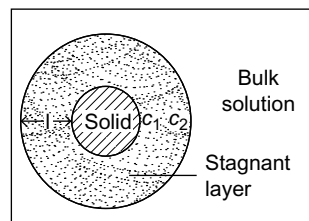


Figure 13-1. Alternating current Wheatstone bridge for measuring conductivity.

Mathematically, then, the rate of solution of the solid is given by

$$\text{Rate of solution} = \frac{DA}{l}(C_1 - C_2) \quad (1)$$

where D is a proportionality constant called the *diffusion coefficient* in cm^2/s . In measuring the rate of solution experimentally, the concentration C_2 is maintained at a low value compared to C_1 and hence is considered to have a negligible effect on the dissolution rate. Furthermore, C_1 most often is the saturation solubility of the solute. Hence Equation 1 is simplified to

$$\text{Rate of solution} = \frac{DA}{l} (\text{saturation solubility}) \quad (2)$$

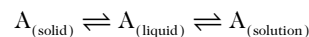
Equation 2 quantitatively explains many of the phenomena commonly observed that affect the rate at which materials dissolve.

1. Small particles go into solution faster than large particles. For a given mass of solute, as the particle size becomes smaller, the surface area per unit of mass of solid increases; Equation 2 shows that as area increases, the rate must increase proportionately. Hence, if a pharmacist wishes to increase the rate of solution of a drug, its particle size should be decreased.
2. Stirring a solution increases the rate at which a solid dissolves. This is because the thickness of the stagnant layer depends on how fast the bulk solution is stirred; as stirring rate increases, the length of the diffusional path decreases. Because the rate of solution is proportional inversely to the length of the diffusional path, the faster the solution is stirred, the faster the solute will go into solution.
3. The more soluble the solute, the faster is its rate of solution. Again, Equation 2 predicts that the larger the saturation solubility, the faster the rate.
4. With a viscous liquid, the rate of solution is decreased. This is because the diffusion coefficient is proportional inversely to the viscosity of the medium; the more viscous the solvent, the slower the rate of dissolution.

Heat of Solution and Temperature Dependency

Turning from the kinetic aspects of dissolution, this discussion will be concerned with the situation in which there is thermodynamic equilibrium between solute in its solid phase and the solute in solution. (It is assumed that there is an amount of solid material in excess of the amount that can go into solution; hence, a solid phase is always present.) As defined earlier, the concentration of solute in solution at equilibrium is the saturation solubility of the substance.

When a solid (Solute A) dissolves in some solvent, two steps may be considered as occurring: the solid absorbs energy to become a liquid, then the liquid dissolves.

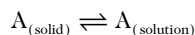


For the overall dissolution, the equilibrium existing between solute molecules in the solid and solute molecules in solution

Table 13-1. Descriptive Terms for Solubility

Descriptive Terms	Parts of Solvent for 1 Part of Solute
Very soluble	Less than 1
Freely soluble	From 1 to 10
Soluble	From 10 to 30
Sparingly soluble	From 30 to 100
Slightly soluble	From 100 to 1000
Very slightly soluble	From 1000 to 10000
Practically insoluble, or insoluble	More than 10000

may be treated as an equilibrium. Thus, for Solute A in equilibrium with its solution,



Using the Law of Mass Action, an equilibrium constant for this system can be defined, just as any equilibrium constant may be written, as

$$K_{\text{eq}} = a_{(\text{solution})}$$

$$K_{\text{eq}} = \frac{a_{(\text{solution})}}{a_{(\text{solid})}}$$

where a denotes the activity of the solute in each phase. Because the activity of a solid is defined as unity,

Because the activity of a compound in dilute solution is approximated by its concentration, and because this concentration is the saturation solubility, K_s , the van't Hoff equation (for a more complete treatment, see Sinko¹) may be used, which defines the relationship between an equilibrium constant (here, solubility) and absolute temperature.

$$\frac{d \log K_s}{dT} = \frac{\Delta H}{2.3RT^2} \quad (3)$$

where $d \log K_s/dT$ is the change of $\log K_s$ with a unit change of absolute temperature, T ; ΔH is a constant that, in this situation, is the heat of solution for the overall process (solid $\rightleftharpoons$ liquid $\rightleftharpoons$ solution); and R is the gas constant, 8.314 Joules/K/mol. Equation 3, a differential, may be solved to give:

$$\log K_s = -\frac{\Delta H}{2.3RT} + J \quad (4)$$

where J is a constant. A more useful form of this equation is:

$$\log \frac{K_{s,T_2}}{K_{s,T_1}} = \frac{\Delta H(T_2 - T_1)}{2.3RT_1T_2} \quad (5)$$

where K_{s,T_1} is the saturation solubility at absolute temperature T_1 and K_{s,T_2} is the solubility at temperature T_2 . Through the use of Equation 5, if ΔH and the solubility at one temperature are known, the solubility at any other temperature can be calculated.

Effect of Temperature

As is evident from Equation 4, the solubility of a solid in a liquid depends on the temperature. In the process of solution, if heat is absorbed (as evidenced by a reduction in temperature), Δ is by convention positive, and the solubility of the solute will increase with increasing temperature. Such is the case for most salts, as is shown in Figure 13-2, in which the solubility of the solute is plotted as the dependent variable and the temperature as the independent variable, and the line joining the experimental points represents the solubility curve for that solute.

If a solute gives off heat during the process of solution (as evidenced by an increase in temperature), by convention Δ is negative, and solubility decreases with an increase in temperature. This is the case with calcium hydroxide and, at higher temperatures, with calcium sulfate. (Because of the slight solubility of these substances, their solubility curves are not included.) When heat is neither absorbed nor given off, the solubility is not affected by variation of temperature, as is nearly the case with sodium chloride.

Solubility curves usually are continuous as long as the chemical composition of the solid phase in contact with the solution remains unchanged, but if there is a transition of the solid phase from one form to another, a break will be found in the curve. Such is the case with $\text{Na}_2\text{SO}_4 \cdot 10\text{H}_2\text{O}$, which dissolves with absorption of heat up to a temperature of 32.4°C, at which

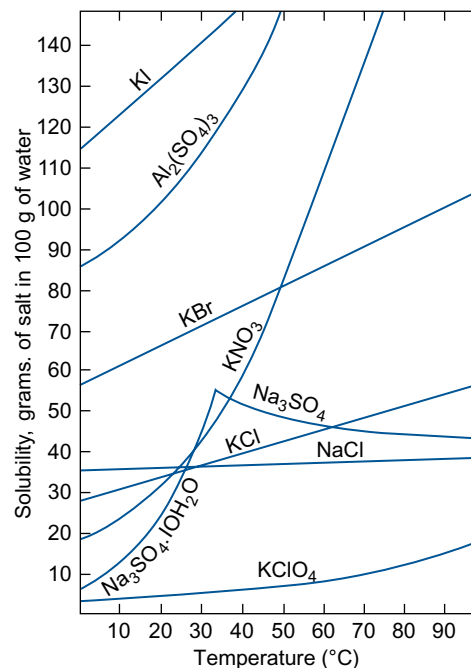


Figure 13-2. Effect of heat on solubility.

point there is a transition of the solid phase to anhydrous sodium sulfate, Na_2SO_4 , which dissolves with evolution of heat. This change is evidenced by increased solubility of the hydrated salt up to 32.4°C, but above this temperature the solubility decreases.

These temperature effects are what would be predicted from Equation 4. When the heat of solution is negative, signifying that energy is released during dissolution, the relation between $\log K_s$ and $1/T$ is typified in Figure 13-3 (Curve A): as $1/T$ increases, $\log K_s$ increases. It can be seen that with increasing temperature (T itself actually increases proceeding left in Figure 13-3), curve A there is a decrease in solubility. On the other hand, when the heat of solution is positive—that is, when heat is absorbed in the solution process—the relation between $\log K_s$ and $1/T$ is typified in Figure 13-3, curve B. Hence, as temperature increases ($1/T$ decreases), the solubility increases.

Effect of Salts

Solubility of a nonelectrolyte in water is generally either decreased or increased by the addition of an electrolyte; it is only

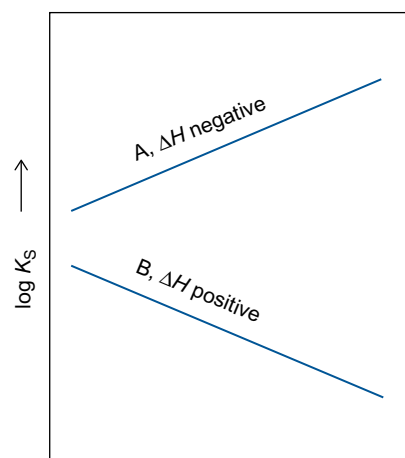


Figure 13-3. Typified relationship between the logarithm of the saturation solubility and the reciprocal of the absolute temperature.

rarely that the solubility is not altered. When the solubility of a nonelectrolyte is decreased, the effect is referred to as *salting-out*; if it is increased, it is described as *salting-in*. Inorganic electrolytes commonly decrease solubility, though there are some exceptions to the generalization.

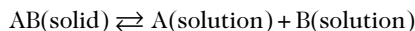
Salting-out occurs because the ions of the added electrolyte interact with water molecules, and thus, in a sense, reduce the amount of water available for dissolution of the nonelectrolyte. (See Thermodynamics of the Solution Process for another view.) The greater the degree of hydration of the ions, the more the solubility of the nonelectrolyte is decreased. If, for example, one compares the effect of equivalent amounts of lithium chloride, sodium chloride, potassium chloride, rubidium chloride, and cesium chloride (all of which belong to the family of alkali metals and are of the same valence type), lithium chloride decreases the solubility of a nonelectrolyte to the greatest extent, and the salting-out effect decreases in the order given.

This is also the order of the degree of hydration of the cations; lithium ion—being the smallest ion and, therefore, having the greatest density of positive charge per unit of surface area—is the most extensively hydrated of the cations, whereas cesium ion is hydrated the least. Salting-out is encountered frequently in pharmaceutical operations.

Salting-in commonly occurs when either the salts of various organic acids or organic-substituted ammonium salts are added to aqueous solutions of nonelectrolytes. In the first case, the solubilizing effect is associated with the anion; in the second, it is associated with the cation. In both cases the solubility increases as the concentration of added salt is increased. The solubility increase may be relatively great, sometimes amounting to several times the solubility of the nonelectrolyte in water.

Solubility of Solutes Containing Two or More Species

In cases where the solute phase consists of two or more species (as in an ionizable inorganic salt), when the solute goes into solution, the solution phase often contains each of these species as discrete entities. For some such substance, AB, the following relationship for the solution process may be written:



As there is an equilibrium between the solute and saturated solution phases, the Law of Mass Action defines an equilibrium constant, K_{eq} :

$$K_{\text{eq}} = \frac{\alpha_{A(\text{solution})} \alpha_{B(\text{solution})}}{\alpha_{AB(\text{solid})}} \quad (6)$$

where $\alpha_{A(\text{solution})}$, $\alpha_{B(\text{solution})}$, and $\alpha_{AB(\text{solid})}$ are the activities of A and B in solution and of AB in the solid phase. Recall from the earlier discussion that the activity of a solid is defined as unity, and that in a very dilute solution (e.g., for a slightly soluble salt), concentrations may be substituted for activities. Equation 6 then becomes:

$$K_{\text{eq}} = C_A C_B$$

where C_A and C_B are the concentrations of A and B in solution. In this situation K_{eq} has a special name, the *solubility product*, K_{sp} . Thus:

$$K_{\text{sp}} = C_A C_B \quad (7)$$

This equation will hold true theoretically only for slightly soluble salts.

As an example of this type of solution, consider the solubility of silver chloride:

$$K_{\text{sp}} = [\text{Ag}^+][\text{Cl}^-]$$

where the brackets [] designate molar concentrations.

At 25°C the solubility product has a value of 1.56×10^{-10} , the concentration of silver and chloride ions being expressed in mol/liter. The same numerical value applies also to solutions of silver chloride containing an excess of either silver or chloride ions. If the silver-ion concentration is increased by the addition of a soluble silver salt, the chloride-ion concentration must decrease until the product of the two concentrations again is equal numerically to the solubility product. To effect the decrease in chloride-ion concentration, silver chloride is precipitated, and hence its solubility is decreased. In a similar manner, an increase in chloride-ion concentration, by the addition of a soluble chloride, effects a decrease in the silver-ion concentration until the numerical value of the solubility product is attained. Again, this decrease in silver-ion concentration is brought about by the precipitation of silver chloride. This phenomenon of decrease in solubility due to the presence of one of the ions in solution is known as the *common-ion effect*.

The solubility of silver chloride in a saturated aqueous solution of the salt may be calculated by assuming that the concentration of silver ion is the same as the concentration of chloride ion, both expressed in mol/liter, and that the concentration of dissolved silver chloride is numerically the same as each silver chloride molecule gives rise to one silver ion and one chloride ion because:

$$[\text{dissolved AgCl}] = [\text{Ag}^+] = [\text{Cl}^-]$$

The solubility of AgCl is equal to 1.56×10^{-5} mol/liter. Multiplying this by the molecular weight of silver chloride (143), we obtain a solubility of approximately 1.8 mg/liter.

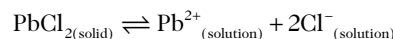
For a salt of the type PbCl_2 the solubility product expression takes the form:

$$[\text{Pb}^{2+}][\text{Cl}^-]^2 = K_{\text{sp}}$$

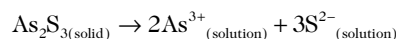
while for As_2S_3 it would be:

$$[\text{As}^{3+}]^2[\text{S}^{2-}]^3 = K_{\text{sp}}$$

because from the Law of Mass Action:

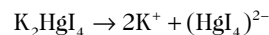


and:



For further details of methods of using solubility-product calculations, see textbooks on qualitative or quantitative analyses or physical chemistry.

Recall that the solubility-product principle is valid for aqueous solutions of slightly soluble salts, provided that the concentration of added salt is not too great. Where the concentrations are high, deviations from the theory occur, and these have been explained by assuming that in such solutions the nature of the solvent has been changed. Frequently, deviations also may occur as the result of the formation of complexes between the two salts. An example of increased solubility, by virtue of complex-ion formation, is seen in the effect of solutions of soluble iodides on mercuric iodide. According to the solubility-product principle, it might be expected that soluble iodides would decrease the solubility of mercuric iodide, but because of the formation of the more soluble complex salt K_2HgI_4 , which dissociates as:



the iodide ion no longer functions as a common ion.

It is possible to formulate some general rules regarding the effect of the addition of soluble salts to slightly soluble salts in which the added salt does not have an ion common to the slightly soluble salt. If the ions of the added soluble salt are not

highly hydrated (see the previous section, Effect of Salts), the solubility product of the slightly soluble salt will increase because the ions of the added salt tend to decrease the interionic attraction between the ions of the slightly soluble salt. On the other hand, if the ions of the added soluble salt are hydrated, water molecules become less available, and the interionic attraction between the ions of the slightly soluble salt increases with a resultant decrease in solubility product. Another way of considering this effect is discussed later (see *Thermodynamics of the Solution Process*).

In general, the effect of temperature is what would be expected: increasing the temperature of the solution results in an increase of the solubility product.

Solubility Following A Chemical Reaction

Thus far, the discussion has been concerned with solubility that comes about because of interplay of entirely physical forces. The dissolution of some substance resulted from overcoming the physical interactions between solute molecules and solvent molecules by the energy produced when a solute molecule interacted physically with a solvent molecule. The solution process, however, can be facilitated also by a chemical reaction. Almost always the chemical enhancement of solubility in aqueous systems is due to the formation of a salt following an acid–base reaction.

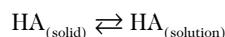
An alkaloidal base, or any other nitrogenous base of relatively high molecular weight, generally is slightly soluble in water, but if the pH of the medium is reduced by addition of acid, the solubility of the base is considerably increased as the pH continues to be reduced. The reason for this increase in solubility is that the base is converted to a salt, which is relatively soluble in water. Conversely, the solubility of a salt of an alkaloid or other nitrogenous base is reduced as pH is increased by addition of alkali.

The solubility of slightly soluble acid substances is, on the other hand, increased as the pH is increased by addition of alkali, the reason again being that a salt, relatively soluble in water, is formed. Examples of acid substances whose solubility is thus increased are aspirin, theophylline and the penicillins, cephalosporins, and barbiturates. Conversely, the solubility of salts of the same substances is decreased as the pH decreases.

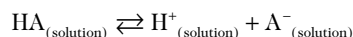
Among some inorganic compounds a somewhat similar behavior is observed. Tribasic calcium phosphate, $\text{Ca}_3(\text{PO}_4)_2$, for example, is almost insoluble in water, but if an acid is added, its solubility increases rapidly with a decrease in pH. This is because hydrogen ions have such a strong affinity for phosphate ions forming nonionized phosphoric acid that the calcium phosphate is dissolved in order to release phosphate ions. Or, stated in another way, the solubilization is an example of a reaction in which a strong acid (the source of the hydrogen ions) displaces a weak acid.

In all of these examples, solubilization occurs as the result of an interaction of the solute with an acid or a base, and thus the species in solution is not the same as the undissolved solute. Compounds that do not react with either acids or bases are slightly, or not at all, influenced in their aqueous solubility by variations of pH. Such effects, if observed, are generally due to ionic *salt effects*.

It is possible to analyze quantitatively the solubility following an acid–base reaction by considering it as a two-step process. The first example is an organic acid, designated as HA, that is relatively insoluble in water. Its two-step dissolution can be represented as:



followed by:



The equilibrium constant for the first step is the solubility of HA ($K_s = [\text{HA}]_{\text{solution}}$), just as was developed earlier when no

chemical reaction took place, and the equilibrium constant for the second step is the dissociation constant of the acid, which is:

$$K_a = \frac{[\text{H}^+][\text{A}^-]}{[\text{HA}]}$$

Since the total amount of compound *in solution* is the sum of non-ionized and ionized forms of the acid, the total solubility may be designated as $S_{t(\text{HA})}$, or:

$$S_{t(\text{HA})} = [\text{HA}] + [\text{A}^-] = [\text{HA}] + K_a \frac{[\text{HA}]}{[\text{H}^+]} \quad (8)$$

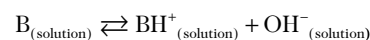
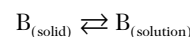
and because $K_s = [\text{HA}]$, Equation 8 becomes:

$$S_{t(\text{HA})} = K_s \left(1 + \frac{K_a}{[\text{H}^+]} \right) \quad (9)$$

Equation 9 is very useful because it equates the total solubility of an acid drug with the hydrogen-ion concentration of the solvent. If the water solubility, K_s , and the dissociation constant, K_a , are known, the total solubility of the acid can be calculated at various hydrogen-ion concentrations.

Equation 9 demonstrates quantitatively how the total solubility of the acid increases as the hydrogen-ion concentration decreases (i.e., as the pH increases).

It is possible to develop an equation similar to Equation 9 for the solubility of a basic drug B, such as a relatively insoluble nitrogenous base (e.g., an alkaloid), at various hydrogen-ion concentrations. The solubility of the base in water may be represented in two steps as:



Again, if K_s is the solubility of the free base in water and K_b is its dissociation constant:

$$K_b = \frac{[\text{BH}^+][\text{OH}^-]}{[\text{B}]}$$

the total solubility of the base in water $S_{t(\text{B})}$ is given by:

$$S_{t(\text{B})} = [\text{B}] + [\text{BH}^+] = [\text{B}] + \frac{K_b[\text{B}]}{[\text{OH}^-]} = K_s \left(1 + \frac{K_b}{[\text{OH}^-]} \right) \quad (10)$$

It is convenient to rewrite Equation 10 in terms of hydrogen-ion concentration by making use of the dissociation constant for water:

$$K_w = [\text{H}^+][\text{OH}^-] = 1 \times 10^{-14}$$

Equation 10 then becomes:

$$S_{t(\text{B})} = K_s \left(1 + \frac{K_b}{K_w / [\text{H}^+]} \right) = K_s \left(1 + \frac{K_b [\text{H}^+]}{K_w} \right) \quad (11)$$

Equation 11 quantitatively shows how the total solubility of the base increases as the hydrogen-ion concentration of the solvent increases. If K_s and K_b are known, it is possible to calculate the total solubility of a basic drug at various hydrogen-ion concentrations using this equation.

Equations 9 and 11 have assumed that the salt formed following a chemical reaction is infinitely soluble. This, of course, is not an acceptable assumption, as suggested and demonstrated by Kramer and Flynn.² Rather, for an acidic or basic drug, there should be a pH at which maximum solubility occurs, where this solubility remains the sum of the solution concentrations of the

free and salt forms of the drug at that pH. Using a basic drug B as the example, this would mean that a solution of B, at pH values greater than the pH of maximum solubility, would be saturated with free-base form but not with the salt form, and the use of Equation 11 would be valid for the prediction of solubility. On the other hand, at pH values less than the pH of maximum solubility, the solution would be saturated with salt form, and Equation 11 is no longer really valid. Because in this situation the total solubility of the base, $S_{t(B)}$, is:

$$S_{t(B)} = [B] + [BH^+]_s$$

where the subscript 's' designates a solution saturated with salt, the correct equation to use at pH values less than the pH maximum would be:

$$S_{t(B)} = [BH^+]_s \left(1 + \frac{[OH^-]}{K_b} \right) = [BH^+]_s \left(1 + \frac{K_w}{K_b [H^+]} \right) \quad (12)$$

A relationship similar to Equation 12 likewise can be developed for an acidic drug at a pH greater than its pH of maximum solubility.

Effecting Solution of Solids in the Prescription Laboratory

The method usually employed by the pharmacist when soluble compounds are to be dissolved in water in compounding a prescription requires the use of the mortar and pestle. The ordinary practice is to crush the substance into fragments in the mortar with the pestle and pour the solvent on it, meanwhile stirring with the pestle until solution is effected. If definite quantities are used and the entire volume of the solvent is required to dissolve the given weight of the salt, only a portion of the solvent should be added first, and when this is saturated, the solution is poured off and a fresh portion of solvent added. This operation is repeated until the solid is dissolved entirely and all the portions combined. Other methods of effecting solution are to shake the solid with the liquid in a bottle or flask, or to apply heat to the substances in a suitable vessel. Substances vary greatly in the rate at which they dissolve; some are capable of producing a saturated solution quickly while others require several hours to attain saturation.

With hygroscopic substances like pepsin, silver protein compounds, and some others, the best method of effecting solution in water is to place the substance directly upon the surface of the water and then stir vigorously with a glass rod. If the ordinary procedure, such as using a mortar and pestle, is employed with these substances, gummy lumps form that are exceedingly difficult to dissolve.

The *solubility* of chemicals and the *miscibility* of liquids are important physical factors for the pharmacist to know, for they often have a bearing on intelligently and properly filling prescriptions. For the information of the pharmacist, the USP provides tabular data indicating the degree of solubility or miscibility of many official substances.

Determination of Solubility

For the pharmacist and pharmaceutical chemist, the question of solubility is of paramount importance. Not only is it necessary to know solubilities when preparing and dispensing medicines, but such information is also necessary to effect separation of substances in qualitative and quantitative analysis. Furthermore, the accurate determination of the solubility of a substance is one of the best methods for determining its purity.

The details of the determination of the solubility are affected markedly by the physical and chemical characteristics of the solute and solvent and also by the temperature at which the solubility is to be determined. Accordingly, it is not possible to describe a universally applicable method, but in general, the following rules must be observed in solubility determinations:

1. It is essential that both the dissolved substance and the solvent are pure because impurities in either affect the solubility.
2. A constant temperature must be maintained accurately during the course of the determination.
3. Complete saturation must be attained.
4. Accurate analysis of the saturated solution and correct expression of the results are imperative.

Consideration should be given also to the varying rates of dissolution of different compounds and to the marked effect of particle size on the time required for the saturation of the solution.

The Phase Rule and Phase-Solubility Analysis

Phase-solubility analysis is a useful and accurate method for the determination of the purity of a substance. It involves the application of precise solubility methods to the principle that constancy of solubility, in the same manner as constancy of melting point, indicates that a material is pure or free from foreign admixture. It is important to recognize that the technique can be used to obtain the exact solubility of the pure substance without the necessity of the experimental material itself being pure.

The method is based on the thermodynamic principles of heterogeneous equilibria that are among the soundest of theoretical concepts of chemistry. Thus, it does not depend on any assumptions regarding kinetics or structure of matter, but is applicable to all species of molecules, and is sufficiently sensitive to distinguish between optical isomers. The requirements for an analysis are simple, for the equipment needed is basic to most laboratories and the quantities of substances required are small.

The standard solubility method consists of five steps:

1. Mixing, in separate systems, increasing amounts of a substance with measured amounts of a solvent.
2. Establishment of equilibrium for each system at identical constant temperature and pressure.
3. Separation of the solid phase from the solutions.
4. Determination of the concentration of the material dissolved in the various solutions.
5. Plotting the concentration of the dissolved material of interest per unit of solvent (y-axis, or solution concentration) against the mass of total material per unit of solvent (x-axis, or system concentration).

The solubility method has been established on the sound theoretical principles of the Gibbs phase rule: $F = C - P + 2$, which relates C , the number of components; F , the degrees of freedom (pressure, temperature, and concentration); and P , the number of phases for a heterogeneous equilibrium.

Solubility analyses are carried out at constant temperature and pressure, so a pure solid in solution would show only one degree of freedom because only one phase is present at concentrations below saturation. This is represented by section AB in Figure 13-4. For a pure solid in a saturated solution at equilibrium (Fig 13-4, BC), two phases are present, solid and solution; there is no variation in concentration and, thus, at constant temperature and pressure, no degrees of freedom.

The curve ABC of Figure 13-4 represents the type of solubility diagram obtained for (1) a pure material, (2) equal amounts of two or more materials having identical solubilities, or (3) a mixture of two or more materials present in the unique ratio of their solubilities. These latter two cases are rare and often may be detected by a change in solvent system.

Line segment BC of Figure 13-4 indicates purity because it has no slope. If, however, this section does exhibit a slope, its numerical value indicates the fraction of impurity present. Line segment BC , extrapolated to the y-axis at D , is the actual solubility of the pure substance.

A representative type of solubility curve, which is obtained when a substance contains one impurity, is illustrated in

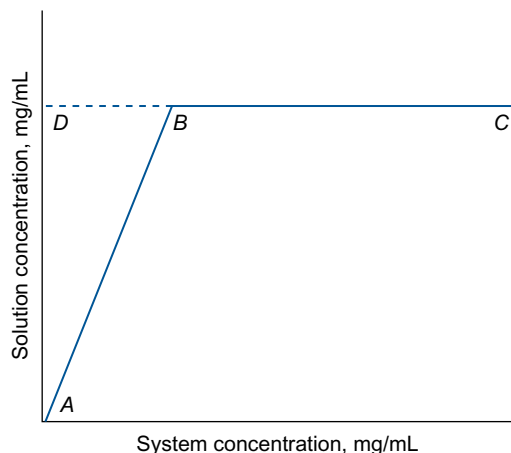


Figure 13-4. Phase-solubility diagram for a pure substance.

Figure 13-5. Here, at *B* the solution becomes saturated with one component. From *B* to *C* there are two phases present: a solution saturated with Component I (usually the major component), containing also some Component II (usually the minor component), and a solid phase of Component I. The one degree of freedom revealed by the slope of the line segment *BC* is the concentration of Component II, which is the impurity (usually the minor component). A mixture of *d* and *l* isomers could have such a curve, as would any simple mixtures in which the solubilities are independent of each other.

The section *CD* indicates that the solvent is saturated with both components of the two-component mixture. Here, three phases are present: a solution saturated with both components and the two solid phases. No variation of concentration is possible; hence, no degree of freedom is possible (indicated by the lack of slope of section *CD*). The distance *AE* on the ordinate represents the solubility of the major component, and the distance *EF* represents the solubility of the minor component.

The equilibration process is time consuming, requiring as long as 3 weeks in certain cases, but this is offset by the fact that all of the sample can be recovered after a determination. This adds to the general usefulness of the method, particularly in cases where the substance is expensive or difficult to obtain. A use for the method other than the determination of purity or of solubility is to obtain especially pure samples by recovering the solid residues at system concentration, corresponding to points on section *BC* in Figure 13-5. Thus, the method is useful not only as a quantitative analytical tool but also for purification.

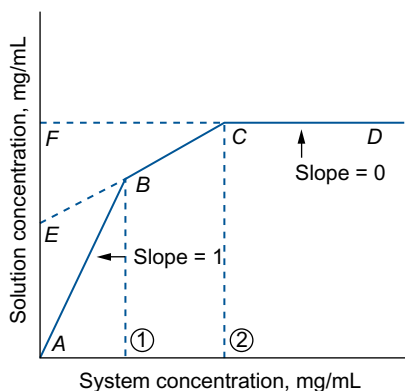


Figure 13-5. Type of solubility curve obtained when a substance contains one impurity.

SOLUTIONS OF LIQUIDS IN LIQUIDS

Binary Systems

The following types of liquid-pairs may be recognized as binary systems:

1. Those that are soluble completely in each other in all proportions. Examples: alcohol and water, glycerin and water, alcohol and glycerin.
2. Those that are soluble in each other in definite proportions. Examples: phenol and water, ether and water, nicotine and water.
3. Those that are imperceptibly soluble in each other in any proportion. Examples: castor oil and water, liquid petrolatum and water.

The mutual solubility of liquid pairs of Type 2 has been studied extensively and found to show interesting regularities. If a series of tubes containing varying, but known, percentages of phenol and water are heated (or cooled, if necessary) just to the point of formation of a homogeneous solution, and the temperature at such points is noted, upon plotting the results a curve is obtained similar to that in Figure 13-6. On this graph the area inside the curve represents the region where mixtures of phenol and water will separate into two layers, while in the region outside of the curve, homogeneous solutions will be obtained. The maximum temperature on this curve is called the *critical solution temperature*, that is, the temperature above which a homogeneous solution occurs, regardless of the composition of the mixture. For phenol and water the critical solution temperature occurs at a composition of 34.5% phenol in water.

Temperature versus composition curves, as depicted in Figure 13-6, provide much useful information in the preparation of homogeneous mixtures of substances showing mutual-solubility behavior. At room temperature (here assumed to be 25°C), by drawing a line parallel to the abscissa at 25°C, we find that we actually can prepare two sets of homogeneous solutions, one containing from 0% to about 7.5% phenol and the other containing phenol from 72% to about 95% (its limit of solubility). At compositions between 7.5% and 72% phenol at 25°C, two liquid layers or phases will separate. In sample tubes containing

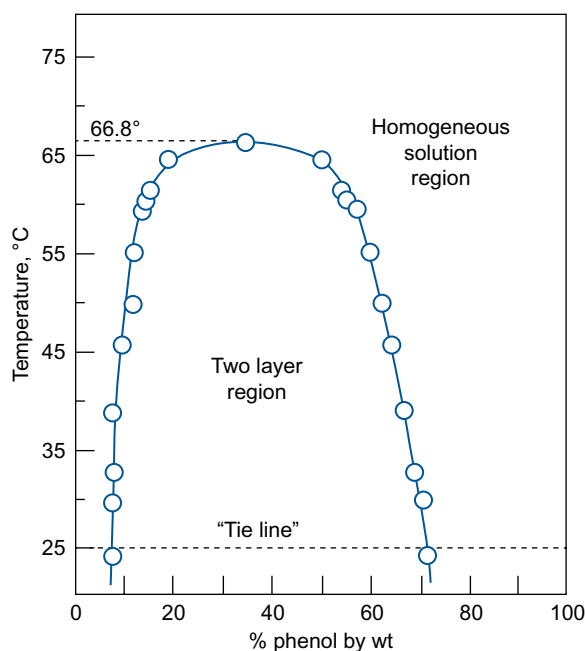


Figure 13-6. Phenol-water solubility. (From Campbell AN, Campbell AJR. *J Am Chem Soc* 1937; 59: 2481.)

a concentration of phenol in this two-layer region at 25°C, one layer always will be phenol-rich and always contain 72% phenol while the other layer will be water-rich and always contain 7.5% phenol. These values are obtained by interpolation of the two points of intersection of the line drawn at 25°C with the experimental curve.

As it may be deduced, at other temperatures, the composition of the two layers in the two-layer region is determined by the points of intersection of the curve with a line (called the *tie line*) drawn parallel to the abscissa at that temperature. The relative amounts of the two layers or phases, phenol-rich and water-rich in this example, will depend on the concentration of phenol added. As expected, the proportion of phenol-rich layer relative to the water-rich layer increases as the concentration of phenol added increases. For example, at 20% phenol in water at 25°C, there would be more of the water-rich layer than of the phenol-rich layer, whereas at 50% phenol in water, there would be more of the phenol-rich layer. The relative portion of each layer may be calculated from such tie lines at any temperature and compositions, as well as the amount of phenol present in each of the two phases. To determine how these calculations are made and for further discussion of this topic, the student should consult Sinko.¹

A simple and practical advantage in the use of phase diagrams is pointed out in Martin *et al.*¹ Based on diagrams such as Figure 13-6, they point out that the most concentrated stock solution of phenol, which should perhaps be used by pharmacists, is one containing 76% w/w phenol in water (equivalent to 80% w/v). At room temperature this mixture is a homogeneous solution and will remain homogeneous to around 3.5°C, at which temperature freezing occurs. It should be noted that Liquefied Phenol USP contains 90% w/w phenol and freezes at 17°C. This means that if the storage area in the pharmacy falls to about 17°C (63°F), the preparation will freeze, resulting in a stock solution no longer convenient to use.

In the case of phenol and water, the mutual solubility increases with an increase in temperature, and the critical solution temperature occurs at a relatively high point. In a certain number of cases, however, the mutual solubility increases with decrease in temperature, and the critical solution temperature occurs at a relatively low value. Most of the substances that show lower critical solution temperatures are amines such as, for example, triethylamine with water.

In addition to pairs of liquids that show *either* upper or lower critical solution temperatures, there are other pairs that show *both* upper and lower critical solution temperatures, and the mutual solubility curve is of the closed type. An example of this type of liquid pair is found in the case of nicotine and water (Fig 13-7). Mixtures of nicotine and water represented by points within the curve will separate into two layers, but mixtures represented by points outside of the curve are perfectly miscible with each other.

In a discussion of solutions of liquids in liquids, it is evident that the distinction between the terms solute and solvent loses its significance. For example, in a solution of water and glycerin, which shall be considered to be the solute and which the solvent? Again, when two liquids are only partially soluble in each other, the distinction between solute and solvent might be reversed easily. In such cases, the term solvent usually is given to the constituent present in larger quantity.

Ternary Systems

The addition of a third liquid to a binary liquid system to produce a ternary or three-component system can result in several possible combinations.

If the third liquid is soluble in only one of the two original liquids, or if its solubility in the two original liquids is markedly different, the mutual solubility of the original pair will be decreased. An upper critical solution temperature will be elevated and a lower critical solution temperature lowered. On the other hand, the addition of a liquid having roughly the same

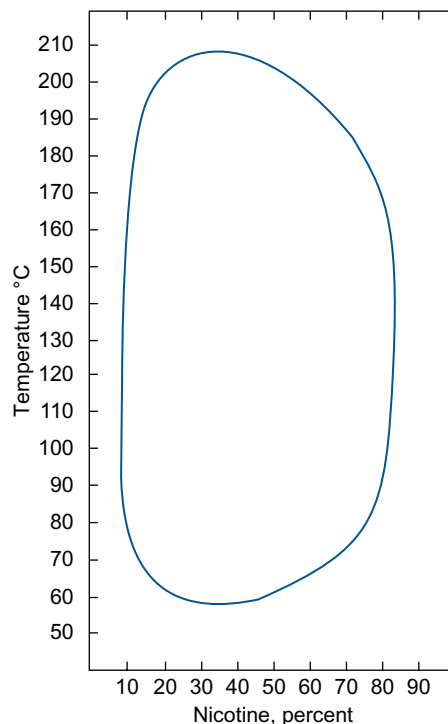


Figure 13-7. Nicotine–water solubility.

solubility in both components of the original pair will result in an increase in their mutual solubility. An upper critical solution temperature then will be lowered and a lower critical solution temperature elevated.

An equilateral-triangle graph may be used to represent ternary systems. In this type of graph, each side of the triangle represents 0% of one of the components, and the apex opposite that side represents 100% of that component. This is illustrated using a particularly common ternary system involving two solvents that are completely miscible and a third that is miscible with only one of the two. In Figure 13-8, water and alcohol are the miscible solvents, and castor oil is the third solvent that is soluble in alcohol but not in water. Such diagrams could be applied, for example, to surfactant/oil/water systems, flavor/water/alcohol systems, drug/propellant mixture systems, drug/water/propylene glycol systems, or any other such system you might think of that would fit into this category.

The data in Figure 13-8 were obtained by determining the amount of water needed to just cloud solutions of oil in alcohol at different concentrations and at room temperature. The percentage of each solvent just clouding the system was then calculated and plotted as shown in the figure. For example, a cloudy solution developed at a mixture of about 67% alcohol, 27% oil, and 6% water. Note that the percentages of the three components must always equal 100%. In the region labeled *miscible*, any combination of the three components will result in a solution. The pharmacist can pick any combination in this region for reasons of taste, safety, stability, or cost. Figure 13-8 is constructed for room temperature; any other temperature would have its own phase diagram. Including temperature as a variable would create a three-dimensional relationship with ternary diagrams, such as Figure 13-8, stacked in the *x-y* plane as a function of temperature on the *z*-axis.

Other possibilities exist in ternary liquid systems—for example, those in which two components are completely miscible and the third is partially miscible with each, and that in which all combinations of two of the three components are only partially miscible.

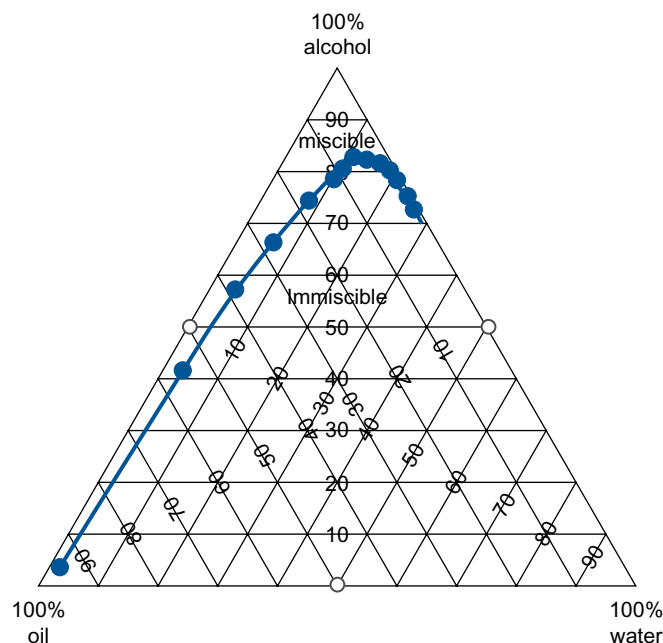


Figure 13-8. Phase diagram at constant temperature for a ternary system: two liquids completely miscible in one another with a third liquid soluble in only one of the two. (Data from Loran MR, Guth EP. *J APhA Sci Ed* 1951; 40: 465.)

SOLUTIONS OF GASES IN LIQUIDS

Nearly all gases are more or less soluble in liquids. One has but to recall the solubility of carbon dioxide, hydrogen sulfide, or air in water as common examples.

The amount of gas dissolved in a liquid in general follows Henry's Law, which states that the weight of gas dissolved by a given amount of a liquid at a given temperature is proportional to its pressure. Thus, if the pressure is doubled, twice as much gas will dissolve as at the initial pressure. The extent to which a gas is dissolved in a liquid, at a given temperature, may be expressed in terms of the solubility coefficient, which is the volume of gas measured under the conditions of the experiment that is absorbed by one volume of the liquid. The degree of solubility also is expressed sometimes in terms of the *absorption coefficient*, which is the volume of gas, reduced to standard conditions, dissolved by one volume of liquid under a pressure of one atmosphere.

Although Henry's Law expresses fairly accurately the solubility of slightly soluble gases, it deviates considerably in the case of very soluble gases, such as hydrogen chloride and ammonia. Such deviations most frequently are due to chemical interaction of solute and solvent.

The solubility of gases in liquids decreases with a rise in temperature and, in general, also when salts are added to the solvent, the latter effect being referred to as the salting-out of the gas.

SOLUTIONS OF SOLIDS IN SOLIDS

Various mixtures of one solid in another are being considered in the pharmaceutical sciences, primarily as a means to increase bioavailability. For example, melts of solid mixtures of drugs with excipients and eutectic mixtures are being investigated (see Chapter 11). It is possible to have a true solution of one solid in another to give rise to a continuum of one solid dispersed in another, as depicted in Figure 13-9. Such a system is referred to as a *continuous* dispersion; it is very rarely found. To achieve this would mean that two materials would have to be of similar size, structure, and interaction energy so that they might enter and occupy a mutual crystalline structure at the molecular level. Hence, such solid solutions may occur only among racemic mixtures of chiral compounds. If it were possible to form a solid

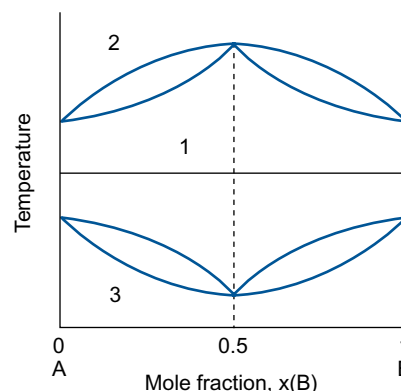


Figure 13-9. Phase diagram for a continuous solution of a Solid A in a Solid B (or of B in A: 1 is an ideal solution; 2 and 3, nonideal solutions. (Data from Duddu S. PhD thesis. University of Minnesota, 1993.)

solution of a drug in a water-soluble excipient, bioavailability could increase dramatically because the drug would transfer into water as individual molecules.

There are three types of continuous dispersions depicted in Figure 13-9: 1 shows an *ideal* dispersion of constant melting point, while 2 and 3 (*nonideal* solutions) show dispersions having a maximum or minimum, respectively. Each of the latter dispersions show an upper or *liquidus* line and a lower or *solidus* line that might be viewed as representing the direction (cooling or heating) used to arrive at the temperature of melting or solidification in each mixture. The composition of the liquid and solid phases in the region between the two lines can be quantified in a way that is related to the tie-line treatment for phenol-water systems, although more complicated.

More common are the *discontinuous* solid dispersions illustrated in Figure 13-10, in which two true solid solutions α and β are separated by a eutectic phase. Such a system is found for urea/acetaminophen, which exists as solid solutions in very small regions at very high urea concentration and very high acetaminophen concentration.

At this point it is worthwhile to briefly consider solid complexes. The interaction of a drug with an excipient to form a new solid phase through strong hydrogen-bond formation can give a solid phase that is not precisely a solid solution, but nonetheless potentially important in its effects on bioavailability—both in a positive and negative sense. The phase diagram in

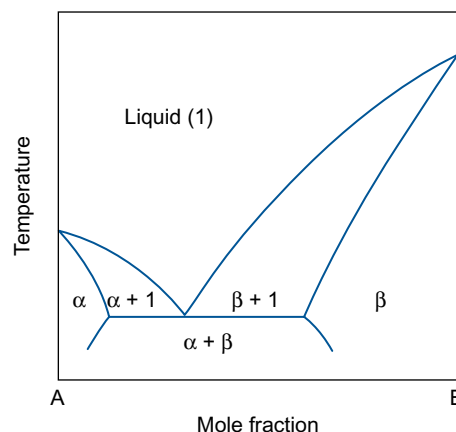


Figure 13-10. Phase diagram for a discontinuous solid solution for Solid α and Solid β : true solid solutions α and β are separated by a eutectic phase. (From Grant DJW, Abougela IKA. Physicochemical interactions in pharmaceutical formulations. *Anal Proc* 1982;19:545–549. DOI: 10.1039/AP9821900545. Reproduced by permission of The Royal Society of Chemistry.)

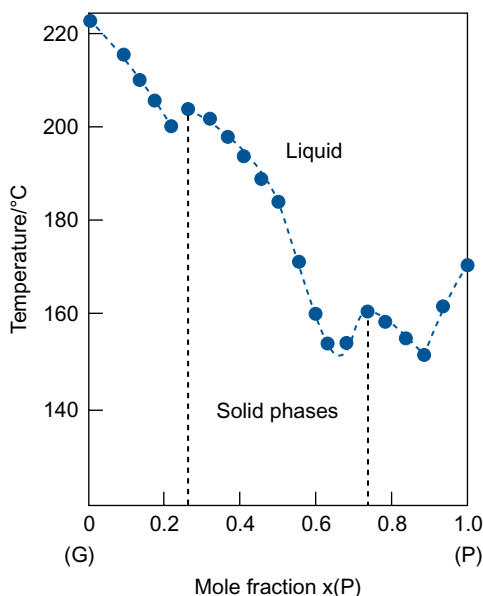
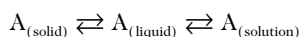


Figure 13-11. Temperature–concentration relationship for fused mixtures of griseofulvin (G) and phenobarbital (P). (From Grant DJW, Abougela IKA. Physicochemical interactions in pharmaceutical formulations. *Anal Proc* 1982; 19:545–549. DOI: 10.1039/AP9821900545. Reproduced by permission of The Royal Society of Chemistry.)

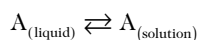
Figure 13-11 was obtained by fusing and cooling mixtures of griseofulvin (G) and phenobarbitone (P). When a complex is stable up to its melting point, the liquidus curve shows a peak referred to as a *congruent melting point*. Two congruently melting complexes, PG_3 and P_3G (at $x = 0.25$ and 0.75 in Fig 13-11), are found for the griseofulvin solidus-phenobarbital system.

THERMODYNAMICS OF THE SOLUTION PROCESS

In this discussion of the thermodynamics of the solution process, the solute is assumed to be in the liquid state; hence, the heat of solution (ΔH) is a term different from that in Equation 3 (ΔH). The heat of solution for a solid solute going into solution, as defined in Equation 3, is the net heat effect for the overall dissolution:



considering only the process:



and assuming that the solute is a liquid (or a supercooled liquid in the case of a solid) at a temperature close to room temperature, at which the energy needed for melting (heat of fusion) is not being considered.

For a physical or chemical reaction to occur spontaneously at a constant temperature and pressure, the net free-energy change, ΔG , for the reaction should be negative (see Thermodynamics, Chapter 12). Furthermore, it is known that the free-energy change depends on heat-related enthalpy ($\Delta H'$) and order-related entropy (ΔS) factors, as seen in:

$$\Delta G = \Delta H' - T\Delta S \quad (13)$$

where T is the temperature. Recall also that the relation between free energy and the equilibrium constant, K , for a reaction is given by:

$$\Delta G = -RT \ln K \quad (14)$$

Equations 13 and 14 certainly apply to the solution of a drug. Because the solubility is, in reality, an equilibrium constant, Equation 14 indicates that the greater the negative value of ΔG , the greater the solubility.

The interplay of these two factors, $\Delta H'$ and ΔS in Equation 13, determines the free-energy change, and hence whether dissolution of a drug will occur spontaneously. Thus, if in the solution process $\Delta H'$ is negative and ΔS positive, dissolution is favored because ΔG will be negative.

Because the heat of solution is quite significant in the dissolution process, one must look at its origin. (For an excellent and more complete discussion of the interactions and driving forces underlying the dissolution process, see Higuchi.⁶) The mechanism of solubility involves severing of the bonds that hold together the ions or molecules of a solute, the separation of molecules of solvent to create a space in the solvent into which the solute can be fitted, and the ultimate response of solute and solvent to whatever forces of interaction may exist between them. In order to sever the bonds between molecules or ions of solute in the liquid state, energy must be supplied, as is the case also when molecules of solvent are to be separated. If heat is the source of energy, it is apparent that both processes require the absorption of heat.

Solute

Solute–solvent interaction, on the other hand, generally is accompanied by the evolution of heat as the process occurs spontaneously. In effecting solution there is, accordingly, a heat-absorbing effect and a heat-releasing effect to be considered beyond those required to melt a solid. If there is no, or very little, interaction between solute and solvent, the only effect will be that of absorption of heat to produce the necessary separations of solute and solvent molecules or ions. If there is a significant interaction between solute and solvent, the amount of heat in excess of that required to overcome the solute–solute and the solvent–solvent forces is liberated. If the opposing heat effects are equal, there will be no change of temperature.

When $\Delta H'$ is zero, and there is no volume change, an *ideal solution* is said to exist because the solute–solute, solvent–solvent, and solute–solvent interactions are the same. For such an ideal solution, the solubility of a solid can be predicted from its heat of fusion (the energy needed to melt the solid) at temperatures below its melting point. The student is referred to Sinko¹ to see how this calculation is made.

When the heat of solution has a positive (energy absorbed) or negative (energy liberated) value, the solution is said to be a *nonideal solution*. A negative heat of solution favors solubility while a positive heat works against dissolution.

The magnitude of the various attractive forces involved between solute, solvent, and solute–solvent molecules may vary greatly and thus could lead to varying degrees of positive or negative enthalpy changes in the solution process. The reason for this is that the molecular structure of the various solutes and solvents determining the interactions can themselves vary greatly. For a discussion of these effects, see Sinko.¹

The solute–solute interaction that must be overcome can vary from the strong ion–ion interaction (as in a salt), to the weaker dipole–dipole interaction (as in nearly all organic materials that are not salts), to the weakest induced dipole–induced dipole interaction (as with naphthalene).

The attractive forces in the solvent that must be overcome are, most frequently, the dipole–dipole interaction (as found in water or acetone) and the induced dipole–induced dipole interaction (as in liquid petrolatum).

The energy-releasing solute–solvent interactions that must be taken into account may be one of four types. In decreasing energy of interaction these are ion–dipole interactions (e.g., a sodium ion interacting with water), dipole–dipole interactions (e.g., an organic acid dissolved in water), dipole–induced dipole interaction, to be discussed later (e.g., an organic acid dissolved in carbon tetrachloride), and induced dipole–induced dipole interactions (e.g., naphthalene dissolved in benzene).

Since the energy-releasing solute–solvent interaction should approximate the energy needed to overcome the solute–solute and solvent–solvent interactions, it should be apparent why it is

not possible to dissolve a salt like sodium chloride in benzene. The interaction between the ions and benzene does not supply enough energy to overcome the interaction between the ions in the solute and therefore gives rise to a positive heat of solution. On the other hand, the interaction of sodium and chloride ions with water molecules does provide an amount of energy approximating the energy needed to separate the ions in the solute and the molecules in the solvent.

Consideration must next be given to entropy effects in dissolution processes. Entropy is an indicator of the disorder or randomness of a system. The more positive the entropy change (ΔS) is, the greater the degree of randomness or disorder of the reaction system and the more favorably disposed is the reaction. Unlike $\Delta H'$, the entropy change (an entropy of mixing) in an ideal solution, is not zero, but has some positive value, for there is an increase in the disorderliness or entropy of the system upon dissolution. Thus, in an ideal solution with $\Delta H'$ zero and ΔS positive, ΔG would have a negative value, and the process would therefore be spontaneous.

In a nonideal solution, on the other hand, where $\Delta H'$ is not zero, ΔS can be equal to, greater than, or less than the entropy of mixing found for the ideal solution. A nonideal solution with an entropy of mixing equal to that of the ideal solution is called a *regular solution*. These solutions usually occur with non-polar or weakly polar solutes and solvents. Such solutions are accompanied by a positive enthalpy change, implying that the solute-solvent molecular interaction is less than the solute-solute and solvent-solvent molecular interactions. Regular solutions are amenable to rigorous physical chemical analysis, which will not be covered in this chapter but which can be found in outline form in Sinko.¹

The possibility exists in a nonideal solution that the entropy change is greater than for an ideal solution. Such a solution occurs when there is an association among solute or solvent molecules. In essence, the dissolution process occurs when starting at a relatively ordered (low entropy) state and progressing to a disorderly (high entropy) state.

The overall entropy change is positive, greater than that of the ideal case, and favorable to dissolution. As may be expected, the enthalpy change in such a solution is positive because association in a solute or solvent must be overcome. The facilitated solubility of citric acid (an unsymmetrical molecule), as compared to inositol (a symmetrical molecule), may be explained on the basis of such a favorable entropy change.⁶

The solubility of citric acid is greater than that of inositol, yet on the basis of their heats of solution, inositol should be more soluble. One may regard this phenomenon in another way. The reason for the higher solubility of citric acid is that, although there is no hindrance in the transfer of a citric acid molecule as it goes from the solute to the solution phase, when the structurally unsymmetrical citric acid attempts to return to the solute phase from solution, it must assume an orientation that will allow ready interaction with polar groups already oriented. If it does not have the required orientation, it will not return readily to the solute but rather will remain in solution, thus bringing about a solubility larger than expected on the basis of heat of solution.

On the other hand, the structurally symmetrical inositol, as it leaves the solution phase, can interact with the solute phase without requiring a definite orientation; all orientations are equivalent. Hence, inositol can enter the solute phase without hindrance, and therefore, no facilitation of its solubility is observed. In general, unsymmetrical molecules tend to be more soluble than symmetrical molecules.

Another type of nonideal solution occurs when there is an entropy change less than that expected of an ideal solution.

Such nonideal behavior can occur with polar solutes and solvents. In a non-ideal solution of this type, there is significant interaction between solute and solvent. As may be expected, the enthalpy change ($\Delta H'$) in such a solution is negative and favors dissolution, but this effect is tempered by the unfavorable

entropy change occurring at the same time. The reason for the lower-than-ideal entropy change can be visualized where the equilibrium system is more orderly and has a lower entropy than that expected for an ideal solution. The overall entropy change of solution thus would be less and not favorable to dissolution.

One may rationalize the lower-than-expected solubility of lithium fluoride on the basis of this phenomenon. Compared with other alkali halides, it has a solubility lower than would be expected based solely on enthalpy changes. Because of the small size of ions in this salt, there may be considerable ordering of water molecules in the solution. This effect must, of course, lead to a lowered entropy and an unfavorable effect on solubility. The effect of soluble salts on the solubility of non-electrolytes may be considered as a result of an unfavorable entropy effect (see Solubility of solutes containing two or more species, above).

PHARMACEUTICAL SOLVENTS

The discussion will focus now on solvents available to pharmacists and on the properties of these solvents. Pharmacists must obtain an understanding of the possible differences in solubility of a given solute in various solvents because they are often called on to select a solvent that will dissolve the solute. Knowledge of the properties of solvents will allow the intelligent selection of suitable solvents.

On the basis of the forces of interaction occurring in solvents, one may broadly classify solvents as one of three types:

1. *Polar solvents*—those made up of strong dipolar molecules having hydrogen bonding (water or hydrogen peroxide).
2. *Semipolar solvents*—those also made up of strong dipolar molecules but that do not form hydrogen bonds (acetone or pentyl alcohol).
3. *Non-polar solvents*—those made up of molecules having a small or no dipolar character (benzene, vegetable oil, or mineral oil).

Naturally, there are many solvents that may fit into more than one of these broad classes; for example, chloroform is a weak dipolar compound but generally is considered non-polar in character, and glycerin could be considered a polar or semipolar solvent even though it is capable of forming hydrogen bonds.

SOLVENT TYPES

Water

Water is a unique solvent. Besides being a highly associated liquid, giving rise to its high boiling point, it has another very important property, a high dielectric constant. The *dielectric constant* (ϵ) indicates the effect that a substance has, when it acts as a medium, on the ease with which two oppositely charged ions may be separated. The ease of solubilizing salts in solvents like water and glycerin can be explained on the basis of their high dielectric constant. Also, in general, the more polar the solvent, the greater its dielectric constant.

An important concept has been introduced to pharmaceutical systems: pharmacists frequently are concerned with dissolving relatively non-polar drugs in aqueous or mixed polar aqueous solvents.¹¹ To understand what may be happening in such cases, factors concerned with the entropic effects arising from interactions originating with the non-polar solutes must be considered. Previously it had been noted that the favorable entropic effect on dissolution was due to the disruption of associations occurring among solute or solvent molecules. Now consider the effects on solubility due to solute interactions in the solution phase—because the solutes under discussion are

relatively non-polar, the interactions are of the London Force type or a hydrophobic association.

This hydrophobic association in aqueous solutions may cause significant structuring of water with a resultant ordered or low-entropy system that is unfavorable to solution. Therefore, the solution of an essentially non-polar molecular in water is not a favorable process. It should be stressed that this is due to not only an unfavorable enthalpy change but also an unfavorable entropy change generated by water structuring.

Such an unfavorable entropy change, known as the *hydrophobic effect*, is quite significant in the solution process. As an example of this effect, the aqueous solubility of a series of alkyl *p*-aminobenzoates shows a 10-million-fold decrease in solubility in going from the 1-carbon analog to the 12-carbon analog. These findings demonstrate clearly the considerable effect that hydrophobic associations can have.

Alcohols

Ethanol, as a solvent, is next in importance to water. An advantage of ethanol is that growth of microorganisms does not occur in solutions containing alcohol in a reasonable concentration.

Resins, volatile oils, alkaloids, glycosides, etc., are dissolved by alcohol, but many therapeutically inert principles, such as gums, albumin, and starch, are insoluble, which makes them more useful as *selective* solvents. Mixtures of water and alcohol, in proportions varying to suit specific cases, are used extensively. They are often referred to as hydroalcoholic solvents.

Glycerin is an excellent solvent although its range is not as extensive as that of water or alcohol. In higher concentrations it has preservative action. It dissolves the fixed alkalis, a large number of salts, vegetable acids, pepsin, tannin, and some active principles of plants, but it also dissolves gums, soluble carbohydrates, and starch. It also is of special value as a simple solvent (as in phenol glycerite) or where the major portion of the glycerin simply is added as a preservative and stabilizer of solutions that have been prepared with other solvents (see *Glycerines*, Chapter 26).

Propylene glycol, which has been used widely as a substitute for glycerin, is miscible with water, acetone, or chloroform in all proportions. It is soluble in ether and will dissolve many essential oils but is immiscible with fixed oils. It is claimed to be as effective as ethyl alcohol in its power of inhibiting mold growth and fermentation.

Isopropyl alcohol possesses solvent properties similar to those of ethanol and is used instead of the latter in a number of pharmaceutical manufacturing operations. It has the advantage in that the commonly available product contains not over 1% of water, whereas ethyl alcohol contains about 5% water, often a disadvantage. Isopropyl alcohol is employed in some liniment and lotion formulations. It cannot be taken internally.

General Properties

Low-molecular-weight and polyhydroxy alcohols form associated structures through hydrogen bonds just as in water. When the carbon-atom content of an alcohol rises above five, generally only monomers then are present in the pure solvent. Although alcohols have high dielectric constants compared to other types of solvents, they are small compared to water. As has been discussed, the solubility of salts in a solvent should be paralleled by its dielectric constant. That is, as the dielectric constant of a series of solvents increases, the probability of dissolving a salt in the solvent increases. This behavior is observed for the alcohols. Table 13-2, taken from Higuchi,⁶ shows how the solubility of salts follows the dielectric constant of the alcohols.

As mentioned earlier, absolute alcohol rarely is used pharmaceutically. However, hydroalcoholic mixtures such as elixirs and spirits frequently are encountered. A very useful generalization is that the dielectric properties of a mixed solvent, such as water and alcohol, can be approximated as the weighted average of the properties of the pure components. Thus, a mixture

Table 13-2. Solubilities of Potassium Iodide and Sodium Chloride in Several Alcohols and Acetone^a

Solvent	g KI/100 g Solvent	g NaCl/100 g Solvent
Water	148	35.9
Glycerin	...	8.3 (20°)
Propylene glycol	50	7.1 (30°)
Methanol	17	1.4
Acetone	2.9	
Ethanol	1.88	0.065
1-Propanol	0.44	0.0124
2-Propanol	0.18	0.003
1-Butanol	0.2	0.005
1-Pentanol	0.089	0.0018

^a All measurements are at 25°C unless otherwise indicated.

(Data from Duddu S. PhD thesis. University of Minnesota, 1993.)

of 60% alcohol (by weight) in water should have a dielectric constant approximated by:

$$\epsilon_{(\text{mixture})} = 0.6(\epsilon_{(\text{alcohol})}) + 0.4(\epsilon_{(\text{water})})$$

$$\epsilon_{(\text{mixture})} = 0.6(25) + 0.4(80) = 47$$

The dielectric constant of 60% alcohol in water is found experimentally to be 43, which is in close agreement with that just calculated. The dielectric constant of glycerin is 46, close to the 60% alcohol mixture. One would therefore expect a salt like sodium chloride to have about the same solubility in glycerin as in 60% alcohol. The solubility of sodium chloride in glycerin is 8.3 g/100 g of solvent and in 60% alcohol about 6.3 g/100 g of solvent. This agreement would be even closer if comparisons were made on a volume rather than weight basis. At least qualitatively it can be said that the solubility of a salt in a solvent or a mixed solvent closely follows the dielectric constant of the medium or, conversely, that the polarity of mixed solvents is paralleled by their dielectric constant, based on salt solubility.

Although the dielectric constant is useful in interpreting the effect of mixed solvents on salt solubility, it cannot be applied properly to the effect of mixed solvents on the solubility of nonelectrolytes. It was seen earlier that unfavorable entropic effects can occur upon dissolution of relatively non-polar nonelectrolytes in water. Such an effect due to hydrophobic association considerably affects solubility. Yalkowsky⁸ studied the ability of cosolvent systems to increase the solubility of nonelectrolytes in polar solvents where the cosolvent system essentially brings about a reduction in structuring of solvent. Thus, by increasing, in a positive sense, the entropy of solution by using cosolvents, it was possible to increase the solubility of the non-polar molecule. Using as an example the solubility of alkyl *p*-aminobenzoates in propylene glycol–water systems, Yalkowsky reported that it is possible to increase the solubility of the nonelectrolyte by several orders of magnitude by increasing the fraction of propylene glycol in the aqueous system.⁸ Sometimes, it is found that, as a good first approximation, the logarithm of the solubility is related linearly to the fraction of propylene glycol added by:

$$\log S_f = \log S_{f=0} + \epsilon f$$

where S_f is the solubility in the mixed aqueous system containing the volume fraction f of nonaqueous cosolvent, $S_{f=0}$ is the solubility in water, and is a constant (not dielectric constant) characteristic of the system under study. Specifically, when a 50% solution of propylene glycol in water is used, there is a 1000-fold increase in solubility of dodecyl *p*-aminobenzoate, in comparison to pure water.

Another empirical equation sometimes used to estimate solubility of a poorly water-soluble substance in a mixed-solvent system is written as:

$$\log S_t = \log S_w f_w + \log S_1 f_1 +$$

where S_t is the total solubility, S_w and S_1 are solubilities in pure water and cosolvent 1, respectively, and f_w and f_1 are the fractions of water and cosolvent 1, respectively.

In a series of studies, Martin *et al.*⁹ have made attempts to predict solubility in mixed solvent systems through an extension of the *regular solution* theory. The equations are logarithmic in nature and can reduce in form to the equations of Yalkowsky.⁸

Acetone and Related Semipolar Materials

Even though acetone has a very high dipole moment (2.8×10^{-18} esu), as a pure solvent it does not form associated structures. This is evidenced by its low boiling point (57°C) in comparison with the boiling point of the lower molecular weight water (100°C) and ethanol (79°C). The reason why it does not associate is that the positive charge in its dipole does not reside in a hydrogen atom, precluding the possibility of its forming a hydrogen bond. However, if some substance that is capable of forming hydrogen bonds, such as water or alcohol, is added to acetone, a very strong interaction through hydrogen bonding will occur (see Mechanism of solvent action, below). Some substances that are semipolar and similar to acetone are aldehydes, low-molecular-weight esters, other ketones, and nitro-containing compounds.

Non-polar Solvents

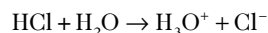
Non-polar class of solvents includes fixed oils, such as vegetable oil and petroleum ether (ligroin), carbon tetrachloride, benzene, and chloroform. On a relative basis there is a wide range of polarity among these solvents; for example, benzene has no dipole moment, whereas that of chloroform is 1.05×10^{-18} esu.

It should be emphasized that when a solvent (such as chloroform) has highly electronegative halogen atoms attached to a carbon atom that also contains at least one hydrogen atom, such a solvent will be capable of forming strong hydrogen bonds with solutes which are polar in character. Thus, through the formation of hydrogen bonds, such solvents will dissolve polar solutes. For example, it is possible to dissolve alkaloids in chloroform.

MECHANISM OF SOLVENT ACTION

A solvent may function in one or more ways. When an ionic salt is dissolved (e.g., by water), the process of solution involves separation of the cations and anions of the salt with attendant orientation of molecules of the solvent about the ions. Such orientation of solvent molecules about the ions of the solute—a process called *solvation* (*hydration*, if the solvent is water)—is possible only when the solvent is highly polar, whereby the dipoles of the solvent are attracted to and held by the ions of the solute. The solvent also must possess the ability to keep the solvated, charged ions apart with minimal energy.

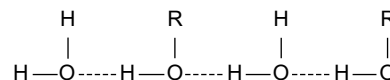
A polar liquid such as water may exhibit solvent action, also, by virtue of its ability to break a covalent bond in the solute and bring about ionization of the latter. For example, hydrogen chloride dissolves in water and functions as an acid as a result of:



The ions formed by this preliminary reaction of breaking the covalent bond are subsequently maintained in solution by the same mechanism as ionic salts.

Still another mechanism by which a polar liquid may act as a solvent is that involved when the solvent and solute are capable of being coupled through hydrogen-bond formation.

The solubility of the low-molecular-weight alcohols in water, for example, is attributed to the ability of the alcohol molecules to become part of a water-alcohol association complex.

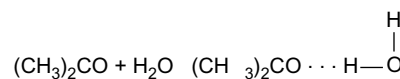


As the molecular weight of the alcohol increases, it becomes progressively less polar and less able to compete with water molecules for a place in the lattice-like arrangement formed through hydrogen bonding; high-molecular-weight alcohols are, therefore, poorly soluble or insoluble in water. When the number of carbon atoms in a normal alcohol reaches five, its solubility in water is reduced significantly.

When the number of hydroxyl groups in the alcohol is increased, its solubility in water generally is increased greatly; it is principally, if not entirely, for this reason that such high-molecular-weight compounds as sugars, gums, many glycosides, and synthetic compounds, such as the polyethylene glycols, are very soluble in water.

The solubility of ethers, aldehydes, ketones, acids, and anhydrides in water and in other polar solvents also is attributable largely to the formation of an association complex between solute and solvent by means of the hydrogen bond. The molecules of ethers, aldehydes, and ketones, unlike those of alcohols, are not associated themselves because of the absence of a hydrogen atom that is capable of forming the characteristic hydrogen bond. Notwithstanding, these substances are more or less polar because of the presence of a strongly electronegative oxygen atom, which is capable of association with water through hydrogen-bond formation.

Acetone, for example, dissolves in water, in all likelihood principally because of the following type of association:



The maximum number of carbon atoms that may be present per molecule possessing a hydrogen-bondable group, while still retaining water solubility, is approximately the same as for the alcohols.

Although nitrogen is less electronegative than oxygen and thus tends to form weaker hydrogen bonds, amines are at least as soluble as alcohols containing an equivalent chain length. The reason for this is that alcohols form two hydrogen bonds with a net interaction of 50,208 joules/mol. Primary amines can form three hydrogen bonds; two amine protons are shared with the oxygens of two water molecules, and the nitrogen accepts one water proton. The net interaction for the primary amine is between 50,208 and 54,392 joules/mol; hence, it shows an equal or greater solubility compared with corresponding alcohols.

The solvent action of non-polar liquids involves a somewhat different mechanism. Because they are unable to form dipoles with which to overcome the attractions between ions of an ionic salt, or to break a covalent bond to produce an ionic compound or form association complexes with a solute, non-polar liquids are incapable of dissolving polar compounds. They only can dissolve, in general, other non-polar substances in which the bonds between molecules are weak. The forces involved usually are of the induced dipole-induced dipole type. Such is the case when one hydrocarbon is dissolved in another, or an oil or a fat is dissolved in petroleum ether.

Sometimes it is observed that a polar substance, such as alcohol, will dissolve in a non-polar liquid, such as benzene. This apparent exception to the preceding generalization may be explained by the assumption that the alcohol molecule induces a temporary dipole in the benzene molecule, which forms an association complex with the solvent molecules. A binding force of this kind is referred to as a *permanent dipole-induced dipole force*.

Some Useful Generalizations

The preceding discussion indicates that enough is known about the mechanism of solubility to be able to formulate some generalizations concerning this important physical property of substances. Because of the greater importance of organic substances in the field of medicinal chemistry, certain of the more useful generalizations about organic chemicals are presented here in summary form. However, it should be remembered that the phenomenon of solubility usually involves several variables, and there may be exceptions to general rules.

One general maxim that holds true in most instances is, the greater the structural similarity between solute and solvent, the greater the solubility. As often stated, *like dissolves like*. Thus, phenol is almost insoluble in petroleum ether but is very soluble in glycerin.

Organic compounds containing polar groups capable of forming hydrogen bonds with water are soluble in water, provided that the molecular weight of the compound is not too great. It is demonstrated easily that the polar groups OH, CHO, COH, CHOH, CH₂OH, COOH, NO₂, CO, NH₂, and SO₃H tend to increase the solubility of an organic compound in water. On the other hand, non-polar or very weak polar groups, such as the various hydrocarbon radicals, reduce solubility; the greater the number of carbon atoms in the radical, the greater the decrease in solubility. Introduction of halogen atoms into a molecule in general tends to decrease solubility because of an increased molecular weight without a proportionate increase in polarity.

The greater the number of polar groups contained per molecule, the greater the solubility of a compound, provided that the size of the rest of the molecule is not altered; thus, pyrogallol is much more soluble in water than phenol. The *relative positions* of the groups in the molecule also influence solubility; thus, in water, resorcinol (*m*-dihydroxybenzene) is more soluble than catechol (*o*-dihydroxybenzene), and the latter is more soluble than hydroquinone (*p*-dihydroxybenzene).

Polymers and compounds of high molecular weight can be poorly soluble.

High melting points frequently are indicative of low solubility for organic compounds. One reason for high melting points is the association of molecules, and this cohesive force tends to prevent dispersion of the solute in the solvent.

The *cis* form of an isomer is more soluble than the *trans* form (Table 13-3).

Solvation, which is evidence of the existence of a strong attractive force between solute and solvent, enhances the

solubility of the solute, provided there is not a marked ordering of the solvent molecules in the solution phase.

Acids, especially strong acids, usually produce water-soluble salts when reacted with nitrogen-containing organic bases.

COLLIGATIVE PROPERTIES OF SOLUTIONS

Up to this point our concern has been with dissolving a solute in a solvent. Once the dissolution has been brought about, naturally the solution has a number of properties that are different from that of the pure solvent. Of very great importance are the colligative properties that a solution possesses.

The colligative properties of a solution are those that depend on the number of solute particles in solution, irrespective of whether these are molecules or ions, large or small. Ideally, the effect of a solute particle of one species is considered to be the same as that of an entirely different kind of particle, at least in dilute solution. Practically, there may be differences that may become substantial as the concentration of the solution is increased.

The colligative properties that will be considered are

1. Osmotic pressure.
2. Vapor-pressure lowering.
3. Boiling-point elevation.
4. Freezing-point depression.

Of these four, all of which are related, osmotic pressure has the greatest direct importance in the pharmaceutical sciences. It is the property that largely determines the physiological acceptability of a variety of solutions used for therapeutic purposes.

OSMOTIC-PRESSURE ELEVATION

Osmosis

The phenomenon of osmosis is based on the fact that substances tend to move or diffuse from regions of higher concentration to regions of lower concentration. When a solution is separated from the solvent by means of a membrane that is permeable to the solvent but not to the solute (such a membrane is referred to as a *semipermeable membrane*), it is possible to demonstrate visibly the diffusion of solvent into the concentrated solution, as volume changes will occur. In a similar manner, if two solutions of different concentration are separated by a membrane, the solvent will move from the solution of lower solute concentration to the solution of higher solute concentration. This diffusion of solvent through a membrane is called osmosis.

There is a difference between the activity or escaping tendency of the water molecules found in the solvent and salt solution separated by the semipermeable membrane. Because *activity*, which is related to water concentration, is higher on the pure solvent side, water moves from solvent to solution in order to equalize escaping-tendency differences. The difference in escaping-tendency gives rise to what is referred to as the osmotic pressure of the solution, which might be visualized as follows. A semipermeable membrane is placed over the end of a tube, and a small amount of salt solution placed over the membrane in the tube. The tube then is immersed in a trough of pure water so that the upper level of the salt solution initially is at the same level as the water in the trough. With time, solvent molecules will move from solvent into the tube. The height of the solution will rise until the hydrostatic pressure exerted by the column of solution is equal to the osmotic pressure.

Osmotic Pressure of Nonelectrolytes

Quantitative studies using solutions of varying concentration of a solute that does not ionize have demonstrated that osmotic pressure is proportional to the concentration of the solute; that is, twice the concentration of a given nonelectrolyte will

Table 13-3. Demonstration of Solubility Rules

Chemical Compound	Solubility ^a
Aniline, C ₆ H ₅ NH ₂	28.6
Benzene, C ₆ H ₆	1430.0
Benzoic acid, C ₆ H ₅ COOH	275.0
Benzyl alcohol, C ₆ H ₅ CH ₂ OH	25.0
1-Butanol, C ₄ H ₉ OH	12.0
<i>t</i> -Butyl alcohol, (CH ₃) ₃ COH	Miscible
Carbon tetrachloride, CCl ₄	2000.0
Chloroform, CHCl ₃	200.0
Fumaric acid (<i>trans</i> -butenedioic acid)	150.0
Hydroquinone, C ₆ H ₄ (OH) ₂	14.0
Maleic acid, <i>cis</i> -butenedioic acid	5.0
Phenol, C ₆ H ₅ OH	15.0
Pyrocatechol, C ₆ H ₄ (OH) ₂	2.3
Pyrogallol, C ₆ H ₃ (OH) ₃	1.7
Resorcinol, C ₆ H ₄ (OH) ₂	0.9

^a Milliliters of mL of water required to dissolve 1 g of solute.

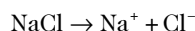
produce twice the osmotic pressure in a given solvent. (This is not strictly true in solutions of fairly high solute concentration but does hold quite well for dilute solutions.)

Furthermore, the osmotic pressures of solutions of different nonelectrolytes are proportional to the number of molecules in each solution. Stated in another manner, the osmotic pressures of two nonelectrolyte solutions of the same molal concentration are identical. Thus, a solution containing 34.2 g of sucrose (mol wt 342) in 1000 g of water has the same osmotic pressure as a solution containing 18.0 g of anhydrous dextrose (mol wt 180) in 1000 g of water. These solutions are said to be *isoosmotic* (*is-osmotic*) with each other because they have identical osmotic pressures.

Osmotic Pressure of Electrolytes

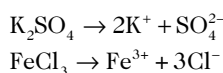
In discussing the generalizations concerning the osmotic pressure of solutions of nonelectrolytes, it was stated that the osmotic pressures of two solutions of the same molal concentration are identical. This generalization, however, cannot be made for solutions of electrolytes—acids, alkalis, and salts (see Chapter 15).

For example, sodium chloride is assumed to ionize as:



It is evident that each molecule of sodium chloride that ionizes produces two ions; if sodium chloride is completely ionized, there will be twice as many particles as would be the case if it were not ionized at all. Furthermore, if each ion has the same effect on osmotic pressure as a molecule, it might be expected that the osmotic pressure of the solution would be twice that of a solution containing the same molal concentration of nonionizing substance.

For solutions that yield more than two ions—for example:



it is expected that the complete dissociation of the molecules would give rise to osmotic pressures that are three and four times, respectively, the pressure of solutions containing an equivalent quantity of a nonionized solute. Accordingly, the equation $PV = nRT$, which may be employed to calculate the osmotic pressure of a dilute solution of a nonelectrolyte, also may be applied to dilute solutions of electrolytes if it is changed to $PV = inRT$, where the value of i approaches the number of ions produced by the ionization of the stronger electrolytes cited in the preceding examples. For weak electrolytes i represents the total number of particles, ions, and molecules together in the solution, divided by the number of molecules that would be present if the solute did not ionize. The experimental evidence indicates that, at least in dilute solutions, the osmotic pressures approach the predicted values. It should be emphasized, however, that in more concentrated solutions of electrolytes, the deviations from this simple theory are considerable, due to interionic attraction, solvation, and other factors.

Biological Aspects of Osmotic Pressure

Osmotic pressure experiments were made as early as 1884 by the Dutch botanist Hugo de Vries in his study of plasmolysis, the term applied to the contraction of the contents of plant cells placed in solutions of comparatively high osmotic pressure. The phenomenon is caused by the osmosis of water out of the cell through the practically semipermeable membrane surrounding the protoplasm. If suitable cells (e.g., the epidermal cells of the leaf of *Tradescantia discolor*) are placed in a solution of higher osmotic pressure than that of the cell contents, water flows out of the cell, causing the contents to draw away from the cell wall. On the other hand, if the cells are placed in solutions of lower osmotic pressure, water enters the cell, producing an expansion that is limited by the rigid cell wall. By immersing cells

in a series of solutions of varying solute concentration, a solution may be found in which plasmolysis is barely detectable or absent. The osmotic pressure of such a solution is then the same, or very nearly the same, as that of the cell contents, and it is then said that the solution is *isotonic* with the cell contents. Solutions of greater concentration than this are said to be *hypertonic*, and solutions of lower concentration are called *hypotonic*.

Red blood cells, or erythrocytes, have been studied similarly by immersion into solutions of varying concentration of different solutes.¹⁰ When introduced into water or into sodium chloride solutions containing less than 0.90 g of solute per 100 mL, human erythrocytes swell and often burst because of the diffusion of water into the cell and the fact that the cell wall is not sufficiently strong to resist the pressure. This phenomenon is referred to as hemolysis. If the cells are placed in solutions containing more than 0.90 g of sodium chloride per 100 mL, they lose water and shrink. By immersing the cells in a solution containing exactly 0.90 g of sodium chloride in 100 mL, no change in the size of the cells is observed; because in this solution the cells maintain their *tone*, the solution is said to be *isotonic* with human erythrocytes. For the reasons indicated it is desirable that solutions to be injected into the blood should be made isotonic with erythrocytes.

Distinction Between Iso-smotic (isomotic) and Isotonic

The terms *isoosmotic* and *isotonic* are not to be considered as equivalent, although a solution often may be described as being both isoosmotic and isotonic. If a plant or animal cell is in contact with a solution that has the same osmotic pressure as the cell contents, there will be no net gain or loss of water by either solution, provided the cell membrane is impermeable to all the solutes present. As the volume of the cell contents remains unchanged, the *tone*, or normal state, of the cell is maintained, and the solution in contact with the cell may be described not only as being isoosmotic with the solution in the cell but also as being isotonic with it. If, however, one or more of the solutes in contact with the membrane can pass through the latter, it is evident that the volume of the cell contents will change, thus altering the *tone* of the cell; in this case the two solutions may be isoosmotic, yet not be isotonic.

VAPOR-PRESSURE LOWERING

When a nonvolatile solute is dissolved in a liquid solvent, the vapor pressure of the solvent is lowered. This easily can be described qualitatively by visualizing solvent molecules on the surface of the solvent, which normally could escape into the vapor, being replaced by solute molecules, which have little, if any, vapor pressure of their own. For ideal solutions of nonelectrolytes, the vapor pressure of the solution follows Raoult's Law:

$$P_A = X_A P_A^0 \quad (15)$$

where P_A is the vapor pressure of the solution, P_A^0 is the vapor pressure of the pure solvent, and X_A is the mole fraction of solvent. This relationship states that the vapor pressure of the solution is proportional to the number of molecules of solvent in the solution. Rearranging Equation 15 gives:

$$\frac{P_A^0 - P_A}{P_A^0} = (1 - X_A) X_B \quad (16)$$

where X_B is the mole fraction of the solute. This equation states that the lowering of vapor pressure in the solution relative to the vapor pressure of the pure solvent—called simply the relative vapor-pressure lowering—is equal to the mole fraction of the solute. The absolute lowering of vapor pressure of the solution is defined by:

$$P_A^0 - P_A = X_B P_A^0 \quad (17)$$

Example Calculate the lowering of vapor pressure, and the vapor pressure at 20°C, of a solution containing 50 g of anhydrous dextrose (mol wt 180.16) in 1000 g of water (mol wt 18.02). The vapor pressure of water at 20°C, in absence of air, is 2.311 kPa.

First, calculate the lowering of vapor pressure, using Equation 17, in which X_B is the mol fraction of dextrose, defined by:

$$X_B = \frac{n_B}{n_A + n_B}$$

where n_A is the number of mols of solvent and n_B is the number of mols of solute. Substituting numerical values:

$$\begin{aligned} n_B &= \frac{50}{180.2} = 0.278 \\ n_A &= \frac{1000}{18.02} = 55.5 \\ n_B &= \frac{0.278}{55.5 + 0.278} = 0.00498 \end{aligned}$$

the lowering of vapor pressure is:

$$\begin{aligned} P_A^0 - P_A &= 0.00498 \times 17.535 \\ &= 0.0873 \text{ mmHg} \end{aligned}$$

The vapor pressure of the solution is:

$$\begin{aligned} P_A &= 17.535 - 0.0873 \\ &= 17.448 \text{ mmHg} \end{aligned}$$

BOILING-POINT ELEVATION

In consequence of the fact that the vapor pressure of any solution of a nonvolatile solute is less than that of the solvent, the *boiling point* of the solution—the temperature at which the vapor pressure is equal to the applied pressure (commonly 760 mmHg)—must be higher than that of the solvent. This is clearly evident in Figure 13-12.

FREEZING-POINT DEPRESSION

The *freezing point of a solvent* is defined as the temperature at which the solid and liquid forms of the solvent coexist in equilibrium at a fixed external pressure, commonly 1 atmosphere (1 atm = 760 mm [torr] of mercury). At this temperature the solid and liquid forms of the solvent must have the same vapor

pressure, for if this were not so, the form having the higher vapor pressure would change into that having the lower vapor pressure.

The *freezing point of a solution* is the temperature at which the solid form of the pure solvent coexists in equilibrium with the solution at a fixed external pressure, again commonly 1 atm. Because the vapor pressure of a solution is lower than that of its solvent, it is obvious that solid solvent and solution cannot coexist at the same temperature as solid solvent and liquid solvent; only at some lower temperature, where solid solvent and solution do have the same vapor pressure, is equilibrium established. A schematic pressure–temperature diagram for water and an aqueous solution, not drawn to scale and exaggerated for the purpose of more effective illustration, shows the equilibrium conditions involved in both freezing-point depression and boiling-point elevation (see Fig 13-12).

The freezing-point lowering of a solution may be quantitatively predicted for ideal solutions, or dilute solutions that obey Raoult's Law, by mathematical operations similar to (though somewhat more complex than) those used in deriving the boiling-point elevation constant. The equation for the freezing-point lowering, ΔT_f , is:

$$\Delta T_f = \frac{RT_0^2 M_A m}{1000 \Delta H_{\text{fus}}} = K_f m \quad (18)$$

where:

$$K_f = \frac{RT_0^2 M_A}{1000 \Delta H_{\text{fus}}} \quad (19)$$

The value of K_f for water, which freezes at 273.1° K and has a heat of fusion of 79.7 cal/g, is:

$$K_f = \frac{8.314}{1000} \frac{273.1^2}{18.02} \frac{1}{333.46} = 0.0068^\circ \quad (20)$$

The molal freezing-point depression constant is not intended to represent the freezing-point depression for a 1-molal solution, which is too concentrated for the premise of ideal behavior to be applicable. In dilute solutions the freezing-point depression, calculated to a 1-molal basis, approaches the theoretical value—the more dilute the solution, the better the agreement between experiment and theory.

To calculate the molecular weight of the solute, the freezing point of a dilute solution of a nonelectrolyte solute may be used (as was the boiling point). The applicable equation is:

$$M_B = \frac{(K_f \times 1000 \times W_B)}{(W_A \times \Delta T_f)} \quad (21)$$

The molecular weight of organic substances soluble in molten camphor may be determined by observing the freezing point of a mixture of the substance with camphor. This procedure, called the *Rast method*, uses camphor because it has a very large molal freezing-point-depression constant, about 40. Because the *constant* may vary with different lots of camphor and with variations of technique, the method should be standardized using a solute of known molecular weight.

Freezing-point determinations of molecular weights have the advantage over boiling-point determinations of greater accuracy and precision by virtue of the larger magnitude of the freezing-point depression compared to boiling-point elevation. Thus, in the case of water the molal freezing-point depression is approximately 3.5 times greater than the molal boiling-point elevation.

IDEAL BEHAVIOR AND DEVIATIONS

In setting out to derive mathematical expressions for colligative properties, such phrases as for *ideal solutions* or for *dilute solutions* were used to indicate the limitations of the expressions.

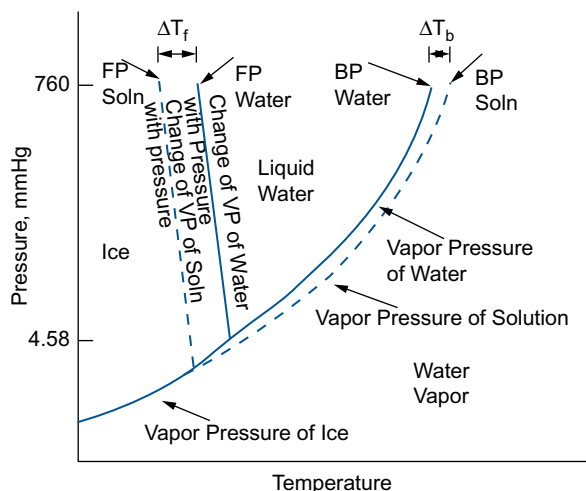


Figure 13-12. Vapor-pressure–temperature diagram for water and an aqueous solution, illustrating elevation of boiling point and lowering of freezing point of the latter.

Samuel Glasstone defines an ideal solution as “one which obeys Raoult’s Law over the whole range of concentration and at all temperatures” and gives as specific characteristics of such solutions their formation only from constituents that mix in the liquid state without heat change and without volume change. These characteristics reflect the fact that addition of a solute to a solvent produces no change in the forces between molecules of the solvent. Thus, the molecules have the same escaping-tendency in the solution as in the pure solvent, and the vapor pressure above the solution is proportional to the ratio of the number of solvent molecules in the surface of the solution to the number of the molecules in the surface of the solvent—which is the basis for Raoult’s Law.

Any change in intermolecular forces produced by mixing the components of a solution may result in deviation from ideality; such a deviation may be expected particularly in solutions containing both a polar and a non-polar substance. Solutions of electrolytes, except at high dilution, are especially prone to depart from ideal behavior, even though allowance is made for the additional particles that result from ionization. When solute and solvent combine to form solvates, the escaping-tendency of the solvent may be reduced in consequence of the reduction in the number of free molecules of solvent; thus, a negative deviation from Raoult’s Law is introduced. On the other hand, the escaping-tendency of the solvent in a solution of nonvolatile solute may be increased because the cohesive forces between molecules of solvent are reduced by the solute; this results in a positive deviation from Raoult’s Law.

Although few solutions exhibit ideal behavior over a wide range of concentration, most solutions behave ideally at least in high dilution, where deviations from Raoult’s Law are negligible.

Colligative Properties of Electrolyte Solution

Earlier in this chapter, attention was directed to the increased osmotic pressure observed in solutions of electrolytes, the enhanced effect being attributed to the presence of ions, each of which acts, in general, in the same way as a molecule in developing osmotic pressure. Similar magnification of vapor-pressure lowering, boiling-point elevation, and freezing-point depression occurs in solutions of electrolytes. Thus, at a given constant temperature, the abnormal effect of an electrolyte on osmotic pressure is paralleled by abnormal lowering of vapor pressure; the other colligative properties are (subject to variation of effect with temperature) comparably intensified. In general, the magnitude of each colligative property is proportional to the total number of particles (molecules and/or ions) in solution.

While in very dilute solutions, the osmotic pressure, vapor-pressure lowering, boiling-point increase, and freezing-point depression of solutions of electrolytes would approach values two, three, and four times greater for NaCl, Na₂SO₄, and Na₃PO₄ than in solutions of the same molality of a nonelectrolyte, two other effects are observed as the concentration of electrolyte is increased. The first effect results in less than two-, three-, or four-fold intensification of a colligative property. This reduction is ascribed to interionic attraction between the positively and negatively charged ions. Consequently, the ions are not completely dissociated from each other and do not exert their full effect in lowering vapor pressure, etc. This deviation generally increases with increasing concentration of electrolyte. The second effect intensifies the colligative properties and is attributed to the attraction of ions for solvent molecules, which holds the solvent in solution and reduces its escaping-tendency, with consequent enhancement of the vapor-pressure lowering. Solvation also may reduce interionic attraction and thereby further lower the vapor pressure.

These factors (and possibly others) combine to effect a progressive reduction in the molal values of colligative properties as the concentration of electrolyte is increased from 0.5 to 1.0 molal, beyond which the molal quantities either increase (sometimes quite abruptly) or remain almost constant.

ACTIVITY AND ACTIVITY COEFFICIENT

Various mathematical expressions are employed to relate properties of chemical systems (equilibrium constants, colligative properties, pH, etc.) to the stoichiometric concentration of one or more molecular, atomic, or ionic species. In deriving such expressions it is either stated or implied that they are valid only so long as intermolecular, interatomic, and/or interionic forces may be ignored or remain constant, under which restriction the system may be expected to behave ideally. But intermolecular, interatomic and/or interionic forces do exist, and not only do they change as a result of chemical reaction, but they also change with variation in the concentration or pressure of the molecules, atoms, or ions under observation. In consequence, mathematical expressions involving stoichiometric concentrations or pressures generally have limited applicability. The conventional concentration terms provide a count of molecules, atoms, or ions per unit volume but afford no indication of the physical or chemical activity of the species measured, and it is this activity that determines the physical and chemical properties of the system.

In recognition of this, G. N. Lewis introduced both the quantitative concept and methods for evaluation of activity as a true measure of the physical or chemical activity of molecular, atomic, or ionic species, whether in the state of gas, liquid, or solid, or whether present as a single species or in a mixture. Activity may be considered loosely as a corrected concentration or pressure that takes into account not only the stoichiometric concentration or pressure but also any intermolecular attractions, repulsions, or interactions between solute and solvent in solution, association, and ionization. Thus, activity measures the net effectiveness of a chemical species.

Because only relative values of activity may be determined, a *standard state* must be chosen for quantitative comparisons to be made. Indeed, because activity measurements are needed for many different types of systems, several standard states must be selected. Because this discussion is concerned mainly with solutions, the standard state for the solvent is pure solvent, while for the solute it is a hypothetical solution with free energy corresponding to unit molality under conditions of ideal behavior of the solution. The relationship of activity to concentration is measured in terms of an activity coefficient, which is discussed in Chapter 15.

PRACTICAL APPLICATIONS OF COLLIGATIVE PROPERTIES

One of the most important pharmaceutical applications of colligative properties is in the preparation of isotonic intravenous and isotonic lacrimal solutions, the details of which are discussed in Chapter 16.

Other applications of the colligative properties are found in experimental physiology. One such application is in the immersion of tissues in salt solutions, which are isotonic with the fluids of the tissue, in order to prevent changes or injuries that may arise from osmosis.

The colligative properties of solutions also may be used in determining the molecular weight of solutes, or in the case of electrolytes, the extent of ionization. The method of determining molecular weight depends on the fact that each of the colligative properties is altered by a constant value when a definite number of molecules of solute is added to a solvent (see Chapter 15). For example, in dilute solutions the freezing point of water is lowered at the rate of 1.855 for each mole of a nonelectrolyte dissolved in 1000 g of water.

The boiling-point elevation may be used similarly for determining molecular weights. The boiling point of water is raised at the rate of 0.52°C for each mole of solute dissolved in 1000 g of water; the corresponding values for benzene, carbon tetrachloride, and phenol are 2.57°C, 4.88°C, and 3.60°C, respectively. The observed vapor-pressure lowering and osmotic pressure likewise may be used to calculate molecular weights.

To determine the extent to which an electrolyte is ionized, it is necessary to know its molecular weight, as determined by some other method, and then to measure one of the four colligative properties. The deviation of the results from similar values for nonelectrolytes then is used in calculating the extent of ionization.

QUANTITATIVE TREATMENT OF SOLUBILITY

The focus of discussion so far has been on the qualitative aspects of solubility. It is, however, important to understand some quantitative relationships that can help pharmaceutical scientists predict the solubility of new-drug entities in various solvents and allow them to choose the best solvent system for a given drug. The observation that structurally similar chemical entities have better solubility in each other is based on the fact that cohesive forces operating in such molecules are of the same order of magnitude. One measure of these cohesive forces is a quantity known as internal pressure (P_i). It is given by¹:

$$P_i = \frac{(\Delta H_v - RT)}{V} \quad (22)$$

where ΔH_v is the heat of vaporization of a substance and V is its molar volume at temperature T . Since Equation 22 contains ΔH_v which depends on the amount of energy required to break intermolecular (cohesive) bonds, P_i is a measure of cohesive forces among the molecules. This value is high in polar substances; for example, water has a P_i value of 2301.2 joules/mL. Therefore, drugs with high internal pressure show higher solubility in water. The term P_i usually is reserved for solubility of liquids in liquids.

For a quantitative estimate of solubility of solids in liquids, it is assumed that in an ideal solution the heat of solution is equal to the heat of fusion (heat required to melt one mole of solid to liquid without changing its temperature). Because ideal solubility does not depend on the nature of solvent, it can be expressed by equation 23:

$$-\log X_2^i = \frac{\Delta H_f}{2.303R} \left(\frac{T_0 - T}{T_0 T} \right) \quad (23)$$

where X_2^i is the mol fraction solubility in an ideal solution, ΔH_f is the molar heat of fusion of solute, T_0 is the melting point of solute, and T is the solution temperature such that $T < T_0$.

In a nonideal solution, the mol fraction solubility (X) has to be replaced by thermodynamic activity (a) of the solute. This activity can be expressed in terms of mol fraction solubility as:

$$a_2 = X_2 \gamma_2 \quad (24)$$

in which a_2 is a proportionality constant called the *activity coefficient*. The value of a_2 in ideal solution is equal to its maximum value of unity. By taking the log of the above equation and substituting in Equation 23, one obtains the equation of non-ideal solubility as:

$$-\log X_2 = \frac{\Delta H_f}{2.303R} \left(\frac{T_0 - T}{T_0 T} \right) + \log a_2 \quad (25)$$

It can be seen that when $a_2 = 1$, $\log a_2$ is zero, and the equation reduces to the ideal solubility equation.

In general, ideal solutions are rare. Solutions of non-polar solutes in non-polar solvents usually come close to being ideal. However, solutions involving polar solutes or solvents almost always show significant deviation from ideality. The value of a_2 is hard to determine and varies with concentration of solution. It can be, however, estimated by:

$$\log a_2 = \left[(\omega_{11})^{1/2} - (\omega_{22})^{1/2} \right]^2 \frac{V_2 \Phi_1^2}{2.303RT} \quad (26)$$

where ω_{11} is the amount of work involved in separating solvent molecules to create space for a solute molecule, ω_{22} is the work involved in breaking a solute molecule from its bulk, V_2 is the molar volume of solute at temperature T , Φ_1 is the volume fraction of the solvent, and R is the gas constant. The terms ω_{11} and ω_{22} are a measure of the internal energy or cohesive forces of the solvent and solute, respectively. It can be seen from Equation 26 that deviation from ideality is high if values of ω_{11} and ω_{22} are different from each other or if the molar volume of the solute is high. The ω terms are also known as the *solubility parameters*, denoted as δ . Thus the equation of nonideal solubility can be written as:

$$-\log X_2 = \frac{\Delta H_f}{2.303R} \left(\frac{T_0 - T}{T_0 T} \right) + \frac{V_2 \Phi_1^2}{2.303RT} (\delta_1 - \delta_2)^2 \quad (27)$$

The following observations can be made from Equation 27.

1. For dilute solutions Φ_1 is approximately equal to unity and thus may be disregarded in estimating solubility in dilute solutions
2. The closer the values of Δ_1 and Δ_2 , the greater the solubility for a given pair of solute and solvent. In fact, when $\Delta_1 = \Delta_2$, the equation reduces to the equation for ideal solution, in which case the solubility is at its maximum value and depends only on the molar heat of fusion of the solute.
3. Solutions of larger solute molecules (high value of V_2) show higher deviation from ideality. It is not surprising, therefore, that solutions of polymers and other high-molecular-weight compounds show a very different behavior than ideal solution (see Solutions of polymers, below).

The solubility parameters can be measured using property of the material that involves molecular or cohesive interactions. These include the molar heat of vaporization, surface tension, internal pressure, and several others. One method suggested by Hilderbrand *et al.*¹² is to use the expression for internal pressure to estimate the value of solubility parameter as follows:

$$\delta = \left(\frac{\Delta H_v - RT}{V} \right)^{1/2} \quad (28)$$

The meanings of the symbols are the same as defined earlier. The values of solubility parameters are available in several references for many commonly used drugs. As intermolecular forces are composed of many kinds of forces, including polar and non-polar forces, the individual contribution of these forces can be included in quantitative estimate of solubility parameter. Hilderbrand and Scott¹³ suggested Equation 29 for this purpose:

$$\delta^2 = \delta_D^2 + \delta_P^2 + \delta_H^2 \quad (29)$$

where Δ_D^2 is the partial solubility parameter arising from non-polar interactions, Δ_P^2 is the partial solubility parameter from polar interactions, and Δ_H^2 is the partial solubility parameter from the hydrogen-bonding tendency among the molecules. The value of Δ_D^2 is fairly constant for all types of molecules, polar as well as non-polar, because non-polar forces operate in all of these molecules. This value ranges from 29.288 to 41.84 joules/mL. Because Δ_P^2 is due to polar forces, which are essentially absent in non-polar compounds, its value range is broader, 0 to 54.392 joules/mL. The value of Δ_H^2 , on the other hand, has the highest contribution where present and has a range of 0 to 104.6 joules/mL. Therefore, for non-polar compounds, such as linear hydrocarbons, the total value of δ^2 is composed entirely of Δ_D^2 and is close to about 7. For this reason, most hydrocarbons show a similar behavior of solubility. In the nonhydrogen-bonding compounds that are relatively polar, Δ_P^2 has significant contribution.

SOLUTIONS OF POLYMERS

The solubility behavior of polymers is usually significantly different from that of small molecules. Although there is no well-defined value of molecular weight cutoff point between polymers and regular molecules, polymer solutions included in the discussion here will focus on molecules whose size approaches the colloidal range.

Depending upon the manner in which the monomers are connected to each other, polymers can be of several types. From the solubility standpoint, however, the nature of the monomers is of great significance. In general, the solubility behavior of homopolymers (consisting of monomers repeated n times) mimics the solubility behavior of the monomers. This implies that the homopolymers consisting of relatively hydrophobic monomers will be poorly soluble in water. Examples of such polymers include polystyrene and polyamines.

However, if the hydrophobic monomers form parts of the block polymers (consisting of blocks of one repeating monomer unit followed by a block of different monomer) or heteropolymers (several monomers attached in random manner), their contribution to solubility may not be as negative as one would expect from their structure. This is because polymers are long molecules and generally have the flexibility to fold themselves in a manner that allows their hydrophobic areas to be folded away from water, much in the way that amphiphiles aggregate to form a hydrophobic core. This arrangement allows the hydrophilic monomers to stay in contact with water, thereby allowing substantial solubility. Examples of such polymers include proteins (which may contain hydrophobic amino acid residues).

Many of the so-called *biological polymers* consist of monomers that carry a net negative or positive charge at near neutral pH. These are known as *polyelectrolytes*, and they are generally very soluble in water. Their solubility is driven by the electrostatic interactions between water and the charged monomers. Examples of such polymers include DNA, proteins, certain derivatized cellulose polymers, and carrageenans. Such polymers are of significant importance in pharmaceutical dosage forms as thickeners, additives, stabilizers, and controlled-release matrices.

Many biological polymers exist as random coil structures in aqueous solution. If the structure is treated as an approximate sphere, then its radius, known as the *radius of gyration* (R_g), is a function of its molecular weight. In polymers of very high molecular weight (typically 100 kDa or higher) this radius may be so large that the polymer in solution behaves like a particle, approaching the size of the colloidal range. The volume of this particle is given by ¹⁴:

$$V_{\text{coil}} = \frac{4}{3} \pi R_g^3 \quad (30)$$

where V_{coil} is the volume of a single polymer chain and R_g is the radius of gyration. When the value of this volume is large, the system no longer behaves as a dilute solution, even when the molar concentration is small and polymer-polymer interactions are significant. Depending on the polymer molecular weight, significant overlapping between the polymer chains may occur at concentration as low as 0.1%.¹⁴ At higher concentration the swollen polymer and free solvent may occupy comparable volumes in the solution.

Unlike in regular solutions, the solubility of polymers is driven primarily by entropic changes. Upon mixing a polymer with a solvent, which is generally water in pharmaceutical solutions, two different kinds of entropic effects occur. One is the increase in entropy due to mixing of two molecular species. This effect is small in a dilute solution. The second effect is that the entropy of the polymer configuration increases, due to swelling of the molecules and also to greater flexibility in solution. Based on these entropic changes, Flory^{15,16} derived Equation 31 to describe the overall entropic change (ΔS_{mix}) in a polymer solution:

$$\Delta S_{\text{mix}} = -R(n_s \ln \Phi_s + n_p \ln \Phi_p) \quad (31)$$

where n_s and n_p are the number of molecules of solvent and polymer, respectively, and Φ_s and Φ_p represent their volume fraction, respectively. The free-energy change (ΔG_{mix}) in the process of solubility can be written as:

$$\Delta G_{\text{mix}} = RT(n_s \ln \Phi_s + n_p \ln \Phi_p) + (n_s + N_p n_p) \varpi \Phi_p \Phi_s \quad (32)$$

The first term on the right side of Equation 32 is the entropy of mixing, and the second term is the enthalpy of mixing. n_p in the second term is the degree of polymerization, and ϖ is the effective molar interaction parameter (effectively, ϖ is the square of the difference between solubility parameters of the polymer and solvent, multiplied by Avogadro's number). It is clear from the above equation that the value of ΔG_{mix} and, therefore, the polymer solubility are driven primarily by the volume fraction of the polymer in solution.

METHODS TO INCREASE SOLUBILITY OF POORLY SOLUBLE DRUGS

A large number of promising drug candidates do not make it to the market because of poor bioavailability, due primarily to their poor solubility in aqueous medium. Recently, several strategies have been used to improve solubility profile of these drugs and include the following:

1. Use of buffers
2. Use of cosolvents
3. Use of surfactants
4. Complexation
5. Solid dispersions.

Use of Buffers

The idea behind the use of buffers to improve solubility is to create and maintain pH conditions in a system that causes the drug to be in its ionized state. As discussed previously in this chapter, the ionized fraction of a drug is much more soluble in water, due to its increased polarity relative to the unionized fraction. Buffers can also help in reducing the likelihood of drug precipitation when drug solution is diluted in an aqueous medium. Consistent with the principles of solubility changes with pH, acidic drugs are formulated under relatively basic conditions while the opposite is true for the basic drugs. Some examples of drugs that are formulated with buffer systems are Amikacin sulfate (pH 3.5–5.5, citrate buffer) and Midazolam hydrochloride (pH 3).^{17–19} The drugs that make good candidates for use of pH variation or buffers are the ones that have the ability to ionize within a pH range of 2–8.

Use of Cosolvents

A common way to increase drug solubility is through the use of a water-miscible organic solvent. This strategy is based on the fact that poor solubility of drugs in water results from the great difference in polarity of the two components, water being of very high polarity and the drug having low polarity. Addition of a cosolvent with a polarity value of less than that of water reduces the difference between polarity of the drug and water cosolvent system, thereby improving solubility. Commonly used cosolvents for this purpose are the hydrogen bonding organic solvents such as ethyl alcohol, propylene glycol, and glycerin.

The polarity scale of solvents is defined by a property known as the dielectric constant. This value for water is 80, and for ethyl alcohol, propylene glycol, and glycerin, it is 24, 32, and 42, respectively. Most poorly soluble drugs have dielectric constant values of less than 20. Examples of some parenteral solution that contain cosolvents include chlorthalidoxepoxide (25% propylene glycol), diazepam (10% ethyl alcohol and 40% propylene glycol), and digoxin (10% ethyl alcohol and 40% propylene

glycol). Non-polar and nonionizable drugs are good candidates for cosolvent systems.^{17–19}

Use of Surfactants

Surfactants are molecules with well defined polar and non-polar regions that allow them to aggregate in solution to form micelles. Non-polar drugs can partition into these micelles and be solubilized. Depending on the nature of the polar area, surfactants can be nonionic (e.g., polyethylene glycol), anionic (e.g., sodium dodecyl sulfate), cationic (e.g., trialkylammonium), and zwitterionic (e.g., glycine and proteins). Among these, the most commonly used ones are the anionic and non-ionic surfactants. Since the process of solubilization occurs due to presence of micelles, generally high concentrations of surfactants are needed to significantly improve drug solubility. One example of surfactant based solution is Taxol (paclitaxel), an anti-cancer drug that is solubilized in 50% solution of Cremophor. Other examples include valrubicin in 50% Cremophor, and cyclosporin in 65% Cremophor.^{17–19}

Complexation

Complexation is the association between two or more molecules to form a noncovalent-based complex that has higher solubility than the drug itself. From the solubility standpoint, complexes can be put into two categories the stacking complexes and inclusion complexes. Stacking complexation is driven by association of non-polar areas of the drug and complexing agent. This results in exclusion of the non-polar areas from contact with water, thereby reducing the total energy of the system. This aggregation is favored by large, planar, non-polar regions on the molecules. Stacking can be homogeneous or mixed but results in a clear solution.

Inclusion complexes are formed by insertion of drug molecule into a cavity formed by the complexing agent. In this arrangement the non-polar area of the drug molecule is excluded from water, due to its insertion in the complexing agent. One requirement for the complexing agent in such systems is that it has a non-polar core and a polar exterior. The most commonly used inclusion complexing molecules are cyclodextrins. The cyclic oligomers of glucose are relatively soluble in water and have cavities large enough to accept non-polar portions of many drug molecules. Cyclodextrins can consist of six, seven, or eight sugar residues and are classified as α , β and γ , respectively. Due to geometric considerations, steroid molecules are very suitable for inclusion into cyclodextrin complexes.

Solid Dispersions

Solid dispersion refers to the dispersion of one or more active ingredients in an inert carrier or matrix at solid state, prepared by the melting (fusion), solvent, or the melting-solvent method. It has also been defined as the product formed by converting a fluid drug-carrier combination to the solid state. The term co-precipitate or co-evaporate has also been used frequently when a solid dispersion is prepared by the solvent method.

CLASSIFICATION OF SOLID DISPERSIONS

Solid dispersions can be classified as follows:

- Simple eutectic mixtures
- Solid solutions
- Glass solutions of suspensions
- Compound or complex formation between the drug and the carriers
- Amorphous precipitations of drug in crystalline carrier.

SIMPLE EUTECTIC MIXTURES

A simple eutectic mixture consists of two compounds that are completely miscible in the liquid state but only to a very limited extent in the solid state. A eutectic mixture of a sparingly water-soluble drug and a highly water-soluble carrier may be regarded

thermodynamically as an intimately blended physical mixture of its two crystalline components. These components are assumed to crystallize simultaneously in very small particulate sizes. The increase in specific surface area, therefore, is mainly responsible for the increased rate of dissolution of a poorly water-soluble drug.

Differential thermal analysis (DTA) of binary mixtures normally exhibits two endotherms, but a binary mixture of eutectic composition usually exhibits a single major endotherm. In the case of a simple eutectic system, the thaw points of binary mixtures of varying compositions are equal to the eutectic temperature of the system.

SOLID SOLUTIONS

A solid solution consists of a solid solute dissolved in a solid solvent. The particle size in solid solution is reduced to molecular level. Successful solubilization of itraconazole has been achieved using solid solution techniques. Solid solutions of lower drug concentrations generally give faster dissolution rate, and drug dissolution improves considerably with an increase in molecular weight of a water-soluble polymer, such as polyethylene glycol.

GLASS SOLUTIONS OF SUSPENSIONS

A glass solution is a homogeneous system in which a glassy or a vitreous form of the carrier solubilizes drug molecules. PVP has been used as a carrier in several formulations. In its matrix form, PVP dissolved in an organic solvent, undergoes a transition to a glassy state upon evaporation of the solvent.

COMPOUND OR COMPLEX FORMATION BETWEEN DRUG AND CARRIERS

This system is characterized by complexation of two components in a binary system during solid dispersion preparation. The availability of a drug from the complex is dependent on the solubility, dissociation constant, and intrinsic absorption rate of the complex. α , β and γ Cyclodextrins in combination with polyethylene glycol (PEG) 6000, have been used to formulate such systems.

AMORPHOUS PRECIPITATION

Amorphous precipitation occurs when the drug precipitates as an amorphous form in the inert carrier. The high-energy state of the drug in this system generally produces much greater dissolution rates than the corresponding crystalline forms of the drug.

REFERENCES

1. Sinko PJ. *Martin's Physical Pharmacy and Pharmaceutical Sciences*. Philadelphia: Lippincott Williams & Wilkins, 2006; pp 231–66.
2. Kramer SF, Flynn GL. Solubility of organic hydrochlorides. *J Pharm Sci* 1972; 61: 1896–904.
3. Campbell AN, Campbell AJR. Concentrations, total and partial vapor pressures, surface tensions and viscosities, in the systems phenol-water and phenol-water-4% succinic acid. *J Am Chem Soc* 1937; 59: 2481–8.
4. Loran MR, Guth EP. A phase study of castor oil, alcohol, and water miscibility. *J APhA Sci Ed* 1951; 40: 465–6.
5. Grant DJW, Abougela IKA. Physicochemical interactions in pharmaceutical formulations. *Anal Proc* 1982; 19: 545–9.
6. Higuchi T. In: Lyman R, ed. *Pharmaceutical Compounding and Dispensing*. Philadelphia: Lippincott, 1949; p 159.
7. Duddu S. PhD thesis. University of Minnesota, 1993.
8. Yalkowsky SH. *Techniques of Solubilization of Drugs*. New York, NY: Dekker, 1981; p 91.
9. Martin A *et al.* Extended Hildebrand solubility approach and the log linear solubility equation. *J Pharm Sci* 1982; 71: 849–56.
10. Setnikar I, Temelcou O. Osmotic concentration and osmotic pressure in injectable solutions. *J APhA Sci Ed* 1959; 48: 628–30.

11. Hilderbrand JH, Wood SE. The derivation of equations for regular solutions. *J Chem Phys* 1933; 1: 817–22.
12. Hilderbrand JH et al. *Regular and Related Solutions*. New York, NY: Van Nostrand Reinhold, 1970; pp 22–3.
13. Hilderbrand JH, Scott RL. *Solubility of Nonelectrolytes*. New York, NY: Dover, 1964; Chap 23.
14. Evans DF, Wennerstrom H. *The Colloidal Domain—Where Physics, Chemistry, Biology, and Technology Meet*. New York, NY: VCH Publishers, 1994; pp 289–303.
15. Flory PJ. *Principles of Polymer Chemistry*. Ithaca, NY: Cornell University Press, 1953.
16. Flory PJ. *Statistical Mechanics of Chain Molecules*. New York, NY: Interscience, 1969.
17. Strickley RG. Parenteral formulations of small molecules therapeutics marketed in the United States (1999)—Part I. *PDA J Pharm Sci Technol* 1999; 53: 324–49.
18. Strickley RG. Parenteral formulations of small molecule therapeutics marketed in the United States (1999). Part II. *PDA J Parenteral Sci Technol* 2000; 54: 69–96.
19. Strickley RG. Parenteral formulations of small molecules therapeutics marketed in the United States (1999). Part III. *PDA J Pharm Sci Technol* 2000; 54: 152–69.

Chapter 14

Separation Methods

Lloyd V. Allen Jr, PhD, RPh

FILTRATION	241
MATHEMATICS OF FILTRATION	242
FILTERING MEDIA	242
FILTRATION AIDS	244
RAPID FILTRATION APPARATUS	245
CENTRIFUGATION	246

PRECIPITATION	249
SEPARATION OF IMMISCIBLE LIQUIDS	249
EXPRESSION	249
COUNTERCURRENT DISTRIBUTION	250
OTHER SEPARATION TECHNIQUES	252

Separation may be defined as an operation that brings about isolation and/or purification of a single chemical constituent or a group of chemically related substances. Most medicinal agents require some degree of purification before being incorporated into desirable dosage forms. Many times the analysis of pharmaceutical preparations requires separation of the chief constituent from other formulation constituents before quantitative measurement can be made.

Although the problems of separation are the concern chiefly of pharmaceutical manufacturers, at times they may be encountered by the pharmacist in the prescription laboratory; hence, all pharmacy practitioners should have knowledge of the underlying principles and the techniques employed in the basic processes of separation.

The processes of separation may be divided into two general categories—simple and complex—depending on the complexity of the method used.

Simple processes bring about separation of constituents through a single mechanical manipulation. Processes in this category are limited usually to separations of relatively simple mixtures or solutions. Some examples of this type are the use of:

- a separatory funnel or pipette to separate two immiscible liquids such as water and ether
- a distillation process to separate two miscible liquids such as benzene and chloroform
- a garbling process to separate solids
- centrifugation, filtration, and expression processes to separate solids from liquids.

Complex processes usually require formation of a second phase by the addition of either a solid, liquid, or gas plus mechanical manipulation to bring about effective separation. One example is the separation of aspirin (acetylsalicylic acid) from salicylic acid. In this mixture, salicylic acid is considered to be an impurity, and to separate the impurity from the desired constituent, a suitable solvent is added to the mixture for the purpose of recrystallizing only the acetylsalicylic acid. The contaminant remains in solution and is removed in the filtrate during the filtration process.

Only selected processes involving separations will be covered in this chapter. Other methods are discussed in such chapters as *Complex Formation* (Chapter 18), *Colloidal Dispersions* (Chapter 20), and *Coarse Dispersions* (Chapter 21).

FILTRATION

Filtration is the process of clarifying, harvesting, or separating particulate matter from a liquid or gas. Filtration is used to separate solid impurities from liquids and gases and to prepare a filtrate free of unwanted suspended substances. The desired

component may be the retained particulates on the filter or the liquid passing through the filter, as shown in Figure 14-1. Molecular filtration involves a process of ultrafiltration.

This is accomplished by the intervention of a porous substance, called the *filter* or the *filtering medium*. The liquid that has passed through the filter is called the *filtrate*. The *retentate* or *residue* is the portion of the sample that does not pass through the filter.

Filters are available in many different pore sizes, physical configurations and chemical compositions. Commonly, filters are divided into two broad classes: Depth and Membrane.

Depth filters consist of a random fiber matrix, bonded together to form a maze of flow channels. Particulate removal results from entrapment by, or adsorption to, the filter matrix. Generally, 95% of the particles larger than the manufacturers stated pore size will be retained by depth filters when gravity or vacuum is used. Examples of depth filters include glass fiber, paper (cellulose) fiber and fritted glass.

Membrane (screen, surface) filters remove all particles larger than the specified pore size, thus removing 100% of these materials from the filtrate. They are sometimes called “absolute filters.” Their major limitation is the low particle holding capacity. Membrane filters are composed of either natural or synthetic materials such as various cellulose derivatives and polymers.

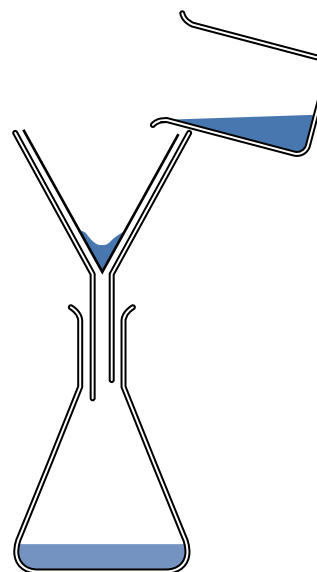


Figure 14-1. The process of filtration.

MATHEMATICS OF FILTRATION

In 1842, Poiseuille proposed a relationship for streamlined flow of liquids under pressure through capillaries. This equation in its simplified form is represented by

$$V = \frac{\pi \Delta p r^4}{8 L \eta}$$

where V = flow velocity, r = flow capillary radius, L = capillary length, η = viscosity of the fluid, and Δp = pressure differential at the two ends of the capillary.

The modified Poiseuille equation has been shown to be valid for liquid flow through sand, glass beads, and various porous media. It represents the foundation for all mathematical models of filtration that were developed subsequently. Of critical importance in this equation is the powerful effect of capillary radius; i.e., by reducing it to 1/8 its original size, the pressure differential must be increased more than 4000 times in order to obtain the same flow velocity, all other factors remaining constant.

On the basis of the Poiseuille formula, the Kozeny-Carman relationship was established. This may be expressed as

$$V = \left[\frac{e^3}{KS^2(1-e)^2} \right] \left[\frac{A \Delta p g}{\eta L} \right] \quad (1)$$

where A = cross-sectional area of porous bed (filter medium), e = porosity of bed, S = surface area of medium, K = constant, and the remaining symbols are the same as in the Poiseuille equation.

The Kozeny-Carman relationship, like Poiseuille's law states that the rate of flow is directly proportional to the pressure drop across the medium and to the area of the bed, and inversely proportional to the viscosity of the liquid and the thickness of the bed. To characterize the material composing the bed, two new quantities, e and S , are introduced, replacing capillary radius.

The use of a non-definite constant K , rather than the definite constant in Poiseuille's equation, $\pi/8$, offers greater utility in the use of this equation in accounting for the geometry of the medium. The constant, K , generally ranges in value from 3 to 6. The Kozeny-Carman equation finds its greatest limitation in complex systems such as filter paper, but provides excellent correlation in filter beds composed of porous material.

In applying Poiseuille's law to filtration processes, one must recognize the capillaries found in the filter bed are highly irregular and non-uniform. Therefore, if the length of a capillary is taken as the thickness of the bed or medium and the correction factor for the radius is applied, the flow rate is more closely approximated. These factors have been taken into account in the formulation of the Darcy equation

$$V = (k \Delta p) / (L \eta) \quad (2)$$

where k is the permeability coefficient and depends on the nature of the precipitate to be filtered and the filter medium itself.

FILTERING MEDIA

The filtering medium, whether a filter paper, synthetic fiber, or porous bed of glass, sand, or stone, is composed of countless channels that impart porosity to the medium. Almost without exception these channels or pores are nonuniform and possess a rather tortuous nature.

The mechanism of filtration basically involves a two-step process:

1. The filter medium itself resists the flow of solid material while permitting the passage of liquid.
2. During the course of the filtration the suspended, solid material builds up on the filter medium and thereby forms a filter bed, which acts as a second, and often more efficient, filter medium.

The ability of a filter medium to eliminate solid matter from a liquid is termed *retention*. It must be borne in mind that the filtration process must compromise retention with filtration rate, the speed at which the purified liquid (the filtrate) is recovered. To illustrate this point, it will be noted that a slab of marble will most effectively retain the solid material contained in a suspension; unfortunately, it would require a few centuries to collect the purified filtrate.

Both the retentive ability of a filter medium and filtration rate of a liquid through the medium depend on the porosity of the medium. Each factor, however, is influenced significantly by the viscosity and nature of the liquid, the proportion of solid matter in the liquid, and the size, shape, and physical nature of the suspended solids.

The flow of a liquid through a filter bed follows the same basic rules that govern the flow of any liquid through a medium offering resistance. The *flow rate* through the medium will vary directly with the area of the medium, as well as the pressure drop or driving force across the bed.

$$\text{Rate of flow} \propto \frac{(\text{driving force})(\text{cross-sectional area})}{\text{resistance}} \quad (3)$$

The flow rate is retarded by the viscosity of the liquid being filtered and by any obstruction to flow. These obstructions include the resistance of the filter medium itself and the second filter bed or *filter cake* that builds up on the medium at a rate dependent on the solids content of the liquid. In considering the nature of the precipitate, it is known that large particles are easier to filter than are small particles because of the tendency of the latter to enter into and occlude the pores of the bed, thus hindering the passage of the filtrate. In addition, the buildup of small particles on the filter tends to form a nonporous, densely packed bed that also resists passage of the filtrate. The resistance offered by the medium itself will not vary significantly during the filtration process. It depends on the thickness of the medium as well as its porosity. The resistance of the filter cake, on the other hand, is not constant and generally increases continuously during the operation. The resistance offered by the cake depends both on its thickness and physical nature. The thickness of the cake is dictated by the amount of filtrate passing through the filter and on the solids content of the liquid. The physical nature of the cake—whether it is loose, compacted, coarse, fine, granular, or gelatinous—determines whether or not it will readily allow the flow of liquid.

Considerations in selecting a filter include (1) chemical composition, (2) surface area, (3) filter holder, (4) extractables, and (5) sterility requirements. After considering these factors, the selection of the proper filter should be greatly narrowed and the final selection may need to be experimentally determined with respect to resolution, capacity, speed, and recovery.

The key factor in choosing a separation method is its ability to discriminate on the basis of one or more molecular properties of the sample components. When choosing a separation method, give primary consideration to those processes that emphasize molecular properties in which the components differ to the greatest extent.

There are four common driving mechanisms used in filtration.

1. gravity (atmospheric pressure)
2. low pressure vacuum (1–15 psi)
3. high pressure syringe or pump (15–100 psi)
4. centrifugation.

TYPES OF FILTRATION MEDIA

Filter Paper

Filter paper most frequently is employed in clarification processes required of the pharmacy practitioner. Only high-quality filter paper should be used to ensure maximum filtering efficiency (See Figure 14-2). When possible the first few milliliters



Figure 14-2. Example of filter paper used in laboratory filtrations. (Courtesy of PCCA-Professional Compounding Centers of America.)

of filtrate should be discarded to eliminate (insofar as possible) contamination of the pharmaceutical product by free fibers associated with most filter paper.

Membrane Filters

Membrane filter media are produced from pure cellulose, cellulose derivatives, and polymeric materials. All have an extremely uniform micropore structure as well as an exceptionally smooth surface. The integral structure contains no fibers or particles that can work loose and contaminate a filtrate. This is a particular advantage in the filtration of ophthalmic solutions. The presence of these fibers is difficult to prevent when using many other filter media, including paper filters.

The efficiency of membrane filters is due to the uniform pore system that functions like a highly effective sieve. The pore size, of different types of these filters, ranges from 10 nm to 100 microns. All particles in liquids or gases that are larger than the pore of a given filter are retained on the surface. The thickness of these membrane filters ranges from 50 to 200 microns.

The pores that penetrate these filters pass directly through the entire thickness of the membrane, with a minimum of cross-linkage. Porosity or pore volume is estimated as 80% of the total fiber volume. The high porosity of these filters, coupled with the *straight-through* configuration of the pores, results in flow rates through the membrane filters that are at least 40 times faster than flow rates through conventional filter media that possess the same particle size retention capabilities.

Membrane (surface, screen) filters remove all particles larger than its specified pore size. It is not true that all particles smaller than the pore size go through the filter. Since particles larger than the pore size deposit on the filter surface, this new layer can act as a depth filter trapping particles smaller than the rated pore size. Agitation can reduce this problem for pore sizes down to 0.1 micron; but with smaller pore sizes, the adhesion forces between particles are too strong to dislodge them from the filter surface or each other.

Since the pore size of membrane filters can be closely controlled, it is possible to assign an absolute pore size rating. The major limitation of these filters is their low particle holding capacity. Prefilters (depth filters) may be used in series with membrane filters to increase the total particle capacity.

Membrane manufacturers have standardized certain diameters which range from 13 to 293 mm. Some common diameter sizes include the 25, 47, 90, 142, and 293 mm filters. Different types are available for use in the filtration of either aqueous or nonaqueous liquids. The discs generally are used in conjunction with specialized holders of either plastic, metal or glass composition. With small volumes (i.e., less than 500 mL), solutions usually are filtered using vacuum techniques. Larger volumes require filtration under pressure provided by an inert gas such as nitrogen. An example of a membrane filter is shown in Figure 14-3.

In addition to their obvious utility in routine filtration processes on both a laboratory and industrial scale, these filters have been used for a wide range of purposes, including chemical analysis, microbiological analysis, and bacterial filtration. The latter process provides an economical and rapid method for sterilizing heat-labile material (see Chapter 25).

Handling guidelines with membrane filters:

1. Never handle the filter with fingers. Always use a blunt, curve-typed, nonserrated forcep.
2. Always remove the “separator papers” (blue or yellow usually) before using.

Other Filtering Media

Many devices have been advanced to replace filter paper, which has many disadvantages, particularly for large-scale operations. A great many variations of filtering processes, each designed to fit the needs of special cases, are found in the modern pharmaceutical laboratory. The filter press, the centrifugal filter, the vacuum filter, sand-bed filter, charcoal filter, paper-pulp filter, and porous porcelain filter are all examples of specialized filtration methods. Each one of these possesses some advantageous quality, and it is the experience of the laboratory operators that guides them in their selection of appropriate filtering devices. Reference is made later in the text to many of these special-scale filters.

However, it would not be inappropriate to refer briefly to special filtering devices that may be useful in the prescription or research laboratory.

Cotton Filters—A small pledget of absorbent cotton, loosely inserted in the neck of a funnel, adequately serves to remove large particles of extraneous material from a clear liquid. Although this properly might be termed colation, the cotton also can be used to serve as a fairly efficient filter. It is sometimes necessary to return the liquid a number of times to secure perfect transparency.



Figure 14-3. Example of a filter membranand holder (Luer lock syringe membrane filter). (Courtesy of PCCA-Professional Compounding Centers of America.)

Glass-Wool Filters—When solutions of highly reactive chemicals, such as strong acids, are to be filtered, filter paper cannot be used. In its place glass wool may be used just as one uses absorbent cotton for filtering. This material is resistant to ordinary chemical action, and when properly packed into the neck of a funnel it constitutes a very effective filtering medium.

Sintered-Glass Filters—These filters have as the filtering medium a flat or convex plate consisting of particles of Jena glass powdered and sifted to produce granules of uniform size that are molded together. The plates can be fused into a glass apparatus of any required shape (Figure 14-4). These filters vary in porosity, depending on the size of the granules used in the plate. They were formerly used in the filtration of solutions such as those intended for parenteral injection but have been replaced by 0.22 micron membrane filters. A vacuum attachment is necessary to facilitate the passage of the liquid through the filter plate (see Chapter 25).

Fritted glass filter—designations are Extra Coarse (170–222 microns), Coarse (40–60 microns), Medium (10–15 microns), Fine (4–5.5 microns), Very fine (2–2.5 microns), and Ultrafine (0.9–1.4 microns).

Fiber glass filters are made with borosilicate glass and they possess the highest retention capacity of all depth filters. They also have a wider chemical compatibility than paper (cellulose) filters. An important precaution is “Do Not Fold” since folding may jeopardize the filter integrity. These filters are suitable for use in Gooch, Buchner, or membrane filtering apparatus.

Funnels

Funnels are conical-shaped utensils intended to facilitate the pouring of liquids into narrow-mouthed vessels. They also are used widely in pharmacy for supporting filter media. Funnels may be made of glass, polyethylene, metal, or any other material that serves a specific purpose. The community pharmacist will find the glass funnel to be quite adequate for all processes of clarification in prescription practice. Most funnels used by the pharmacy practitioner are conical in shape and may be fluted, grooved, or ribbed for the purpose of facilitating the downward flow of the filtrate, as shown in Figure 14-5.

The *Büchner* type of funnel is used today largely in pharmaceutical laboratories. A piece of round filter paper is laid on the perforated porcelain diaphragm and the filtration conducted. This funnel is especially applicable to vacuum filtration, as shown in Figure 14-6 (see the discussion, *Vacuum Filtration*).

Filtration of Volatile Liquids

It is evident that the ordinary methods of filtering liquids will not be practical for very volatile liquids because of the loss through evaporation, and the liability to explosion in the case of flammable volatile liquids. Funnels must be covered, the receiving vessel closed, and provision made for the escape of the confined air in the receiving vessel. The following method is quite useful. A rubber cover, perforated to admit a tube, is placed on

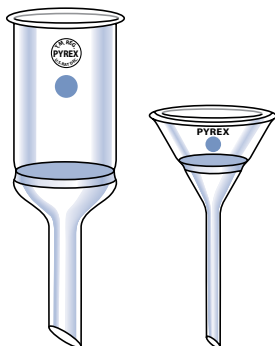


Figure 14-4. Sintered-glass filters.

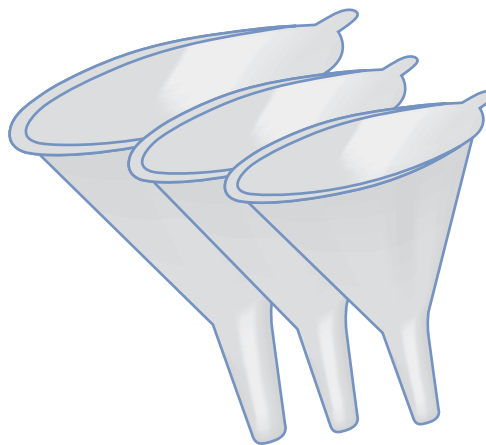


Figure 14-5. Typical funnel apparatus.

top of the funnel; connection between the bottle and funnel is effected as shown in Figure 14-7.

FILTRATION AIDS

It has long been known that addition of an insoluble adsorbent powder to a liquid prior to its filtration greatly increases the efficiency of the process. Purified talc, siliceous earth (kieselguhr), clays, charcoal, paper pulp, chalk, magnesium carbonate, bentonite, silica gel, and others have been used for this purpose. It must not be overlooked, however, that powdered substances employed for such purposes must be insoluble and inert, so not all of those in the foregoing list are applicable for general filtration.

Talc is nonadsorbent to materials in solution and is a chemically inert medium for filtering any liquid, provided it has been purified for this purpose and it is not the impalpably fine variety that will pass through the filter paper.

Kieselguhr is almost pure silica (SiO_2). It is as applicable as talc for general filtration purposes, with no danger of removing active constituents by adsorption.

Siliceous earths or clays, such as fuller's earth or kaolin in the hydrated form which is produced when they are brought into contact with aqueous liquids, are safe for general use only in filtering fixed oils. Liquids containing coloring matter or

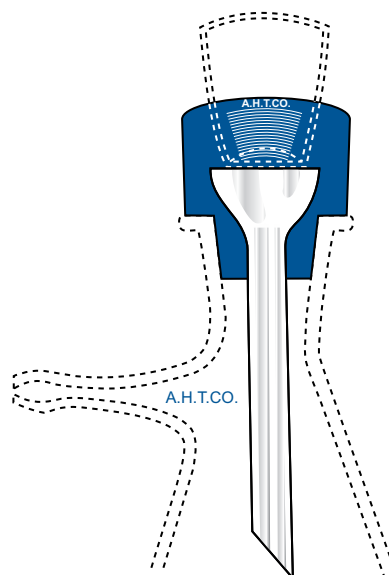


Figure 14-6. Gooch crucible and Büchner funnel apparatus. (Courtesy of Thomas.)

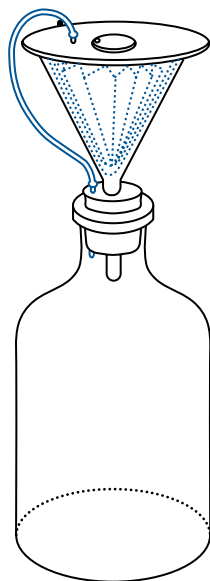


Figure 14-7. Filtration of volatile liquids.

alkaloidal principles must not be filtered through these media, for adsorption of both color and alkaloids occurs and the filtrate is altered in comparison.

Charcoals, as a rule, possess adsorptive properties not only toward color but for many active constituents of medicinal preparations, such as alkaloids and glycosides. Consequently, charcoal should never be used as a filtering medium unless the removal of such constituents is desirable.

Chalk and *magnesium carbonate* readily react with acids and possess a finite solubility in water and aqueous fluids, with the production of alkalinity in the filtrate. This is particularly true of magnesium carbonate; the degree of alkalinity imparted to the filtrate is sufficiently great to cause precipitation of alkaloids. Either of these media, when added to an alkaloidal preparation prior to filtration, will precipitate and remove all of the alkaloidal constituents. Neither is suitable for general use.

A *prefilter* (a depth filter) is used prior to a membrane filter to prevent premature clogging and blocking of the membrane filter. In some cases, a pure and inert powder-like material can be used to form a porous film or cake on the surface of depth filters. Filter aids include diatomaceous earth, silica or activated charcoal. Filter aids will reduce filter pore clogging and thus maintain an adequate filtration speed.

RAPID FILTRATION APPARATUS

Much attention has been given to methods for increasing the rapidity of filtration. This may be accomplished by applying pressure on the filter or by creating a vacuum in the receiving vessel.

VACUUM FILTRATION

One of the first practical efforts made to create a vacuum to aid filtration was by means of the Bunsen pump. Its action depends on the principle that a column of water descending through a tube from a height is capable of carrying with it the air contained in a lateral tube, if the latter is placed properly. This form of aspirator is practicable where water pressure is available.

Pumps Acting by Water Pressure—The various aspirator or vacuum pumps that operate under the influence of water pressure are all based on the same principle. The following are selected for illustration from the great variety in use. Figure 14-8 shows Chapman's vacuum pump. Valve *a* prevents the water from flowing into the bottle which carries the filter when the pressure of water ceases or is reduced and *b* is an inline restrictor.

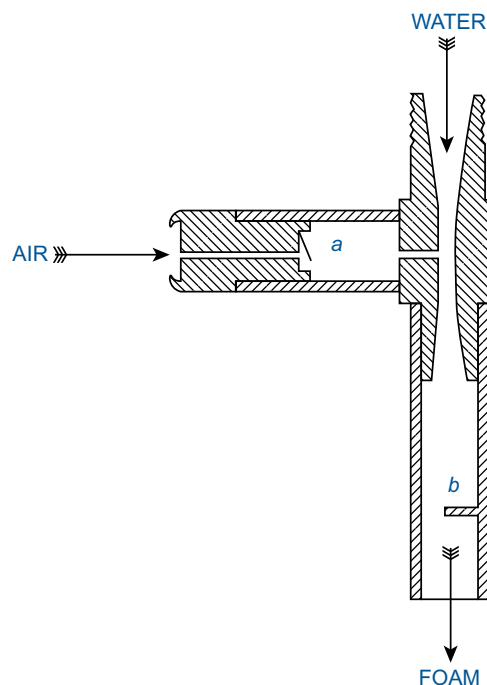


Figure 14-8. Chapman's Vacuum Pump.

On a larger scale, the vacuum for filtration is produced by one of the many types of vacuum pumps now available. The pump should be protected from vapors by placing a suitable vapor trap between the filter unit and the pump. The trap usually is cooled to very low temperatures by means of dry ice and acetone when very high vacuum is needed.

In assembling a filtering apparatus using the vacuum principle, it is necessary that there be no leaks in the connections from the filter to the aspirator. If filter paper is used in connection therewith, a plainly folded paper must be used and its tip must be protected against breakage by reinforcing it with a filter paper support or some other device. A Büchner filter also may be used, employing a specially strong filter paper.

In analytical work it is customary to use the Gooch crucible and flask (Figure 14-6) for rapid filtration. The flask, of especially thick glass, is provided with a side tube that is connected to a water aspirator pump. The perforated crucible bottom is converted into a filter bed of the required thickness by means of a filter mat placed over the perforations in the porcelain base.

PRESSURE FILTRATION

Figure 14-9 illustrates a sectional drawing of a plate-and-frame filter press. The material to be filtered enters the apparatus under pressure through a pipe at the bottom and is forced into one of the many chambers. A filter cloth is positioned on both sides of each chamber. As the material passes through the filtering cloth, solids remain behind in the chamber and the clear filtrate passes through and out of an opening located on top of the apparatus.

Rotary-drum vacuum filters are used widely in the pharmaceutical industry, especially in the preparation of antibiotics by the fermentation process. In this type of filtration a perforated drum, wrapped with a cloth or other suitable substance holding a filter medium, is immersed partially in a tank holding the material to be filtered (Figure 14-10).

The drum is rotated through the slurry of material and a vacuum within the drum draws the material into and through the filter medium. During this step of the process, the filtrate is taken into the drum and collected, while the solid material remains deposited on the outer surface of the drum. This material is then removed by a scraper in the last step of the operating cycle, just before the rotating drum repeats another cycle.

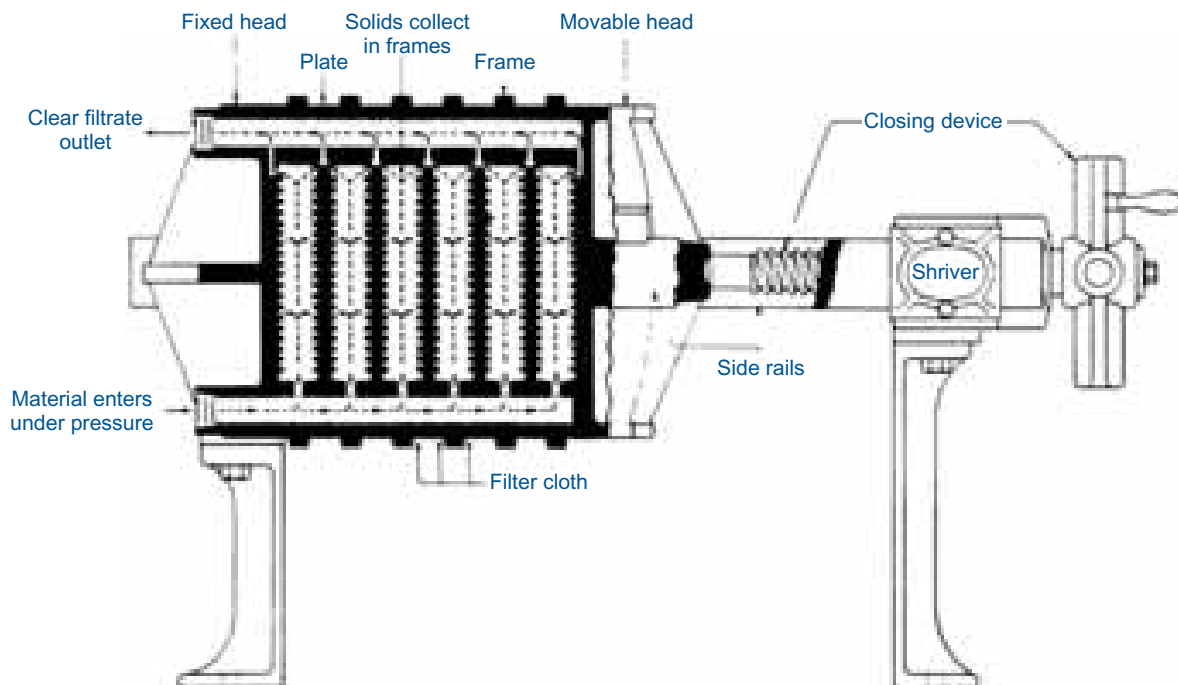


Figure 14-9. A plate-and-frame filter press. (Courtesy of Shriver.)

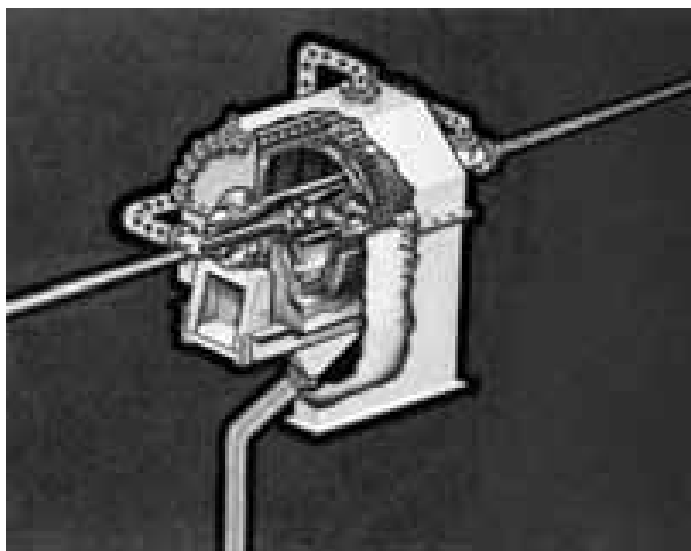


Figure 14-10. Rotary filter. (Courtesy of Andritz Separation, Inc.)

CENTRIFUGATION

Centrifugation is useful particularly when separation by ordinary filtration is difficult, as in separating a highly viscous mixture. A diagram is shown in Figure 14-11. Separations may be accomplished more rapidly in a centrifuge than under the action of gravity. In addition, the degree of separation that is attainable may be greater because the forces available are of a far higher order of magnitude.

The apparatus consists essentially of a container in which a mixture of solid and liquid, or of two liquids, is rotated at high speeds so that the mixture is separated into its constituent parts by the action of centrifugal force. A solid or liquid, mixed with a liquid of lesser density, may be separated because the substance of higher specific gravity is thrown outward with greater force—it will be impelled to the bottom of the container, leaving a clear supernatant layer of pure liquid.

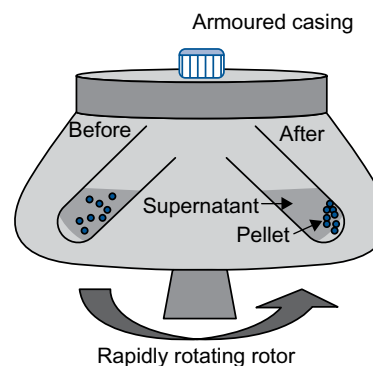


Figure 14-11. Example of the before and after centrifugation process.

Two basic types of centrifuges are available: *sedimentation* and *filtration*. The *sedimentation* type of centrifuge depends on differences in the densities of the two or more phases comprising the mixture. This instrument is capable of separating both solid–liquid and liquid–liquid mixtures. *Filtration centrifuges*, however, are limited to the separation of solid–liquid mixtures.

DIFFERENTIAL CENTRIFUGATION

Separation is achieved primarily based on the size of the particles in differential centrifugation. This type of separation is commonly used in simple pelleting and in obtaining partially-pure preparation of subcellular organelles and macromolecules. For the study of subcellular organelles, tissue or cells are first disrupted to release their internal contents. This crude disrupted cell mixture is referred to as a homogenate. During centrifugation of a cell homogenate, larger particles sediment faster than smaller ones and this provides the basis for obtaining crude organelle fractions by differential centrifugation. A cell homogenate can be centrifuged at a series of progressively higher g-forces and times to generate pellets of partially-purified organelles.

When a cell homogenate is centrifuged at $1000 \times g$ for 10 minutes, unbroken cells and heavy nuclei pellet to the bottom of the tube. The supernatant can be further centrifuged at $10,000 \times g$ for 20 minutes to pellet subcellular organelles of intermediate velocities such as mitochondria, lysosomes, and microbodies. Some of these sedimenting organelles can be obtained in partial purity and are typically contaminated with other particles. Repeated washing of the pellets by resuspending in isotonic solvents and repelleting may result in removal of contaminants that are smaller in size. Obtaining partially-purified organelles by differential centrifugation serves as the preliminary step for further purification using other types of centrifugal separation (density gradient separation).

DENSITY GRADIENT CENTRIFUGATION

Density gradient centrifugation is the preferred method to purify subcellular organelles and macromolecules. Density gradients can be generated by placing layer after layer of gradient media such as sucrose in a tube with the heaviest layer at the bottom and the lightest at the top in either a discontinuous or continuous mode. The cell fraction to be separated is placed on top of the layer and centrifuged. Density gradient separation can be classified into two categories, rate-zonal (size) separation and isopycnic (density) separation.

Rate-zonal (size) separation

Rate-zonal separation takes advantage of particle size and mass instead of particle density for sedimentation. Figure 14-12 illustrates a rate-zonal separation process and the criteria for successful rate-zonal separation. Examples of common applications include separation of cellular organelles such as endosomes or separation of proteins, such as antibodies. For instance, Antibody classes all have very similar densities, but different masses. Thus, separation based on mass will separate the different classes, whereas separation based on density will not be able to resolve these antibody classes.

Certain types of rotors are more applicable for this type of separation than others. The criteria for successful rate-zonal

centrifugation include the following factors: (1) the density of the sample solution must be less than that of the lowest density portion of the gradient, (2) the density of the sample particle must be greater than that of the highest density portion of the gradient, (3) the pathlength of the gradient must be sufficient for the separation to occur, and (4) the time is important. If you perform too long runs, particles may all pellet at the bottom of the tube.

Isopycnic (density) separation

In this type of separation, a particle of a particular density will sink during centrifugation until a position is reached where the density of the surrounding solution is exactly the same as the density of the particle. Once this quasi-equilibrium is reached, the length of centrifugation does not have any influence on the migration of the particle. A common example for this method is separation of nucleic acids in a CsCl gradient. Figure 14-13 illustrates the isopycnic separation and criteria for successful separation. The criteria for successful isopycnic separations include the following factors: (1) the density of the sample particle must fall within the limits of the gradient densities, (2) any gradient length is acceptable, and (3) the run time must be sufficient for the particles to band at their isopycnic point. Excessive run times have no adverse effect.

SEDIMENTATION CENTRIFUGES

Bottle Centrifuge

The design of the bottle centrifuge and the disc centrifuge are based on the sedimentation principle (i.e., separation by density difference). The bottle centrifuge, which consists of a vertical spindle that rotates the containers in a horizontal plane, commonly is used to separate materials of different densities. Separation in a centrifugal field is brought about because denser particles in a mixture require greater forces to hold them in a circular path of a given radius than do lighter particles. Thus, the lighter particles are displaced toward the axis of the centrifuge by the heavier particles. During the centrifugation of

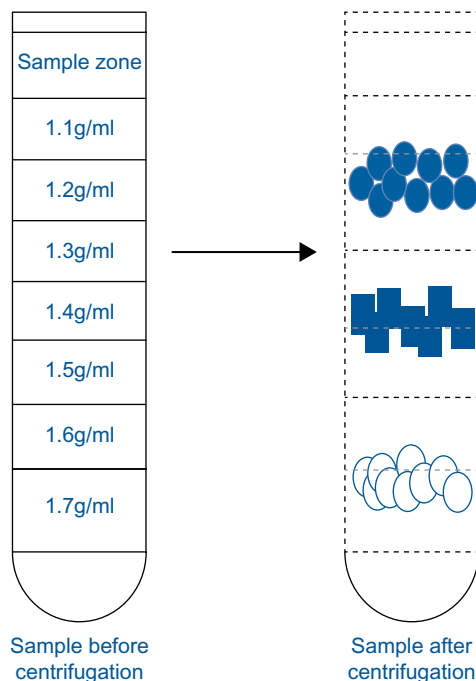


Figure 14-12. Example of size separation by centrifugation. (From Cole-Parmer Technical Library, Basics of Centrifugation. ©Thermo Fisher Scientific. <http://www.coleparmer.com/techinfo/techinfo.asp?htmlfile=basic-centrifugation.htm&ID=30>, accessed 30 November, 2011)

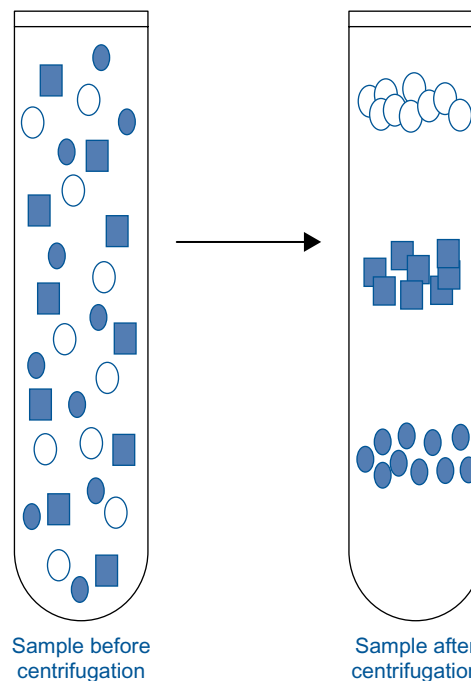


Figure 14-13. Example of density separation by centrifugation. (From Cole-Parmer Technical Library, Basics of Centrifugation. ©Thermo Fisher Scientific. <http://www.coleparmer.com/techinfo/techinfo.asp?htmlfile=basic-centrifugation.htm&ID=30>, accessed 30 November, 2011)

blood, for example, a speed of 3000 rpm is required to separate blood corpuscles from serum. If the radius of the centrifuge is assumed to be 10 cm, the acceleration, α , acting on a particle can be approximated to be 10^6 cm/sec^2 ; or about 1000 times the acceleration due to gravity, g

$$\alpha = 4\pi^2 N^2 r = \frac{4(3.14)^2 (3000)^2 (10)}{3600} = 10^6 \text{ cm/sec}^2$$

$$N = \text{revolution/sec}; r = \text{radius in cm} \quad (4)$$

$$\frac{10^6 \text{ cm/sec}^2}{10^3 \text{ cm/sec}^2} = 1000(g)$$

$$10^3 \text{ cm/sec}^2 = \text{approximate acceleration due to gravity}$$

Under these conditions, the blood corpuscles eventually migrate under the influence of centrifugal force to the tip of the centrifuge tube.

The separation of particles in a liquid medium also depends on the nature of the medium. A solid particle settling under the influence of acceleration due to gravity in a liquid phase accelerates until a constant terminal velocity is reached. The terminal velocity is known as the settling velocity of the particle and is described mathematically by Stokes' Law. It can be shown that Stokes' Law can be extended to those cases where settling takes place in a centrifugal field,

$$v_s = v_g \frac{\omega^2 r}{g} \quad (5)$$

where v_s is the settling velocity of a particle in a centrifugal field, v_g is the settling velocity of a particle in a gravitational field (Stokes' Law), ω is the angular velocity of the particle in the settling zone, and r is the radius at which the settling velocity is determined.

Consider a solid particle at an initial position in a liquid medium and a distance r from the axis of rotation. Under these conditions,

$$v_s = dr/dt \quad (6)$$

Substituting Equation 6 into Equation 5 gives

$$dr/dt = v_g \frac{\omega^2 r}{g} \quad (7)$$

Rearranging and integrating between limits gives

$$\int_r^{r_c} \frac{dr}{r} = \int_0^t v_g \frac{\omega^2 r}{g} dt \quad (8)$$

$$\ln \frac{r_c}{r} = v_g \frac{\omega^2 t}{g} \quad (9)$$

where r_c is the distance between the surface of the sedimented cake in the tip of the tube and the axis of rotation, and t is the time during which the particle is subjected to centrifugal acceleration while the particle travels the distance from r to r_c . Equation 9 shows that if centrifuging conditions for a given suspension are to be compared in different centrifuges, the speed, bottle size, centrifuge dimensions, and centrifuging time must be taken into consideration.

Filtration Centrifuge

The filtration centrifuge is restricted to the separation of solid-liquid mixtures. It is similar in principle to the sedimentation type, but rather than containers it possesses a porous wall through which the liquid phase may pass but upon which the solid phase is retained. Analogous to filtration, this process requires consideration of the flow of liquid through the solid bed that accumulates on the porous plate. The plate may be either solid or semisolid (gel).

Ultracentrifuge

When extremely fine solid matter must be separated from a liquid, such as in colloid or biological research, the ultracentrifuge is employed. In this instrument a relatively small rotor is operated at speeds exceeding 100,000 rpm and forces up to one million times gravity are exerted. High speeds are attained with air or oil turbines and bearings lubricated with a film of compressed air. Friction heat may be minimized by the use of high vacuum.

By placing the samples in specially constructed cells and spinning them in the ultracentrifuge, it is possible to separate the dispersed phase from the continuous phase rather rapidly. To aid the investigator, optical attachments may be employed to photograph the settling while the centrifuge is in operation.

Only small batches of material can be handled in these instruments during a single run. Ultracentrifuges are employed in the determination of particle size and molecular weight of polymeric and other high-molecular-weight materials such as proteins and nucleic acids by direct or indirect observation of the rate of separation of particles in solution or suspension.

Rotors

Rotors can be broadly classified into three common categories namely swinging-bucket rotors, fixed-angle rotors, and vertical rotors (Figure 14-14, Table 14-1). Note that each type of rotor has strengths and limitations depending on the type of separation. Other rotors include continuous flow and elutriation rotors.

In swinging bucket rotors, the sample tubes are loaded into individual buckets that hang vertically while the rotor is at rest. When the rotor begins to rotate the buckets swing out to a horizontal position (Figure 14-14). This rotor is particularly useful when samples are to be resolved in density gradients. The longer pathlength permits better separation of individual particle types from a mixture. However, this rotor is relatively inefficient for pelleting. Also, care must be taken to avoid

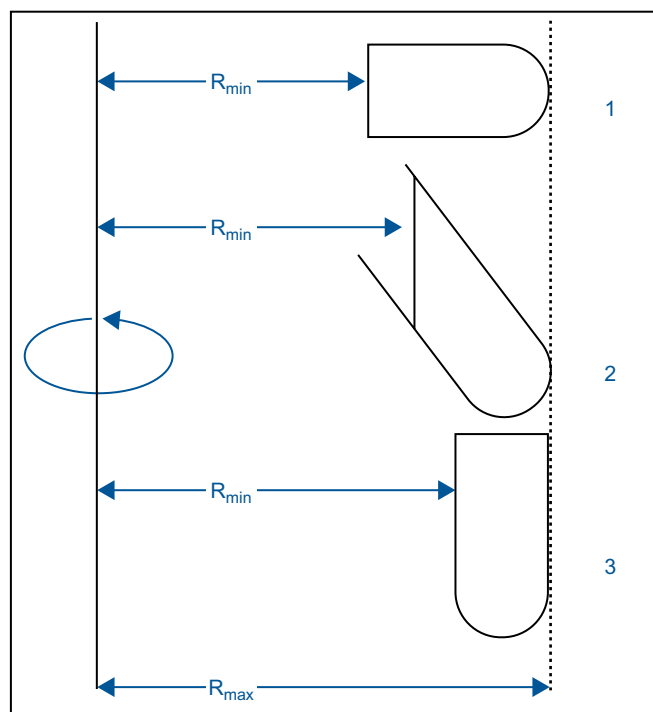


Figure 14-14. Rotor types used in centrifugation. (From Cole-Parmer Technical Library, Basics of Centrifugation. ©Thermo Fisher Scientific. <http://www.coleparmer.com/techinfo/techinfo.asp?htmlfile=basic-centrifugation.htm&ID=30>, accessed 30 November, 2011)

Table 14-1. Types and Characteristics of Centrifuge Rotors

Type of rotor	Pelleting	Rate-zonal sedimentation	Isopycnic
Fixed-angle	Excellent	Limited	Variable*
Swinging-Bucket	Inefficient	Good	Good**
Vertical	NS	Good	Excellent
Zonal	NS	Excellent	Good

NS= not suitable

* Good for macromolecules, poor for cells, and organelles

** Good for cells and organelles, caution needed if used with CsCl (Data from Cole-Parmer Technical Library, Basics of Centrifugation. <http://www.coleparmer.com/techinfo/techinfo.asp?htmlfile=basic-centrifugation.htm&ID=3> accessed 28 July, 2011)

“point loads” caused by spinning CsCl or other dense gradient materials that can precipitate.

In fixed-angle rotors, the sample tubes are held fixed at the angle of the rotor cavity. When the rotor begins to rotate, the solution in the tubes reorients. This rotor type is most commonly used for pelleting applications. Examples include pelleting bacteria, yeast, and other mammalian cells. It is also useful for isopycnic separations of macromolecules such as nucleic acids.

In vertical rotors, sample tubes are held in vertical position during rotation. This type of rotor is not suitable for pelleting applications but is most efficient for isopycnic (density) separations due to the short pathlength. Applications include plasmid DNA, RNA, and lipoprotein isolations.

Selection of Centrifuge Tubes

The selection of the appropriate centrifuge tube is one that prevents sample leakage or loss, ensures chemical compatibility, and allows easy sample recovery. The selection should also consider factors such as clarity, chemical resistance, and the sealing mechanism (if needed).

One should check the product guide pages or tube packaging for notes on recommended sample volume and maximum speed. If using thin-walled sealed tubes, they should be run in a fixed angle or vertical rotor. If necessary to autoclave the tubes, it should be only at 121°C for 15 minutes. One should avoid cleaning plastic tubes in automated dishwashers or glassware washers, which may produce excessively hot temperatures. Also, only a mild laboratory detergent in warm water should be used, followed by a rinse and air dry.

PRECIPITATION

Precipitation is the process of separating solid particles from a previously clear liquid—a solution—by physical or chemical changes. The separated solid is termed a *precipitate*; the cause of precipitation is the *precipitant*; and the liquid that remains in the vessel above the precipitate is called the supernatant liquid.

In pharmacy, precipitation may be useful for many purposes. It provides a convenient method of obtaining solid substances in the form of fine particles, such as the precipitation of calcium carbonate (precipitated chalk). White Lotion is an example of a preparation prepared by precipitation, in this case by mixing aqueous solutions of zinc sulfate and sulfurated potash to form an insoluble, finely divided zinc sulfide, free sulfur, and various polysulfides.

One of the most important uses of precipitation is in the purification of solids. The process as applied to purification is termed *recrystallization*. The impure solid usually is dissolved in a suitable solvent at elevated temperatures. On cooling, the bulk of the impurities remain solubilized while the purified solid product precipitates. This procedure is repeated as many times as necessary, using a number of solvents if required.

SEPARATION OF IMMISCIBLE LIQUIDS

The separation of liquids that are mutually soluble usually is effected by distillation, if one or both of the liquids are volatile. The separation of liquids that are immiscible is generally a simpler process. Separations of this kind are necessary in analytical procedures, manufacturing operations, distillation of volatile oils, and accidental contaminations and admixtures, and are usually best made using a separatory funnel. When very small amounts of liquids are floating on the surface of another liquid, separation is accomplished most easily by using a pipet, medicine dropper, or glass syringe with an attached needle.

The Florentine Receiver can be used for the separation of volatile oils from the water that accompanies them during steam distillation. Where the volatile oil is lighter than water, the principle shown in Figure 14-15 may be used. The oil and water collect in the glass receiver during distillation, the oil floating on the top, while the water ascends the bent tube from the bottom; further addition of distillate causes the water to overflow from the side tube. The reverse action is produced in the receiver for light or heavy oils (Figure 14-16), in which either a lighter or a heavier fraction may be collected continuously.

EXPRESSION

Expression is a process of *forcibly* separating liquids from solids. A number of mechanical principles have been recognized in the operation of expression, namely the use of the spiral twist press, the screw press, the roller press, the filter press, and the hydraulic press.

SPIRAL TWIST PRESS

The principle of this press is best and most practically illustrated in the usual process of manually expressing a substance contained in a cloth.

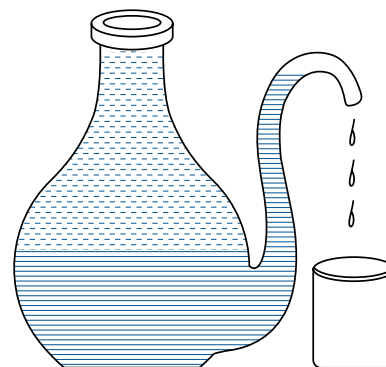


Figure 14-15. Florentine Receiver apparatus.

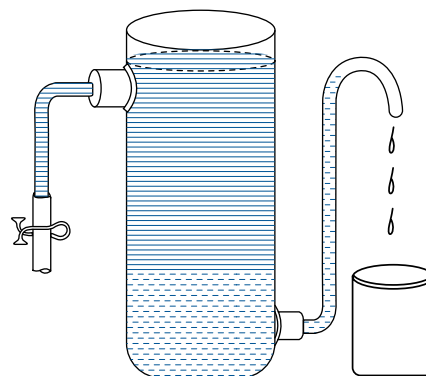


Figure 14-16. Receiver for light or heavy oils.

ROLLER PRESS

This is used for large-scale pressing of oily seeds, fatty substances, and so on. Care must be taken to apply the force gradually to the bag containing the material to be pressed, and not to use it on substances that will be corrosive to the rubber rollers.

HYDROSTATIC OR HYDRAULIC PRESS

Of the presses heretofore mentioned, each has some special advantage of use, but each also has some objectionable feature. The spiral twist is not powerful and its action is limited. The screw presses have friction with which to contend; the friction of a screw increases with the intensity of the pressure applied, and when a certain limit is reached all further force applied is wasted, and if continued may result in destruction of the press. The roller press is very limited in its action. Although the hydraulic press is expensive, after the first coat it is the most economical because the greatest power is obtained at the expense of the least labor. The principle of a hydraulic press is based on the fact that pressure exerted upon an enclosed liquid is transmitted equally in all directions. Tremendous pressures can be developed with hydraulic presses. An example is shown in Figure 14-9.

COUNTERCURRENT DISTRIBUTION

Countercurrent Distribution (CCD) may be defined as a series of liquid-liquid extractions (immiscible solvents) conducted in a multiple-tube apparatus in which one phase is permitted to advance to the next tube in the series independently of the other phase.¹ The separation of the components in the mixture depends on the distribution coefficient of each of the components, the volume of the solvents used, and the number of transfers taken.

Some important applications of CCD in the pharmaceutical sciences are:

- the isolation and purification of chemicals and biochemicals that might otherwise be damaged by the extremes of temperature or pH that occur during the separation processes
- the separation of a crude plant extract into its various chemically related fractions as a preparative step
- the determination of purity and homogeneity of chemicals and medicinal agents
- the characterization of substances extracted from biochemical systems in studies determining the metabolic or biologic disposition of drugs.

Separation using CCD is based on Nernst Law. According to this law, when two practically immiscible solvents are in contact with each other and a substance that is soluble in each is added, the substance distributes itself in such a way that at equilibrium and at a given temperature the ratio of the concentrations of the two solutions is a constant. Strictly speaking, it is the activity ratio rather than the concentration ratio that remains constant. For most purposes, however, concentration values give satisfactory approximations.

When the ratio of concentrations expresses a distribution value for a single chemical species, the constant is designated as a partition coefficient or distribution coefficient, K , and may be expressed mathematically as

$$K = C_u / C_l \quad (10)$$

In this expression C_u and C_l represent concentrations in the upper and lower phases, respectively. There is no accepted convention to date, and the distribution coefficient could just as well be expressed as the reciprocal: C_l/C_u . In actual practice one deals with and measures total analytical concentrations; thus, more than one chemical species usually is present in each phase. An example would be the distribution of benzoic acid between benzene and water. In the aqueous phase, benzoic

acid would be present both in the ionized (A^-) and un-ionized form (HA). In benzene, benzoic acid would be present in the un-ionized form (HA) and in the dimerized form $(HA)_2$. The ratio expressing total benzoic acid in the organic phase and total benzoic acid in the aqueous phase is the partition ratio or the apparent distribution coefficient, K .

Although the purpose of using CCD is to bring about the separation of two or more substances, the basic principles of operation are best introduced by first considering the distribution pattern of a single solute in the two immiscible solvents.

1. Assume that the solute under consideration has a distribution coefficient of unity when distributed between chloroform and buffer solution and that there are no deviations from Nernst's law of distribution due to molecular association, dissociation, ionization, or chemical reactions.
2. Consider six containers such as 250-mL glass-stoppered Erlenmeyer flasks, each holding 50 mL of chloroform (lower phase) as diagrammed in Figure 14-17 (Row A). Add to container No 0, 100 mg of solute under consideration dissolved in 50 mL of buffer solution, and shake until equilibrium has been established. Because equal volumes of solvent are used and the distribution coefficient of solute in these two solvents is unity, the solute at equilibrium will distribute itself in such a way that one-half is found in each of the upper and lower phases (Row B). Because 100 mg was originally present, 50 mg will be found in both layers of Container 0 (Row B).
3. Transfer the upper phase of Container 0 holding 50 mg of solute to Container 1 (Row B) and add fresh buffer solution to Container 0 (Row B). Shake both containers until equilibrium has been established. At equilibrium the quantity of solute in each phase of Containers 0 and 1 (Row C) will be 25 mg.
4. Transfer the upper phase of Container 1 (Row C) to Container 2 (Row C), and the upper phase of Container 0 (Row C) to Container 1. Add fresh buffer solution to Container 0 (Row C) and shake all three containers until equilibrium has been established. At equilibrium the quantity of solute (25 mg) in Container 2 (Row D) will have distributed itself so that one-half (12.5 mg) is in the upper phase and one-half (12.5 mg) is in the lower phase. Because 25 mg of solute was transferred to Container 1 from Container 0, 25 mg of solute will be present in each phase of Container 1 (Row D). The quantity (25 mg) of solute in Container 0 will distribute itself between the chloroform layer and freshly added buffer solution so that one-half (12.5 mg) will be present in each layer (Row D).

Continue this general procedure of transferring the upper phases of Containers 0, 1, and 2 to Containers 1, 2, and 3, respectively; then add fresh buffer to Container 0. Shake the four flasks until equilibrium is established. A distribution is obtained as shown in Row E. Continuing in a like manner will give a distribution as shown in Row F.

A plot of the fraction of solute in each container versus container number is shown in Figure 14-18. The significance of this curve is that the distribution of the solute shows a peak in which the maximum is located in a specific container and the location of the peak container is a function of the partition coefficient. Hence, it can be seen that two or more solutes with different K values can be separated effectively after the passage of a mixture through many tubes (usually 25 or more, depending upon K values) in a CCD apparatus.

Figure 14-18 illustrates the distribution of a solute after only four transfers. In actual practice between 8 and 2000 containers or tubes usually are used in multiple extractions of this kind. The tubes are connected in series in a train and are rocked simultaneously rather than individually to bring about distribution of solutes between the two phases. The device also permits

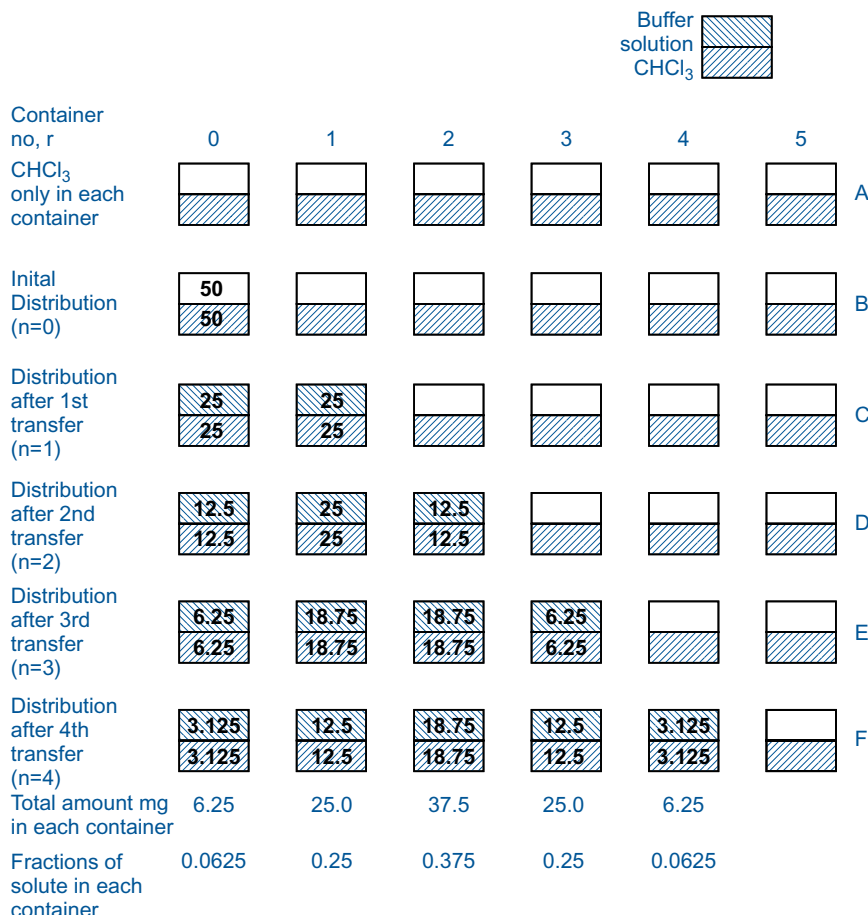


Figure 14-17. Theoretical distribution of solute after varying numbers of transfer.

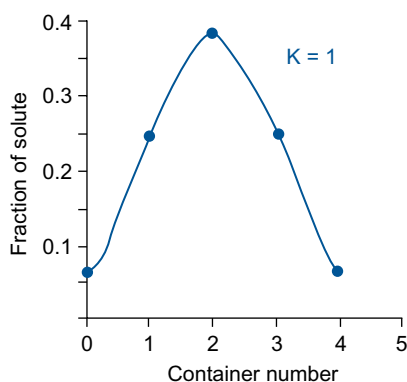


Figure 14-18. Distribution of solute after four transfers.

the transfer of upper phases to the next tube in series, in one operation. A device of this type is called a countercurrent distribution apparatus.

To study the fraction of a given solute present in each tube r , after n number of transfers, it is convenient to use:

$$f_{n,r} = \frac{n!}{r!(n-r)!} \left(\frac{1}{1+KR} \right)^n (KR)^r \quad (11)$$

where K is defined as the partition coefficient and R is defined as the ratio of the volume of the upper phase to the volume of the lower phase, (V_u/V_l).

This equation can be illustrated as follows: Calculate the fraction of solute in tubes 0, 1, 2, 3, and 4 after four transfers are made in a CCD apparatus using equal volumes of upper and

lower phases. The K value for the solute in the solvent system is assumed to be 1.0 in this example.

For Tube 3,

$$f_{4,3} = \frac{4!}{3!(4-3)!} \left(\frac{1}{1+1} \right)^4 (1)^3 = 0.25 \quad (12)$$

By similar calculations the fraction of solutes in Tube 0, 1, 2, and 4 is found to equal

$$f_{4,0} = 0.0625; f_{4,1} = 0.25; f_{4,2} = 0.375; f_{4,4} = 0.0625$$

The distribution of solute using Equation 2 is shown in Figure 14-18.

When a large number of transfers (50) are made and K is near unity it is more convenient to use a Gaussian treatment to calculate the fraction of solute in a particular tube. The appropriate equations are

$$y_x = \frac{1.00}{\sqrt{2\pi nKR/(KR+1)^2}} \exp \left\{ - \left(\frac{x^2}{2nKR/(KR+1)^2} \right) \right\} \quad (13)$$

$$T_{\max} = \frac{nKR}{KR+1}$$

where y_x represents the fraction of solute with distribution coefficient K in the tube that is x distant from the peak tube; $\exp$ is the exponent of the base e , $\exp 2 = e^2$; $\pi = 3.14$; K , R , and n are terms that have been defined previously and $r_{\max}$ represents the number of the tube containing the maximum amount of solute.

Distribution curves may be prepared from the hypothetical data or from a computer program using these equations. Figure 14-19 illustrates a series of curves for a solute in which $K = 1.0$

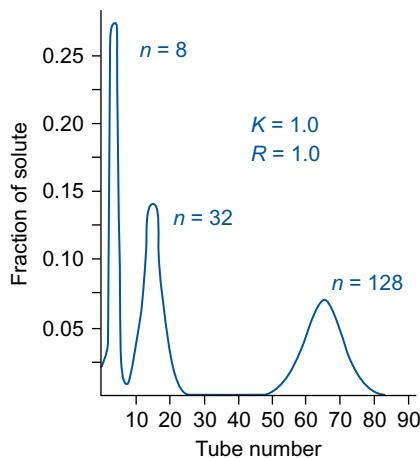


Figure 14-19. Distribution of solute after varying number of transfers.

and $R = 1.0$ following 8, 32, and 128 transfers. It is interesting to observe that as the number of transfers increases, the amplitude of the curve decreases and the solute spreads through more and more tubes. At first thought, this would seem undesirable, but the significant point is that the fraction of vessels containing solute after 128 transfers is now much less than after 10 transfers.

Therefore, two solutes with different but similar K values can be separated in 128 transfers because each solute occupies a smaller fraction of total tubes. If this separation were attempted with 10 to 20 transfers, both solutes would occupy nearly all of the tubes and no separation would be obtained.

Figure 14-20 illustrates the distribution patterns obtained in a 16-transfer experiment for solutes having distribution coefficients that differ by one order of magnitude. Under no circumstances can a separation be obtained if the distribution coefficients of the solutes are equal.

The procedure of operation that has been considered thus far is known as the *fundamental procedure*. Here, the solute is distributed through a specified number of tubes and nothing is withdrawn from the system until the entire operation is completed. Then the tube contents are withdrawn and analyzed for the purpose of determining solute concentrations, or the solutes are withdrawn simply for the purpose of isolating them from a mixture.

Another procedure of operation that is of interest primarily due to its analogy to elution chromatography is known as *end withdrawal*. In this operation the fundamental procedure is

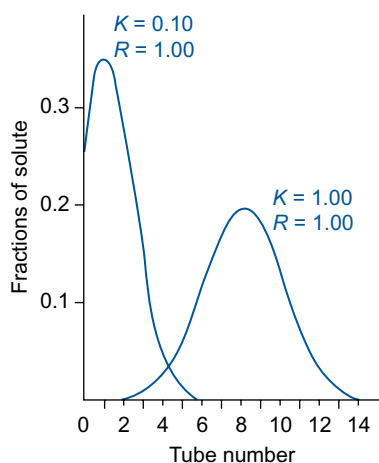


Figure 14-20. Distribution of two solutes with different K values.

followed for a predetermined number of transfers as previously described. Then the upper phase only of the last tube in the train is collected. All other upper phases are advanced to the next tube in succession and after equilibration the upper phase of the last tube, n , is again collected.

This process is continued until all upper phases have passed through n tubes containing lower phase. In elution chromatography the analogy is similar. However, fresh upper phase is added continuously to the first *tube* (called a *plate* in elution chromatography) until only upper phase is eluted from the column.

In summary, the degree of separation of two or more solutes using CCD depends upon the distribution coefficients of the solutes, nature and volume of the solvents used, and number of transfers taken.

OTHER SEPARATION TECHNIQUES

CLARIFICATION

Clarification is the process by which finely divided solids and colloidal materials are separated from liquids without the use of filters. The process is employed to remove suspended oil from aqueous solutions, such as aromatic waters, and for the removal of undesirable solids that interfere with the transparency of such natural products as honey and fruit juices.

Clarification generally is resorted to when the contaminating material is finely subdivided, amorphous, or colloidal in nature and tends to plug a filtration medium rapidly. A number of methods are available to handle this difficult problem.

This may involve varying the temperature or pH of the medium. When a viscid liquid is heated, its viscosity and specific gravity are decreased and particles that are suspended in it will separate. Those particles that are more dense than the liquid will fall to the bottom, while those that are less dense will rise to the surface. In the latter case the minute bubbles of steam formed in the heating process become enveloped in the viscid particles, rise through their buoyancy, and a scum is formed that may be separated readily.

The dewaxing of oils at a reduced temperature offers a further example of the possibilities of contaminant modification. Oil that is chilled rapidly often produces an amorphous wax that will plug a straining medium. Slow chilling, on the other hand, produces a wax with a more crystalline nature, which has good filtration characteristics.

The simplest method of clarification, although not always feasible, is gravitational sedimentation. This method involves the least amount of labor and expense and is used frequently, particularly on a large scale, when haste is unnecessary. The deposit formed is called a *sediment* or *sludge*. These terms are not synonymous with *precipitate*. A sediment is solid matter separated merely by the action of gravity from a liquid in which it has been suspended. A precipitate, on the other hand, is solid matter separated from a previously clear solution by physical or chemical change. Fixed oils usually are clarified by gravitational sedimentation. In vegetable oils the sediment consists principally of albuminous and gummy substances, cellular tissue, and water, all of which have been separated with the oil during the expression process.

The clarification process generally is carried out by adding a clarifying agent such as paper, pulp, tale, infusorial earth, as well as a number of other materials to the turbid liquid. These agents usually act to reduce turbidity by physical adsorption of the contaminating material, although a large number of specific, physicochemical coagulants also are in use. After the addition of the clarifying agent, the mixture is agitated and the agents, along with the adsorbed impurities, are removed by filtration or any other suitable means. Albumin and gelatin are examples of clarifying agents obtained from natural sources. Substances of a synthetic nature, such as polyamines, also are used for this purpose.

COLATION

Colation or straining (from Latin *colare*, to strain) is the process of separating a solid from a fluid by pouring the mixture on a cloth or porous substance that will permit the fluid to pass through, but will retain the solid. This operation frequently is used for separating sediment or mechanical impurities of various kinds from liquids. Colation should not be considered as a separate process but simply as a crude form of filtration, with larger pores in the straining medium than usually are employed for filtration.

The essential apparatus is a straining medium and a strainer support or frame. The straining medium is usually a cloth material such as flannel, muslin, wool, or cheesecloth. The material should be colorless and washed before use. Fabrics, particularly those of cotton, usually are treated or impregnated with a material called *sizing* to improve their appearance and quality for certain purposes; however, for use as a strainer, the fabric must be free of sizing because it causes contamination. Many different substances are used for sizing, some being soluble in cold water, others only in hot water. Thus, the proper method for their removal is to soak the fabric for a few hours in cold distilled water, rinse thoroughly; then cover with distilled water, boil for a few minutes, and rinse well in distilled water to remove the last traces of the gelatin, albumin, glue, or starch that may have been present in the sizing.

CONTINUOUS WASHING

The use of the wash bottle is limited to small operations. A simple method of automatically supplying the wash liquid in larger quantities is shown in Figure 14-21. This requires attention from the operator only at the beginning of the operation. The inverted bottle containing the washing solvent is furnished with a perforated stopper and a short glass tube. All that is necessary is to fill the bottle and adjust it over the funnel so that the end of the tube is at the height at which the level of liquid in the funnel is to be maintained. When the bottle is tilted slightly (if the tube selected is not too narrow in diameter), the liquid runs into the funnel until it rises to the orifice of the tube, whereupon the flow ceases. As the liquid gradually passes through the solid substance in the funnel, the level falls below the orifice, bubbles of air pass through the tube into the bottle, the liquid once more flows, and the operation continues until the upper bottle is empty. Many elaborate methods of continuous washing have been suggested, but the simple apparatus just described is quite satisfactory if a tube of proper diameter has been selected,

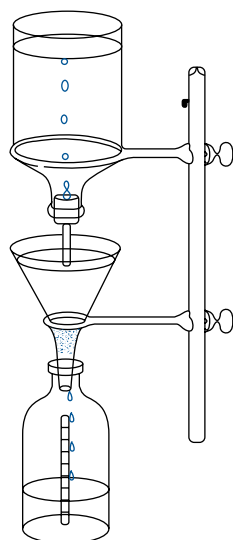


Figure 14-21. Continuous washing apparatus.

one of such size that the force of capillary attraction will not be strong enough to prevent the passage of air.

DECANTATION

The simplest method available for the separation of a solid from its soluble impurities is the technique of decantation. This method involves washing and subsequent agitation of the solid with an appropriate solvent, allowing the solid to settle and removing the supernatant solvent. These three steps are repeated as often as required to attain the desired purity of the solid. This method also is applicable to the simple separation of solids and liquids, such as after precipitation of a material from a mother liquor. Decantation provides an effective method for washing magmas and other gelatinous products.

Some degree of skill is required to decant liquids effectively. It is most convenient to decant from a lipped vessel that is not filled to capacity. In addition, the use of a stirring rod over the lip and rim of the vessel is suggested as a guide to steady the hand of the operator. Also, very low vacuum with a glass micro-pipette can be used to remove small quantities of supernatant close to the interface boundary.

DECOLORATION

Decoloration, or decolorization as it sometimes is called, is the process of depriving solutions of color by use of an appropriate adsorptive medium. In many respects it is closely related to the clarification process. Decoloration is used for removal of coloring matter from a number of raw materials, both natural and synthetic, and from many finished products. Animal charcoal (also called bone black), wood charcoal, or activated charcoal frequently are used as decolorizing agents. Clays such as bentonite, kaolin, and fuller's earth also are used for this purpose.

DIFFUSION AND DIALYSIS

Diffusion is the spontaneous penetration of one substance into another under the potential of a concentration gradient. Simply stated, material will tend to move from a region of higher concentration to one of lower concentration. The driving force or potential of such a process may be enhanced by the application of an electric field.

If the two regions of concentration noted are separated by a selective membrane, certain species will diffuse through the membrane, while other molecular species will be held back. When this selectivity is dictated by the porosity of the membrane, the process is termed *dialysis*. Dialysis is used principally for the separation of small molecules and ions contained in a mixture with colloidal material. The latter substances diffuse with difficulty or not at all. Materials such as gums, starch, albumin, and proteins fall into this colloidal, nondiffusible category.

The rate of diffusion across a semipermeable membrane is directly proportional to the concentration gradient between the two surfaces of the membrane and to the area of the membrane, but is inversely proportional to the membrane thickness. These factors are expressed in Fick's law of diffusion

$$ds/dt = [kA(C_i - C_o)]/[h]$$

where S is the amount of substance diffused at time t , k is a permeability constant, A is the membrane area, h is the membrane thickness, dS/dt is the diffusion rate, C_i is concentration on one side, and C_o is concentration on the other side of the membrane.

GEL FILTRATION

Gel filtration is a chromatographic method, also called size-exclusion chromatography (SEC), where molecules in solution are separated by their size, not by their molecular weight. SEC is usually applied to large molecules or macromolecular complexes such as proteins. When an aqueous solution is used to transport the sample through the column, the technique is known as gel-filtration chromatography, versus gel permeation

chromatography, which is used when an organic solvent is used as a mobile phase. The technique of SEC is widely used for polymer characterization.

One primary application of SEC is fractionation of proteins and other water-soluble polymers, while gel permeation chromatography is used to analyze the molecular weight distribution of organic-soluble polymers. SEC typically uses a gel medium, such as polyacrylamide, dextran or agarose, and filtration under low pressure.

Changyin *et al.* used different types of Sephadex gels for separation. Their study investigated various reagents necessary to perform the separation in an ultimate purification of the compound. The results indicated that optimization was capable of being done to separate the impurities from the active compound. The nature of the mobile phase, the ionic type, pH value, and molarity were important for the optimization. Cephalosporin gel chromatography was shown to be important in the separation of high-molecular-weight impurities which frequently are associated with allergic responses in patients. This method has been demonstrated to serve as an excellent quality-control procedure for the impurities in cephalosporin preparations.²

A feasibility study of liposome separation that was undertaken to explore the use of size-exclusion chromatography, such as gel filtration of a large-scale process, demonstrated that it could separate liposomes from freeze-dried material in a chromosome preparation.³ The chromatographic step was intended to improve the drug encapsulation by removing free (unencapsulated) drugs from external media. The selected stationary phase was G-50 Sephadex. The model drug used in the study was orciprenaline sulfate. The technique was able to produce a suitable size exclusion that efficiently removed the free drug from the liposome preparation.

In a study of liposomes loaded with calcitonin, it was necessary to observe the location of the protein to protect it from enzymatic digestion.⁴ The analysis of the liposome produced from this protein was extracted using suitable gel separation of the liposome mixture to ensure the location of the protein within the system. It established the stability and the ultimate formation of the liposome product. This ensured the appropriate loading of the protein within the liposome product.

A process for purifying bovine pancreatic glucagon as a by-product of insulin production has been described.⁵ The glucagon-containing supernatant from the alkaline crystalline crystallization of insulin was precipitated using ammonium sulfate and isoelectric precipitation. The precipitate was then purified by ion-exchange chromatography on Q-Sepharose FF gel filtration on Sephadex G-25 and ion-exchange chromatography on S-Sepharose FF. Successful yields were obtained using this technique, which was successful because of the gel filtration procedure.

A report was presented on the characterization of adenosine receptors in porcine striatal membranes and their solubilization by detergent digitonin.⁶ Once the drug was solubilized, the material was bound to sites after the removal of receptors from the lipid environment. Gel filtration on Superdex 200 accomplished the separation into appropriate molecular weights. Suitable purification was achieved by this means.

In another report of the use of gel filtration, the expression and purification of human gamma glutamylcysteine synthetase were studied.⁷ Specific proteins and polypeptides were isolated and their amounts characterized by the use of Superdex 200 along with ATP-affinity resins. Cyclosporin A has potential for wide clinical use, limited only by the very narrow therapeutic index.⁸ Potentiation of its clinical efficacy is thus very desirable. Preliminary data had indicated that the mixture of cyclosporin A, with hyaluronate, could increase its efficiency. In this study, it was found that cyclosporin A could reduce the hypersensitivity in test animals when administered along with hyaluronate. To demonstrate the association of this mixture, gel filtration was required, which showed the protection of the molecule from being

bound to red blood cells. This association would improve the clinical response and was proven only by the use of gel filtration.

Lotion

Lotion (displacement washing) is the process by which soluble impurities are removed from insoluble material by the addition of a suitable washing solvent. The wash liquid usually is separated from the purified solid by decantation or filtration. An expedient method of adding the washing solvent to the solid in a fine, controlled spray is by the use of wash bottles or spray bottles and a Buchner funnel apparatus.

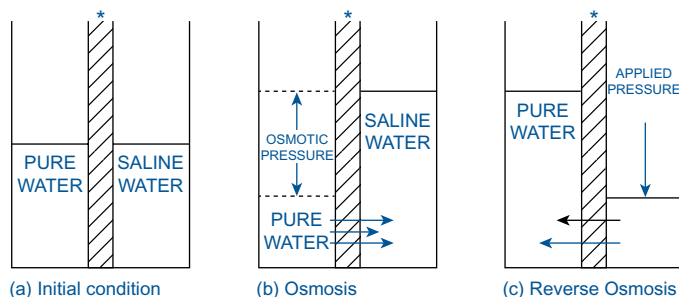
RECRYSTALLIZATION

Recrystallization is a process in which the crystal structure of the sample is completely disrupted by dissolution and then crystals are allowed to regrow leaving impurities in the solution. The mechanism involved is that impurities can seldom fit into the crystal structure of another compound. The chosen solvent generally is one where the impurities are more soluble than the substance being purified.

Solute recrystallization as a technique involves several steps: (1) selection of solvent, (2) dissolution of the solid to be purified in the solvent at or near its boiling point, (3) filtration of the hot solution to remove impurities, (4) cooling of the solution to form crystals, (5) separation of the crystals from the supernatant solution, (6) washing of the crystals to remove adhering solution, and (7) drying the crystals.

REVERSE OSMOSIS

Reverse osmosis is the separation, concentration and fractionation of inorganic or organic substances in aqueous or non-aqueous solutions in the liquid or the gaseous phase; it involves a semipermeable membrane and a driving force or pressure. As reverse osmosis (Figure 14-22) is used it is necessary to evaluate new composite reverse osmosis membranes that have been developed with significant improved performance over older commercially available conventional composite membranes. The Energy Saving Polyamide (ESPA) membrane chemistry provides a high flux at low operating pressure while maintaining a very good salt and organic rejection. The membranes have been demonstrated to operate for several years. Appropriate transmission and field emission electron micrographs of the membrane demonstrated the structure of the membrane skin layer is the reason for the improved performance. This surface charge of the various membranes was demonstrated qualitatively using zeta-potential measurements. Newer membranes have a low surface charge and operate at a lower pressure. In an effort to further improve the available reverse osmosis water-treatment membranes, other studies have been conducted to evaluate specific ultra-low-pressure membranes. Newer membranes have been designed with a 30% increase in productivity over conventional membranes. These improvements are particularly important to multistage systems for water purification.



* semipermeable membrane

Figure 14-22. Principles of reverse osmosis.

Recommendations have been made by many to improve the systems by using ultra-low-pressure membranes.

TANGENTIAL FLOW FILTRATION

Tangential flow filtration permits rapid flow of the small molecules and solvent to pass through the filter. The “sweeping” action of the liquid moving over the membrane decreases the concentration of retentate on the filter surface preventing concentration polarization. This method permits high filtration flow rates without shearing fragile molecules or cells. A variable restrictor is used to provide pressure drop across the filter.

ULTRAFILTRATION

Ultrafiltration (UF), also termed molecular filtration, is a technique for separating dissolved molecules on the basis of effective Stoke's Radius (size) under applied pressures. The molecular filter is a thin, selectively permeable membrane that retains most macro-molecules above a certain size while permitting smaller molecules to pass into the filtrate. It is difficult to assign a molecular weight cutoff (MWCO) of an UF membrane since many factors affect it, including the pore size distribution of the membrane and the size, shape and electrical charge of the dissolved analytes to be filtered. At best the MWCO of an UF can be regarded as a “general” guide to the molecular weight or size (Stoke's radius) in which each type of filter is most efficient. MWCO should not be interpreted as a sharp cut-off point but rather a range. UF is not a high resolution technique but is useful for certain types of fractionation.

REFERENCES

1. Rogers LB. Principles and techniques. In: Kolthoff IM, Elving PJ, eds. *Treatise on Analytical Chemistry, Part 1 Theory and Practice, Vol 2*. New York: Interscience, 1961: Chapter 22.
2. Changqin H *et al.* The chromatographic behaviour of cephalosporins in gel filtration chromatography, a novel method to separate high molecular weight impurities. *J Pharm Biomed Anal* 1994; 12: 533–541.
3. Vemuri S, Rhodes C. Separation of liposomes by a gel filtration chromatographic technique: a preliminary evaluation. *Pharm Acta Helv* 1994; 69: 107–113.
4. Arien A *et al.* Cholate-induced disruption of calcitonin-loaded liposomes: formation of trypsin-resistant lipid-calcitonin-cholate complexes. *Pharm Res* 1995; 12: 1289–1292.
5. Andrade A *et al.* Purification of bovine pancreatic glucagon as a by-product of insulin production. *J Med Biol Res* 1997; 30: 1421–1426.
6. Costa B *et al.* A2a Adenosine receptors: guanine nucleotide derivative regulation in porcine striatal membranes and digitonin soluble fraction. *Neurochem Int* 1998; 33: 121–141.
7. Misra I, Griffith O. Expression and purification of human gamma-glutamylcysteine synthetase. *Prot Express Purif* 1998; 13: 268–276.
8. Gowland G. Fourfold increase in efficacy of cyclosporine A when combined with hyaluronan: evidence for mode of drug transport and targeting. *Int J Immunother* 1998; 14(1): 1–7.

BIBLIOGRAPHY

- Cheremisinoff NP. *Liquid Filtration*, 2nd Ed. Butterworth-Heinemann, 1998.
- Graham J. *Biological Centrifugation (The Basics)*. Garland Science, 2001.
- Lachman L *et al.* *The Theory and Practice of Industrial Pharmacy*, 3rd ed. Philadelphia: Lea & Febiger, 1986.
- Perry JH *et al.* *Chemical Engineer's Handbook*, 6th ed. New York: McGraw-Hill, 1984.
- Records A, Sutherland K. *Decanter Centrifuge Handbook*. Elsevier Science, 2001.
- Regel LL, Wilcox WR. *Processing by Centrifugation*. Springer. 2001.
- Rickwood D, Graham JM. *Biological Centrifugation*. Springer Verlag, 2001.
- Rickwood R. *Preparative Centrifugation: A Practical Approach*. Oxford University Press, 1993.
- Striegel A *et al.* *Modern Size-Exclusion Liquid Chromatography: Practice of Gel Permeation and Gel Filtration Chromatography*, 2nd Ed. Wiley, 2009.
- Sutherland K. *Filters and Filtration Handbook*, 5th Edition. Elsevier Science, 2008.
- Swarbrick J, Boylan J. *Encyclopedia of Pharmaceutical Technology*. New York: Dekker, 1990.
- Tarleton S, Wakeman R. *Solid/Liquid Separation: Principles of Industrial Filtration*. Elsevier Science, 2005.
- Townsend A. *Encyclopedia of Analytical Science*. New York: Academic, 1995.

Chapter 15

Ionic Solutions and Electrolytic Equilibria

Barbara R. Conway, PhD

ELECTROLYTES 257

CHEMICAL EQUILIBRIA 260

PHARMACEUTICAL SIGNIFICANCE 274

ELECTROLYTES

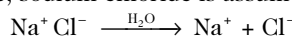
Electrolytes are substances containing free ions, thus rendering the substance electrically conductive. The most typical electrolyte is an **ionic solution** typified by solutions of acids, bases, or salts. As most drugs are weak acids or bases, it is essential to understand the properties of ionic solutions. **Colligative properties** are properties of solutions that depend on the number of molecules in a given volume of solvent rather than the weight concentration of the molecules. The colligative properties of solutions of electrolytes depend on the total number of entities in solution, including ions. In order to determine which species are present in ionic solutions, it is imperative to understand **ionic equilibria** and their impact on subsequent drug activity. An understanding of these fundamental concepts and an ability to manipulate and predict the subsequent drug properties is crucial in pharmaceutical disciplines.

COLLIGATIVE PROPERTIES

In general, for nonelectrolytes, a given colligative property of two equimolar solutions will be identical. This generalization, however, cannot be made for solutions of electrolytes.

Van't Hoff pointed out that the osmotic pressure of a solution of an electrolyte is considerably greater than the osmotic pressure of a solution of a nonelectrolyte of the same molal concentration. This anomaly remained unexplained until 1887 when Arrhenius proposed a hypothesis that forms the basis for our modern theories of electrolyte solutions.

This theory postulated that when electrolytes are dissolved in water they split up into charged particles known as **ions**. Each of these ions carries one or more electrical charges, with the total charge on the positive ions (**cations**) being equal to the total charge on the negative ions (**anions**). Such solutions conduct electricity by virtue of the movement of these ions under application of an electrical charge; cations move to the negative terminal and anions to the positive terminal. Thus, although these solutions may contain charged particles, it remains neutral. The increased osmotic pressure of such solutions is due to the increased number of particles formed in the process of ionization. For example, sodium chloride is assumed to dissociate as



It is evident that each molecule of sodium chloride that is dissociated produces two ions, and if dissociation is complete, there will be twice as many particles as would be the case if it were not dissociated at all. Furthermore, if each ion has the same effect on osmotic pressure as a molecule, it might be expected that the osmotic pressure of the solution would be twice that of a solution containing the same molal concentration of a non-ionizing solute.

Osmotic-pressure data indicate that, in very dilute solutions of salts that yield two ions, the pressure is very nearly double that of solutions of equimolar concentrations of nonelectrolytes. Similar magnification of vapor-pressure lowering, boiling-point elevation and freezing-point depression occurs in dilute solutions of electrolytes.

The Van't Hoff factor, i , is the ratio of the colligative effect produced by a concentration, m , of electrolyte, divided by the effect observed for the same concentration of nonelectrolyte, or

$$i = \frac{\pi}{(\pi)_0} = \frac{\Delta P}{(\Delta P)_0} = \frac{\Delta T_b}{(\Delta T_b)_0} = \frac{\Delta T_f}{(\Delta T_f)_0} \quad (1)$$

in which π , ΔP , ΔT_b , ΔT_f refer to the osmotic pressure, vapor-pressure lowering, boiling-point elevation and freezing-point depression, respectively, of the electrolyte. The terms $(\pi)_0$ and so on refer to the nonelectrolyte of the same concentration. In general, with strong electrolytes (those assumed to be 100% ionized), the van't Hoff factor is equal to the number of ions produced when the electrolyte goes into solution (2 for NaCl and MgSO_4 , 3 for CaCl_2 and Na_2SO_4 , 4 for FeCl_3 and Na_3PO_4 , etc.).

In very dilute solutions the osmotic pressure, vapor-pressure lowering, boiling-point elevation and freezing-point depression of solutions of electrolytes approach values two, three, four, or more times greater (depending on the type of strong electrolyte) than in solutions of the same molality of nonelectrolyte, thus confirming the hypothesis that an ion has the same primary effect as a molecule on colligative properties. Interionic attraction and solvation effects can lead to deviations from this:

Interionic attraction between the positive and negatively charged ions, in consequence of which the ions are not dissociated completely from each other and thus do not exert their full effect on vapor pressure and other colligative properties. This results in less than 2-, 3-, or 4-fold intensification of a colligative property and this deviation generally increases with increasing concentration of electrolyte.

The second effect intensifies the colligative properties and is attributed to the attraction of ions for solvent molecules (called **solvation**, or, if water is the solvent, **hydration**), which holds the solvent in solution and reduces its escaping tendency, with a consequent enhancement of the vapor-pressure lowering. Solvation also reduces interionic attraction and, thereby, further lowers the vapor pressure.

CONDUCTIVITY

The ability of metals to conduct an electric current results from the mobility of electrons in the metals. This type of conductivity is called **metallic conductance**. On the other hand, various chemical compounds—notably acids, bases, and salts—conduct electricity by virtue of ions present or formed, rather than by electrons. This is called **electrolytic conductance**, and the conducting compounds are electrolytes. Although the fact that certain electrolytes conduct electricity in the molten state is important, their behavior when dissolved in a solvent, particularly in water, is of greater concern in pharmaceutical science.

The electrical conductivity (or conductance) of a solution of an electrolyte is merely the reciprocal of the resistance of the solution. Therefore, to measure conductivity is actually to measure electrical resistance, commonly with a **Wheatstone bridge apparatus**, and then to *calculate* the conductivity. Figure 15-1 is a representation of the component parts of the apparatus.

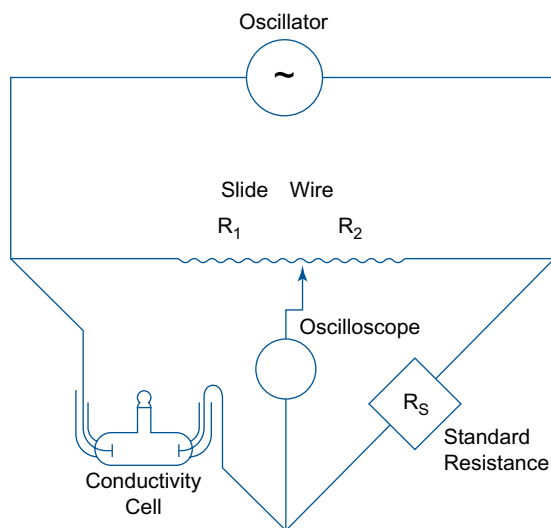


Figure 15-1. Alternating current Wheatstone bridge for measuring conductivity.

The solution to be measured is placed in a glass or quartz cell having two inert electrodes, commonly made of platinum or gold and coated with spongy platinum to absorb gases, across which passes an alternating current generated by an oscillator at a frequency of about 1000 Hz. The reason for using alternating current is to reverse the electrolysis that occurs during flow of current that would cause polarization of the electrodes and lead to abnormal results. The size of the electrodes and their distance apart may be varied to reduce very high resistance or increase very low resistance to increase the accuracy and precision of measurement. Thus, solutions of high conductance (low resistance) are measured in cells having small electrodes relatively far apart, whereas solutions of low conductance (high resistance) are measured in cells with large electrodes placed close to each other.

Electrolytic resistance, like metallic resistance, varies directly with the length of the conducting medium and inversely with its cross-sectional area. The known resistance required for the circuit is provided by a resistance box containing calibrated coils. Balancing of the bridge may be achieved by sliding a contact over a wire of uniform resistance until no (or minimum) current flows through the circuit, as detected either visually with a cathode-ray oscilloscope or audibly with earphones.

The resistance, in ohms, is calculated by the simple procedure used in the Wheatstone bridge method. The reciprocal of the resistance is the conductivity, the units of which are **reciprocal ohms** (also called *mho*). As the numerical value of the conductivity will vary with the dimensions of the conductance cell, the value must be calculated as **specific conductance**, L ,

which is the conductance in a cell having electrodes of 1 cm² cross-sectional area and 1 cm apart. If the dimensions of the cell used in the experiment were known, calculating the specific conductance would be possible. Nevertheless, this information actually is not required, because calibrating a cell by measuring in it the conductivity of a standard solution of known specific conductance is possible—and much more convenient—and then calculating a **cell constant**. Because this constant is a function only of the dimensions of the cell, it can be used to convert all measurements in that cell to specific conductivity. Solutions of known concentration of pure potassium chloride are used as standard solutions for this purpose.

Equivalent conductance

In studying the variation of conductance of electrolytes with dilution it is essential to make allowance for dilution so that the comparison of conductances may be made for identical amounts of solute. This may be achieved by expressing conductance measurements in terms of **equivalent conductance**, Λ , which is obtained by multiplying the specific conductance, L , by the volume in milliliters, V_e , of a solution containing 1 g-eq of solute. Thus,

$$\Lambda = LV_e = \frac{1000L}{C} \quad (2)$$

where C is the concentration of electrolyte in the solution in g-eq/L, that is, the normality of the solution. For example, the equivalent conductance of 0.01 *N* potassium chloride solution, which has a specific conductance of 0.001413 mho/cm, may be calculated in either of the following ways:

$$\begin{aligned} \Lambda &= 0.001413 \times 100,000 \times 141.3 \text{ mho cm}^2/\text{eq} \\ \text{or} \quad \Lambda &= \frac{1000 \times 0.001413}{0.01} = 141.3 \end{aligned}$$

Strong and weak electrolytes

Electrolytes are classified broadly as strong electrolytes and **weak electrolytes**. The usual criterion for distinguishing between strong and weak electrolytes is the extent of **ionization**. The former category includes solutions of strong acids, strong bases, and most salts; the latter includes weak acids and bases, primarily organic acids, amines, and a few salts. Such strong electrolytes completely, or almost completely, dissociate into their ions in solution, and as such, it is difficult to determine an ionization constant for these. Weak electrolytes, which exist both in the molecular and ionized state in solution, can be categorized by the extent of this dissociation by an equilibrium constant. For the purposes of this discussion, classification of electrolytes as strong or weak will be based on certain conductance characteristics exhibited in aqueous solution. The equivalent conductances of some electrolytes, at different concentrations, are given in Table 15-1 and for certain of these electrolytes again in Figure 15-2, where the equivalent conductance is plotted against the square root of concentration. By plotting the

Table 15-1. Equivalent Conductances^a, at 25°C

GEQ/L	HCl	HOAc	NaCl	KCl	NaI	KI	NaOAc
Inf dil	426.1	390.6 ^a	126.5	149.9	126.9	150.3	91.0
0.0005	422.7	67.7	124.5	147.8	125.4	—	89.2
0.0010	421.4	49.2	123.7	146.9	124.3	—	88.5
0.0050	415.8	22.9	120.6	143.5	121.3	144.4	85.7
0.0100	412.0	16.3	118.5	141.3	119.2	142.2	83.8
0.0200	407.2	11.6	115.8	138.3	116.7	139.5	81.2
0.0500	399.1	7.4	111.1	133.4	112.8	135.0	76.9
0.1000	391.3	5.2	106.7	129.0	108.8	131.1	72.8

^aThe equivalent conductance at infinite dilution for acetic acid, a weak electrolyte, is obtained by adding the equivalent conductances of hydrochloric acid and sodium acetate and subtracting that of sodium chloride.

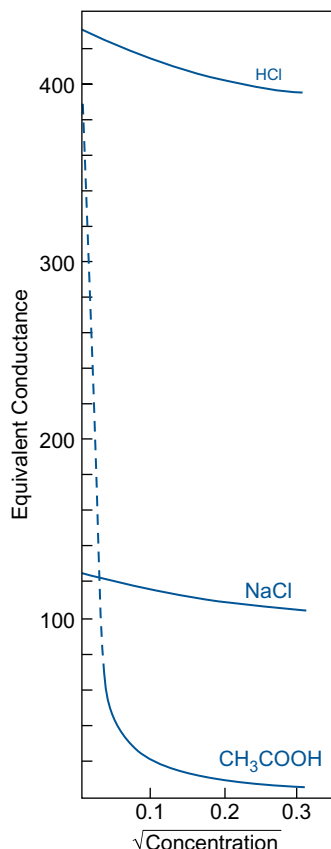


Figure 15-2. Variation of equivalent conductance with square root of concentration.

data in this manner a linear relationship is observed for strong electrolytes, whereas a steeply rising curve is noted for weak electrolytes; this difference is a characteristic that distinguishes strong and weak electrolytes. The interpretation of the steep rise in the equivalent conductance of weak electrolytes is that the degree of ionization increases with dilution, becoming complete at infinite dilution.

Interionic interference effects generally have a minor role in the conductivity of weak electrolytes. With strong electrolytes, which are usually completely ionized, the increase in equivalent conductance results not from increased ionization but from diminished ionic interference as the solution is diluted, in consequence of which ions have greater freedom of mobility (i.e., increased conductance).

The value of the equivalent conductance extrapolated to infinite dilution (i.e., zero concentration), designated by the symbol Λ_0 , has special significance. It represents the equivalent conductance of the completely ionized electrolyte when the ions are so far apart that there is no interference with their migration due to interionic interactions. It has been shown, by Kohlrausch, that the equivalent conductance of an electrolyte at infinite dilution is the sum of the equivalent conductances of its component ions at infinite dilution, expressed symbolically as

$$\Lambda_0 = l_0(\text{cation}) + l_0(\text{anion}) \quad (3)$$

The significance of Kohlrausch's law is that each ion, at infinite dilution, has a characteristic value of conductance that is independent of the conductance of the oppositely charged ion with which it is associated. Thus, if the equivalent conductances of various ions are known, the conductance of any electrolyte may be calculated simply by adding the appropriate ionic conductances.

As the fraction of current carried by cations (**transference number** of the cations) and by anions (**transference number**

of anions) in an electrolyte may be determined readily by experiment, ionic conductances are known. Table 15-2 gives the equivalent ionic conductances at infinite dilution of some cations and anions. It is not necessary to have this information to calculate the equivalent conductance of an electrolyte, for Kohlrausch's law permits the latter to be calculated by adding and subtracting values of Λ_0 for appropriate electrolytes. For example, the value of Λ_0 for acetic acid may be calculated as

$$\Lambda_0(\text{CH}_3\text{COOH}) = \Lambda_0(\text{HCl}) + \Lambda_0(\text{CH}_3\text{COONa}) + \Lambda_0(\text{NaCl})$$

which is equivalent to

$$l_0(\text{H}^+) + l_0(\text{CH}_3\text{COO}^-) + l_0(\text{H}^+) + l_0(\text{Cl}^-) + l_0(\text{Na}^+) + l_0(\text{CH}_3\text{COO}^-) + l_0(\text{Na}^+) + l_0(\text{Cl}^-)$$

This method is especially useful for calculating conductance for weak electrolytes such as acetic acid. As evident from Figure 15-2, the Λ_0 value for acetic acid cannot be determined accurately by extrapolation because of the steep rise of conductance in dilute solutions. For strong electrolytes, on the other hand, the extrapolation can be made very accurately. Thus, in the example above, the values of Λ_0 for HCl, CH_3COONa , and NaCl are determined easily by extrapolation as the substances are strong electrolytes. Substitution of these extrapolated values, as given in Table 15-2, yields a value of 390.6 for the value of Λ_0 for CH_3COOH .

Ionization of weak electrolytes

When Arrhenius introduced his theory of ionization he proposed that the degree of ionization, α , of an electrolyte is measured by the ratio

$$\alpha = \frac{\Lambda}{\Lambda_0} \quad (4)$$

where Λ is the equivalent conductance of the electrolyte at any specified concentration of solution and Λ_0 is the equivalent conductance at infinite dilution. As strong electrolytes were not yet recognized as being 100% ionized, and interionic interference effects had not been evaluated, he believed the equation to be applicable to both strong and weak electrolytes. It now is known that the apparent variation of ionization of strong electrolytes arises from a change in the mobility of ions at different concentrations, rather than from varying ionization, so the equation is not applicable to strong electrolytes. It does provide, however, a generally acceptable approximation of the degree of ionization of weak electrolytes, for which deviations resulting from neglect of activity coefficients and of some change of ionic mobilities with concentration are, for most purposes, negligible. The following example in Box 15-1 illustrates the use of the equation to calculate the degree of ionization of a typical weak electrolyte.

The degree of dissociation also can be calculated using the van't Hoff factor, i , and

$$\alpha = \frac{i - 1}{v - 1} \quad (5)$$

where v is the number of ions into which the electrolyte dissociates.

Table 15-2. Equivalent Ionic Conductivities at Infinite Dilution, at 25°C

CATIONS	l_0	ANIONS	l_0
H^+	349.8	OH^-	198.0
Li^+	38.7	Cl^-	76.3
Na^+	50.1	Br^-	78.4
K^+	73.5	I^-	76.8
NH_4^+	61.9	AcO^-	40.9
$\frac{1}{2} \text{Ca}^{2+}$	59.5	$\frac{1}{2} \text{SO}_4^{2-}$	79.8
$\frac{1}{2} \text{Mg}^{2+}$	53.0		

Box 15-1

1. Calculate the degree of ionization of 1×10^{-3} N acetic acid, the equivalent conductance of which is 48.15 mho cm^2/eq . The equivalent conductance at infinite dilution is 390.6 mho cm^2/eq .

$$\alpha = \frac{48.15}{390.6} = 0.12$$

The percentage ionization is therefore 12%.

2. A 1.0×10^{-3} N solution of acetic acid has a van't Hoff factor equal to 1.12. Calculate the degree of dissociation of the acid at this concentration.

$$\alpha = \frac{i - 1}{v - 1} = \frac{1.12 - 1}{2 - 1} = 0.12$$

This result agrees with that obtained using equivalent conductance and Equation 4.

Ionic strength

Ionic strength is a measure of the intensity of the electrical field in a solution and may be expressed as

$$\mu = \frac{1}{2} \sum C_i z_i^2$$

where z_i is the valence of ion i . The **mean ionic activity coefficient** is a function of ionic strength as are such diverse phenomena as solubilities of sparingly soluble substances, rates of ionic reactions, effects of salts on pH of buffers, electrophoresis of proteins, and so on.

The greater effectiveness of ions of higher charge on a specific property, compared with the effectiveness of the same number of singly charged ions, generally coincides with the ionic strength calculated by Equation 6. The variation of ionic strength with the valence (charge) of the ions comprising a strong electrolyte should be noted.

For univalent cations and univalent anions (called **uniunivalent** or 1-1) electrolytes, the ionic strength is identical with molarity. For bivalent cation and univalent anion (**biunivalent** or 2-1) electrolytes, or univalent cation and bivalent anion (**unibivalent** or 1-2) electrolytes, the ionic strength is three times the molarity. For bivalent cation and bivalent anion (**bibivalent** or 2-2) electrolytes, the ionic strength is four times the molarity. These relationships are evident from the following example in Box 15-2.

The ionic strength of a solution containing more than one electrolyte is the sum of the ionic strengths of the individual salts comprising the solution. For example, the ionic strength of a solution containing NaCl, Na_2SO_4 , MgCl_2 , and MgSO_4 , each at a concentration of 0.1 M, is 1.1.

Debye-Huckel theory

The **Debye-Huckel equations**, which are applicable only to very dilute solutions (about 0.02 M), may be extended to some-

what more concentrated solutions (about 0.1 M) in the simplified form

$$\log f_i = \frac{-0.51 z_i^2 \sqrt{\mu}}{1 + \sqrt{\mu}} \quad (7)$$

The mean ionic activity coefficient for aqueous solutions of electrolytes at 25° can be expressed as

$$\log f_{\pm} = \frac{-0.51 z_+ z_- \sqrt{\mu}}{1 + \sqrt{\mu}} \quad (8)$$

in which z_+ is the valence of the cation and z_- is the valence of the anion. When the ionic strength of the solution becomes high (approximately 0.3 to 0.5), these equations become inadequate and a linear term in μ is added. This is illustrated for the mean ionic activity coefficient,

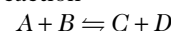
$$\log f_{\pm} = \frac{-0.51 z_+ z_- \sqrt{\mu}}{1 + \sqrt{\mu}} + k_s \mu \quad (9)$$

in which K_s is a **salting-out constant** chosen empirically for each salt. This equation is valid for solutions with ionic strength up to approximately 1.

CHEMICAL EQUILIBRIA

Equilibrium is the balance between two opposing reactions or forces. Most chemical reactions proceed in both forward and reverse directions. Chemical equilibrium maintains constant concentration of reactants and products as both processes are occurring at the same rate. Thus the concentration of the reactants and the products may not necessarily be equal but the rate of the forward reaction equals the rate of the reverse reaction. In practice some reactions can proceed almost to completion and thus can be considered as irreversible. The ionization of weak electrolytes, however, is a reversible reaction and if the overall system remains static, that is, the products are not removed, then an equilibrium expression can be written.

For the generalised reaction



the **equilibrium constant (K)** for the reaction can be written as the product of the concentration of the reactants divided by the product of the concentration of the products, i.e.

$$K = \frac{[C]x[D]}{[A]x[B]} \quad (10)$$

For reactions where the concentration of the products is high, the equilibrium will be driven to the right hand side of the equation. Thus the numerator will exceed the denominator and K will have a large value. Conversely, if the reaction only proceeds to a very limited extent, the equilibrium will lie mainly to the left and K will be less than unity. Le Chatelier's Principle states that if a dynamic equilibrium is disturbed by changing the conditions, then the position of equilibrium will move to counteract the change. The equilibrium constant, K, will therefore remain constant under fixed conditions.

BRØNSTED-LOWRY ACID-BASE THEORY

Due to shortcomings in the Arrhenius theory, it was recognised that a broader definition was needed and this led to parallel theories being proposed in 1923 by Brønsted and Lowry. It defines an acid as a substance that can ionise in solution to give a solvated hydrogen ion (i.e., a proton stabilised by interaction with either the solvent or a substance in solution). Conversely a base is a substance that will accept a hydrogen atom. This led to the simplified concept that an acid is a **proton donor** and a base is a **proton acceptor**. This led to the concept of a complementary relationship between acids and bases (similar to that between oxidants and reductants). This ionization process will always involve both species and they are known as **conjugate acid-base pairs**.

Box 15-2

Calculate the ionic strength of 0.1 M solutions of NaCl, Na_2SO_4 , MgCl_2 , and MgSO_4 , respectively, for

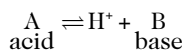
$$\text{NaCl } \mu = \frac{1}{2} (0.1 \times 1^2 + 0.1 \times 1^2) = 0.1$$

$$\text{Na}_2\text{SO}_4 \mu = \frac{1}{2} (0.2 \times 1^2 + 0.1 \times 2^2) = 0.3$$

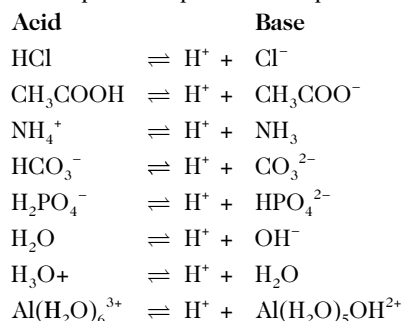
$$\text{MgCl}_2 \mu = \frac{1}{2} (0.1 \times 2^2 + 0.2 \times 1^2) = 0.3$$

$$\text{MgSO}_4 \mu = \frac{1}{2} (0.1 \times 2^2 + 0.1 \times 2^2) = 0.4$$

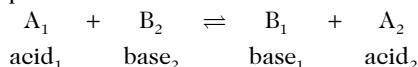
This conjugate acid-base relationship may be expressed by



The pair of substances is thus related through their mutual ability to gain or lose a proton. Specific examples of such pairs are

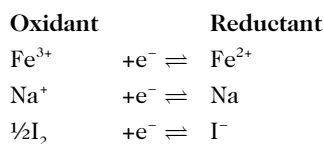


However, these examples of acid-base pairs represent reactions that are possible in principle only. Protons, (and electrons), can be transferred only from one substance (an ion, atom, or molecule) to another. Thus an acid will not release a proton unless a base capable of accepting it is present simultaneously. This means that any actual manifestation of acid-base behavior must involve interaction between two sets of conjugate acid-base pairs, represented as

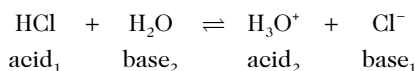


In such a reaction, which is called **protolysis** or a **protolytic reaction**, A_1 and B_1 constitute one conjugate acid-base pair, and A_2 and B_2 the other; the proton given up by A_1 (which thereby becomes B_1) is transferred to B_2 (which becomes A_2).

The complementary nature of the acid-base pairs listed is reminiscent of the complementary relationship of pairs of oxidants and reductants where, however, the ability to gain or lose one or more electrons—rather than protons—is the distinguishing characteristic.

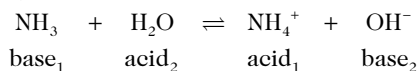


When an acid, such as hydrochloric, is dissolved in water, a protolytic reaction occurs.

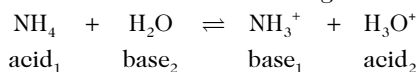


The ionic species H_3O^+ , called **hydronium** or **oxonium** ion, always is formed when an acid is dissolved in water. Often, for convenience, this is written simply as H^+ and is called the hydrogen ion, although the “bare” ion practically is nonexistent in solution.

When a base (e.g., ammonia) is dissolved in water, the reaction of protolysis is



It is apparent that not only molecules, but also cations and anions, may function as acids or bases. The conjugate acid, NH_4^+ is also capable of protolysis in water, thus when ammonium chloride is dissolved in water the following reactions occurs

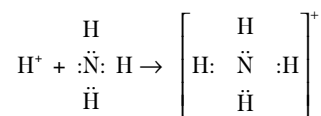


The proton theory of acid-base function makes the concept of hydrolysis superfluous. In the preceding examples water sometimes behaves as an acid and at other times as a base. Such an amphoteric substance is called, in Brønsted's terminology, an **amphiprotic** substance.

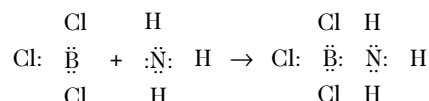
LEWIS ACID-BASE THEORY

The proton concept of acids and bases provides a more general definition for these substances, but it does not indicate the basic reason for proton transfer, nor does it explain how such substances as sulfur trioxide, which is not capable of donating a proton—can behave as acids. A number of theories have been proposed to counteract this deficiency, and a more inclusive definition of acids and bases, the electronic theory, was proposed by Lewis in 1923. In 1916 he proposed that sharing of a pair of electrons by two atoms established a bond (covalent) between the atoms; therefore, an acid is a molecule or ion capable of accepting a pair of electrons made available by another substance called a base, thereby forming a **coordinate covalent bond**. The base is the substance that donates a share in its electron pair to the acid. Thus substances such as boron trichloride, stannic chloride and carbon dioxide can be classified as acids whereas amines, ethers and carboxylic acid anhydrides, which do not contain any hydroxyl group, can be act as bases.

The following equation illustrates how Lewis' definitions explain the transfer of a proton (hydrogen ion) to ammonia to form ammonium ion.



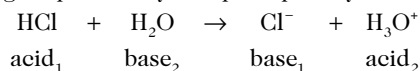
The reaction of boron trichloride, which according to the Lewis theory is an acid, with ammonia is similar, for the boron lacks an electron pair if it is to attain a stable octet configuration, whereas ammonia has a pair of electrons that may be shared, thus,



For acid-base and ionic equilibria, it is the Brønsted-Lowry theory that is most useful and is the definition generally quoted with respect to pharmaceuticals.

Dissociation of strong acids and bases

When the strong acids such as HClO_4 , H_2SO_4 , HCl , or HNO_3 are dissolved in water, the solutions—if they are of identical normality and are not too concentrated—all have about the same hydrogen-ion concentration, indicating the acids to be of about the same strength. The reason for this is that each one of these acids undergoes practically complete protolysis in water.

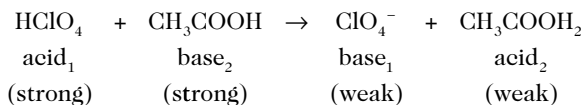


This phenomenon, called the **leveling effect of water**, occurs whenever the added acid is stronger than the hydronium ion. Such a reaction manifests the tendency of proton-transfer reactions to proceed spontaneously in the direction of forming a weaker acid or weaker base.

Because the strongest acid that can exist in an amphiprotic solvent is the conjugate acid form of the solvent, any stronger acid will undergo protolysis to the weaker solvent acid. HClO_4 , H_2SO_4 , HCl , or HNO_3 are all stronger acids than the hydronium ion, so they are converted in water to the hydronium ion.

When the strong bases sodium hydride, sodium amide, or sodium ethoxide are dissolved in water, each reacts with water to form sodium hydroxide. These reactions illustrate the leveling effect of water on bases. Because the hydroxide ion is the strongest base that can exist in water, any base stronger than the hydroxide ion undergoes protolysis to hydroxide.

Intrinsic differences in the acidity of acids become evident if they are dissolved in a relatively poor proton acceptor such as anhydrous acetic acid. Perchloric acid (HClO_4), a strong acid, undergoes practically complete reaction with acetic acid to produce the **acetonium** ion (acid_2):



but sulfuric acid and hydrochloric acid behave as weak acids. It is because perchloric acid is a very strong acid when dissolved in glacial acetic acid that it has found many important applications in analytical chemistry as a titrant for a variety of substances that behave as bases in acetic acid. Because of its ability to differentiate the acidity of various acids, it is called a **differentiating solvent for acids**; this property results from its relatively weak proton-acceptor tendency. A solvent that differentiates basicity of different bases must have a weak proton-donor tendency; it is called a **differentiating solvent for bases**. Liquid ammonia is typical of solvents in this category.

Solvents that have both weak proton-donor and proton-acceptor tendencies are called **aprotic solvents** and may serve as differentiating solvents for both acids and bases; they have little if any action on solutes and serve mainly as inert dispersion media for the solutes. Useful aprotic solvents are benzene, toluene, or hexane.

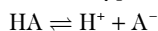
IONIZATION OF ACIDS AND BASES

Ionization of acids

Acids and bases commonly are classified as strong or weak acids and strong or weak bases depending on whether they are ionized extensively or slightly in aqueous solutions. If, for example, 1 N aqueous solutions of hydrochloric acid and acetic acid are compared, it is found that the former is a better conductor of electricity, reacts much more readily with metals, catalyzes certain reactions more efficiently, and possesses a more acid taste than the latter. Both solutions, however, will neutralize identical amounts of alkali. A similar comparison of 1 N solutions of sodium hydroxide and ammonia reveals the former to be more active than the latter, although both solutions will neutralize identical quantities of acid.

The differences in the properties of the two acids are attributed to differences in the concentration of hydrogen (more accurately hydronium) ion, the hydrochloric acid being ionized to a greater extent and thus containing a higher concentration of hydrogen ion than acetic acid. Similarly, most of the differences between the sodium hydroxide and ammonia solutions are attributed to the higher hydroxyl-ion concentration in the former.

The ionization of incompletely ionized acids may be considered a reversible reaction of the type

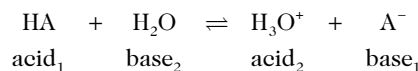


where HA is the molecular acid and A⁻ is its anion. An equilibrium expression based on the law of mass action may be applied to the reaction

$$K_a = \frac{[\text{H}^+][\text{A}^-]}{[\text{HA}]} \quad (11)$$

where K_a is the **ionization constant** or **dissociation constant**, and the brackets signify the concentration of each species. It should be borne in mind, that in some pharmaceutical preparations, especially when formulating poorly soluble drugs, it can be common to add a cosolvent such as ethanol or propylene glycol. The resulting change in dielectric constant causes a shift in the equilibrium and thus the pK_a is altered, decreasing acid and basic strength.

Although the strength of an acid commonly is measured in terms of the ionization or dissociation constant defined in Equation 11, the process of ionization is not that simple given that the proton will not detach itself from one molecule unless it is accepted simultaneously by another molecule. When an acid is dissolved in water, the latter acts as a base, accepting a proton (Brønsted's definition of a base) by donating a share in a pair of electrons (Lewis' definition of a base). This reaction may be written as



Application of the law of mass action gives

$$K = \frac{[\text{H}_3\text{O}^+][\text{A}^-]}{[\text{HA}][\text{H}_2\text{O}]} \quad (12)$$

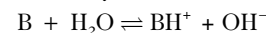
For ionizations in dilute solutions, it is acceptable to neglect the role of water as it can be considered as remaining constant (at approximately 55.3 moles/liter at 25 °C). The constant K can be rewritten to include the concentration of water and the equation becomes

$$K_a = \frac{[\text{H}_3\text{O}^+][\text{A}^-]}{[\text{HA}]} \quad (13)$$

where K_a is the ionization or dissociation constant (or the **acidity constant**). This is numerically identical to equation 11 as [H₃O⁺] is numerically equal to [H⁺]. For exact work, the **thermodynamic activity** should be used in place of concentration, especially for more concentrated solutions. K_a remains relatively constant for any given weakly ionized acid in any given solvent and at any given temperature. With increasingly stronger acids, some deviation may occur but in simple terms, the ionization constant increased with increasing acid strength (due to a shift in the equilibrium to the right-hand side of the equation).

Ionization of bases

The ionization of a base may be written



The equilibrium expression for this reaction is

$$K = \frac{[\text{BH}^+][\text{OH}^-]}{[\text{H}_2\text{O}][\text{B}]} \quad (14)$$

With [H₂O] constant, as before, the expression becomes

$$K_b = \frac{[\text{BH}^+][\text{OH}^-]}{[\text{B}]} \quad (15)$$

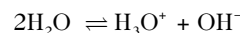
Ionization of a base is often exemplified by ammonia, NH₃. Thus for ammonia, the expression can be written

$$K_b = \frac{[\text{NH}_4^+][\text{OH}^-]}{[\text{NH}_3]} \quad (16)$$

and K_b for ammonia is 1.8 × 10⁻⁵. K_b is not used extensively in modern terminology as it is more useful to relate both acids and bases on the same terms, i.e. their affinity for the hydrogen ion.

Ionization of water

Although it is a poor conductor of electricity, pure water does ionize through a process known as **autoprotolysis**, in the following manner:



Application of the law of mass action to this reaction gives

$$K = \frac{[\text{H}_3\text{O}^+][\text{OH}^-]}{[\text{H}_2\text{O}]^2} \quad (17)$$

where K is the equilibrium constant for the reaction. Because the concentration of H₂O (molecular water) is much higher the hydronium ion or hydroxyl ion concentrations, it can be considered to be constant and [H₂O]² can be combined with K to give a new constant, K_w, known as the **ion product of water**, the **dissociation constant of water** or **autoprotolysis constant**. K_w is approximately equal to 1 × 10⁻¹⁴ at 25°, however, it does vary with temperature. For example it is 2.57 × 10⁻¹⁴ at 37°C, and this should be considered in any calculations pertaining to deviations from the standard.

$$K_w = [\text{H}_3\text{O}^+][\text{OH}^-] = K[\text{H}_2\text{O}]^2 \quad (18)$$

Because the autoprotolysis of pure water yields one hydronium ion for each hydroxyl ion produced, [H₃O⁺] equal to [OH⁻]. At

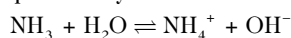
25°C each has a value of 1×10^{-7} mol/L ($1 \times 10^{-7} \times 1 \times 10^{-7}$ so $K_w = 1 \times 10^{-14}$). A solution in which $[H_3O^+]$ is equal to $[OH^-]$ is termed a **neutral solution**.

If an acid is added to water, the hydronium ion concentration will be increased and the equilibrium between hydronium and hydroxyl ions will be disturbed momentarily. To restore equilibrium, some of the hydroxyl ions, originally present in the water, will combine with a part of the added hydronium ions to form nonionized water molecules, until the product of the concentrations of the two ions has been reduced to 10^{-14} . When equilibrium again is restored, the concentrations of the two ions no longer will be equal. If, for example, the hydronium-ion concentration is 1×10^{-3} N when equilibrium is established, the concentration of hydroxyl ion will be 1×10^{-11} (the product of the two concentrations being equal to 10^{-14}). As $[H_3O^+]$ is much greater than $[OH^-]$ the solution is said to be an **acid** or **acidic**.

In a similar manner, the addition of an alkali to pure water momentarily disturbs the equilibrium between hydronium and hydroxyl ions. To restore equilibrium, some of the hydronium ions originally present in the water will combine with part of the added hydroxyl ions to form nonionized water molecules. The process continues until the product of the hydronium and hydroxyl ion concentrations again is equal to 10^{-14} . Assuming that the final hydroxyl-ion concentration is 1×10^{-4} N, the concentration of hydronium ion in the solution will be 1×10^{-10} . Because $[OH^-]$ is much greater than $[H_3O^+]$, the solution is said to be **basic** or **alkaline**.

Relationship of K_a and K_b

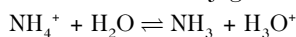
A particularly interesting and useful relationship between the strength of an acid and its conjugate base, or a base and its conjugate acid, exists. For illustration, consider the strength of the base NH_3 and its conjugate acid NH_4^+ in water. The behavior of NH_3 as a base is expressed by



The equilibrium expression (as stated in Equation 16) is

$$K_b = \frac{[NH_4^+][OH^-]}{[NH_3]} \quad (19)$$

If we consider the behavior of the conjugate acid, NH_4^+ , in water,



then the equilibrium expression is

$$K_a = \frac{[NH_3][H_3O^+]}{[NH_4^+]} \quad (20)$$

Multiplying Equations 19 and 20

$$K_a K_b = \frac{[NH_3][H_3O^+][NH_4^+][OH^-]}{[NH_4^+][NH_3]} \quad (21)$$

It is obvious that

$$K_a K_b = [H_3O^+][OH^-] = K_w \quad (22)$$

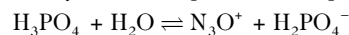
where K_w is the ion product of water as defined in Equation 18. The utility of this relationship, which is a general one for any conjugate acid/base pair, is evident from the following deductions: (1) The strength of an acid may be expressed in terms either of the K_a or the K_b of its conjugate base, or vice versa; (2) the K_a of an acid may be calculated if the K_b of its conjugate base is known, or vice versa; and (3) the stronger an acid is, the weaker its conjugate base, or vice versa.

The pH of the purest water obtainable, so-called **conductivity water**, is 7 when the measurement is made carefully under conditions to exclude carbon dioxide and prevent errors inherent in the measuring technique (such as acidity or alkalinity of the indicator). Upon agitating this water in the presence of carbon dioxide in the atmosphere (equilibrium water), the value drops rapidly to 5.7. This is the pH of nearly all distilled water that has been exposed to the atmosphere for even a short time and often is called **equilibrium water**.

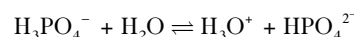
It should be emphasized strongly that the generalizations stated concerning neutrality, acidity, and alkalinity hold exactly only when (1) the solvent is water, (2) the temperature is 25°C, and (3) there are no other factors to cause deviation from the simply formulated equilibria underlying the definition of pH given in the preceding discussion.

Ionization of polyprotic electrolytes

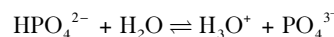
Acids that are capable of donating more than one proton and bases capable of accepting more than one proton are termed **polyprotic**. The ionization of a polyprotic acid occurs in stages, two stages to diprotic (or dibasic) acids, three stages for triprotic acids, and can be illustrated by considering the equilibria involved in the ionization of phosphoric acid, a tribasic acid. In general for a polyprotic substance (e.g., HnA), there are $n+1$ possible species in solution. The three ionization processes are considered separately and the equilibrium expressions are:



$$K_1 = \frac{[H_3O^+][H_2PO_4^-]}{[H_3PO_4]} \quad (23)$$



$$K_2 = \frac{[H_3O^+][HPO_4^{2-}]}{[H_2PO_4^-]} \quad (24)$$

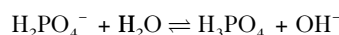
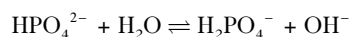
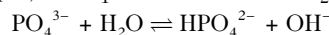


$$K_3 = \frac{[H_3O^+][PO_4^{3-}]}{[HPO_4^{2-}]} \quad (25)$$

If the three expressions for the ionization constants are multiplied together, an overall ionization, K , can be obtained

$$K = K_1 K_2 K_3 = \frac{[H_3O^+]^3 [PO_4^{3-}]}{[H_3PO_4]} \quad (26)$$

Each of the successive ionizations is suppressed by the hydronium ion formed from preceding stages according to Le Chatelier's principle. The successive dissociation constants always decrease in value, as successive protons must be removed from species that always are charged more negatively and this are more difficult. The third and final ionization of phosphoric acid is very weak and the amount of PO_4^{3-} is very low. This can be seen from the data in Table 15-3, in which K_1 for phosphoric acid is approximately 100,000 times greater than K_2 , which is in turn approximately 100,000 times greater than K_3 . Although successive dissociation constants are always smaller, the difference is not always as great as it is for phosphoric acid. Tartaric acid, for example, has $K_1 = 9.12 \times 10^{-4}$ and $K_2 = 4.27 \times 10^{-5}$.



Applying the law of mass action to this series of reactions, and using the concepts outlined in Equations 14–16, the relationship between the various K_a and K_b values for phosphoric acid are

$$K_w = K_{a1} \times K_{b3} = K_{a2} \times K_{b2} \times K_{a3} \times K_{b1} \quad (27)$$

where K_{a1} , K_{a2} , and K_{a3} refer to the equilibria given by Equations, respectively; K_{b1} , K_{b2} , and K_{b3} refer to the reaction of PO_4^{3-} , HPO_4^{2-} , and $H_2PO_4^-$, respectively, with water.

Electronegativity and dissociation constants

Table 15-3 gives the dissociation constants of several weak acids and weak bases, in water, at 25°C. Strong acids and strong bases do not obey the law of mass action, so dissociation constants cannot be formulated for these strong electrolytes. Table 15-3 shows that great variations occur in the strength of weak

Table 15-3. Dissociation Constants in Water at 25°C

Substance		K
Weak acids		
Acetic		1.75×10^{-3}
Acetylsalicylic		3.27×10^{-4}
Barbital		1.23×10^{-8}
Barbituric		1.05×10^{-4}
Benzoic		6.30×10^{-5}
Benzyl penicillin		1.74×10^{-3}
Boric	K_1	5.8×10^{-10}
Caffeine		1×10^{-14}
Carbonic	K_1	4.31×10^{-7}
	K_2	4.7×10^{-11}
Citric (1H ₂ O)	K_1	7.0×10^{-4}
	K_2	1.8×10^{-5}
	K_3	4.0×10^{-7}
Dichloroacetic		5×10^{-2}
Ethylenediaminetetraacetic acid (EDTA)	K_1	1×10^{-2}
	K_2	2.14×10^{-3}
	K_3	6.92×10^{-7}
	K_4	5.5×10^{-11}
Formic		1.77×10^{-4}
Glycerophosphoric	K_1	3.4×10^{-2}
	K_2	6.4×10^{-7}
Glycine	K_1	4.5×10^{-3}
	K_2	1.7×10^{-10}
Lactic		1.39×10^{-4}
Mandelic		4.29×10^{-4}
Monochloroacetic		1.4×10^{-3}
Oxalic (2H ₂ O)	K_1	5.5×10^{-2}
	K_2	5.3×10^{-5}
Phenobarbital		3.9×10^{-8}
Phenol		1×10^{-10}
Phosphoric	K_1	7.5×10^{-3}
	K_2	6.2×10^{-8}
	K_3	2.1×10^{-13}
Picric		4.2×10^{-1}
Propionic		1.34×10^{-5}
Saccharin		2.5×10^{-2}
Salicylic		1.06×10^{-3}
Succinic	K_1	6.4×10^{-5}
	K_2	2.3×10^{-6}
Sulfadiazine		3.3×10^{-7}
Sulfamerazine		8.7×10^{-8}
Sulfapyridine		3.6×10^{-9}
Sulfathiazole		7.6×10^{-8}
Tartaric	K_1	9.6×10^{-4}
	K_2	4.4×10^{-5}
Trichloroacetic		1.3×10^{-1}
Weak bases		
Acetanilide		4.1×10^{-14} (40°)
Ammonia		1.74×10^{-5}
Apomorphine		1.0×10^{-7}
Atropine		4.5×10^{-5}
Benzocaine		6.0×10^{-12}
Caffeine		4.1×10^{-14} (40°)
Cocaine		2.6×10^{-6}
Codeine		9×10^{-7}

Substance		K
Ephedrine		2.3×10^{-5}
Morphine		7.4×10^{-7}
Papaverine		8×10^{-9}
Physostigmine	K_1	7.6×10^{-7}
	K_2	5.7×10^{-13}
Pilocarpine	K_1	7×10^{-8}
	K_2	2×10^{-13}
Procaine		7×10^{-6}
Pyridine		1.4×10^{-9}
Quinine	K_1	1.0×10^{-6}
	K_2	1.3×10^{-10}
Reserpine		4×10^{-8}
Strychnine	K_1	1×10^{-6}
	K_2	2×10^{-12}
Theobromine		4.8×10^{-14} (40°)
Thiourea		1.1×10^{-15}
Urea		1.5×10^{-14}

acids and weak bases. Dissociation constants for pharmaceuticals are often found in the appropriate monographs, published literature or some useful lists.¹ The effect of various substituents on the strength of acids and bases depends on the electronegativity of the substituent atom or radical. For example, the substitution of one chlorine atom into the molecule of acetic acid increases the degree of ionization of the acid. Substitution of two chlorine atoms further increases the degree of ionization, and introduction of three chlorine atoms produces a still stronger acid. Acetic acid ionizes primarily because the oxygen atom adjacent to the hydrogen atom of the carboxyl group has a stronger affinity for electrons than the hydrogen atom. Thus, when acetic acid is dissolved in water, the polar molecules of the water have a stronger affinity for the hydrogen of acetic acid than the hydrogen atoms of water. The acetic acid ionizes as a consequence of this difference in affinities.

When an atom of chlorine is introduced into the acetic acid molecule, forming ClCH_2COOH , the electrons in the molecule are attracted very strongly to the chlorine because of its relatively high electronegativity; the bond between the hydrogen and the oxygen in the carboxyl group is thereby weakened, and the degree of ionization increased. Introduction of two or three chlorine atoms weakens the bond further and increases the strength of the acid. On the other hand, substitution of chlorine into the molecule of ammonia reduces the strength of the base because of its decreased affinity for the hydrogen ion.

Ionic strength and dissociation constants

Most solutions of pharmaceutical interest are in a concentration range such that the ionic strength of the solution may have a marked effect on ionic equilibria and observed dissociation constants. One method of correcting dissociation constants for solutions with an ionic strength up to about 0.3 is to calculate an **apparent dissociation constant**, pK'_a as

$$pK'_a = pK_a + \frac{0.51 (2Z - 1) \sqrt{\mu}}{1 + \sqrt{\mu}} \quad (28)$$

in which pK_a is the tabulated thermodynamic dissociation constant, Z is the charge on the acid, and 0 is the ionic strength. See Box 15-3.

DETERMINATION OF PH AND PKA

The numerical values of hydronium ion concentration may vary enormously; for a normal solution of a strong acid the value is nearly 1, whereas for a normal solution of a strong base it is approximately 1×10^{-14} ; there is a variation of 100,000,000,000,000

Box 15-3

Calculate pK'_a , for succinic acid at an ionic strength of 0.1. Assume that pK_2 is 5.63. The charge on the acid species is -1 .

$$\begin{aligned} pK'_2 &= 5.63 \frac{0.51(-2-1)\sqrt{0.1}}{1 + \sqrt{0.1}} \\ &= 5.63 - 0.37 = 5.26 \end{aligned}$$

between these two limits. Because of the inconvenience of dealing with such large numbers, in 1909 Sørensen proposed that hydronium ion concentration be expressed in terms of the logarithm (log) of its reciprocal. To this value he assigned the symbol pH. Mathematically this is written as

$$pH = \log \frac{1}{[H_3O^+]} \quad (29)$$

This equation can also be displayed as

$$pH = \log 1 - \log[H_3O^+] \quad (30)$$

$$pH = -\log[H_3O^+] \quad (31)$$

as the logarithm of 1 is zero.

Thus the pH also may be defined as the negative logarithm of the hydronium ion concentration. In general, this type of notation is used to indicate the negative logarithm of the term that is preceded by the p, which gives rise to the following

$$pOH = \log[OH^-] \quad (32)$$

and similarly

$$pK_a = -\log K_a \quad (33)$$

and

$$pK_b = -\log K_b \quad (34)$$

This enables the pH of a solution to be considered on a numerical scale from 0–14 and is more convenient in terms of speech, writing and data manipulation. Acidic solutions having a predominance of $[H_3O^+]$ have pH values between 0 and 7. The hydronium ion concentration of pure water, at 25°C, is $1 \times 10^{-7} N$, corresponding to a pH of 7. This figure, therefore, is designated as the neutral point, and all values below a pH of 7 represent acidity—the smaller the number, the greater the acidity. Values above 7 represent alkalinity—the larger the number, the greater the alkalinity. Mathematically there is no reason why negative numbers or numbers above 14 should not be used. In practice, however, such values are never encountered because solutions that might be expected to have such values are too concentrated to be ionized extensively or the interionic attraction is so great as to materially reduce ionic activity.

Ionization constants are small, often inconvenient figures whereas the pK_a is convenient in speech and notation. This, the K_a of acetic acid is 1.75×10^{-5} , which can be converted to the more convenient 4.76.

The relationship of pH to hydronium ion and hydroxyl ion concentrations may be seen in Table 15-4. A calculation of pK_a is given in Box 15-4.

When bases are expressed as K_b , it is often convenient to convert them to K_a . The ionic product of water can be used to facilitate this and the “p” notation is useful. K_a and K_b are related to K_w

$$K_a \times K_b = K_w = [H_3O^+][OH^-] = 1 \times 10^{-14} \text{ (at 25°C)} \quad (35)$$

Taking negative logarithms

$$pK_a + pK_b = pK_w = 14.00 \text{ (at 25°C)} \quad (36)$$

The pK_b value for ammonia is 4.74 at 25°C but it is more common and convenient to use the pK_a value of 9.26. See Box 15-5 for a calculation.

Box 15-4

The following examples illustrate the conversion from exponential to p notation.

1. Calculate the pH corresponding to a hydronium ion concentration of 1×10^{-4} g-ion/L.

Solution:

$$\begin{aligned} pH &= \log \frac{1}{1 \times 10^{-4}} \\ &= \log 10,000 \text{ or } \log (1 \times 10^{-4}) \\ (1 \times 10^{-4}) &= +4 \\ pH &= 4 \end{aligned}$$

2. Calculate the pH corresponding to a hydronium ion concentration of 0.000036 N (or g-ion/L). (Note: This more frequently is written as a number multiplied by a power of 10, thus, 3.6×10^{-5} for 0.000036.)

$$\begin{aligned} pH &= \log \frac{1}{3.6 \times 10^{-5}} \\ &= \log 28,000 \text{ or } \log (2.8 \times 10^{-4}) \\ \log (1 \times 10^{-4}) &= \log 2.8 + 10^{-4} \\ \log 2.8 &= +0.44 \\ \log 10^{-4} &= +4.00 \\ pH &= 4.44 \end{aligned}$$

This problem also may be solved as follows:

$$\begin{aligned} pH &= -\log (3.6 \times 10^{-5}) \\ \log 3.6 &= +0.56 \\ \log 10^{-5} &= -5.00 \\ &= -4.00 = \log (3.6 \times 10^{-5}) \\ pH &= -(-4.44) = +4.44 = 4.44 \end{aligned}$$

The procedure should be reversed to convert pH to hydronium ion-concentration.

Thus:

$$\begin{aligned} pH &= \log \frac{1}{[H_3O^+]} \\ 4.44 &= \log \frac{1}{[H_3O^+]} \\ \frac{1}{[H_3O^+]} &= \text{antilog of } 4.44 = 28,000 \text{ (rounded off)} \\ [H_3O^+] &= \frac{1}{28,000} = 0.000036 \text{ or } 3.6 \times 10^{-5} \end{aligned}$$

Tip:

In finding the antilog of -4.44 it should be kept in mind that the mantissa (the number to the right of the decimal point) of a log to the base 10 (the common or Briggsian logarithm base) is **always positive** but that the characteristic (the number to the left of the decimal point) may be **positive or negative**. As the entire log -4.44 is negative, it is obvious that one cannot look up the antilog of -0.44 . Using modern calculators, the order of entry of functions can be checked by remembering that a neutral solution (pH 7) has a hydronium ion-concentration of 1×10^{-7} g-ion/L.

Table 15-4. Hydronium ion and hydroxyl ion concentrations

	pH	Normality in Terms of Hydroniumion	Normality in Terms of Hydroxylion
	0	10 ⁰ (=1)	10 ⁻¹⁴
	1	10 ⁻¹	10 ⁻¹³
	2	10 ⁻²	10 ⁻¹²
Increasing acidity	3	10 ⁻³	10 ⁻¹¹
	4	10 ⁻⁴	10 ⁻¹⁰
	5	10 ⁻⁵	10 ⁻⁹
Neutral point	6	10 ⁻⁶	10 ⁻⁸
	7	10 ⁻⁷	10 ⁻⁷
	8	10 ⁻⁸	10 ⁻⁶
	9	10 ⁻⁹	10 ⁻⁵
	10	10 ⁻¹⁰	10 ⁻⁴
	11	10 ⁻¹¹	10 ⁻³
Increasing alkalinity	12	10 ⁻¹²	10 ⁻²
	13	10 ⁻¹³	10 ⁻¹
	14	10 ⁻¹⁴	10 ⁰ (=1)

Box 15-5

Calculate the pK_a of atropine given that the K_b is 4.5 × 10⁻⁵.

$$K_a \times K_b = K_w$$

$$K_a = \frac{1 \times 10^{-14}}{4.5 \times 10^{-5}}$$

$$K_a = 222 \times 10^{-10}$$

$$pK_a = 9.65$$

OR

$$K_b \text{ is } 4.5 \times 10^{-5} \text{ so } pK_b = 4.35$$

$$pK_a = 14.00 - pK_b = 9.65$$

Species concentration

When a weak acid, H_nA is added to water, n+1 species, including the unionized acid, can exist. After equilibrium is established, the sum of the concentrations of all species must be equal to C_a, the stoichiometric (added) concentration of acid. Thus, for a triprotic acid H₃A,

$$C_a = [H_3A] + [H_2A^-] + [HA^{2-}] + [A^{3-}] \quad (37)$$

In addition, the concentrations of all acidic and basic species in solution vary with pH, and can be represented solely in terms of equilibrium constants and the hydronium concentration. These relationships may be expressed as

$$[H_nA] = [H_3O^+]^n C_a / D \quad (38)$$

$$[H_{n-j}A^{-j}] = [H_3O^+]^{n-j} K_1, \dots, K_j C_a / D \quad (39)$$

in which *n* represents the total number of dissociable hydrogens in the parent acid, *j* is the number of protons dissociated, C_a is the stoichiometric concentration of acid, and *K* represents the acid dissociation constants. The term *D* is a power series in [H₃O⁺] and *K*, starting with [H₃O⁺] raised to the *n*th power. The last term is the product of all the dissociation constants. The intermediate terms can be generated from the last term by substituting [H₃O⁺] for K_n to obtain the next-to-last term, then substituting [H₃O⁺] for K_{n-1} to obtain the next term, and onward until the first term is reached. The following examples show the denominator, *D*, to be used for various types of acids:

$$H_3A: D = [H_3O^+]^3 + K_1[H_3O^+]^2 + K_1K_2[H_3O^+] + K_1K_2K_3 \quad (40)$$

$$H_2A: D = [H_3O^+]^2 + K_1[H_3O^+] + K_1K_2 \quad (41)$$

$$HA: D = [H_3O^+] + K_a \quad (42)$$

The numerator in all instances is C_a multiplied by the term from the denominator that has [H₃O⁺] raised to the *n-j* power. Thus, for diprotic acids such as carbonic, succinic, tartaric, and so on,

$$[H_2A] = \frac{[H_3O^+]^2 C_a}{[H_3O^+]^2 + K_1[H_3O^+] + K_1K_2} \quad (43)$$

$$[HA^-] = \frac{K_1[H_3O^+] C_a}{[H_3O^+]^2 + K_1[H_3O^+] + K_1K_2} \quad (44)$$

$$[A^{2-}] = \frac{K_1K_2[H_3O^+] C_a}{[H_3O^+]^2 + K_1[H_3O^+] + K_1K_2} \quad (45)$$

See Box 15-6.

Box 15-6

Calculate the concentrations of all succinic acid species in a 1.0 × 10⁻³ M solution of succinic acid at pH 6. Assume that K₁ = 6.4 × 10⁻⁵ and K₂ = 2.3 × 10⁻⁶.

As all the equations have the same denominator, *D*, it can be calculated as

$$\begin{aligned} D &= [H_3O^+]^2 + K_1[H_3O^+] + K_1K_2 \\ &= 1.0 \times 10^{-12} + 6.4 \times 10^{-5} \times 1.0 \times 10^{-6} + 6.4 \times 10^{-5} \times 2.3 \times 10^{-6} \\ &= 1.0 \times 10^{-12} + 6.4 \times 10^{-11} + 14.7 \times 10^{-11} \\ &= 21.2 \times 10^{-11} \end{aligned}$$

Therefore,

$$\begin{aligned} [H_2A] &= \frac{[H_3O^+]^2 C_a}{D} \\ &= \frac{1.0 \times 10^{-12} \times 1.0 \times 10^{-3}}{21.2 \times 10^{-11}} = 4.7 \times 10^{-6} M \end{aligned}$$

$$\begin{aligned} [HA^-] &= \frac{K_1[H_3O^+] C_a}{D} \\ &= \frac{6.4 \times 10^{-5} \times 1.0 \times 10^{-3}}{21.2 \times 10^{-11}} = 3.0 \times 10^{-4} M \end{aligned}$$

$$\begin{aligned} [A^{2-}] &= \frac{K_1K_2 C_a}{D} \\ &= \frac{14.7 \times 10^{-11} \times 1.0 \times 10^{-3}}{21.2 \times 10^{-11}} = 6.9 \times 10^{-4} M \end{aligned}$$

PROTON-BALANCE EQUATION

In the Brønsted–Lowry system, the total number of protons released by the acidic species must equal the total number of protons consumed by basic species. This results in a very useful relationship known as the proton-balance equation (PBE), in which the sum of the concentration terms for species that form by proton consumption is equated to the sum of the concentration terms for species that are formed by the release of protons. The PBE forms the basis of a unified approach to pH calculations, as it is an exact accounting of all proton transfers occurring in solution.

When HCl is added to water, for example, it dissociates yielding one Cl⁻ for each proton released. Thus, Cl⁻ is a species formed by the release of a proton. In the same solution, and actually in all aqueous solutions



where H₃O⁺ is formed by proton consumption and OH⁻ is formed by proton release. Thus, the PBE is

$$[\text{H}_3\text{O}^+] = [\text{OH}^-] + [\text{Cl}^-] \quad (46)$$

In general, the PBE can be formed in the following manner:

1. Start with the species added to water.
2. Place all species that can form when protons are released on the right side of the equation.
3. Place all species that can form when protons are consumed on the left side of the equation.
4. Multiply the concentration of each species by the number of protons gained or lost to form that species.
5. Add $[\text{H}_3\text{O}^+]$ the left side of the equation and $[\text{OH}^-]$ to the right side of the equation. These result from the interaction of two molecules of water as shown above.

See Box 15-7.

Box 15-7

1. When H_3PO_4 is added to water, the species H_2PO_4^- forms with the release of one proton; HPO_4^{2-} forms with the release of two protons; and PO_4^{3-} forms with the release of three protons, which gives the following PBE:

$$[\text{H}_3\text{O}^+] = [\text{OH}^-] + [\text{H}_2\text{PO}_4^-] + 2[\text{HPO}_4^{2-}] + 3[\text{PO}_4^{3-}]$$

2. When Na_2HPO_4 is added to water, it dissociates into two Na^+ and one HPO_4^{2-} . The sodium ion is neglected in the PBE because it is not formed from the release or consumption of protons. The species HPO_4^{2-} , however, may react with water to give H_2PO_4^- with the consumption of one proton, H_3PO_4 with the consumption of two protons, and PO_4^{3-} with the release of one proton to give the following PBE:

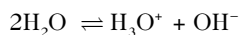
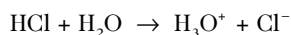
$$[\text{H}_3\text{O}^+] + [\text{H}_2\text{PO}_4^-] + 2[\text{H}_3\text{PO}_4] = [\text{OH}^-] + [\text{PO}_4^{3-}]$$

CALCULATION OF PH OF SOLUTIONS: ACIDS, BASES, AND SALTS

Calculation of pH for strong acids

For our purposes, a strong acid is one which is completely dissociated in water (e.g., HCl or H_2SO_4). When a strong acid such as HCl is added to water, the following reactions occur:

(note this reaction will proceed to completion)



The PBE for this system would be

$$[\text{H}_3\text{O}^+] = [\text{OH}^-] + [\text{Cl}^-] \quad (47)$$

As fully ionised, $[\text{Cl}^-] = C_a$ (the stoichiometric concentration of the acid)

$$\text{Thus } [\text{H}_3\text{O}^+] = [\text{OH}^-] + C_a \quad (48)$$

From previous equations, we know

$$K_w = [\text{H}_3\text{O}^+][\text{OH}^-] \text{ so } [\text{OH}^-] = \frac{K_w}{[\text{H}_3\text{O}^+]} \quad (49)$$

Substituting into Equation 48

$$[\text{H}_3\text{O}^+] = \frac{K_w}{[\text{H}_3\text{O}^+]} + C_a \quad (50)$$

Multiplication by $[\text{H}_3\text{O}^+]$ to remove the denominator gives

$$[\text{H}_3\text{O}^+]^2 - C_a[\text{H}_3\text{O}^+] - K_w = 0 \quad (51)$$

This is a quadratic equation* that yields the solution (because $[\text{H}_3\text{O}^+]$ can never be negative)

$$[\text{H}_3\text{O}^+] = \frac{C_a + \sqrt{C_a^2 + 4K_w}}{2} \quad (52)$$

*The solution to a quadratic equation is

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

In most instances, the concentration of the acid (C_a) will be greater than $1 \times 10^{-6}\text{M}$ which is much greater than $4K_w$ (4×10^{-14}). Thus equation 52 simplifies to $[\text{H}_3\text{O}^+] = C_a$ and pH is therefore equal to the negative of the logarithm of C_a .

Calculate the pH of 0.1 M HCl

$$0.1\text{M HCl} = 0.1\text{M H}_3\text{O}^+ = 10^{-1}\text{M pH} = 1$$

Similarly for $1 \times 10^{-6}\text{M HCl}$, the pH can be calculated to be 6. For dilute solutions of weak acids, this simplification is not valid and equation 52 must be used. See Box 15-8.

Box 15-8

Calculate the pH of $1 \times 10^{-8}\text{M HCl}$ (pH cannot equal 8)

$$[\text{H}_3\text{O}^+] = \frac{1 \times 10^{-8} + \sqrt{(1 \times 10^{-8})^2 + 4 \times 10^{-14}}}{2}$$

$$[\text{H}_3\text{O}^+] = \frac{1 \times 10^{-8} + \sqrt{4.01 \times 10^{-14}}}{2}$$

$$[\text{H}_3\text{O}^+] = \frac{2.1 \times 10^{-7}}{2}$$

$$[\text{H}_3\text{O}^+] = 1.05 \times 10^{-7}$$

$$\text{pH} = 6.98$$

In a similar manner a relationship can be derived for strong bases

$$[\text{OH}^-] = \frac{C_b + \sqrt{C_b^2 + 4K_w}}{2} \quad (53)$$

where C_b = concentration of base. For non-dilute solutions of a base such as NaOH (where $C_b > 1 \times 10^{-6}$), then

$$[\text{OH}^-] = [\text{Na}^+] = C_b$$

and the pH of a 0.005M solution of NaOH would be

$$\text{pOH} = -\log 0.005 = 2.30$$

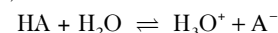
$$\text{pH} = \text{p}K_w - \text{pOH}$$

$$\text{pH} = 14 - 2.30 = 11.70$$

pH of solutions of weak acids and bases

SOLUTIONS CONTAINING WEAK ACID ONLY

If a weak acid, HA , is added to water, it will equilibrate with its conjugate base, A^- , as



Accounting for the ionization of water gives the following PBE for this system:

$$[\text{H}_3\text{O}^+] = [\text{OH}^-] + [\text{A}^-] \quad (54)$$

The concentration of A^- as a function of hydronium ion concentration can be obtained as shown previously to give

$$[\text{H}_3\text{O}^+] = [\text{OH}^-] + \frac{K_a C_a}{[\text{H}_3\text{O}^+] + K_a} \quad (55)$$

Algebraic simplification yields

$$[\text{H}_3\text{O}^+] = K_a \frac{(C_a - [\text{H}_3\text{O}^+] + [\text{OH}^-])}{([\text{H}_3\text{O}^+] - [\text{OH}^-])} \quad (56)$$

In most instances for solutions of weak acids, $[\text{H}_3\text{O}^+] \gg [\text{OH}^-]$, and the equation simplifies to give

$$[H_3O^+]^2 + K_a[H_3O^+] - K_a C_a = 0$$

This is a quadratic equation as before that yields

$$[H_3O^+] = \frac{-K_a + \sqrt{K_a^2 + 4K_a C_a}}{2} \quad (57)$$

because $[H_3O^+]$ can never be negative. Furthermore, if $[H_3O^+]$ is less than 5% of C_a , Equation 57 is simplified further to give

$$[H_3O^+] = \sqrt{K_a C_a} \quad (58)$$

It generally is preferable to use the simplest equation to calculate $[H_3O^+]$. However, when $[H_3O^+]$ is calculated, it must be compared to C_a in order to determine whether the assumption $C_a \gg [H_3O^+]$ is valid. If the assumption is not valid, the quadratic equation should be used. See Box 15-9.

Box 15-9

Calculate the pH of a $5.00 \times 10^{-5} M$ solution of a weak acid having a $K_a = 1.90 \times 10^{-5}$.

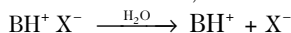
$$\begin{aligned} [H_2O^+] &= \sqrt{K_a C_a} \\ &= \sqrt{1.90 \times 10^{-5} \times 5.00 \times 10^{-5}} \\ &= 3.08 \times 10^{-5} M \end{aligned}$$

As C_a ($5.00 \times 10^{-5} M$) is not much greater than $[H_3O^+]$, the quadratic equation (Equation 57) should be used.

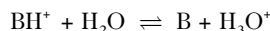
$$\begin{aligned} [H_2O^+] &= \frac{-1.90 \times 10^{-5} + \sqrt{(1.90 \times 10^{-5})^2 + (5.00 \times 10^{-5})}}{2} \\ &= 7.06 \times 10^{-3} \\ \text{pH} &= -\log(7.06 \times 10^{-3}) = 2.15 \end{aligned}$$

Note that the assumption $[H_3O^+] \gg [OH^-]$ is valid. The hydronium ion concentration calculated from Equation 58 has a relative error of about 100% when compared to the correct value obtained from Equation 57.

When a salt obtained from a strong acid and a weak base—such as ammonium chloride, morphine sulfate, or pilocarpine hydrochloride—is dissolved in water, it dissociates as



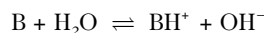
in which BH^+ is the protonated form of the base B, and X^- is the anion of a strong acid. Because X^- is the anion of a strong acid, it is too weak a base to undergo any further reaction with water. The protonated base, however, can act as a weak acid to give



Thus, Equations 57 and 58 are valid, with C_a being equal to the concentration of the salt in solution. If K_a for the protonated base is not available, it can be obtained by dividing K_b for the base B, into K_w . See Box 15-10.

SOLUTIONS CONTAINING WEAK BASE ONLY

When a weak base, B, is dissolved in water it ionizes to give the conjugate acid as



The PBE for this system is

$$[BH^+] + [H_3O^+] = [OH^-] \quad (59)$$

Substituting $[BH^+]$ as a function of hydronium ion concentration and simplifying, in the same manner as shown for a weak acid, gives

$$[OH^-] = K_b \frac{(C_b - [OH^-] + [H_3O^+])}{([OH^-] - [H_3O^+])} \quad (60)$$

If $[OH^-] \gg [H_3O^+]$, as is true generally, then

$$[OH^-]^2 + K_b[OH^-] + K_b C_b = 0 \quad (61)$$

Box 15-10

Calculate the pH of a $0.026 M$ solution of ammonium chloride. Assume that K_b for ammonia is 1.74×10^{-5} and K_w is 1.00×10^{-14} .

$$\begin{aligned} K_a &= \frac{K_w}{K_b} = \frac{1.00 \times 10^{-14}}{1.74 \times 10^{-5}} = 5.75 \times 10^{-10} \\ [H_3O^+] &= \sqrt{K_a C_a} \\ &= \sqrt{5.75 \times 10^{-10} \times 2.6 \times 10^{-2}} \\ &= 3.87 \times 10^{-6} M \\ \text{pH} &= -\log(3.87 \times 10^{-6}) = 5.41 \end{aligned}$$

As C_a is much greater than $[H_3O^+]$ and $[H_3O^+]$ is much greater than $[OH^-]$, the assumptions are valid and the value calculated for pH is sufficiently accurate.

which is a quadratic with the following solution:

$$[OH^-] = \frac{-K_b + \sqrt{K_b^2 + 4K_b C_b}}{2} \quad (62)$$

If $C_b \gg [OH^-]$, the quadratic equation simplifies to

$$[OH^-] = \sqrt{K_b C_b} \quad (63)$$

or the more useable

$$[H_3O^+] = \sqrt{\frac{K_a K_w}{C_b}} \quad (64)$$

See Box 15-11.

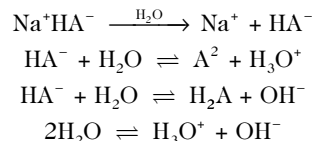
Box 15-11

Calculate the pH of a $0.05 M$ solution of sodium acetate. Assume that K_a for acetic acid $= 1.75 \times 10^{-5}$ and $K_w = 1 \times 10^{-14}$.

$$\begin{aligned} [H_3O^+] &= \sqrt{\frac{1.75 \times 10^{-5} \times 1.00 \times 10^{-14}}{0.05}} \\ [H_3O^+] &= \sqrt{3.5 \times 10^{-18}} \\ [H_3O^+] &= 1.87 \times 10^{-9} \\ \text{pH} &= 8.73 \end{aligned}$$

SOLUTIONS CONTAINING AMPHOLYTES

Substances such as NaHCO_3 and NaH_2PO_4 are termed **ampholytes** and are capable of functioning both as acids and bases. When an ampholyte of the type NaHA is dissolved in water, the following series of reactions can occur:



The total PBE for the system is

$$[H_3O^+] + [H_2A] = [OH^-] + [A^{2-}] \quad (65)$$

Substituting both $[H_2A]$ and $[A^{2-}]$ as a function of $[H_3O^+]$ yields

$$\begin{aligned} [H_3O^+] &= \frac{[H_3O^+]^2 C_s}{[H_3O^+]^2 + K_1 [H_3O^+] + K_1 K_2} \\ &= \frac{K_u}{[H_3O^+]} + \frac{K_1 K_2 C_s}{[H_3O^+]^2 + K_1 [H_3O^+] + K_1 K_2} \end{aligned} \quad (66)$$

where C_s is the concentration of salt. This gives a fourth-order equation in $[H_3O^+]$, which can be simplified using certain judicious assumptions to

$$[H_3O^+] = \sqrt{\frac{K_1 K_2 C_s}{K_1 + C_s}} \quad (67)$$

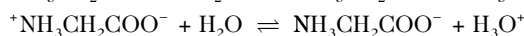
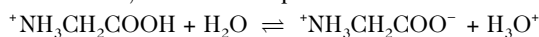
In most instances, $C_s \gg K_1$, and the equation further simplifies to

$$[H_3O^+] = \sqrt{K_1 K_2} \quad (68)$$

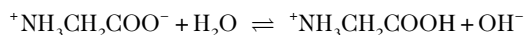
and $[H_3O^+]$ becomes independent of the concentration of the salt. A special property of ampholytes is that the concentration of the species HA^- is maximal at the pH corresponding to Equation 68.

Zwitterions

When the simplest amino acid salt, glycine hydrochloride, is dissolved in water, it acts as a diprotic acid and ionizes as



The form, $^+NH_3CH_2COO^-$, is an ampholyte because it also can act as a weak base:



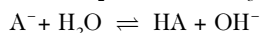
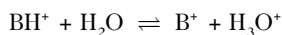
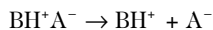
Glycine when titrated in water shows two pK_a values. One at 2.4 corresponding to the proton gained and one at 9.8 for the proton loss.

The term **zwitterion** was introduced on 1897 to describe this type of “internal salt”, which carries both a charged acidic and a charged basic moiety on the same molecule. Because the two charges balance each other, the molecule acts essentially as a neutral molecule. The pH at which the zwitterion concentration is maximal is known as the **isoelectric point**, which can be calculated from Equation 68. For a simple amino acid, such as glycine, the pI is simply the average of the two pK_a values. For more complex amino acids, such as glutamic acid or arginine which have ionisable groups in the side-chains, the pI is given by averaging the two pK_a values that lie on either side of the zwitterion.

On the acid side of the isoelectric point, amino acids and proteins are cationic and incompatible with anionic materials such as the naturally occurring gums used as suspending and/or emulsifying agents. On the alkaline side of the isoelectric point, amino acids and proteins are anionic and incompatible with cationic materials such as benzalkonium chloride. However, not all amphoteric substances having both acidic and basic groups are zwitterions. *m*-Aminophenol has 2 pK_a values, 4.4 and 9.8 and although it is an ampholyte, it does not exist as a zwitterion. A substance is often suspected to be a zwitterion if it is more soluble in water, less soluble in organic solvents, has a higher melting point and related substances with one ionizable group.

Solutions containing salts of weak acids and weak bases

When a salt such as ammonium acetate (which is derived from a weak acid and a weak base) is dissolved in water, it undergoes the following reactions:



The total PBE for this system is

$$[HA] + [H_3O^+] = [OH^-] + [B] \quad (69)$$

Replacing $[HA]$ and $[B]$ as a function of $[H_3O^+]$, gives

$$[H_3O^+] + \frac{[H_3O^+]C_s}{[H_3O^+] + K_a} = [OH^-] + \frac{K'_a C_s}{[H_3O^+] + K'_a} \quad (70)$$

in which C_s is the concentration of salt, K_a is the ionization constant of the conjugate acid formed from the reaction between A^- and water, and K'_a is the ionization constant for the

protonated base, BH^+ . In general, $[H_3O^+]$, $[OH^-]$, K_a , and K'_a usually are smaller than C_s and the equation simplifies to

$$[H_3O^+] = \sqrt{K_a K'_a} \quad (71)$$

See Box 15-12.

Box 15-12

Calculate the pH of a 0.01 M solution of ammonium acetate. The ammonium ion has a K_a equal to 5.75×10^{-10} , which represents K'_a in Equation 71. Acetic acid has a K_a of 1.75×10^{-5} , which represents K_a in Equation 71:

$$[H_3O^+] = \sqrt{1.75 \times 10^{-5} \times 5.75 \times 10^{-10}}$$

$$[H_3O^+] = 1 \times 10^{-7}$$

$$pH = 7$$

All of the assumptions are valid.

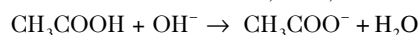
BUFFERS

The terms **buffer**, **buffer solution**, and **buffered solution**, when used with reference to hydrogen-ion concentration or pH, refer to the ability of a system, particularly an aqueous solution, to resist a change of pH on adding acid or alkali, or on dilution with a solvent.

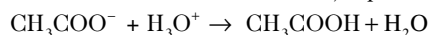
If an acid or base is added to water, the pH of the latter is changed markedly, for water has no ability to resist change of pH; it is completely devoid of buffer action. Even a very weak acid such as carbon dioxide changes the pH of water, decreasing it from 7 to 5.7 when the small concentration of carbon dioxide present in air is equilibrated with pure water. This extreme susceptibility of distilled water to a change of pH upon adding very small amounts of acid or base is often of great concern in pharmaceutical operations where the majority of drugs are either weak acids or weak bases. Solutions of neutral salts, such as sodium chloride, similarly lack ability to resist change of pH on adding acid or base; such solutions are called **unbuffered**.

When partially neutralized weak acids or bases are present in aqueous solution, the addition of small amount of strong acid or strong base causes little or no detectable change in pH. This resistance was described in 1900 by Fernbach and Hubert as buffering. In Brønsted-Lowry terms, the common feature of these solutions is the presence of a weak acid and a salt of the weak acid, or a weak base and a salt of the weak base. An example of the former system is acetic acid and sodium acetate; and of the latter, ammonium hydroxide and ammonium chloride. From the proton concept of acids and bases discussed earlier, it is apparent that such buffer action involves a conjugate acid-base pair in the solution. It will be recalled that acetate ion is the conjugate base of acetic acid, and that ammonium ion is the conjugate acid of ammonia (the principal constituent of what commonly is called ammonium hydroxide).

The mechanism of action of the acetic acid-sodium acetate buffer pair is that the acid, which exists largely in molecular (nonionized) form, combines with hydroxyl ion that may be added to form acetate ion and water; thus,

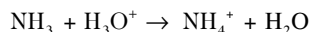


The acetate ion, which is a base, combines with the hydrogen (or more exactly hydronium) ion that may be added to form essentially nonionized acetic acid and water, represented as

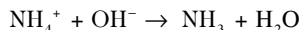


As will be illustrated later by an example, the change of pH is slight as long as the amount of hydronium or hydroxyl ion added does not exceed the capacity of the buffer system to neutralize it.

The ammonia-ammonium chloride pair functions as a buffer because the ammonia combines with hydronium ion that may be added to form ammonium ion and water; thus,

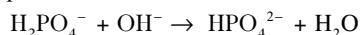


Ammonium ion, which is an acid, combines with added hydroxyl ion to form ammonia and water, as

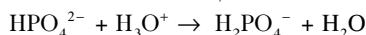


Again, the change of pH is slight if the amount of added hydronium or hydroxyl ion is not in excess of the capacity of the system to neutralize it.

Besides these two general types of buffers, a third appears to exist. This is the buffer system composed of two salts, as monobasic potassium phosphate, KH_2PO_4 , and dibasic potassium phosphate, K_2HPO_4 . This is not, however, a new type of buffer; it is actually a weak-acid/conjugate-base buffer in which an ion, H_2PO_4^- , serves as the weak acid, and HPO_4^{2-} is its conjugate base. When hydroxyl ion is added to this buffer the following reaction takes place:



and when hydronium ion is added,



It is apparent that the mechanism of action of this type of buffer is essentially the same as that of the weak-acid/conjugate-base buffer composed of acetic acid and sodium acetate.

A buffer system composed of a conjugate acid-base pair, NaA-HA (such as sodium acetate and acetic acid), would have a PBE of

$$[\text{H}_3\text{O}^+] + [\text{HA}] = [\text{OH}^-] + [\text{A}^-] \quad (72)$$

Replacing $[\text{HA}]$ and $[\text{A}^-]$ as a function of hydronium ion concentration gives

$$[\text{H}_3\text{O}^+] + \frac{[\text{H}_3\text{O}^+]\text{C}_b}{[\text{H}_3\text{O}^+] + K_a} = [\text{OH}^-] + \frac{K_a\text{C}_a}{[\text{H}_3\text{O}^+] + K_a} \quad (73)$$

where C_b is the concentration of the salt, NaA , and C_a is the concentration of the weak acid, HA . This equation can be rearranged to give

$$[\text{H}_3\text{O}^+] + K_a \left(\frac{\text{C}_a - [\text{H}_3\text{O}^+] + [\text{OH}^-]}{\text{C}_b + [\text{H}_3\text{O}^+] - [\text{OH}^-]} \right) \quad (74)$$

Equation 74 is an exact equation and can be used to determine the pH of a weak acid, a weak base or a conjugate pair. An activity correction can be applied to this equation (usually when concentrations exceed 0.01M).

$$[\text{H}_3\text{O}^+] + K_a \left(\frac{\text{C}_a - [\text{H}_3\text{O}^+] + [\text{OH}^-]}{\text{C}_b + [\text{H}_3\text{O}^+] - [\text{OH}^-]} \right) \cdot \left(\frac{\gamma_{\text{HA}}}{\gamma_{\text{H}_3\text{O}^+} + \gamma_{\text{A}^-}} \right)$$

Assumptions and simplifications

If only an acid is present, then $b = 0$ and $[\text{H}_3\text{O}^+] \gg [\text{OH}^-]$, the equations simplify to

$$[\text{H}_3\text{O}^+]^2 = K_a[\text{H}_3\text{O}^+] - K_a\text{C}_a = 0 \quad (75)$$

This obviously solves to the same approximation as before, that is, Equation 57. Similarly, if only a base is present, then $a = 0$ and $[\text{OH}^-] \gg [\text{H}_3\text{O}^+]$, the equation simplifies to

$$[\text{H}_3\text{O}^+] = \frac{K_a[\text{OH}^-]}{\text{C}_b + [\text{OH}^-]} = \frac{K_a K_w}{[\text{H}_3\text{O}^+]\text{C}_b - K_w} \quad (76)$$

The equation can be solved for either $[\text{H}_3\text{O}^+]$ or $[\text{OH}^-]$ but it is most useful on terms of $[\text{H}_3\text{O}^+]$, giving the quadratic equation

$$\text{C}_b[\text{H}_3\text{O}^+]^2 - K_w[\text{H}_3\text{O}^+] - K_a K_w = 0 \quad (77)$$

This solves to

$$[\text{H}_3\text{O}^+] = K_w + \sqrt{\frac{K_w^2 + 4\text{C}_b K_a K_w}{2\text{C}_b}} \quad (78)$$

Application of judicious assumptions leads to the equation 79, which has previously been described for calculating pH of a solution of a weak base

$$[\text{H}_3\text{O}^+] = \sqrt{\frac{K_a K_w}{\text{C}_b}} \quad (79)$$

If a single conjugate pair is present, then in general, both C_a and C_b are much greater than $[\text{H}_3\text{O}^+]$, which is in turn much greater than $[\text{OH}^-]$ and the equation simplifies to

$$[\text{H}_3\text{O}^+] = \frac{K_a \text{C}_a}{\text{C}_b} \quad (80)$$

Expressed in terms of pH, this becomes

$$\text{pH} = \text{pH}_a + \log \frac{\text{C}_b}{\text{C}_a} \quad (81)$$

This equation generally is called the **Henderson-Hasselbalch equation** or the **Buffer equation**. It applies to all buffer systems formed from a single conjugate acid-base pair, regardless of the nature of the salts. For example, it applies equally well to the following buffer systems: ammonia-ammonium chloride, monosodium phosphate-disodium phosphate, and phenobarbital-sodium phenobarbital. In the ammonia-ammonium chloride system, ammonia is obviously the base and the ammonium ion is the acid (C_a equal to the concentration of the salt). In the phosphate system, monosodium phosphate is the acid and disodium phosphate is the base. For the phenobarbital buffer system, phenobarbital is the acid and the phenobarbital anion is the base (C_b equal to the concentration of sodium phenobarbital). See Box 15-13.

Box 15-13

As an example of the application of this equation, the pH of a buffer solution containing acetic acid and sodium acetate, each in 0.1 M concentration, may be calculated. The K_a of acetic acid, as defined above, is 1.8×10^{-5} , at 25°C.

What is the pH of an acetate buffer containing 0.1M acetic acid and 0.05M sodium acetate ($\text{pK}_a = 4.76$)?

$$\text{pH} = 4.76 + \log \left(\frac{0.05}{0.1} \right) = 4.76 - 0.3 = 4.46$$

Calculate the pH of 0.1M acetic acid

$$\begin{aligned} [\text{H}_3\text{O}^+] &= \sqrt{K_a \text{C}_a} \\ [\text{H}_3\text{O}^+] &= \sqrt{1.74 \times 10^{-5} \times 0.1} \\ [\text{H}_3\text{O}^+] &= 1.32 \times 10^{-3} \end{aligned}$$

pH = 2.88

Add sodium acetate to make a solution of 0.1M with respect to the salt. What is the pH of the resultant solution?

$$\text{pH} = 4.76 + \log \left(\frac{0.1}{0.1} \right) = 4.76$$

The Henderson Hasselbalch equation predicts that any solutions containing the same molar concentration of acetic acid as of sodium acetate will have the same pH. Thus, a solution of 0.01 M concentration of each will have the same pH, i.e., 4.74, as one of 0.1 M concentration of each component. Actually, there will be some difference in the pH of the solutions, for the **activity coefficient** of the components varies with concentration. For most practical purposes, however, the approximate values of pH calculated by the equation are satisfactory. It should be pointed out that the buffer of higher concentration of each component will have a much greater capacity for neutralizing added acid or base and this point will be discussed further in the discussion of buffer capacity.

An acid exists in water in equilibrium between its ionised and molecular form. If a strong base such as sodium hydroxide is added, the hydroxyl ions combine with the hydrogen ions, disrupting the equilibrium. This increases the ratio of ionised to unionised species. Thus, if one mole equivalent of sodium hydroxide is added, the acid, HA, will be quantitatively converted to its anion (A^-) and the pH will increase. It should be noted that when the concentrations of acid and conjugate base are equal, then the $pH = pK_a$. When looking at the titration curve of a weak acid with a strong base (Figure 15-3), this can be used to estimate the pK_a , that is, at 0.5 the pH equals the pK_a . This is because if $pH = pK_a$ then the log term in the Henderson-Hasselbalch equation is 1 and [unionised] = [ionised].

Derivations of the Henderson-Hasselbalch equation are used to quantify the **percentage ionised** (or α , **fraction ionised**). They can, however easily be derived from first principles, for the example HA below.

$$\alpha = \frac{[A^-]}{[HA] + [A^-]} = \frac{C_b}{C_a + C_b} \quad (82)$$

Know

$$[H_3O^+] = K_a \frac{C_a}{C_b} \quad (83)$$

Therefore

$$C_a = \frac{C_b [H_3O^+]}{K_a} \quad (84)$$

$$C_a + C_b = C_b + \frac{C_b [H_3O^+]}{K_a} \quad (85)$$

$$C_a + C_b = C_b \left(1 + \frac{[H_3O^+]}{K_a} \right) \quad (86)$$

$$C_a + C_b = C_b \left(\frac{K_a + [H_3O^+]}{K_a} \right) \quad (87)$$

$$\alpha = \frac{C_b}{C_a + C_b} = \left(\frac{K_a}{K_a + [H_3O^+]} \right) \quad (88)$$

When $C_a = C_b$, $pH = pK_a$, either can be substituted into equations 88 and

$$\alpha = \frac{K_a}{K_a + K_a} = \frac{1}{2} = 50\% \quad (89)$$

A similar relationship can be derived for a weak base

$$\alpha = \frac{[H_3O^+]}{K_a + [H_3O^+]} \quad (90)$$

If the pH of the solution is increased to one unit above the pK_a of the acid (or one unit below the pK_a of the conjugate acid), then the percentage of the compound ionised increases to approximately 90% (90.9%). If the pH increases to two units above the pK_a (or two units below for a base), the percentage ionised

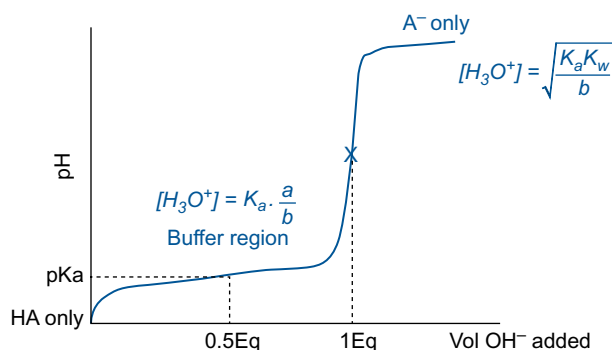


Figure 15-3. Titration curve for a weak acid with a strong base.

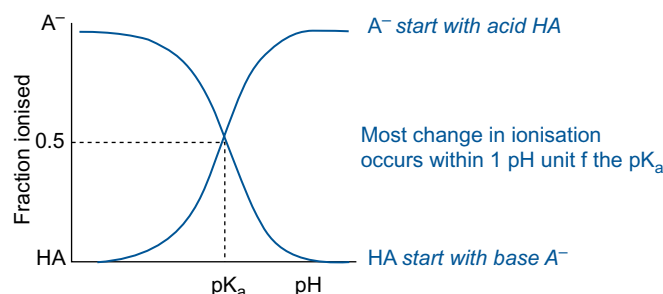


Figure 15-4. Variation in fraction ionised with pH.

increases to 99%, because both pH and pK_a are logarithmic relationships, and so on to 99.9%, 99.99% etc. This approximation is useful when estimating the fraction ionised (α) in Figure 15-4.

For weak acids:

$pH = pK_a$ compound is approximately 50% ionised

$pH = pK_a - 1$ compound is approximately 90% ionised

$pH = pK_a - 2$ compound is approximately 99% ionised

$pH = pK_a - 3$ compound is approximately 99.9% ionised

For weak bases:

$pH = pK_a$ compound is approximately 50% ionised

$pH = pK_a + 1$ compound is approximately 90% ionised

$pH = pK_a + 2$ compound is approximately 99% ionised

$pH = pK_a + 3$ compound is approximately 99.9% ionised

The Henderson Hasselbalch equation is useful also for calculating the ratio of molar concentrations of a buffer system required to produce a solution of specific pH. As an example, suppose that an acetic acid-sodium acetate buffer of pH 4.5 must be prepared. What ratio of the buffer components should be used?

Solution

Rearranging Equation 81, which is used to calculate the pH of weak acid-salt type buffers, gives

$$pH - pK_a = \log \frac{C_b}{C_a}$$

$$= 4.5 - 4.76 = -0.24$$

Antilog of -0.24 is 0.575

The interpretation of this result is that the **proportion** of sodium acetate to acetic acid should be 0.575 mol of the former to 1 mol of the latter to produce a pH of 4.5. A solution containing 0.0575 mol of sodium acetate and 0.1 mol of acetic acid per liter would meet this requirement, as would also one containing 0.00575 mol of sodium acetate and 0.01 mol of acetic acid per liter. The actual concentration selected would depend chiefly on the desired buffer capacity.

Buffers solutions are frequently found in pharmaceutical formulations, for example aqueous eye drops and oral solutions. Many different buffer systems are available, such as phosphate buffer saline and Sørensen's buffers. Sørensen's buffer comprises a mixture of the salts of sodium phosphate. For solutions intended for parenteral administration or application to mucous membranes, discomfort (and even injury), is minimized by adjusting the solution to approximate isotonicity with body fluid (equivalent to 0.90% sodium chloride at 37°C). Pharmacopeiae and reference sources contain details of buffers, and their ionic strengths, that can be used to control pH.

The body contains many buffer systems, which function to control the pH of body compartments and fluids. For example, there are three main buffer systems which function to maintain plasma pH at 7.4. Dissolved carbon dioxide, giving rise to carbonic acid (H_2CO_3) in solution, and its sodium salt (usually sodium bicarbonate, $NaHCO_3$) is responsible for most of the buffering capacity. The other two buffers are dihydrogenphosphate ($H_2PO_4^-$), also with its sodium salt, and protein macromolecules.

Buffer capacity

When determining which buffer form is suitable, buffer capacity should be maximised by selection of a buffer with a pK_a close to the pH of interest. **Buffer capacity** is the ability of a buffer solution to resist changes in pH upon addition of acid or alkali.

The concentration of acid in a weak-acid/conjugate-base buffer determines the capacity to “neutralize” added base, whereas the concentration of salt of the weak acid determines the capacity to neutralize added acid. Similarly, in a weak-base/conjugate-acid buffer the concentration of the weak base establishes the buffer capacity toward added acid, whereas the concentration of the conjugate acid of the weak base determines the capacity toward added base. When the buffer is equimolar in the concentrations of weak acid and conjugate base, or of weak base and conjugate acid, it has equal buffer capacity toward added strong acid or strong base.

Van Slyke, the biochemist, introduced a quantitative expression for evaluating buffer capacity. This may be defined as the amount, in gram-equivalents (geq) per liter, of strong acid or strong base required to be added to a solution to change its pH by 1 unit; a solution has a buffer capacity of 1 when 1 L requires 1 geq of strong base or acid to change the pH 1 unit. (In practice, considerably smaller increments are measured, expressed as the ratio of acid or base added to the change of pH produced.) From this definition it is apparent that the smaller the pH change in a solution caused by the addition of a specified quantity of acid or alkali, the greater the buffer capacity of the solution. See Box 15-14.

As is in part evident from the example in Box 14, and may be further evidenced by calculations of pH changes in other systems, the degree of buffer action and, therefore, the buffer capacity, depend on the kind and concentration of the buffer components, the pH region involved and the kind of acid or alkali added.

Buffer capacity can be calculated more exactly using the equation developed by Koppel and Spiro

$$\beta = 2.303C \cdot \frac{K_a [H_3O^+]}{(K_a + [H_3O^+])^2} \quad (91)$$

where C is the total buffer concentration, that is, the sum of the molar concentrations $C_a + C_b$. The maximum buffer capacity exists when there are equimolar concentrations of acid and conjugate base, that is $pH = pK_a$. On substitution of K_a for $[H_3O^+]$, or vice versa, into the equation above, a relationship can be derived for maximal buffer capacity,

$$\begin{aligned} \beta &= 2.303C \cdot \frac{K_a^2}{(2K_a)^2} & \beta &= 2.303C \cdot \frac{[H_3O^+]^2}{(2[H_3O^+])^2} \\ \beta &= \frac{2.303C}{4} & (92) \\ \beta_{MAX} &= 0.576C \end{aligned}$$

THE DETERMINATION OF IONIZATION CONSTANTS

There are a variety of ways to determine the ionization constants including conductivity measurements and partition coefficients, but the most convenient method to determine pK_a is by **potentiometric titration** and is normally the method of choice if applicable.

It involves the measurement of the pH during stepwise titration of a known weight of substance with accurately standardised base or acid. In this method the **mole-ratio** of the acid-base conjugate pairs is calculated from the titrant added. The method involves the universal equation

$$[H_3O^+] = K_a \left(\frac{C_a - [H_3O^+] + [OH^-]}{C_b + [H_3O^+] - [OH^-]} \right) \quad (93)$$

In **acidic** conditions $pH < 4$, $[H_3O^+] \gg [OH^-]$ and the $[OH^-]$ term can be ignored. Similarly, in **basic** conditions $pH > 10$, $[OH^-] \gg [H_3O^+]$; ignore $[H_3O^+]$ term.

Box 15-14

These examples illustrate certain basic principles and calculations concerning buffer action and buffer capacity.

1. What is the change of pH on adding 0.01 mol of NaOH to 1 L of 0.10 M acetic acid?

- i. Calculate the pH of a 0.10 molar solution of acetic acid:

$$\begin{aligned} [H_3O^+] &= \sqrt{K_a C_a} = \sqrt{1.75 \times 10^{-4} \times 1.0 \times 10^{-1}} = 4.18 \times 10^{-3} \\ pH &= -\log 4.18 \times 10^{-3} = 2.38 \end{aligned}$$

- ii. On adding 0.01 mol of NaOH to a liter of this solution, 0.01 mol of acetic acid is converted to 0.01 mol of sodium acetate, thereby decreasing C_a to 0.09 M, and $C_b = 1.0 \times 10^{-2} M$. Using the Henderson-Hasselbach equation gives

$$pH = 4.76 + \log \frac{0.01}{0.09} = 4.76 - 0.95 = 3.81$$

The pH change is, therefore, 1.43 unit. The buffer capacity as defined above is calculated to be

$$\frac{\text{mols of NaOH added}}{\text{change in pH}} = 0.011$$

2. What is the change of pH on adding 0.1 mol of NaOH to 1 L of buffer solution 0.1 M in acetic acid and 0.1 M in sodium acetate?

- a. The pH of the buffer solution before adding NaOH is

$$\begin{aligned} pH &= \log \frac{[\text{base}]}{[\text{acid}]} + pK_a \\ &= \log \frac{0.1}{0.1} + 4.76 = 4.76 \end{aligned}$$

- b. On adding 0.01 mol of NaOH per liter to this buffer solution, 0.01 mol of acetic acid is converted to 0.01 mol of sodium acetate, thereby decreasing the concentration of acid to 0.09 M and increasing the concentration of base to 0.11 M. The pH is calculated as

$$\begin{aligned} pH &= \log \frac{0.11}{0.09} + 4.76 \\ &= 0.087 + 4.76 = 4.85 \end{aligned}$$

The change of pH in this case is only 0.09 unit, about 1/10 the change in the preceding example. The buffer capacity is calculated as

$$\frac{\text{mols of NaOH added}}{\text{change of pH}} = \frac{0.01}{0.09} = 0.11$$

Thus, the buffer capacity of the acetic acid–sodium acetate buffer solution is approximately 10 times that of the acetic acid solution.

This can be illustrated by the worked example of the determination of the pK_a of boric acid which is a commonly used buffer in ophthalmic solutions. Potentiometric titrations are most accurately carried out in water that has been boiled to remove any dissolved carbon dioxide. Results are commonly portrayed in a table (Table 15-5). In this case, no correction is required for hydronium or hydroxyl ions. The pK_a of boric acid is 9.25 (and can be refined by correcting for $[OH^-]$ at later values).

When the pH values fall outside the range 4–10, it is necessary to include a correction for the contribution of $[H_3O^+]$ and $[OH^-]$ as in the worked example for benzoic acid (Table 15-6) where column 6 is the calculation

$$\frac{a - [H_3O^+]}{b + [H_3O^+]} \quad (94)$$

Table 15-5. pK_a determination of boric acid following potentiometric titration (adapted from Albert & Serjeant²)

Stoichiometric conc						
Vol KOH (ml)	pH	[HA] (C_a)	[A ⁻] (C_b)	C_b/C_a	Log of (C_b/C_a)	pK_a (column 2 +6)
0	6.16	0.01	0			
0.5	8.34	0.009	0.001	9/1	0.95	9.29
1.0	8.68	0.008	0.002	8/2	0.60	9.28
1.5	8.89	0.007	0.003	7/3	0.37	9.26
2.0	9.07	0.006	0.004	6/4	0.18	9.25
2.5	9.26	0.005	0.005	5/5	0	9.26
3.0	9.43	0.004	0.006	4/6	-0.18	9.25
3.5	9.62	0.003	0.007	3/7	-0.37	9.25
4.0	9.84	0.002	0.008	2/8	-0.60	9.24
4.5	10.14	0.001	0.009	1/9	-0.95	9.19
5.0	10.56	0	0.01			

Table 15-6. pK_a determination of benzoic acid following potentiometric titration (adapted from Albert & Serjeant²)

Stoichiometric conc							
Vol KOH (ml)	pH	[HA] (C_a)	[A ⁻] (C_b)	[H ₃ O ⁺]	Column 6	Log of (C_b/C_a)	pK_a (column 2 + 6)
0	3.10	0.01	0	7.94×10^{-4}			
0.5	3.38	0.009	0.001	4.16×10^{-4}	6.062	0.78	4.16
1.0	3.63	0.008	0.002	2.34×10^{-4}	3.476	0.54	4.17
1.5	3.83	0.007	0.003	1.48×10^{-4}	2.177	0.34	4.17
2.0	3.99	0.006	0.004	1.02×10^{-4}	1.438	0.16	4.15
2.5	4.16	0.005	0.005	6.92×10^{-5}	0.973	-0.01	4.15
3.0	4.34	0.004	0.006	4.57×10^{-5}	0.654	-0.18	4.16
3.5	4.53	0.003	0.007	2.95×10^{-5}	0.422	-0.37	4.16
4.0	4.76	0.002	0.008	1.74×10^{-5}	0.247	-0.60	4.16
4.5	5.12	0.001	0.009	7.59×10^{-6}	0.110	-0.95	4.17
5.0	8.18	0	0.01	6.61×10^{-9}			

Errors can arise from impurities in both the titrant and the substance undergoing titration with water being a common impurity if the substance is not thoroughly dried. If this is not possible, or if there is a degree of solvation, then the Z function, as defined by Benet and Goyan,³ can be determined. It is the sum of 4 known concentrations at any part of the titration. For a weak acid (HA) being titrated with, for example KOH, then the Z function would be

$$Z = [A^-] = [K^+] + [H_3O^+] - [OH^-] \quad (95)$$

As C_a is the sum of [HA] and [A⁻] at equilibrium, for an acid, a plot of Z versus $Z/[H_3O^+]$ will yield a straight line with slope equal to the negative reciprocal of the ionization constant and an intercept equal to C_o where

$$Z = C_o - \frac{1}{K_a} Z[H_3O^+] \quad (96)$$

The corresponding relationship for determination of pK_a of a weak base is

$$Z = C_o - K_a \frac{Z}{[H_3O^+]} \quad (97)$$

Thus, in addition to obtaining an accurate value for pK_a , the stoichiometric concentration of the substance being titrated is also obtained.

The dissociation constants for diprotic acids can be obtained by defining P as the average number of protons dissociated per mole of acid, or

$$P = Z / C_a \quad (98)$$

and

$$\frac{[H_3O^+]^2 P}{(2-P)} = K_1 K_2 + \frac{K_1 [H_3O^+] (1-P)}{(2-P)} \quad (99)$$

A plot of Equation 99 should yield a straight line with a slope equal to K_1 and an intercept of $K_1 K_2$. Dividing the intercept by the slope yields K_2 .

The addition of solvents to achieve adequate solubility for determination of pK_a by potentiometric titration is useful to rank a series of related compounds but because changes in ionization result and complications which cannot be easily extrapolated arise, **spectrophotometric methods** are used as an alternative.

Determination of pK_a by UV spectroscopy is also useful when the pK_a is particularly low or high (<2 or >11). It relies on the selection of an **analytical wavelength**, at which the difference between the absorbances of the unionized and the ionized species are maximal. The first step is therefore the determination of pure spectra for the two species involved in the equilibrium and the choice of the suitable wavelength. The absorbances of the ionized species (A_i) and the molecular (unionized species, A_m) are then recorded. Two component mixtures of the species are prepared by adjusting the pH of the buffer used and, assuming Beer-Lambert law is followed, the absorbance of the partially ionized acid or base is solely determined by the ratio of the species).

By rearrangement of the Henderson-Hasselbalch equation, the pK_a for weak acids can be determined as below

$$pK_a = pH + \log \left(\frac{A_i - A_t}{A_t - A_m} \right) \quad (100)$$

Table 15-7. pK_a determination of acridine by UV spectroscopy (adapted from Albert & Serjeant²)

pH	At	$A_t - A_i$	$A_t - A_m$	Log column(4/3)	pK_a (column 1+ 5)
6.30	0.125	0.483	0.100	-0.68	5.62
6.10	0.170	0.438	0.145	-0.48	5.62
5.89	0.235	0.373	0.210	-0.25	5.64
5.68	0.299	0.309	0.274	-0.05	5.63
5.47	0.367	0.241	0.342	0.15	5.62
5.27	0.429	0.179	0.404	0.36	5.63
5.08	0.474	0.134	0.449	0.53	5.61
4.85	0.523	0.085	0.498	0.77	5.62

where A_t is the absorbance of the test mixture. Once an approximate pK_a is known, an exact determination can be made by preparing solutions of the substance in buffers over a range close to the approximate pK_a ($pK_a \pm 0.5$)

If the substance is a weak base, then the following equation may be used

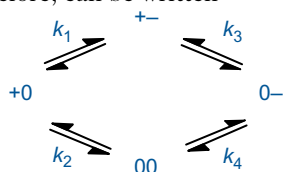
$$pK_a = pH + \log \left(\frac{At - Am}{Ai - At} \right) \quad (101)$$

The data in Table 15-7 were obtained from a 0.0002M solution of the weak base, acridine, illustrating how the pK_a can be determined in this manner. The absorbance of the unionized species (A_m) is 0.025 whereas the absorbance of the ionised species (A_i) is 0.608.

Micro dissociation constants

The dissociation constants for polyprotic acids, as determined by potentiometric titration, are known generally as **macro**, or **titration**, **constants**. As it is known that carboxyl groups are stronger acids than protonated amino groups, there is no difficulty in assigning K_1 and K_2 , as determined by Equation 37, to the carboxyl and amino groups, respectively, of a substance such as glycine hydrochloride.

In other chemicals or drugs such as phenylpropanolamine, in which the two acidic groups are the phenolic and the protonated amino group, the assignment of dissociation constants is more difficult. This is because, in general, both groups have dissociation constants of equal magnitude. Thus, there will be two ways of losing the first proton and two ways of losing the second, resulting in four possible species in solution. This can be illustrated using the convention of assigning a plus (+) to a positively charged group, a 0 to an uncharged group, and a minus (-) to a negatively charged group. Thus, +0 would represent the fully protonated phenylpropanolamine, +/- the dipolar ion, 00 the uncharged molecule, and 0- the anion. The total ionization scheme, therefore, can be written



The micro constants are related to the macro constants as

$$K_1 = k_1 + k_2 \quad (102)$$

$$K_1 K_2 = k_1 k_3 = k_2 k_4 \quad (103)$$

It can be seen from Equation 102 that unless k_1 or k_2 is very much smaller than the other, the observed macro constant is a composite of the two and cannot be assigned to one or the other acidic group in a nonambiguous way. Methods for determining k_1 are given by Riegelman *et al.*⁴ and Niebergall *et al.*⁵ Once k_1 , K_1 , and K_2 have been determined, all of the other micro constants can be obtained from equations 102 and 103.

Other approaches to accurate pK_a determination include capillary electrophoresis and chromatographic, typically HPLC approaches⁵ but, it is often useful to be able to predict the pK_a of a substance prior to synthesis or when only small quantities are available especially with current high-through put methods of drug discovery.

There are several commercially available software packages using linear free energy relations based on the classical Hammett equation-based fragment approaches for the calculation of microscopic and macroscopic ionization constants.⁶ More recent QSPR pK_a prediction methods have used a variety of descriptors including atomic charges, topological and E-states descriptors, chemical reactivity models, atom types, and group philicity. Online chemical modeling systems such as SPARC are now relatively mature and are used widely in academic, government and industrial laboratories and are useful for monoprotic substances but sometimes fail to match with more complex reactivities.^{7,8}

PHARMACEUTICAL SIGNIFICANCE

pK_a is one of the most important physicochemical properties of a molecule and the impact of pH on drug systems is widespread. Both solubility and lipophilicity are both governed by pK_a , therefore understanding and predication of ADME behaviour is impacted. The metabolic profile is highly influenced by the parent compound's pK_a which is also extremely relevant in metabolite identification. Phase II reactions, mechanism-based inhibition are also influenced by pK_a . Receptor binding can also be strongly influenced by pK_a because most drugs are ionized in physiological conditions.

SALT FORMATION

When the first "drugs" or alkaloids were isolated from plant materials they were purified as well-crystallising salts such as morphine hydrochloride, atropine sulphate, codeine phosphate, quinine sulphate and pilocarpine nitrate. In contrast to the free bases, the salts were found to be water soluble and also more stable rendering them more suitable for use as therapeutic agents. Around 50% of drugs are administered as salts.¹⁰ Salt formation is a simple way of modifying the properties of a drug having ionisable functional groups in order to overcome some undesirable characteristic of the parent drug, normally poor solubility. This affords the opportunity to modify other physicochemical characteristics, such as melting point, hygroscopicity, chemical stability, dissolution rate, solution pH, and crystal form; and mechanical properties, such as hardness and elasticity of the potential drug substance and to develop dosage forms with acceptable bioavailability, stability, manufacturability, and patient compliance.

Salts are formed when compound that is ionised in solution forms a strong ionic interaction with an oppositely charged compound leading to the precipitation of the salt form. The counterions are attracted by intermolecular coulombic forces. These interactions change the potential energy landscape and lead to stronger interaction between the charged active pharmaceutical ingredient and polar aqueous solvents, which can result in enhanced dissolution rates and higher apparent solubility on physiological timescales, resulting in increased drug delivery rates *in vivo*. For the salt to be dissolved the solvent must overcome the crystal lattice energy of the solid and create space for the solute. Thus, the solubility of the salt depends on its polarity, lipophilicity, ionization potential, and size. A salt's solubility also depends on the properties of the solvent and solid, such as the crystal packing and the presence of solvates.

The advantages and disadvantages of salt formation for manipulation of drug properties are summarised in Table 15-8. A large number of different salt forms are potentially available for application as the counterion and the following criteria are desirable for a particular salt form:

Table 15-8. Advantages and disadvantages of salt formation (data from Kumar⁹)

Advantages
Altered solubility and dissolution rate
Controlled-release potential
Improved thermal, hydrolytic and photostability
Reduced hygroscopicity
Improved permeability
Improved organoleptic properties
Improved drug efficacy
Reduced pain on injection
Altered melting point resulting in improved milling and formulation properties
Ease of purification and handling
Improved compactability
Extended patent protection
Disadvantages
Only suitable for ionizable compounds
Decreased percentage of active content
Increased potential for formation of solvates and polymorphs
Reduced dissolution rate or solubility for hydrochloride salts in gastric fluid
Increased chance of poor solid-state stability at the microenvironment pH of the salt e.g. precipitation of the free acid in the gastrointestinal environment.
Corrosiveness of salts, resulting in tableting problems
Possible dissociation of hydrochloride and hydrobromide salts, resulting in the release of hydrohalide gas or reaction with excipients or process-related chemicals
Additional step in synthesis of medicinal product

- high aqueous solubility, over a wide pH range, depending on the intended pharmaceutical profile
- high degree of crystallinity
- low hygroscopicity, for consistent performance
- optimal chemical and solid-state stability under accelerated conditions.

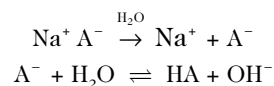
A serious deficiency in any of these characteristics should exclude the salt for further development. Other influential criteria are:

- limited number of polymorphs
- ease of synthesis, handling and formulation development.

The **solubility product** of a salt, K_{sp} , needs to be taken into account when predicting the solubility of a salt in a particular environment that contains other salts with a common counterion.

In spite of the abundance of available counterions, few are used frequently. USP (2006) showed that salt forms (56.15%) are generally preferred over free acid or base forms of a drug (43.85%).¹⁰ Hydrochloride (63%) and sodium (40%) salts remain the favourite counterions in salt formation with basic and acidic drugs, respectively. Because of the low solubility of many basic drugs, where pH_{max} (pH of maximum solubility of the salt) is very low, most common carboxylic acids do not form acceptable salts and it is anticipated that the use of relatively strong counterions will continue in the future. Similarly, 14 out of 19 salt forms of new chemical entities approved by the FDA from 1995 to 2006 are prepared with strong alkalis such as NaOH and KOH, and this trend is also expected to continue in the future.¹¹

If a salt, NaA, is added to water to give a concentration C_s , the following reactions occur:



If the pH of the solution is lowered, more of the A^- would be converted to the unionized acid, HA, in accordance with Le Chatelier's principle. Eventually, a pH will be obtained, below which the amount of HA formed exceeds its **intrinsic aqueous solubility**, S_0 , and the acid will precipitate from solution; this pH can be designated as pH_p . At this point, at which the amount of HA formed just equals S_0 , a mass balance on the total amount of drug in solution yields

$$C_s = [HA] + [A^-] = S_0 + [A^-] \quad (104)$$

Replacing $[A^-]$ as a function of hydroniumion concentration gives

$$C_s = S_0 + \frac{K_a C_s}{[H_3O^+]_p + K_a} \quad (105)$$

where K_a is the ionization constant for the conjugate acid, HA, and $[H_3O^+]_p$ refers to the hydroniumion concentration above which precipitation will occur. This equation can be rearranged to give

$$[H_3O^+]_p = K_a \frac{S_0}{C_s - S_0} \quad (106)$$

Taking logarithms gives

$$pH_p = pK_a + \log \frac{C_s - S_0}{S_0} \quad (107)$$

Thus, the pH below which precipitation occurs is a function of the amount of salt added initially, the pK_a and the solubility of the free acid formed from the salt.

The analogous equation for salts of weak bases and strong acids (such as pilocarpine hydrochloride, cocaine hydrochloride, or codeine phosphate) is

$$pH_p = pK_a + \log \frac{S_0}{C_s - S_0} \quad (108)$$

in which pK_a refers to the protonated form of the weak base.

For both acids and bases, when $pH = pK_a$, $[H_3O^+] = K_a$. If either is substituted the solution is as below,

$$C_s = S_0 \left(\frac{[H_3O^+] + [H_3O^+]}{[H_3O^+]} \right) \quad (109)$$

It can be seen that at 50% ionization, the solubility is always equal to twice the intrinsic solubility (S_0).

Ionic liquids

Ionic liquids (ILs) are salt-like materials that are liquid below 100°C and whose melts are composed of discrete cations and anions, thus there is no molecular species present. Although they have been known for over a century, interest is increasing due to their appealing solvent properties. They are miscible with water or organic solvents their use can be classified as process chemicals (e.g., solvents, separation media) and performance chemicals (e.g., electrolytes, lubricants) and are useful solvents for extraction processes.

Other properties of ILs are that they have practically no vapor pressure and are reported to have a wide window of electrochemical stability, good electrical conductivity, high ionic mobility, and excellent chemical stability. The third and most recent generation of ILs involve active pharmaceutical ingredients, which are being used to produce ILs with biological activity.¹²

DRUG STABILITY

One of the most diversified and fruitful areas of study is the investigation of the effect of hydrogen-ion concentration on the stability or, in more general terms, the reactivity of pharmaceutical systems. The evidence for enhanced stability of systems

when these are maintained within a narrow range of pH, as well as of progressively decreasing stability as the pH de-parts from the optimum range, is abundant. Stability (or instability) of a system may result from gain or loss of a proton (hydrogen ion) by a substrate molecule—often accompanied by an electronic rearrangement—that reduces (or increases) the reactivity of the molecule. **Instability** results when the substance desired to remain unchanged is converted to one or more other, unwanted, substances. In aqueous solution, instability may arise through the catalytic effect of acids or bases—the former by transferring a proton to the substrate molecule, the latter by accepting a proton.

Specific illustrations of the effect of hydrogen-ion concentration on the stability of medicinals are myriad; only a few will be given here, these being chosen to show the importance of pH adjustment of solutions that require sterilization.

Morphine solutions are not decomposed during a 60-min exposure at a temperature of 100°C if the pH is less than 5.5; neutral and alkaline solutions, however, are highly unstable. Minimum hydrolytic decomposition of solutions of cocaine occurs in the range of pH of 2 to 5; in one study a solution of cocaine hydrochloride, initially at a pH of 5.7, remained stable during 2 months (although the pH dropped to 4.2 in this time), whereas another solution buffered to about pH 6 under-went approximately 30% hydrolysis in the same time. Similarly, solutions of procaine hydrochloride containing some hydrochloric acid showed no appreciable decomposition; when dissolved in water alone, 5% of the procaine hydrochloride hydrolyzed, whereas when buffered to pH 6.5, from 19 to 35% underwent decomposition by hydrolysis. Solutions of thiamine hydrochloride may be sterilized by autoclaving without appreciable decomposition if the pH is below 5; above this, thiamine hydrochloride is unstable.

The stability of many disperse systems, and especially of certain emulsions, is often pH dependent. Information concerning specific emulsion systems, and the effect of pH upon them, may be found in Chapter 20.

DRUG ACTIVITY

Drugs that are weak acids or weak bases—and hence may exist in ionized or nonionized form (or a mixture of both)—may be active in one form but not in the other; often such drugs have an optimum pH range for maximum activity. Thus, mandelic acid, benzoic acid, or salicylic acid have pronounced antibacterial activity in their nonionized forms but have practically no such activity in the equivalent ionized form. Accordingly, these substances require an acidic environment to function effectively as antibacterial agents. For example, sodium benzoate is effective as a preservative in 4% concentration at pH 7, in 0.06 to 0.1% concentration at pH 3.5 to 4, and in 0.02 to 0.03% concentration at pH 2.3 to 2.4. Other antibacterial agents are active principally, if not entirely, in cationic form. Included in this category are the acridines and quaternary ammonium compounds.

DRUG ABSORPTION

The degree of ionization and lipid solubility of a drug are two important factors that determine the rate of **absorption** of drugs from the gastrointestinal tract, and indeed their passage through cellular membranes generally. Drugs that are weak organic acids or bases, and that in nonionized form are soluble in lipids, apparently are absorbed through cellular membranes by virtue of the lipoidal nature of the membranes. Completely ionized drugs, on the other hand, are absorbed poorly, if at all. Rates of absorption of a variety of drugs are related to their ionization constants and in many cases may be predicted quantitatively on the basis of this relationship. Thus, not only the degree of the acidic or basic character of a drug, but also consequently the pH of the physiological medium (e.g., gastric or intestinal fluid, plasma, cerebrospinal fluid) in which a drug is dissolved or dispersed—because this pH determines the extent to which the drug will be converted to ionic or nonionic form—become important parameters of drug absorption. Further information on drug absorption is given in Chapter 39.

ACKNOWLEDGMENT: The author acknowledges the extensive contribution of previous versions of this chapter.

REFERENCES

1. Avdeef A. *Absorption and Drug Development: Solubility, Permeability, and Charge State*. Wiley-IEEE, New York, 2003.
2. Albert A, Serjeant EP. *The Determination of Ionisation Constants - A laboratory Manual*, 3rd Edition. London: Chapman and Hall, 1984.
3. Benet LZ, Goyan JE. Nonlogarithmic titration curves for the determination of dissociation constants and purity. *J Pharm Sci* 1965; 54: 1179–82.
4. Riegelman S *et al.* Acid dissociation constants of phenylal kanolamines. *J Pharm Sci* 1962; 51: 129–33.
5. Niebergall PJ *et al.* Spectral determination of microdissociation constants. *J Pharm Sci* 1972; 61: 232–4.
6. Perrin DD *et al.* *pK_a prediction for organic acids and bases*. London; New York: Chapman and Hall, 1981.
7. Van de Waterbeemd H, Gifford E. ADMET in silico modelling: towards prediction paradise? *Nature Rev Drug Discovery* 2003; 2: 192–204.
8. Balogh GT *et al.* Comparative Evaluation of in Silico pK_a Prediction Tools on the Gold Standard Dataset. *QSAR Comb Sci* 2009; 28: 1148–55.
9. Kumar L *et al.* Salt Selection in Drug Development. *Pharm Tech* 2008; 3: 22.
10. *United States Pharmacopeia*, USP 29–NF 24. Rockville, MD: US Pharmacopeial Convention, 2006.
11. Serajuddin AT M. Salt formation to improve drug solubility. *Adv Drug Del Rev* 2007; 59: 603–16.
12. Ferraz R *et al.* Ionic liquids as active pharmaceutical ingredients. *Chem Med Chem* 2011; 6: 975–85.

Chapter 16

Tonicity, Osmoticity, Osmolality, Osmolarity

Andrew Ingham, MRPharmS, PhD and Cathy Y. Poon, PharmD

BASIC DEFINITIONS 277

THERAPEUTIC CONSIDERATIONS 278

OSMOLALITY AND OSMOLARITY 278

COMPUTATION OF OSMOLARITY 279

OSMOMETRY AND THE CLINICAL LABORATORY 280

UNDESIRABLE EFFECTS OF ABNORMAL OSMOTICITY 281

OSMOTICITY AND ENTERAL HYPERALIMENTATION 282

OSMOLALITY DETERMINATION 284

FREEZING-POINT CALCULATIONS 284

TONICITY TESTING BY OBSERVING ERYTHROCYTE CHANGES 284

METHODS OF ADJUSTING TONICITY 284

APPENDIX A 288

APPENDIX B 299

BASIC DEFINITIONS

If a solution is placed in contact with a membrane that is permeable to molecules of the solvent, but not to molecules of the solute, the movement of solvent through the membrane is called *osmosis*. Such a membrane often is called *semipermeable*. As the several types of membranes of the body vary in their permeability, it is well to note that they are *selectively* permeable. Most normal living-cell membranes maintain various solute concentration gradients. A selectively permeable membrane may be defined either as one that does not permit free, unhampered diffusion of all the solutes present or as one that maintains at least one solute concentration gradient across itself. Osmosis, then, is the diffusion of water through a membrane that maintains at least one solute concentration gradient across itself.

Assume that Solution A is on one side of the membrane, and Solution B of the same solute but of a higher concentration is on the other side; the solvent will tend to pass into the more concentrated solution until equilibrium has been established. The pressure required to prevent this movement is the osmotic pressure. It is defined as the excess pressure, or pressure greater than that above the pure solvent, that must be applied to Solution B to prevent passage of solvent through a perfect semipermeable membrane from A to B. The concentration of a solution with respect to effect on osmotic pressure is related to the number of particles (un-ionized molecules, ions, macromolecules, aggregates) of solute(s) in solution and, thus, is affected by the degree of ionization or aggregation of the solute. See Chapter 13 for review of colligative properties of solutions.

Body fluids, including blood and lacrimal fluid, normally have an osmotic pressure that often is described as corresponding to that of a 0.9%w/v solution of sodium chloride. The body also attempts to keep the osmotic pressure of the contents of the gastrointestinal (GI) tract at about this level, but there the normal range is much wider than that of most body fluids. The 0.9%w/v sodium chloride solution is said to be *iso-osmotic* with physiological fluids. In medicine, the term *isotonic*, meaning equal tone, is commonly used interchangeably with *iso-osmotic*. However, terms such as *isotonic* and *tonicity* should be used *only* with reference to a physiological fluid. *Iso-osmotic* actually is a physical term that compares the osmotic pressure (or another colligative property, such as freezing-point depression) of two liquids, neither of which may be a physiological fluid, or which may be a physiological fluid only under certain circumstances. For example, a solution of boric acid that is *iso-osmotic* with

both blood and lacrimal fluid is *isotonic* only with the lacrimal fluid. This solution causes hemolysis of red blood cells because molecules of boric acid pass freely through the erythrocyte membrane, regardless of concentration. Thus, *isotonicity* infers a sense of physiological compatibility where *iso-osmoticity* need not. As another example, a *chemically defined elemental diet* or enteral nutritional fluid can be *iso-osmotic* with the contents of the GI tract but would not be considered a physiological fluid or suitable for parenteral use.

A solution is *isotonic* with a living cell if there is no net gain or loss of water by the cell, or no other change in the cell, when it is in contact with that solution. Physiological solutions with an osmotic pressure lower than that of body fluids, or of 0.9% sodium chloride solution, are referred to commonly as being *hypotonic*. Physiological solutions having a greater osmotic pressure are termed *hypertonic*.

Such qualitative terms are of limited value, and it has become necessary to state osmotic properties in quantitative terms. To do so, a term must be used that will represent all the particles that may be present in a given system. The term used is *osmol*: the weight, in grams, of a solute, existing in a solution as molecules (and/or ions, macromolecules, aggregates, etc.), which is osmotically equivalent to a mole of an ideally behaving nonelectrolyte. Thus, the *osmol* weight of a nonelectrolyte, in a dilute solution, generally is equal to its gram molecular weight. A *milliosmol*, abbreviated *mOsm*, is the weight stated in milligrams.

If one extrapolates this concept of relating an *osmol* and a mole of a nonelectrolyte as being equivalent, then one also may define an *osmol* in the following ways. It is the amount of solute that will provide 1 Avogadro's number (6.02×10^{23}) of particles in solution, and it is the amount of solute that, on dissolution in 1 kg of water, will result in an osmotic pressure increase of 17 000 torr at 0° or 19 300 torr at 37°. One *mOsmol* is 1/1000 of an *osmol*. For example, 1 mol of anhydrous glucose is equal to 180 g. One *osmol* of this nonelectrolyte is also 180 grams. One *mOsmol* would be 180 mg. Thus, 180 mg of this solute dissolved in 1 kg of water will produce an increase in osmotic pressure of 19.3 torr at body temperature.

For a solution of an electrolyte such as sodium chloride, one molecule of sodium chloride represents one sodium and one chloride ion. Hence, 1 mol will represent 2 *osmol* of sodium chloride theoretically. Accordingly, 1 *osmol* NaCl = 58.5 g/2 or 29.25 g. This quantity represents the sum total of 6.02×10^{23} ions as the total number of particles. Ideal solutions infer very dilute solutions or infinite dilution.

However, as pH or concentration is increased, other factors enter. Strong electrolyte, interionic attraction causes a decrease in their effect on colligative properties. In addition, the charge of ions must also follow an electrical gradient which operates to intensify or reduce their colligative effect *in vivo* physiological conditions. Body pH is actively maintained at differing values through out the gastro-intestinal tract with more minor variations in other organs. A combination of electrochemical gradient and chemical concentration maintains pH at different values across membranes, often forcing a change in state of organic acids or bases between ionized (multiple species) and un-ionized (one species) states. The body uses these systems for absorption of nutrition and medications but also excretion of waste products.

Therefore, it is very difficult and often impossible to predict accurately the osmoticity of a solution. It may be possible to do so for a dilute solution of a single pure and well-characterized solute entering a similarly well defined environment, but not for most parenteral and enteral medicinal and/or nutritional fluids; experimental determination is likely required to predict behavior *in vivo*.

THERAPEUTIC CONSIDERATIONS

Generally it is accepted that osmotic effects have a major place in the maintenance of homeostasis (the state of equilibrium in the living body, with respect to various functions and to the chemical composition of the fluids and tissues, e.g., temperature, heart rate, blood pressure, water content, or blood sugar). To a great extent these effects occur within or between cells and tissues where they cannot be measured. One of the most troublesome problems in clinical medicine is the maintenance of adequate body fluids and proper balance between extracellular and intracellular fluid volumes in seriously ill patients. It should be kept in mind, however, that fluid and electrolyte abnormalities are not diseases but are the manifestations of disease.

The physiological mechanisms that control water intake and output appear to respond primarily to serum osmoticity. Renal regulation of output is influenced by variation in rate of release of pituitary antidiuretic hormone (ADH) and other factors in response to changes in serum osmoticity. Osmotic changes also serve as a stimulus to moderate thirst. This mechanism is sufficiently sensitive to limit variations in osmoticity in the normal individual to less than about 1%. Body fluid continually oscillates around a narrow genetically determined set point in the normal range of 280–295 mOsmol/L. An increase of plasma osmoticity of 1% will stimulate ADH release, resulting in a reduction of urine flow, and, at the same time, stimulate thirst that results in increased water intake. Both the increased renal reabsorption of water (without solute) stimulated by circulating ADH and the increased water intake tend to lower serum osmoticity.

The transfer of water through the cell membrane occurs so rapidly that any lack of osmotic equilibrium between the two fluid compartments in any given tissue usually is corrected within a few seconds and, at most, within a minute or so. However, this rapid transfer of water does not mean that complete equilibration occurs between the extracellular and intracellular compartments throughout the entire body within this same short period of time. The reason is that fluid usually enters the body through the gut and then must be transported by the circulatory system to all tissues before complete equilibration can occur. In the normal person it may require 30 to 60 minutes to achieve reasonably good equilibration throughout the body after drinking water. Osmoticity is the property that largely determines the physiological acceptability of a variety of solutions used for therapeutic and nutritional purposes.

Pharmaceutical and therapeutic consideration of osmotic effects has been, to a great extent, directed toward the side effects of ophthalmic and parenteral medicinals due to abnormal osmoticity, and either to formulating to avoid the side effects or

to finding methods of administration to minimize them. More recently this consideration has been extended to total (central) parenteral nutrition, to enteral hyperalimentation ("tube" feeding), and to concentrated-fluid infant formulas.^{1,2} Also, in recent years, the importance of osmometry of serum and urine in the diagnosis of many pathological conditions has been recognized.

There are a number of examples of the direct therapeutic effect of osmotic action, such as the *intravenous* (IV) use of mannitol as a diuretic that is filtered at the glomeruli and thus increases the osmotic pressure of tubular urine. Water must then be reabsorbed against a higher osmotic gradient than otherwise, so reabsorption is slower and diuresis is observed. The same fundamental principle applies to the IV administration of 30% urea used to affect intracranial pressure in the control of cerebral edema. Peritoneal dialysis fluids tend to be somewhat hyperosmotic to withdraw water and nitrogenous metabolites. Two to 5% sodium chloride solutions or dispersions in an oleaginous base (Muro, *Bausch & Lomb*) and a 40% glucose ointment are used topically for corneal edema. Ophthalmalgin (*Wyeth-Ayerst*) is ophthalmic glycerin employed for its osmotic effect to clear edematous cornea to facilitate an ophthalmoscopic or gonioscopic examination. Glycerin solutions in 50% concentration (Osmoglyn, *Alcon*) and isosorbide solution (Ismotic, *Alcon*) are oral osmotic agents for reducing intraocular pressure.

The osmotic principle also applies to plasma extenders, such as polyvinylpyrrolidone and to saline laxatives, such as magnesium sulfate, magnesium citrate solution, magnesium hydroxide (via gastric neutralization), sodium sulfate, sodium phosphate, sodium biphosphate oral solution, and enema (*Fleet*).

An interesting osmotic laxative that is a nonelectrolyte is a lactulose solution. Lactulose is a nonabsorbable disaccharide that is colon-specific, wherein colonic bacteria degrade some of the disaccharide to lactic and other simple organic acids. These, in total, lead to an osmotic effect and laxation. An extension of this therapy is illustrated by Cephulac (*Marion Merrell Dow*) solution, which uses the acidification of the colon via lactulose degradation to serve as a trap for ammonia migrating from the blood to the colon. The conversion of ammonia of blood to the ammonium ion in the colon ultimately is coupled with the osmotic effect and laxation, thus expelling undesirable levels of blood ammonia. This product is employed to prevent and treat frontal systemic encephalopathy.

Osmotic laxation is observed with the oral or rectal use of glycerin and sorbitol. Epsom salt has been used in baths and compresses to reduce edema associated with sprains. Another approach is the indirect application of the osmotic effect in therapy via osmotic pump drug delivery systems.³

OSMOLALITY AND OSMOLARITY

It is necessary to use several additional terms to define expressions of concentration in reflecting the osmoticity of solutions. The terms include *osmolality*, the expression of osmolal concentration, and *osmolality*, the expression of osmolal concentration.

OSMOLALITY (OSMOL/KG)

A solution has an osmolal concentration of 1 when it contains 1 osmol of solute/kg of water. A solution has an osmolality of n when it contains n osmol/kg of water. Osmolal solutions, like their counterpart molal solutions, reflect a weight-to-weight relationship between the solute and the solvent. Because an osmol of any nonelectrolyte is equivalent to 1 mol of that compound, a 1 osmolal solution is synonymous to a 1 molal solution for a typical nonelectrolyte.

With a typical electrolyte like sodium chloride, 1 osmol is approximately 0.5 mol of sodium chloride. Thus, it follows that a 1 osmolal solution of sodium chloride essentially is equivalent to a 0.5 molal solution. Recall that a 1 osmolal solution of glucose or sodium chloride each will contain the same particle

concentration. In the glucose solution there will be 6.02×10^{23} molecules/kg of water, and in the sodium chloride solution, one will have 6.02×10^{23} total ions/kg of water, one-half of which are Na^+ ions and the other half Cl^- ions.

As in molal solutions, osmolal solutions usually are employed where quantitative precision is required, as in the measurement of physical and chemical properties of solutions (i.e., colligative properties). The advantage of the w/w relationship is that the concentration of the system is not influenced by temperature.

OSMOLARITY (osmol/L)

The relationship observed between molality and osmolality is shared similarly between molarity and osmolality. A solution has an osmolar concentration of 1 when it contains 1 osmol of solute per liter of solution. Likewise, a solution has an osmolality of n when it contains n osmols/L of solution. Osmolar solutions, unlike osmolal solution, reflect a weight in volume relationship between the solute and final solution. It is of note that concentrations expressed as Osmolarity are temperature dependent, whereas Osmolality are independent of temperature. For sodium chloride a 1 osmolar solution would contain 1 osmol of sodium chloride per liter, which approximates a 0.5 molar solution. The advantage of employing osmolar concentrations over osmolal concentrations is the ability to relate a specific number of osmols or milliosmols to a volume, such as a liter or milliliter. Thus, the osmolar concept is simpler and more practical. Volumes of solution, rather than weights of solution, are more practical in the delivery of liquid dosage forms.

Many health professionals do not have a clear understanding of the difference between osmolality and osmolality. In fact, the terms have been used interchangeably. A 1 osmolar solution of a solute always will be more concentrated than a 1 osmolal solution. With dilute solutions the difference may be acceptably small. For example, a 0.9% w/v solution of sodium chloride in water contains 9 g of sodium chloride/L of solution, equivalent to 0.308 osmolar; or 9 g of sodium chloride/996.5 g of water, equivalent to 0.309 osmolal, less than a 1% error. For concentrated solutions the percent difference between osmolality and osmolality is much greater and may be highly significant; 3.5% for 5% w/v glucose solution and 25% for 25% w/v glucose solution. One should be alerted to the sizable errors that may occur with concentrated solutions or fluids, such as those employed in total parenteral nutrition, enteral hyperalimentation, and oral nutritional fluids for infants.

Reference has been made to the terms hypertonic and hypotonic. Analogous terms are hyperosmotic and hypo-osmotic. Assuming normal serum osmolality to be 285 mOsmol/kg, as serum osmolality increases due to water deficit, the following signs and symptoms usually are found to accumulate progressively at approximately these values: 294 to 298—thirst (if the patient is alert and communicative); 299 to 313—dry mucous membranes; 314 to 329—weakness, doughy skin; above 330—disorientation, postural hypotension, severe weakness, fainting, CNS changes, stupor, and coma. As serum osmolality decreases due to water excess, the following may occur: 275 to 261—headache; 262 to 251—drowsiness, weakness; 250 to 233—disorientation, cramps; below 233—seizures, stupor, and coma.

Serum osmolality is often loosely stated to be about 300 mOsmol/L, with a variety of normal bands reported across literature: 280 to 295 mOsmol/L, 275 to 300 mOsmol/L, 290 mOsmol/L, 306 mOsmol/L, and 275 to 295 mOsmol/kg.

In recent years much attention has been directed at determining osmoticity of total parenteral nutrition solutions, enteral formulas, and parenteral and enteral medications.⁴⁻⁶ Hyperosmoticity of parenteral and enteral formulas and medications serves as an indicator for potential risks, including thrombophlebitis, pain at injection site, diarrhea, and abdominal cramping. However, the terms osmolality and osmolality often have been used interchangeably and caused much confusion for

practitioners. Often, when the term osmolality is used, one cannot discern whether this simply is incorrect terminology or if osmolality actually has been calculated from osmolality.

Another current practice that can cause confusion is the use of the terms *normal* or *physiological* for isotonic sodium chloride solution (0.9%). The solution surely is iso-osmotic. However, as to being physiological, the concentration of ions is each of 154 mEq/L, whereas serum contains about 140 mEq of sodium and about 103 mEq of chloride.

The range of mOsmol values found for serum raises the question as to what really is meant by the terms hypotonic and hypertonic for medicinal and nutritional fluids. One can find the statement that fluids with an osmolality of 50 mOsmol or more above normal are hypertonic; and if they are 50 mOsmol or more below normal, they are hypotonic. One also can find the statement that peripheral infusions should not have an osmolality exceeding 700 to 800 mOsmol/L.⁷ Examples of osmol concentrations of solutions used in peripheral infusions are (D5W) 5% glucose solution, 252 mOsmol/L; (D10W) 10% glucose solution, 505 mOsmol/L; Hartmann's solution, 278 mOsmol/L and Lactated Ringer's 5% Glucose, 525 mOsmol/L. When a fluid is hypertonic, undesirable effects often can be decreased by using relatively slow rates of infusion and/or relatively short periods of infusion. For example, 25% glucose solution (D25W)—4.25% Amino Acids is a representative of a highly osmotic hyperalimentation solution. It has been stated that when osmol loading is needed, a maximum safe tolerance for a normally hydrated subject would be an approximate increase of 25 mOsmol/kg of water over four hours.⁸

COMPUTATION OF OSMOLARITY

Several methods are used to obtain numerical values of osmolality. The osmolar concentration, sometimes referred to as the *theoretical osmolality*, is calculated from the w/v concentration using the following equation:

$$\frac{\text{g}}{\text{L}} \times \frac{\text{mols}}{\text{g}} \times \frac{\text{osmol}}{\text{mol}} \times \frac{1000 \text{ mOsmol}}{\text{osmol}} = \frac{\text{mOsmol}}{\text{L}} \quad (1)$$

The number of osmol/mol is equal to 1 for nonelectrolytes and is equal to the number of ions per molecule for strong electrolytes.

This calculation omits consideration of factors such as solvation and interionic forces. By this method of calculation, 0.9% sodium chloride has an osmolar concentration of 308 mOsmol/L and a concentration of 154 mOsmol/L in either sodium or chloride ion.

Two other methods compute osmolality from values of osmolality. The determination of osmolality will be discussed later. One method has a strong theoretical basis of physical-chemical principles⁹ using values of the partial molal volume(s) of the solute(s). An 0.9% sodium chloride solution, found experimentally to have an osmolality of 286 mOsmol/kg, was calculated to have an osmolality of 280 mOsmol/L, rather different from the value of 308 mOsmol/L calculated as above. The method, using partial molal volumes, is relatively rigorous, but many systems appear to be too complex and/or too poorly defined to be dealt with by this method.

The other method is based on calculating the weight of water from the solution density and concentration:

$$\frac{\text{g water}}{\text{mL solution}} = \frac{\text{g solution}}{\text{mL solution}} - \frac{\text{g solute}}{\text{mL solution}} \quad (2)$$

then,

$$\text{osmolality} \left(\frac{\text{mOsmol}}{\text{L solution}} \right) = \text{osmolality} \left(\frac{\text{mOsmol}}{1000 \text{ g water}} \right) \times \frac{\text{g water}}{\text{mL solution}} \quad (3)$$

The experimental value for the osmolality of 0.9% sodium chloride solution was 292.7 mOsmol/kg; the value computed for osmolality was 291.4 mOsmol/L. This method uses easily obtained values of density of the solution and of its solute content and can

be used with all systems. For example, the osmolality of a nutritional product was determined by the freezing-point depression method to be 625 mOsmol/kg; its osmolality was calculated as $625 \times 0.839 = 524$ mOsmol/L (Bray AJ, Evansville: Mead Johnson Nutritional Division, personal communication, 1978).

US pharmacopeia monographs may require labeling of osmolality for IV solutions, nutrients, or electrolytes, and for osmotic diuretics such as mannitol.¹⁰ Where, the osmolar concentration is demanded, it is specified in mosmol/L; however, when the contents are less than 100 mL, or when the label states the article is not for direct injection but is to be diluted before use, the label alternatively may state the total osmolar concentration in mOsmol/mL. The European pharmacopoeia, in contrast, has very few monographs requiring osmolality labeling; it has been required only in solutions for Haemofiltration & Peritoneal Dialysis¹¹ but specifying regulatory osmolality test limits for in excess of 30 specific monographs and blood products.

An example of the use of the first method described above is the computation of the approximate osmolar concentration (*theoretical osmolality*) of a Lactated Ringer's 5% Glucose Solution (Abbott), which is labeled to contain, per liter, glucose (hydrous) 50 g, sodium chloride 6 g, potassium chloride 300 mg, calcium chloride 200 mg, and sodium lactate 3.1 g. Also stated is that the total osmolar concentration of the solution is approximately 524 mOsmol/L, in part contributed by 130 mEq of Na⁺, 109 mEq of Cl⁻, 4 mEq of K⁺, 3 mEq of Ca²⁺, and 28 mEq of lactate ion.

The derivation of the osmolar concentrations from the stated composition of the solution may be verified by calculations using Equation 1.

Glucose

$$\frac{50 \text{ g}}{\text{L}} \times \frac{1 \text{ mol}}{198 \text{ g}} \times \frac{1 \text{ osmol}}{\text{mol}} \times \frac{1000 \text{ mOsmol}}{\text{Osmol}} = 252 \text{ mOsmol/L} \quad (4)$$

Sodium Chloride

$$\frac{0.3 \text{ g}}{\text{L}} \times \frac{1 \text{ mol}}{74.6 \text{ g}} \times \frac{2 \text{ osmol}}{\text{mol}} \times \frac{1000 \text{ mOsmol}}{\text{osmol}} = \frac{8.04 \text{ mOsmol}}{\text{L}} \left\{ \begin{array}{l} (4.02 \text{ mOsmol K}^+) \\ (4.02 \text{ mOsmol Cl}^-) \end{array} \right. \quad (5)$$

Potassium Chloride

$$\frac{0.3 \text{ g}}{\text{L}} \times \frac{1 \text{ mol}}{74.6 \text{ g}} \times \frac{2 \text{ osmol}}{\text{mol}} \times \frac{1000 \text{ mOsmol}}{\text{osmol}} = \frac{8.04 \text{ mOsmol}}{\text{L}} \left\{ \begin{array}{l} (4.02 \text{ mOsmol K}^+) \\ (4.02 \text{ mOsmol Cl}^-) \end{array} \right. \quad (6)$$

Calcium Chloride

$$\frac{0.2 \text{ g}}{\text{L}} \times \frac{1 \text{ mol}}{111 \text{ g}} \times \frac{3 \text{ osmol}}{\text{mol}} \times \frac{1000 \text{ mOsmol}}{\text{osmol}} = \frac{5.41 \text{ mOsmol}}{\text{L}} \left\{ \begin{array}{l} (1.80 \text{ mOsmol Ca}^{2+}) \\ (3.61 \text{ mOsmol Cl}^-) \end{array} \right. \quad (7)$$

Sodium Lactate

$$\frac{3.1 \text{ g}}{\text{L}} \times \frac{1 \text{ mol}}{112 \text{ g}} \times \frac{2 \text{ osmol}}{\text{mol}} \times \frac{1000 \text{ mOsmol}}{\text{osmol}} = \frac{55.4 \text{ mOsmol}}{\text{L}} \left\{ \begin{array}{l} (27.7 \text{ mOsmol Na}^+) \\ (27.7 \text{ mOsmol lactate}) \end{array} \right. \quad (8)$$

The total osmolar concentration of the five solutes in the solution is 526, in good agreement with the labeled total osmolar concentration of approximately 524 mOsmol/L.

The mOsmol of sodium in 1 L of the solution is the sum of the mOsmol of the ion from sodium chloride and sodium lactate: $102 + 27.6 = 129.6$ mOsmol. Chloride ions come from the sodium chloride, potassium chloride, and calcium chloride, the total osmolar concentration being $102 + 4.02 + 3.61 = 109.6$ mOsmol. The mOsmol values of potassium, calcium, and lactate are calculated to be 4.02, 1.80, and 27.6, respectively.

The osmolality of a mixture of complex composition, such as an enteral hyperalimentation fluid, cannot be calculated with any acceptable degree of certainty; therefore, the *osmolality* of such preparations should be determined experimentally.

OSMOMETRY AND THE CLINICAL LABORATORY

Serum and urine osmometry may assist in the diagnosis of certain fluid and electrolyte problems. However, osmometry values have little meaning unless the clinical situation is known. Osmometry is used in renal dialysis as a check on the electrolyte composition of the fluid. In the clinical laboratory, as stated above, the term *osmolality* is used generally but usually is reported as mOsmol/L. It may seem unnecessary to mention that osmolality depends not only on the number of solute particles but also on the quantity of water in which they are dissolved. However, it may help one to understand the statement that the normal range of urine osmolality is 50 to 1400 mOsmol/L, and for a random specimen is 500 to 800 mOsmol/L.

SERUM OSMOTICITY

Sodium is by far the principal solute involved in serum osmoticity. Therefore, abnormal serum osmoticity is most likely to be associated with conditions that cause abnormal sodium concentration and/or abnormal water volume.

Thus, hyperosmotic serum is likely to be caused by an increase in serum sodium and/or loss of water. It may be associated with diabetes insipidus, hypercalcemia, and diuresis during severe hyperglycemia, or with early recovery from renal shutdown. Alcohol ingestion is said to be the most common cause of the hyperosmotic state and of coexisting coma and the hyperosmotic state. An example of hyperosmoticity is a comatose diabetic with a serum osmoticity of 365 mOsmol/L.

In a somewhat analogous fashion, hypo-osmotic serum is likely to be due to decrease in serum sodium and/or excess of water. It may be associated with the postoperative state (especially with excessive water replacement therapy), treatment with diuretic drugs and low-salt diet (as with patients with heart failure, cirrhosis, etc.), adrenal disease (e.g., Addison's disease, adrenogenital syndrome), or SIADH (syndrome of inappropriate ADH secretion). There are many diseases that cause ADH to be released inappropriately (i.e., in spite of serum osmoticity and volume having been normal initially). These include oat-cell carcinoma of the lung, bronchogenic carcinoma, congestive heart failure, inflammatory pulmonary lesions, porphyria, severe hypothyroidism, or cerebral disease (such as tumor, trauma, infection, and vascular abnormalities). It also may be found with some patients with excessive diuretic use. Serum and urine osmoticity are measured when SIADH is suspected. In SIADH there is hypo-osmoticity of the blood in association with a relative hyperosmoticity of urine. The usual cause is a malfunction of the normal osmotic response of osmoreceptors, an excess of exogenous vasopressin, or a production of a vasopressin-like hormone that is not under the regular control of serum osmoticity. The diagnosis is made by simultaneous measurement of urine and serum osmolality. The serum osmolality will be lower than normal and much lower than the urine osmolality, indicating inappropriate secretion of a concentrated urine in the presence of a dilute serum.

Cardiac, renal, and hepatic disease characteristically reduce the sodium/osmolality ratio, this being partially attributed to the effects of increased blood sugar, urea, or unknown metabolic products. Patients in shock may develop disproportionately elevated measured osmolality, compared to calculated osmolality, which points toward the presence of circulating metabolic products.

There are several approximate methods for estimating serum osmolality from clinical laboratory values for sodium ion. They may be of considerable value in an emergency situation.

1. Serum osmolality may be estimated from (Na in mEq/L, blood sugar and Blood Urea Nitrogen (BUN) in mg/100 mL).

$$\text{mOsmol} = (1.86 \text{ sodium}) + \frac{\text{blood sugar}}{18} + \frac{\text{BUN}}{2.8} + 5 \quad (9)$$

2. A quick approximation is:

$$\text{mOsmol} = 2\text{Na} + \frac{\text{BS}}{20} + \frac{\text{BUN}}{3} \quad (10)$$

3. The osmolality is usually, *but not always*, very close to two times the sodium reading plus 10.

URINE OSMOTICITY

The two main functions of the kidney are glomerular filtration and tubular reabsorption. Clinically, tubular function is measured best by tests that determine the ability of the tubules to concentrate and dilute the urine. Tests of urinary dilution are not as sensitive in the detection of disease as those that test the concentration of urine. Concentration of urine occurs in the renal medulla (interstitial fluids, loops of Henle, capillaries of the medulla, and collecting tubules), a disease process that disturbs the function or structure of the medulla and produces early impairment of the concentrating power of the kidney. Such diseases include acute tubular necrosis, obstructive uropathy, pyelonephritis, papillary necrosis, medullary cysts, hypokalemic and hypercalcemic nephropathy, and sickle cell disease.

Measurement of urine osmolality is an accurate test for the diluting and concentrating ability of the kidneys. In the absence of ADH, the daily urinary output is likely to be six to eight liters or more. The normal urine osmolality depends on the clinical setting; normally, with maximum ADH stimulation, it can be as much as 1200 mOsmol/kg, and with maximum ADH suppression as little as 50 mOsmol/kg. Simultaneous determination of serum and urine osmolality often is valuable in assessing the distal tubular response to circulating ADH. For example, if the patient's serum is hyperosmolal or in the upper limits of normal ranges, and the patient's urine osmolality measured at the same time is much lower, a decreased responsiveness of the distal tubules to circulating ADH is suggested.

Measurement of urine osmolality during water restriction is an accurate, sensitive test of decreased renal function. For example, under the conditions of one test, normal osmolality would be greater than 800 mOsmol/kg. With severe impairment the value would be less than 400 mOsmol/kg. Knowledge of urine osmolality may point to a problem even though other tests are normal (e.g., the Fishberg concentration test, blood urea nitrogen, PSP excretion, creatinine clearance, or IV pyelogram). Knowledge of its value may be useful, especially in diabetes mellitus, essential hypertension, and silent pyelonephritis. The urine/serum osmolality ratio should be calculated and should be equal to or greater than 3.

UNDESIRABLE EFFECTS OF ABNORMAL OSMOTICITY

OPHTHALMIC MEDICATION

It is generally accepted that ophthalmic preparations intended for instillation into the cul-de-sac of the eye should, if possible,

be approximately isotonic to avoid irritation (see Chapter 28). It also has been stated that the abnormal tonicity of contact lens solutions can cause the lens to adhere to the eye and/or cause burning or dryness and photophobia.

NEONATAL ENTERAL AND TOTAL PERIPHERAL MEDICATION

Adult tolerances of osmoticity cannot be expected in pediatric and neonatal therapy. Paracetamol solutions have an osmolality approaching ~15000 mOsm/L with a target of less than 500 mOsmol/L suitable for enteral feed, modifications, or dilutions of these formulations and must be considered to prevent the risks of *pneumotosis intestinalis*, hypernatraemia,² particularly in the cases of significant imbalance of ADH caused by existing trauma or acute diarrhea states.^{12,13}

Because of the different fluid and protein requirements of neonates and pediatric patients, the final concentration of glucose and amino acids are often different from those of an adult. It is now expected that these variations are available to the clinician and that calculation or active measurement of the osmolality of a system should be made. Many automated compounding systems are now capable of calculation of the full bag osmolality, using previously recorded individual coefficients of variations for the individual components. Large validation cycles allow these calculations to be clinically relied upon but pharmacists must take particular care with regard to the extemporaneous addition of compounds to these systems, for these concentrated systems are not been governed by the more simple calculations.

HYPEROSMOTIC AGENTS

Therapies directly applying osmotic gradients therapeutically are few. However, pediatric and mature cystic fibrosis patients infrequently make use of salt solutions in the range of 295–700 mOsm/L for their direct osmotic effect. Upon nebulization of these solutions, water movement is increased through the lungs' transepithelial surface to dilute any mucus volume. Higher osmolalities produce transepithelial movement through tight junctions,^{14–15} further accelerating tracheobronchial mucociliary clearance.¹⁶

DIAGNOSTIC CONTRAST AGENTS

Radiopharmaceutical and tomography contrast agents demand a pharmaceutical concentration to be held in a specific area of interest allowing appropriate imaging. This requirement can produce localized concentrations in the region of 600 mOsm resulting in tissue shrinkage and the potential for resultant cell death.¹⁷ Since dilution of the agent is not possible only after treatment or reformulation with alternate agents could be considered.

PARENTERAL MEDICATION

Osmoticity is of great importance in parenteral injections, its effects depending on such factors as the degree of deviation from tonicity, the concentration, the location of the injection, the volume injected, the speed of the injection, and the rapidity of dilution and diffusion. When formulating hypotonic parenterals, they usually have tonicity adjusted by the addition of glucose or sodium chloride. Hypertonic parenteral drug solutions cannot be adjusted, but if prepared from a dry powder, they may be reconstituted in a different solvent volume. Hypotonic and hypertonic solutions usually are administered slowly, in small volumes, or into a large vein such as the subclavian, where dilution and distribution occur rapidly. Solutions that differ from the serum in tonicity generally cause tissue irritation, pain on injection, and electrolyte shifts, the effect depending on the degree of deviation from tonicity:

- Excessive infusion of *hypotonic* fluids may cause swelling of red blood cells, hemolysis, and water invasion of the body's cells in general. When this is beyond the body's

tolerance for water, water intoxication results, with convulsions and edema, such as pulmonary edema.

- Excessive infusion of *isotonic* fluids can cause an increase in extracellular fluid volume, which can result in circulatory overload.
- Excessive infusion of *hypertonic* fluids leads to a wide variety of complications. For example, the sequence of events when the body is presented with a large IV load of hypertonic fluid, rich in glucose, is as follows: hyperglycemia, glycosuria and intracellular dehydration, osmotic diuresis, loss of water and electrolytes, dehydration, and coma.

One cause of osmotic diuresis is the infusion of glucose at a rate faster than the ability of the patient to metabolize it (as greater than perhaps 400–500 mg/kg per hour for an adult on total parenteral nutrition). A heavy load of unmetabolizable glucose increases the osmoticity of blood and acts as a diuretic; the increased solute load requires more fluid for excretion, 10 to 20 mL of water being required to excrete each gram of glucose. Solutions such as those for total parenteral nutrition should be administered by means of a metered constant-infusion apparatus over a lengthy period (usually more than 24 hours) to avoid sudden hyperosmotic glucose loads. Such solutions may cause osmotic diuresis; if this occurs, water balance is likely to become negative because of the increased urinary volume, and electrolyte depletion may occur because of excretion of sodium and potassium secondary to the osmotic diuresis. If such diuresis is marked, body weight falls abruptly and signs of dehydration appear. Urine should be monitored for signs of osmotic diuresis, such as glycosuria and increased urine volume.

If the IV injection rate of hypertonic solution is too rapid, there may be catastrophic effects on the circulatory and respiratory systems. Blood pressure may fall to dangerous levels, cardiac irregularities or arrest may ensue, respiration may become shallow and irregular, and there may be heart failure and pulmonary edema. Probably the precipitating factor is a bolus of concentrated solute suddenly reaching the myocardium and the chemoreceptors in the aortic arch and carotid sinus.⁸

Abrupt changes in serum osmoticity can lead to cerebral hemorrhage. It has been shown experimentally that rapid infusions of therapeutic doses of hypertonic saline with osmotic loads produce a sudden rise in cerebrospinal fluid (CSF) pressure and venous pressure (VP), followed by a precipitous fall in CSF pressure. This particularly may be conducive to intracranial hemorrhage, for the rapid infusion produces an increase in plasma volume and venous pressure at the same time the CSF pressure is falling. During the CSF pressure rise, there is a drop in hemoglobin and hematocrit, reflecting a marked increase in blood volume.

Hyperosmotic medications, such as sodium bicarbonate (osmolarity of 1560 at 1 mEq/mL), which are administered intravenously, should be diluted prior to use and should be injected slowly to allow dilution by the circulating blood. Rapid *push* injections may cause a significant increase in blood osmoticity.⁹

Safety, therefore, demands that all IV injections, especially highly osmotic solutions, be performed slowly, usually being given preferably over a period not less than that required for a complete circulation of the blood (one minute). The exact danger point varies with the state of the patient, the concentration of the solution, the nature of the solute, and the rate of administration.

Hyperosmotic solutions also should not be discontinued suddenly. In dogs, marked increase in levels of intracranial pressure have occurred when hyperglycemia produced with glucose infusions was suddenly reversed by stopping the infusion and administering saline. It also has been shown that the CSF pressure in humans rises during treatment of diabetic ketoacidosis in association with a fall in the plasma concentration of glucose and a fall in plasma osmolality. These observations may be explained by the different rates of decline in glucose content of the brain and of plasma. The concentration of glucose in the brain may fall more slowly than in the plasma, causing a shift of fluid from

the extracellular fluid space to the intracellular compartment of the CNS, resulting in increased intracranial pressure.

CLINICAL APPLICATIONS

Although there are many issues with abnormal osmoticity, most pharmacists are concerned with preventable adverse effects, such as thrombophlebitis and pain at the injection site. The understanding of these potential risks from hyperosmotic parenteral medications has fine-tuned IV administration techniques. The site of administration—peripheral versus central venous catheter—plays a significant role in determining the final concentration of parenteral medications infused IV. Attention should be directed toward establishing the optimal osmolality of IV administered parenteral medications via the peripheral venous route that will result in the least adverse effects.

Since the introduction of parenteral nutrition support, hyperosmoticity of these nutrition solutions remains a concern. The commonly accepted osmolality of less than 900 mOsmol/L has been quoted for safe peripheral administration of parenteral nutrition solutions.^{18,19} All attempts should be made to prepare solutions with osmoticity close to that of serum osmoticity or no greater than 900 mOsmol/L. This can be achieved by carefully selecting the diluent for dilution and determining the final concentration of the parenteral medication. Glucose 5% in Water for Injection and Sodium Chloride 0.9% have been used routinely as diluents. When comparing the two diluents, parenteral medications diluted with Glucose 5% in Water for Injection have a lower osmolality than do solutions diluted with Sodium Chloride 0.9% at the same final concentration.

Several studies have been conducted to determine optimal final concentration of commonly used parenteral medications.^{4–6} The published final concentrations for most parenteral medications are recommended for peripheral as well as central venous catheter IV administration for patients with no special needs, such as fluid restriction. In the event that fluid restriction is required or the recommended final concentration is not achievable, the parenteral medication should be administered via a central venous catheter, by which immediate dilution and distribution is achieved rapidly. This will minimize potential for phlebitis and pain at the injection site.

Osmoticity issues associated with parenteral medications are also applicable to *total parenteral nutrition* (TPN) solutions, especially via peripheral venous administration. Peripheral parenteral nutrition support remains an integral part of therapeutic options for hospitalized patients. The peripheral route of administration often is preferred for patients who require short-term therapy or supplemental nutrition support.

OSMOTICITY AND ENTERAL HYPERALIMENTATION

Some aspects of nutrition have already been discussed. Here, we present the potential side effects caused with repeated nutritional replacement. There exists an increasing dialogue on nutrition through professional bodies. The professional organizations of The American Society for Parenteral and Enteral Nutrition (ASPEN) and The British Association for Parenteral and Enteral Nutrition (BAPEN) aim to advance clinical nutrition by enforcing best practice standards among dietitians, doctors, nurses, pharmacists, patients, industry representatives and others concerned with the research, development, manufacture, supply, provision, and consumption.

Osmoticity has been of special importance in the IV infusion of large volumes of highly concentrated nutritional solutions. Hyperosmoticity has been a major factor in the requirement that they must be injected centrally into a large volume of rapidly moving blood in preference to peripheral infusion. Technology expansion has led to routes of administration directly to the gastrointestinal tract through instillation directly into specific regions either orally or by nasogastric tube, feeding gastrostomy, or needle-catheter jejunostomy. These oral routes

are always more beneficial to the patient but do raise their own osmotically related challenges.

Enteral nutritional formulas can be modular, allowing individual supplementation of protein, carbohydrate, or fat. Other formulas are called *defined formula diets* and contain protein, carbohydrate, fat, minerals, and vitamins. These nutritionally complete formulations can be monomeric (or oligomeric), based on amino acids, short peptides, and simple carbohydrates or can be polymeric, based on complex protein and carbohydrates. These complex systems cannot have osmolality calculated, and experimental determination must be performed; however, knowledge of these experimental determinations has allowed clinically significant predictions to be validated for routine use.

Enteral diets are necessarily relatively high in osmoticity because their smaller molecules result in more particles per gram than in normal foods. An example is a fluid consisting of L-amino acids, glucose oligosaccharides, vitamins (including fat-soluble vitamins), fat as a highly purified safflower oil or soybean oil, electrolytes, trace minerals, and water. Components not in solution as part of an oily phase (fat) do not contribute to osmoticity. However, the potential for interactions between these phases can cause some significant changes in total particle concentration and indirectly affect the osmoticity.²⁰

Although it is easily digested, glucose contributes more particles than most other carbohydrate sources, such as starch, and is more likely to cause osmotic diarrhea, especially with bolus feeding. For enteral feeds, osmoticity is improved (decreased) by replacing glucose with glucose oligosaccharides (carbohydrates that yield on hydrolysis 2 to 10 monosaccharides).²¹ Additional flavorings, not required for all patients, will also increase the osmoticity of a product.

Commercial diets are packaged as fluids or as powders for reconstitution with water. Categorized on caloric density (calories/mL), protein content, or osmolality (mOsm/kg of H₂O), parenteral nutritional products differ because they are labeled in terms of osmolality (mOsm/L).

The enteral route for nutrition is frequently overlooked in many diseases or post-trauma states. If the patient is not readily responsive to traditional oral feedings, there is no reason that the gastrointestinal tract or regions of it should not be used.

However, when large volumes of osmotically mismatched fluids are administered, they cause an upset in the normal water balance within the body. The nutritional components below are listed in order of decreasing osmotic effect per gram:²²

- 1. Electrolytes such as sodium chloride
- 2. Relatively small organic molecules, such as dextrose (glucose) and amino acids
- 3. Dextrose oligosaccharides
- 4. Starches
- 5. Proteins
- 6. Fats (no osmotic effect, only producing variation in other components).

Foods with high proportions of electrolytes, amino acids, and simple sugars have the greatest effect on osmolality and, as a result, the level of tolerance that can be expected

in specific patients. The approximate osmolality of common foods and an illustrative parenteral nutrition solution is given in Table 16-1.

When nutrition of high osmoticity is ingested, large amounts of water will transfer to the stomach and intestines from fluid surrounding the organs in an attempt to lower the osmoticity. The higher the osmoticity, the larger the amount of water required; a large amount of water in the gastrointestinal tract can cause distention, cramps, nausea, vomiting, hypermotility, and shock. The food may move through the tract too rapidly for the water to be reabsorbed, and result is diarrhea and dehydration. The hyperosmotic enteral effects have been observed by the administration of undiluted hypertonic oral medication.^{23,24} The table in Appendix A lists average osmolality values of some commercially available drug solutions and suspensions. Thus, there is some analogy to the effect of hyperosmotic IV infusions.

Hyperosmotic feedings may result in mucosal damage in the GI tract. Rats given hyperosmotic feeding showed transient decrease in disaccharidase activity and an increase in alkaline phosphatase activity. They also showed morphological alterations in the microvilli of the small intestines. After a period of severe gastroenteritis, the bowel may be unusually susceptible to highly osmotic formulas, and their use may increase the frequency of diarrhea. Infant formulas that are hyperosmotic may affect preterm infants adversely during the early neonatal period, and they may produce or predispose neonates to necrotizing enterocolitis when the formulas delivered to the jejunum through a nasogastric tube. The body attempts to keep the osmoticity of the contents of the stomach and intestines at approximately the same level as that of the fluid surrounding them. As a fluid of lower osmoticity requires the transfer of less water to dilute it, it should be tolerated more than one of higher osmoticity.

As to tolerance, there is a great variation from one individual to another in sensitivity to the osmoticity of foods. The majority of patients receiving nutritional formulas, either orally or by tube, are able to tolerate feedings with a wide range of osmotivities when the formulas are administered slowly and when adequate additional fluids are given. However, certain patients are more likely to develop symptoms of intolerance when receiving fluids of high osmoticity. These include debilitated patients, patients with GI disorders, pre- and post-operative patients, gastrostomy and jejunostomy-fed patients, and patients whose GI tracts have not been challenged for an extended period of time. Thus, osmoticity always should be considered in the selection of the formula for each individual patient.

With all products additional fluid intake may be indicated for individuals with certain clinical conditions. Frequent feedings of small volume or a continual instillation (pumped) may be of benefit initially in establishing tolerance to a formula. For other than iso-osmotic formulas, feedings of reduced concentration (osmolality less than 400 mOsmol/kg) also may be helpful initially if tolerance problems arise in sensitive individuals. Concentration and size of feeding then can be increased gradually to normal as tolerance is established.

A common disturbance of intake encountered in elderly individuals relates to excess solid intake rather than to reduced water intake. For example, an elderly victim of a cerebral vascular accident who is being fed by nasogastric tube may be given a formula whose solute load requires a greatly increased water intake. Thus, tube feeding containing 120g of protein and 10g of salt will result in the excretion of more than 1000 mOsmol of solute. This requires the obligatory excretion of a volume of urine between 1200 and 1500 mL when the kidneys are capable of normal concentration ability. Because elderly individuals often have significant impairment in renal function, water loss as urine may exceed 2000 to 2500 mL per day. Such an individual would require three to four liters of water per day simply to meet the increased demand created by this high solute intake. Failure of the physician to provide such a patient with the

Table 16-1.

	mOsmol/kg
Whole milk	295
Tomato juice	595
Orange juice	935
Ice cream	1150
Parenteral nutrition	1400

increased water intake needed will result in a progressive water deficit that rapidly may become critical. The importance of knowing the complete composition of the tube feeding formulas used for incapacitated patients cannot be overemphasized.

OSMOLALITY DETERMINATION

The need for experimental determination of osmolality has been established. In regard to this there are four properties of solutions that depend only on the number of *particles* in the solution. They are *osmotic-pressure elevation*, *boiling-point elevation*, *vapor-pressure depression*, and *freezing-point depression*. These are called *colligative properties*, and if one of them is known, the others can be calculated from its value. Osmotic-pressure elevation is the most difficult to measure satisfactorily. The boiling-point elevation may be determined, but the values are rather sensitive to changes in barometric pressure. Also, for an aqueous solution the molal boiling point elevation is considerably less than the freezing-point depression. Thus, it is less accurate than the freezing-point method. Determinations of vapor-pressure lowering are quite easy, rapid, and convenient. A vapor pressure osmometer with a precision of 2 mOsmol/kg is reported by Dickerson *et al.*²³ Another commonly used method is that of freezing-point depression, which can be determined quite readily with a fair degree of accuracy (see *Freezing-Point Depression* in Chapter 13). It should be noted that the data in Appendix A can be converted readily to vapor-pressure lowering, if desired.

The results of investigations by Lund *et al.*,²⁵ indicate that the freezing point of normal, healthy human blood is -0.52°C . In as much as water is the medium in which the various constituents of blood are either suspended or dissolved in this method, it is assumed that *any aqueous solution* freezing at -0.52°C is *isotonic with blood*. Now, it is rare that a simple aqueous solution of the therapeutic agent to be injected parenterally has a freezing point of -0.52°C , and to obtain this freezing point, it is necessary either to add some other therapeutically inactive solute if the solution is hypotonic (freezing point above -0.52°C) or to dilute the solution if it is hypertonic (freezing point below -0.52°C). The usual practice is to add either sodium chloride or glucose to adjust hypotonic parenteral solutions to isotonicity. Certain solutes, including ammonium chloride, boric acid, urea, glycerin, and propylene glycol, cause hemolysis even when they are present in a concentration that is iso-osmotic, and such solutions obviously are not isotonic and do not act as a solvent, semi-permeable membrane. (See Appendix A.)

In a similar manner, solutions intended for ophthalmic use may be adjusted to have a freezing point identical to that of lacrimal fluid, namely -0.52°C . Ophthalmic solutions with higher freezing points usually are made isotonic by the addition of boric acid or sodium chloride.

In laboratories where the necessary equipment is available, the method usually followed for adjusting hypotonic solutions is to determine the freezing-point depression produced by the ingredients of a given prescription or formula, and then to add a quantity of a suitable inert solute calculated to lower the freezing point to -0.52°C , whether the solution is for parenteral injection or ophthalmic application. A final determination of the freezing-point depression may be made to verify the accuracy of the calculation. If the solution is hypertonic, it must be diluted if an isotonic solution is to be prepared, but it must be remembered that some solutions cannot be diluted without impairing their therapeutic activity. For example, solutions to be used for treating varicose veins require a high concentration of the active ingredient (solute) to make the solution effective. Dilution to isotonic concentration is not indicated in such cases.

FREEZING-POINT CALCULATIONS

As explained in the preceding section, freezing-point data often may be employed in solving problems of isotonicity adjustment.

Obviously, the utility of such data is limited to those solutions where the solute does not penetrate the membrane of the tissue (e.g., red blood cells) with which it is in contact. In such cases, Appendix A, which gives the freezing-point depression of solutions of different concentrations of various substances, provides information essential for solving the problem.

For most substances listed in the table in Appendix B, the concentration of an isotonic solution (one that has a freezing point of -0.52°C) is given. If this is not listed in the table, it may be determined with sufficient accuracy by simple proportion using, as the basis for calculation, the figure that most nearly produces an isotonic solution. Actually, the depression of the freezing point of a solution of an electrolyte is not absolutely proportional to the concentration but varies according to dilution; for example, a solution containing 1 g of procaine hydrochloride in 100 mL has a freezing-point depression of 0.12°C , whereas a solution containing 3 g of the same salt in 100 mL has a freezing-point depression of 0.33°C , *not* 0.36°C ($3 \times 0.12^{\circ}\text{C}$). Because typical adjustments to isotonicity need not be absolutely exact, approximation are made. Nevertheless, adjustments to isotonicity should be as exact as practicable or confirmed experimentally.

EFFECT OF SOLVENTS

Besides water, certain other solvents frequently are employed in nose drops, eardrops, and other preparations to be used in various parts of the body. Liquids such as glycerin, propylene glycol, or alcohol may compose part of the solvent. In solving isotonicity adjustment problems for such solutions, it should be kept in mind that these solvent components contribute to the freezing-point depression, but they may or may not have an effect on the tone of the tissue to which they are applied; thus, an *iso-osmotic* solution may not be *isotonic*. In such cases, it is apparent that the utility of the methods described above—or for that matter, of any other method of evaluating *tonicity*—is questionable.

TONICITY TESTING BY OBSERVING ERYTHROCYTE CHANGES

Observation of the behavior of human erythrocytes when suspended in a solution is the ultimate and direct procedure for determining whether the solution is isotonic, hypotonic, or hypertonic. If hemolysis or marked change in the appearance of the erythrocytes occurs, the solution is not isotonic with the cells. If the cells retain their normal characteristics, the solution is isotonic.

Hemolysis may occur when the osmotic pressure of the fluid in the erythrocytes is greater than that of the solution in which the cells are suspended, but the specific chemical reactivity of the solute in the solution often is far more important in producing hemolysis than is the osmotic effect. There is no certain evidence that any single mechanism of action causes hemolysis. The process appears to involve such factors as pH, lipid solubility, molecular and ionic sizes of solute particles, and possibly inhibition of cholinesterase in cell membranes and denaturing action on plasma membrane protein.

Some investigators test the tonicity of injectable solutions by observing variations of red blood cell volume produced by these solutions. This method appears to be more sensitive to small differences in tonicity than those based on observation of a hemolytic effect. Much useful information concerning the effect of various solutes on erythrocytes has been obtained by this procedure.

METHODS OF ADJUSTING TONICITY

There are several methods for adjusting the tonicity of an aqueous solution, provided, of course, that the solution is hypotonic when the drug and additives are dissolved. The most prominent of these methods are the freezing-point depression method, the sodium chloride equivalent method, and the isotonic solution

V-value method. The first two of these methods can be used with a three-step problem-solving process, based on sodium chloride.

1. Identify a reference solution and the associated tonicity parameter.
2. Determine the contribution of the drug(s) and additive(s) to the total tonicity.
3. Determine the amount of sodium chloride needed by subtracting the contribution of the actual solution from the reference solution.

The result of the third step also indicates whether the actual solution is hypotonic, isotonic, or hypertonic. If the actual solution contributes less to the total tonicity than the reference solution, then the actual solution is hypotonic. If, however, the actual solution contributes a greater amount to tonicity than the reference solution, the actual solution is hypertonic and can be adjusted to isotonicity only by dilution. This may not be possible on therapeutic grounds.

The amount of sodium chloride resulting in the third step also can be converted into an amount of other materials, such as glucose, to render the actual solution isotonic.

FREEZING-POINT-DEPRESSION METHOD

The freezing-point method makes use of a *D value* (found in Appendix A) which has the units of degree centigrade/(x% drug). For example, in Appendix A, dexamethasone sodium phosphate has *D* values of 0.050°/(0.5% drug), 0.180°/(2.0% drug), 0.52°/(6.75% drug), etc. It is apparent that the *D* value is nearly proportional to concentration. If a *D* value is needed for a concentration of drug not listed in Appendix A, a *D* value can be calculated from the appendix by direct proportion, using a *D* value closest to the concentration of drug in the actual solution.

The reference solution for the freezing-point-depression method is 0.9% sodium chloride, which has a freezing-point depression of $\Delta T_f = 0.52^\circ$. Using the three steps described above, the dexamethasone sodium phosphate solution in Example 1 can be rendered isotonic as follows:

Example 1

Dexamethasone Sodium Phosphate	0.1%
Purified Water qs	30 mL

Mft Isotonic Solution

Step 1—Reference solution: 0.9% sodium chloride.

$$\Delta T_f = 0.52^\circ$$

$D = 0.050^\circ/0.5\%$ (dexamethasone sodium phosphate)

Step 2—Contribution of drug.

$$\frac{0.050^\circ}{0.5\% \text{ drug}} \times 0.1\% \text{ drug} = 0.010^\circ \quad (11)$$

Step 3—Reference solution – Actual solution.

$$0.52^\circ - 0.01^\circ = 0.51^\circ$$

Sodium chloride needed.

$$\frac{0.9\% \text{ NaCl}}{0.52^\circ} \times 30 \text{ mL} = 0.265 \text{ g NaCl} \quad (12)$$

$$\frac{0.883 \text{ g NaCl}}{100 \text{ mL}} \times 30 \text{ mL} = 0.265 \text{ g NaCl} \quad (13)$$

The above solution could be made isotonic with any appropriate material other than sodium chloride by using the *D* value for that material. For example, to make the solution isotonic with glucose with a *D* value, $D = 0.091^\circ/1\%$;

$$\frac{1\% \text{ Dextrose}}{0.091^\circ} \times 0.51^\circ = 5.60\% \text{ Dextrose} \quad (14)$$

$$\frac{5.60 \text{ g Dextrose}}{100 \text{ mL}} \times 30 \text{ mL} = 1.68 \text{ g Dextrose} \quad (15)$$

Example 2

Naphazoline HCl (N.HCl)	0.02%
Zinc Sulfate	0.25%
Purified Water qs	30 mL

Mft Isotonic solution

Step 1—Reference solution: 0.9% sodium chloride.

$$\Delta T_f = 0.52^\circ$$

$D = 0.14^\circ/1\%$ (naphazoline HCl)

$D = 0.086^\circ/1\%$ (zinc sulfate)

Step 2—Contribution of drugs.

$$\frac{0.14^\circ}{1\% \text{ N.HCl}} \times 0.02\% \text{ N.HCl} = 0.003^\circ \quad (16)$$

$$\frac{0.086^\circ}{1\% \text{ ZnSO}_4} \times 0.25\% \text{ ZnSO}_4 = 0.022^\circ \quad (17)$$

Step 3—Reference solution – actual solution.

$$0.52^\circ - 0.025^\circ = 0.495^\circ$$

Sodium chloride needed.

$$\frac{0.9\% \text{ NaCl}}{0.52^\circ} \times 0.495^\circ = 0.857\% \text{ NaCl} \quad (18)$$

$$\frac{0.857 \text{ g NaCl}}{100 \text{ mL}} \times 30 \text{ mL} = 0.257 \text{ g NaCl} \quad (19)$$

The above solution could be made isotonic with any appropriate material, other than sodium chloride, by using the *D* value for that material.

For example, to make the solution isotonic with glucose with a *D* value, $D = 0.091^\circ/1\%$;

$$\frac{1\% \text{ Dextrose}}{0.091^\circ} \times 0.495^\circ = 5.44\% \text{ Dextrose} \quad (20)$$

$$\frac{5.44 \text{ g Dextrose}}{100 \text{ mL}} \times 30 \text{ mL} = 1.63 \text{ g Dextrose} \quad (21)$$

Sodium Chloride Equivalent Method

A sodium chloride equivalent, *E value*, is defined as the weight of sodium chloride that will produce the same osmotic effect as 1 g of the drug. For example, in Appendix A, dexamethasone sodium phosphate has an *E* value of 0.18 g NaCl/g drug at 0.5% drug concentration, 0.17 g NaCl/g drug at 1% drug concentration and a value of 0.16 g NaCl/g drug at 2% drug. This slight variation in the sodium chloride equivalent with concentration is due to changes in interionic attraction at different concentration of drug; the *E* value is not directly proportional to concentration, as was the freezing-point-depression.

The reference solution for the sodium chloride equivalent method is 0.9% sodium chloride, as it was for the freezing-point-depression method.

The dexamethasone sodium phosphate solution in Example 1 can be rendered isotonic, using the sodium chloride equivalent method as follows:

Example 1

Dexamethasone Sodium Phosphate	0.1%
Purified Water qs	30 mL

Mft Isotonic Solution

Step 1—Reference solution: 0.9% sodium chloride.

$$\frac{0.9 \text{ g NaCl}}{100 \text{ mL}} \times 30 \text{ mL} = 0.270 \text{ g NaCl} \quad (22)$$

$$E = 0.18 \text{ g NaCl/g drug}$$

Step 2—Contribution of drug.

$$\frac{0.18 \text{ g NaCl}}{1 \text{ g drug}} \times \frac{0.1 \text{ g drug}}{100 \text{ mL}} \times 30 \text{ mL} = 0.0054 \text{ g NaCl} \quad (23)$$

Step 3—Reference solution – Actual solution.

$$0.270 \text{ g NaCl} - 0.0054 \text{ g NaCl} = 0.265 \text{ g NaCl}$$

The above solution can be made isotonic with a material other than sodium chloride, such as glucose, by using the E value of that material. For example, to make the solution isotonic with glucose, $E = 0.16 \text{ g NaCl/g glucose}$, the amount of sodium chloride needed in Step 3, can be converted to glucose as follows:

$$\frac{1 \text{ g Dextrose}}{0.16 \text{ g NaCl}} \times 0.265 \text{ g NaCl} = 1.66 \text{ g Dextrose} \quad (24)$$

Example 2

Naphazoline HCl (N.HCl)	0.02%
Zinc Sulfate	0.25%
Purified Water qs	30 mL

Mft Isotonic Solution

Step 1—Reference solution: 0.9% sodium chloride.

$$\frac{0.9 \text{ g NaCl}}{100 \text{ mL}} \times 30 \text{ mL} = 0.270 \text{ g NaCl} \quad (25)$$

$$E = 0.27 \text{ g NaCl/g N.HCl}$$

$$E = 0.15 \text{ g NaCl/g ZnSO}_4$$

Step 2—Contribution of drugs.

$$\frac{0.27 \text{ g NaCl}}{1 \text{ g N.HCl}} \times \frac{0.02 \text{ g N.HCl}}{100 \text{ mL}} \times 30 \text{ mL} = 0.002 \text{ g NaCl} \quad (26)$$

$$\frac{0.15 \text{ g NaCl}}{1 \text{ g ZnSO}_4} \times \frac{0.25 \text{ g ZnSO}_4}{100 \text{ mL}} \times 30 \text{ mL} = 0.011 \text{ g NaCl} \quad (27)$$

$$0.002 \text{ g NaCl} + 0.011 \text{ g NaCl} = 0.013 \text{ g NaCl}$$

Step 3—Reference solution actual solution.

$$0.270 \text{ g NaCl} - 0.013 \text{ g NaCl} = 0.257 \text{ g NaCl}$$

$$\frac{1 \text{ g Dextrose}}{0.16 \text{ g NaCl}} \times 0.257 \text{ g NaCl} = 1.61 \text{ g Dextrose} \quad (28)$$

The above solution can be made isotonic with a material other than sodium chloride, such as glucose, by using the E value of that material. For example, to make the solution isotonic with glucose, $E = 0.16 \text{ g NaCl/g glucose}$, the amount of sodium chloride needed in Step 3 can be converted to glucose as follows:

ISOTONIC SOLUTION V VALUES

The *V value* of a drug is the volume of water to be added to a specified weight of drug (0.3 g or 1.0 g, depending on the table used) to prepare an isotonic solution. Appendix B gives such values for some commonly used drugs. The reason for providing data for 0.3 g of drug is for convenience in preparing 30 mL (approximately 1 fluid ounce) of solution, a commonly prescribed volume. The basic principle underlying the use of V values is to prepare an isotonic solution of the prescribed drug and then dilute this solution to final volume with a suitable isotonic vehicle.

The two solutions in the previous examples can be prepared as follows using the V-value method:

Example 1

Dexamethasone Sodium Phosphate	0.1%
Purified Water qs	30 mL

Mft Isotonic Solution

Step 1—The V value for dexamethasone sodium phosphate can be calculated from the sodium chloride equivalent, *E*, as outlined in the footnote in Appendix B.

$$\frac{100 \text{ mL Soln}}{0.9 \text{ g NaCl}} \times \frac{0.17 \text{ g NaCl}}{1 \text{ g drug}} \times 0.3 \text{ g drug} = 5.67 \text{ mL Soln} \quad (29)$$

for a dilute solution:

$$5.67 \text{ mL Soln} \cong 5.67 \text{ mL H}_2\text{O} \text{ there } V \\ = (5.67 \text{ mL H}_2\text{O}) / (0.3 \text{ g drug})$$

Step 2—Amount of drug needed.

$$\frac{0.1 \text{ g drug}}{100 \text{ mL}} \times 30 \text{ mL} = 0.030 \text{ g drug} \quad (30)$$

Volume of water needed to prepare an isotonic solution.

$$\frac{5.67 \text{ mL H}_2\text{O}}{0.3 \text{ g drug}} \times 0.030 \text{ g} = 0.57 \text{ mL H}_2\text{O} \quad (31)$$

Step 3—To prepare the solution, dissolve 0.030 g of drug in 0.57 mL water, and qs to volume with a suitable isotonic vehicle, such as 0.9% sodium chloride solution, 5.51% glucose, or an isotonic phosphate buffer.

Example 2

Naphazoline HCl (N.HCl)	0.02%
Zinc Sulfate	0.25%
Purified Water qs	30 mL

Mft Isotonic Solution

Step 1—The V value for naphazoline HCl can be calculated from the sodium chloride equivalent, *E*, as outlined in the footnote in Appendix B; the V value for zinc sulfate is taken directly from Appendix B.

$$\frac{100 \text{ mL Soln}}{0.9 \text{ g NaCl}} \times \frac{0.27 \text{ g NaCl}}{1 \text{ g N.HCl}} \times 0.3 \text{ g N.HCl} = 9.00 \text{ mL Soln} \quad (32)$$

for a dilute solution:

$$9.00 \text{ mL Soln} \cong 9.00 \text{ mL H}_2\text{O} \text{ there } V \\ = (9.00 \text{ mL H}_2\text{O}) / (0.3 \text{ g N.HCl}) \\ V = 5.00 \text{ mL H}_2\text{O} / 0.3 \text{ g ZnSO}_4$$

Step 2—Amount of drugs needed.

$$\frac{0.02 \text{ g N.HCl}}{100 \text{ mL}} \times 30 \text{ mL} = 0.006 \text{ g N.HCl}$$

$$\frac{0.25 \text{ g ZnSO}_4}{100 \text{ mL}} \times 30 \text{ mL} = 0.075 \text{ g ZnSO}_4$$

Volume of water needed to prepare an isotonic solution:

$$\frac{9.00 \text{ mL H}_2\text{O}}{0.3 \text{ g N.HCl}} \times 0.006 \text{ g drug} = 0.18 \text{ mL H}_2\text{O}$$

$$\frac{5.00 \text{ mL H}_2\text{O}}{0.3 \text{ g ZnSO}_4} \times 0.075 \text{ g ZnSO}_4 = 1.25 \text{ mL H}_2\text{O}$$

Step 3—To prepare the solution, dissolve 0.006 g of naphazoline HCl and 0.075 g zinc sulfate in 1.43 mL water, and qs to volume with a suitable isotonic vehicle, such as 0.9% sodium chloride solution, 5.51% glucose, or an isotonic phosphate buffer.

REFERENCES

- Kaminski MV. Enteral hyperalimentation. *Surg Gynecol Obstet* 1976; 143: 12–16.
- Terris M. Crean P. Fluid and electrolyte balance in children. *Anaesth Intensive Care Med* 2012; 13(1); 15–19.
- Theeuwes F. Elementary osmotic pump. *J Pharm Sci* 1975; 64: 1987–1991.

4. Wermeling DP *et al.* Osmolality of small-volume intravenous admixtures. *Am J Hosp Pharm* 1985; 42: 1739–1744.
5. Crane VS. Significance of osmoticity in antibiotic small-volume parenterals. *Drug Intell Clin Pharm* 1987; 21: 830–834.
6. Santeiro ML *et al.* Osmolality of small-volume i.v. admixtures for pediatric patients. *Am J Hosp Pharm* 1990; 47: 1359–1364.
7. McDuffee L. *IL Council Hosp Pharm Drug Inf Newsl* 1978; 8(Oct–Nov).
8. Zenk K, Huxtable RF. *Hosp Pharm* 1978; 13: 577–586.
9. Streng WH *et al.* Relationship between osmolality and osmolarity. *J Pharm Sci* 1978; 67: 384–386.
10. Murty BSR *et al.* Compliance with USP osmolarity labeling requirements. *Am J Hosp Pharm* 1976; 33: 546–551.
11. European Directorate to the Quality of Medicines, Council of Europe. *European Pharmacopoeia*, 4th edn. 2002.
12. Fontaine O. Multicentre evaluation of reduced osmolality and rehydration salts solution. *Lancet* 1995; 345: 282–285.
13. Luedtke SA. Pediatric and neonatal therapy In: Helms RA, Quan DS, eds. *Textbook of therapeutics: Drug and disease management*, 8th edn. Philadelphia: Lippincott, Williams and Wilkins, 2006; . 324–340.
14. Nilsson M *et al.* Effects of hyperosmotic stress on cultured airway epithelial cells. *Cell Tissue Res* 2007; 330(2): 257–269.
15. Cole TS, Liedtke CM. Activation of NKCC1 by hyperosmotic stress in human tracheal epithelial cells involves PKC- δ and ERK. *Biochim Biophys Acta* (Molecular Cell Res). 2002; 1589(1): 77–88.
16. Bye BTP, Elkins MR. Other munoactive agents for cystic fibrosis. *Ped Resp Rev* 2007; 8(1): 30–39.
17. Turunen MJ *et al.* Hyperosmolaric contrast agents in cartilage tomography may expose cartilage to overload induced cell death. *J Biomech* 2012; 45: 497–503.
18. Payne-James JJ *et al.* First choice for total parenteral nutrition: the peripheral route. *J Parenter Enter Nutr* 1993; 17: 468–478.
19. Miller SJ. Peripheral parenteral nutrition: theory and practice. *Hosp Pharm* 1991; 26: 796.
20. Andrassy RJ *et al.* Continual catheter administration of an elemental diet in infants and children. *Surgery* 1977; 82: 205–210.
21. Dobbie RP, Hoffmeister JA. Continuous pump-tube enteric hyperalimentation. *Surg Gynecol Obstet* 1976; 143: 273–276.
22. Osmolality. Minneapolis: Doyle Pharmaceutical, 1978.
23. Dickerson RNMelnik G. Osmolality of oral drug solutions and suspensions. *Am J Hosp Pharm* 1988; 45: 832–834.
24. Holtz L *et al.* Compatability of medications with enteral feedings. *J Parenter Enter Nutr* 1987; 11: 183–186.
25. Lund CG *et al.* *The Preparation of Solutions Isoosmotic with Blood, Tears, and Tissue*. Copenhagen: Danish Pharmacopoeial Commission, Einar Munksgaard, 1947.
26. Hammarlund ER *et al.* Additional sodium chloride equivalents and freezing point depressions for various medicinal solutions. *J Pharm Sci* 1965; 54: 160–162.
27. Hammarlund ER, Pedersen-Bjergaard K. A simplified graphic method for the preparation of isotonic solutions. *J APhA Sci Ed* 1958; 47: 107–114.
28. Hammarlund ER, Pedersen-Bjergaard K. Hemolysis of erythrocytes in various iso-osmotic solutions. *J Pharm Sci* 1961; 50: 24–30.
29. Hammarlund ER, Van Pevenage GL. Sodium chloride equivalents, cryoscopic properties, and hemolytic effects of certain medicinals in aqueous solution. *J Pharm Sci* 1966; 55: 1448–1445.
30. Sapp C *et al.* Sodium chloride equivalents, cryoscopic properties, and hemolytic effects of certain medicinal in aqueous solution. III: Supplemental values. *J Pharm Sci* 1975; 64: 1884–1886.
31. British Pharmaceutical Codex. London: Pharmaceutical Press, 1973.
32. Fassett WE *et al.* Sodium chloride equivalents, cryoscopic properties, and hemolytic effects of certain medicinals in aqueous solution. II. Supplemental values. *J Pharm Sci* 1969; 58: 1540–1542.
33. Kagan DG, Kinsey VE. The cornea. I. The transfer of water and sodium chloride by osmosis and diffusion through the ex-cised cornea. *Arch Ophthalmol* 1942; 27: 696.

BIBLIOGRAPHY

- Alberty RA, Daniels F. *Physical Chemistry*, 7th edn. New York: Wiley, 1987.
- Cowan G, Scheetz W, eds. *Intravenous Hyperalimentation*. Philadelphia: Lea & Febiger, 1972.
- Garb S. *Laboratory Tests in Common Use*, 6th edn. New York: Springer, 1976.
- Hall WE. *Am J Pharm Ed* 1970; 34: 204.
- Harvey AM *et al.* *The Principles and Practice of Medicine*, 18th edn. New York: Appleton Century Crofts, 1972.
- Martin AN *et al.* *Physical Pharmacy*, 4th edn. Philadelphia: Lea & Febiger, 1993.
- Plumer AL. *Principles and Practice of Intravenous Therapy*, 4th edn. Boston: Little, Brown, 1987.
- Ravel R. *Clinical Laboratory Medicine*, 5th edn. St Louis: Mosby, 1988.
- Shizgal HM. Parenteral and enteral nutrition. *Ann Rev Med* 1991; 42: 549.
- Tilkian SM, Conover MH. *Clinical Implications of Laboratory Tests*, 4th edn. St Louis: Mosby, 1987.
- Turco S, King RE. *Sterile Dosage Forms*, 3rd edn. Philadelphia: Lea & Febiger, 1987.
- Wallach J. *Interpretation of Diagnostic Tests*, 4th edn. Boston: Little, Brown, 1986.

APPENDIX A: Sodium Chloride Equivalents, Freezing-Point Depressions, and Hemolytic Effects of Certain Medicinals in Aqueous Solution

	Iso-Osmotic Concentration ^a											
	0.5 %			1 %			2 %			3 %		
	E	D		E	D		E	D		E	D	
Acetrizolate methylglucamine	0.09			0.08			0.08			0.08	12.12	0.07
Acetrizolate sodium	0.10	0.027		0.10	0.055		0.10	0.109		0.163	0.273	9.64
Acetylsteyne	0.20	0.055		0.20	0.113		0.20	0.227		0.341		0.52
Adrenaline HCl												0.52
Alphaprodine HCl	0.19	0.053		0.19	0.105		0.18	0.212		0.315		0.52
Alum (potassium)				0.18							0.15	0.14
Amantadine HCl	0.31	0.090		0.310	0.180		0.31	0.354				24*
Aminoacetic acid	0.42	0.119		0.41	0.235		0.41	0.47				2.95
Aminohippuric acid	0.13	0.035		0.13	0.075							2.20
Aminophylline					0.098 ^e							0.41
Ammonium carbonate	0.70	0.202		0.70	0.405							0.52
Ammonium chloride				1.12								0.70
Ammonium lactate	0.33	0.093		0.33	0.185		0.33	0.37				1.12
Ammonium nitrate	0.69	0.200		0.69	0.400							0.33
Ammonium phosphate, dibasic	0.58	0.165		0.55	0.315							0.52
Ammonium sulfate	0.55	0.158		0.55	0.315							0.51
Amobarbital sodium				0.25	0.143 ^e							0.54
d-Amphetamine HCl							0.25					0.52
Amphetamine phosphate				0.34	0.20		0.27			0.47		0.26
Amphetamine sulfate				0.22	0.129 ^e		0.21			0.36		0.52
Amprotropine phosphate												0.21
Amylcaine HCl				0.22			0.19					5.90
Anileridine HCl	0.19	0.052		0.19	0.104		0.18	0.212		0.316	0.509	4.98
Antazoline phosphate												0.18
Antimony potassium tartrate				0.18			0.13					5.13
Antipyrine				0.17	0.10		0.14			0.24	0.40	6.05
Apomorphine HCl				0.14	0.080 ^e							0.13
Arginine glutamate	0.17	0.048		0.17	0.097		0.17	0.195		0.292	0.487	5.37
Ascorbic acid					0.105 ^e							5.05
Atropine methylbromide				0.14			0.13					7.03
Atropine methylnitrate												0.13
Atropine sulfate				0.13	0.075		0.11			0.19	0.32	6.52
Bacitracin				0.05	0.03		0.04			0.07	0.12	8.85

	0.5 %			1 %			2 %			3 %			5 %			Iso-Osmotic Concentration ^a				
	E	D		E	D		E	D		E	D		E	D		%	E	D	H	pH
Barbital sodium			0.30	0.171 ^e						0.29	0.50					3.12	0.29	0.52	0	9.8
Benzalkonium chloride			0.16							0.14			0.13							
Benzotropine mesylate	0.26	0.073	0.21	0.115		0.15	0.170			0.12	0.203		0.09	0.242						
Benzyl alcohol			0.17	0.09 ^e						0.15										
Bethanechol chloride	0.50	0.140	0.39	0.225		0.32	0.368			0.30	0.512					3.05	0.30		0	6.0
Bismuth potassium tartrate			0.09							0.06			0.05							
Bismuth sodium tartrate			0.13							0.12						8.91	0.10		0	6.1
Boric acid	0.50	0.288 ^e											0.11			1.9	0.47	0.52	100	4.6
Brompheniramine maleate	0.10	0.026	0.09	0.05		0.08	0.084													
Bupivacaine HCl	0.17	0.048	0.17	0.096		0.17	0.193			0.17	0.29		0.17	0.484		5.38	0.17	0.52	83	6.8
Butabarbital sodium	0.27	0.078	0.27	0.155		0.27	0.313			0.27	0.47					3.33	0.27	0.52	0	6.8
Butacaine sulfate			0.20	0.12						0.13	0.23		0.10	0.29						
Caffeine and sodium benzoate			0.26	0.15						0.23	0.40					3.92	0.23	0.52	0	7.0
Caffeine and sodium salicylate			0.12	0.12						0.17	0.295		0.16	0.46		5.77	0.16	0.52	0	6.8
Calcium aminosaliclylate																4.80			0	6.0
Calcium chloride			0.51	0.298 ^e												1.70	0.53	0.52	0	5.6
Calcium chloride (6 H ₂ O)			0.35	0.20												2.5	0.36	0.52	0	5.7
Calcium chloride, anhydrous			0.68	0.39												1.3	0.69	0.52	0	5.6
Calcium disodium edetate	0.21	0.061	0.21	0.120		0.21	0.240			0.20	0.357					4.50	0.20	0.52	0	6.1
Calcium gluconate			0.16	0.091 ^e						0.14	0.24									
Calcium lactate			0.23	0.13						0.12	0.36					4.5	0.20	0.52	0	6.7
Calcium lactobionate	0.08	0.022	0.08	0.043		0.08	0.085			0.07	0.126		0.07	0.197						
Calcium levulinate			0.27	0.16						0.25	0.43					3.58			0	7.2
Calcium pantothenate			0.129													5.50			0	7.4
Camphor			0.12 ^d																	
Capreomycin sulfate	0.04	0.011	0.04	0.02		0.04	0.042			0.04	0.063		0.04	0.106						
Carbachol				0.205 ^e												2.82			0	5.9
Carbenicillin sodium	0.20	0.059	0.20	0.118		0.20	0.236			0.20	0.355					4.40	0.20	0.52	0	6.6
Carboxymethylcellulose sodium	0.03	0.007	0.03	0.017		0.145														
Cephaloridine	0.09	0.023	0.07	0.041		0.06	0.074			0.06	0.106		0.05			4.10			100*	9.1
Chloramine-T																				
Chloramphenicol				0.06 ^d																
Chloramphenicol sodium succinate	0.14	0.038	0.14	0.078		0.14	0.154			0.13	0.230		0.13	0.382		6.83	0.13	0.52	partial	6.1

	0.5 %			1 %			2 %			3 %			5 %			Iso-Osmotic Concentration ^a					
	E	D		E	D		E	D		E	D		E	D		%	E	D		H	pH
Chlordiazepoxide HCl	0.24	0.068		0.22	0.125	0.19	0.220	0.18	0.315	0.17	0.487	5.50	0.16	0.52	66						2.7
Chlorobutanol (hydrated)		0.24		0.24	0.14																
Chloroprocaine HCl	0.20	0.054		0.20	0.108	0.18	0.21														
Chloroquine phosphate	0.14	0.039		0.14	0.082	0.14	0.162	0.14	0.242	0.13	0.379	7.15	0.13	0.52	0					4.3	
Chloroquine sulfate	0.10	0.028		0.09	0.050	0.08	0.090	0.07	0.127	0.07	0.195										
Chlorpheniramine maleate	0.17	0.048		0.15	0.085	0.14	0.165	0.13	0.22	0.09	0.265										
Chlortetracycline HCl	0.10	0.030		0.10	0.061	0.10	0.121														
Chlortetracycline sulfate				0.13	0.08			0.10	0.17												
Citric acid				0.18	0.10			0.17	0.295	0.16	0.46	5.52	0.16	0.52	100*					1.8	
Clindamycin phosphate	0.08	0.022		0.08	0.046	0.08	0.095	0.08	0.144	0.08	0.242	10.73	0.08	0.52	58*					6.8	
Cocaine HCl				0.16	0.090 ^e			0.15	0.26	0.14	0.40	6.33	0.14	0.52	47					4.4	
Codeine phosphate				0.14	0.080 ^e			0.13	0.23	0.13	0.38	7.29	0.12	0.52	0					4.4	
Colistimethate sodium	0.15	0.045		0.15	0.085	0.15	0.170	0.15	0.253	0.14	0.411	6.73	0.13	0.52	0					7.6	
Cupric sulfate				0.18	0.100 ^e			0.15		0.14		6.85	0.13		trace*					3.9	
Cyclizine HCl	0.20	0.060																			
Cyclophosphamide	0.10	0.031		0.10	0.061	0.10	0.125														
Cytarabine	0.11	0.034		0.11	0.066	0.11	0.134	0.11	0.198	0.11	0.317	8.92	0.10	0.52	0					8.0	
Deferoxamine mesylate	0.09	0.023		0.09	0.047	0.09	-0.093	0.09	0.142	0.09	0.241										
Demecarium bromide	0.14	0.038		0.12	0.069	0.10	0.108	0.08	0.139	0.07	0.192										
Dexamethasone sodium phosphate	0.18	0.050		0.17	0.095	0.16	0.18	0.15	0.260	0.14	0.410	6.75	0.13	0.52	0					8.9	
Dextroamphetamine HCl	0.34	0.097		0.34	0.196	0.34	0.392														
Dextroamphetamine phosphate				0.25	0.14			0.25	0.44			2.64	0.34	0.52						4.7	
Dextroamphetamine sulfate	0.24	0.069		0.23	0.134	0.22	0.259	0.22	0.380			4.16	0.22	0.52	0					5.9	
Dextrose				0.16	0.091 ^e			0.16	0.28	0.16	0.46	5.51	0.16	0.52	0					5.9	
Dextrose (anhydrous)				0.18	0.101 ^e			0.18	0.31			5.05	0.18	0.52	0					6.0	
Diatrizoate sodium	0.10	0.025		0.09	0.049	0.09	0.098	0.09	0.149	0.09	0.248	10.55	0.09	0.52	0					7.9	
Dibucaine HCl					0.074 ^e																
Dicloxacillin sodium (1 H ₂ O)	0.10	0.030		0.10	0.061	0.10	0.122	0.10	0.182												
Diethanolamine	0.31	0.089		0.31	0.177	0.31	0.358					2.90	0.31	0.52	100					11.3	
Dihydrostreptomycin sulfate				0.06	0.03			0.05	0.09	0.05	0.14	19.4	0.05	0.52	0					6.1	
Dimethylpyrindene maleate	0.13	0.039		0.12	0.07	0.11	0.12														
Dimethyl sulfoxide	0.42	0.122		0.42	0.245	0.42	0.480														
Diperodon HCl	0.15	0.045		0.14	0.079	0.13	0.141					2.16	0.42	0.52	100					7.6	

	0.5 %			1 %			2 %			3 %			5 %			Iso-Osmotic Concentration ^a				
	E	D	E	E	D	E	E	D	E	E	D	E	E	D	E	%	D	E	H	pH
Diphenhydramine HCl	0.16	0.045	0.16	0.09	0.161 ^e	0.16	0.16	0.180								5.70			88*	5.5
Diphenidol HCl	0.12	0.035	0.12	0.070		0.12	0.12	0.140	0.12	0.210										
Doxapram HCl	0.12	0.035	0.12	0.072		0.12	0.12	0.134	0.11	0.186	0.09	0.264								
Doxycycline hyclate	0.10	0.025	0.10	0.052		0.09	0.104	0.09	0.155	0.08	0.245									
Dyphylline	0.16	0.045	0.16	0.090		0.16	0.179													
Echothiophate iodide	0.24	0.070	0.23	0.132		0.22	0.248	0.21	0.360	0.21	0.472	0.20	0.52	0	4.7	4.44	0.20	0.52	0	4.7
Edetate disodium	0.29	0.079	0.29	0.158		0.28	0.316	0.27	0.462	0.16	0.52	0.27	0.52	0	8.0	3.31	0.27	0.52	0	8.0
Edetate trisodium monohydrate																				
Emetine HCl				0.058 ^e				0.17	0.29											
Ephedrine HCl	0.30		0.30	0.165 ^e		0.28										3.2	0.28		96	5.9
Ephedrine sulfate	0.23		0.23	0.13		0.20		0.35	0.20	0.52	0	5.7				4.54	0.20	0.52	0	5.7
Epinephrine bitartrate	0.18		0.18	0.104		0.16		0.28	0.16	0.462	0.16	0.52	100*		3.4	5.7	0.16	0.52	100*	3.4
Epinephrine hydrochloride	0.29		0.29	0.16 ^b		0.26										3.47	0.26			
Ergonovine maleate				0.089 ^e																
Erythromycin lactobionate	0.08	0.020	0.07	0.040		0.078	0.07	0.115	0.06	0.187										
Ethyl alcohol																				
Ethylenediamine				0.253 ^e												1.39			100	6.0
Ethylmorphine HCl	0.16		0.16	0.088 ^e		0.15	0.26	0.15	0.43	0.15	0.52	38	4.7			2.08	0.15	0.52	100*	11.4
Eucatropine HCl				0.11 ^d												6.18	0.15	0.52	38	4.7
Ferric ammonium citrate (green)																6.83			0	5.2
Floxuridine	0.14	0.040	0.13	0.076		0.147	0.12	0.213	0.12	0.335	0.12	0.52	3*	4.5		8.47	0.12	0.52	3*	4.5
Fluorescein sodium	0.31		0.31	0.181 ^e		0.27		0.47	0.27	0.52	0	8.7				3.34	0.27	0.52	0	8.7
Fluphenazine 2-HCl	0.14	0.041	0.14	0.082		0.145	0.09	0.155												
d-Fructose																5.05			0*	5.9
Furtrethonium iodide	0.24	0.070	0.24	0.133		0.250	0.21	0.360	0.21	0.456	0.16	0.52	0	5.4		4.44	0.20	0.52	0	5.4
Galactose																4.92			0	5.9
Gentamicin sulfate	0.05	0.015	0.05	0.030		0.060	0.05	0.093	0.05	0.153										
D-Glucuronic acid																5.02			48*	1.6
Glycerin																2.6			100	5.9
Glycopyrrolate	0.15	0.042	0.15	0.084		0.166	0.14	0.242	0.13	0.381	0.12	0.52	92*	4.0		7.22	0.12	0.52	92*	4.0
Gold sodium thiomalate	0.10	0.032	0.10	0.061		0.111	0.09	0.159	0.09	0.250										
Hetacillin potassium	0.17	0.048	0.17	0.095		0.190	0.17	0.284	0.17	0.474	0.17	0.52	0	6.3		5.50	0.17	0.52	0	6.3
Hexafluorenum bromide	0.12	0.033	0.11	0.065																
Hexamethonium tartrate	0.16	0.045	0.16	0.089		0.181	0.16	0.271	0.16	0.456	0.16	0.52				5.68	0.16	0.52		
Hexamethylene sodium acetaminosalicylate	0.18	0.049	0.18	0.099		0.199	0.17	0.297	0.17	0.485	0.16	0.52	0*	4.0		5.48	0.16	0.52	0*	4.0

	0.5 %			1 %			2 %			3 %			5 %			Iso-Osmotic Concentration ^a				
	E	D	E	E	D	E	E	D	E	E	D	E	E	D	E	%	D	E	H	pH
Hexobarbital sodium						0.15 ^c										4.30			100	4.8
Hexyleaine HCl																			79*	3.7
Histamine 2HCl	0.40	0.115	0.40	0.40	0.40	0.233		0.466								2.24	0.52	0.40		
Histamine phosphate						0.149 ^c										4.10	4.6	0		
Histidine HCl																3.45			40	3.9
Holocaine HCl						0.20														
Homatropine						0.17														
Hydrobromide																			92	5.0
Homatropine						0.19														
methyllbromide						0.11														
4-Homosulfanilamide HCl																				
Hyaluronidase						0.007		0.013	0.01	0.02	0.01	0.01	0.01	0.033		3.69			0	4.9
Hydromorphone HCl	0.01	0.004	0.01	0.01	0.01															
Hydroxyamphetamine HBr						0.15 ^d										6.39			64	5.6
8-Hydroxyquinoline																3.71			92	5.0
sulfate																9.75			59*	2.5
Hydroxystilbamidine																				
isethionate	0.20	0.06	0.16	0.12	0.137	0.090	0.12	0.137	0.10	0.17	0.10	0.07	0.07	0.216						
Hyoseyamine																				
hydrobromide																6.53			68	5.9
Imipramine HCl	0.20	0.058	0.20	0.20	0.13	0.110	0.13	0.143												
Indigotindisulfonate	0.30	0.085	0.30	0.30	0.172	0.172														
sodium																				
Intracaine HCl						0.07 ^e										4.97			85	5.0
Iodophthalein sodium																9.58			100	9.4
Isometheptene mucate	0.18	0.048	0.18	0.18	0.18	0.095	0.18	0.196	0.18	0.302						4.95	0.52	0.18	0	6.2
Isoproterenol sulfate	0.14	0.039	0.14	0.14	0.156	0.078	0.14	0.156	0.14	0.234	0.14	0.14	0.389			6.65	0.52	0.14	trace	4.5
Kanamycin sulfate	0.08	0.021	0.07	0.07	0.083	0.041	0.07	0.083	0.07	0.125	0.07	0.07	0.210							
Lactic acid						0.239 ^e										2.30			100*	2.1
Lactose						0.040 ^e										9.75		0.09	0*	5.8
Levallorphan tartrate	0.13	0.036	0.13	0.13	0.073	0.073	0.13	0.143	0.12	0.210	0.12	0.12	0.329			9.40	0.52	0.10	59*	6.9
Levorphanol tartrate	0.12	0.033	0.12	0.12	0.067	0.067	0.12	0.136	0.12	0.203										
Lidocaine HCl						0.13 ^e										4.42			85	4.3
Lircomycin HCl	0.16	0.045	0.16	0.16	0.090	0.090	0.15	0.170	0.14	0.247	0.14	0.14	0.40			6.60	0.52	0.14	0	4.5
Lobeline HCl						0.090*														
Lyapolate sodium	0.10	0.025	0.09	0.09	0.051	0.051	0.09	0.103	0.09	0.157	0.09	0.09	0.263			9.96	0.52	0.09	0	6.5+
Magnesium chloride						0.45										2.02		0.45	0	6.3
Magnesium sulfate						0.094 ^e										6.3	0.52	0.14	0	6.2
Magnesium sulfate, anhydrous	0.34	0.093	0.32	0.30	0.345	0.184	0.30	0.345	0.29	0.495	0.29	0.15	0.43			3.18	0.52	0.28	0	7.0

	0.5 %			1 %			2 %			3 %			5 %			Iso-Osmotic Concentration ^a				
	E	D	E	E	D	E	E	D	E	E	D	E	E	D	E	%	D	H	H	pH
Mannitol						0.098 ^e														
Maphenide HCl		0.27	0.075	0.27	0.153	0.27	0.27	0.303	0.26	0.448	5.07				0*	6.2	0.52			
Menadiol sodium diphosphate															0.25	3.55	0	0	8.2	
Menadione sodium bisulfite					0.12 ^d											5.07	0			
Menthol																				
Meperidine HCl					0.125 ^e											4.80	98	5.0		
Mepivacaine HCl	0.21	0.060	0.21	0.116	0.20	0.230	0.20	0.342	0.342					0.20	0.52	4.60	45	4.5		
Merbromin				0.08 ^b																
Mercuric cyanide			0.15	0.06*		0.14				0.13										
Mersalyl																				
Mesoridazine besylate	0.10	0.024	0.07	0.04	0.05	0.058	0.04	0.071	0.03	0.087					0.17	5.17	0.52	59	3.8	
Metaraminol bitartrate	0.20	0.06	0.20	0.112	0.19	0.21	0.18	0.308	0.17	0.505						3.21	0	0	4.5	
Methacholine chloride				0.184 ^e												8.59	100*	5.0		
Methadone HCl				0.101 ^e												2.75	97	5.9		
Methamphetamine HCl				0.213 ^e																
Methdilazine HCl	0.12	0.035	0.10	0.056	0.08	0.08	0.06	0.093	0.04	0.112						3.68	100	8.4		
Methenamine			0.23				0.24								0.25					
Methiodal sodium	0.24	0.068	0.24	0.136	0.24	0.274	0.24	0.41							0.24	3.81	0.52	0	5.9	
Methitural sodium	0.26	0.074	0.25	0.142	0.24	0.275	0.23	0.407							0.23	3.85	0.52	78	9.8	
Methocarbamol	0.10	0.03	0.10	0.06																
Methotrimeprazine HCl	0.12	0.034	0.10	0.060	0.07	0.077	0.06	0.094	0.04	0.125										
Methoxyphenamine HCl	0.26	0.075	0.26	0.150	0.26	0.300	0.26	0.450							0.26	3.47	0.52	96	5.4	
p- Methylaminoethanol-phenol tartrate	0.18	0.048	0.17	0.095	0.16	0.19	0.16	0.282	0.16	0.453					0.16	5.83	0.52	0	6.2	
Methyldopate HCl	0.21	0.063	0.21	0.122	0.21	0.244	0.21	0.365							0.21	4.28	0.52	partial	3.0	
Methylergonovine maleate	0.10	0.028	0.10	0.056																
N-Methylglucamine	0.20	0.057	0.20	0.111	0.18	0.214	0.18	0.315	0.18	0.517					0.18	5.02	0.52	4	11.3	
Methylphenidate HCl	0.22	0.065	0.22	0.127	0.22	0.258	0.22	0.388							0.22	4.07	0.52	66	4.3	
Methylprednisolone Na succinate	0.10	0.025	0.09	0.051	0.09	0.102	0.08	0.143	0.07	0.20										
Minocycline HCl	0.10	0.030	0.10	0.058	0.09	0.107	0.08	0.146												
Monoethanolamine	0.53	0.154	0.53	0.306												1.70	0.53	100	11.4	
Morphine HCl			0.15	0.086 ^e			0.14													
Morphine sulfate			0.14	0.079 ^e			0.11		0.19	0.09	0.26									
Nalorphine HCl	0.24	0.07	0.21	0.121	0.18	0.210	0.17	0.288	0.15	0.434					0.14	6.36	0.52	63	4.1	
Naloxone HCl	0.14	0.042	0.14	0.083	0.14	0.158	0.13	0.230	0.13	0.367					0.11	8.07	0.52	35	5.2	
Naphazoline HCl			0.27	0.14 ^d			0.24								0.22	3.99		100	5.3	

	0.5 %			1 %			2 %			3 %			5 %			Iso-Osmotic Concentration ^a				
	E	D	E	E	D	E	E	D	E	D	E	D	E	D	E	%	D	H	H	pH
Neosarphenamine																2.32	17	7.80		
Neomycin sulfate			0.11	0.063 ^e					0.09	0.16	0.08	0.232								
Neostigmine bromide			0.22	0.127 ^e					0.19							4.98		0	4.6	
Neostigmine methylsulfate			0.20	0.115 ^e					0.18		0.17					5.22	0.17			
Nicotinamide			0.26	0.148 ^e					0.21	0.36						4.49	0.20	100	7.0	
Nicotinic acid			0.25	0.144 ^e																
Nikethamide				0.100 ^e												5.94		100	6.9	
Novobiocin sodium	0.12	0.033	0.10	0.057	0.07	0.073														
Oleandomycin phosphate	0.08	0.017	0.08	0.038	0.08	0.084			0.08	0.129	0.08	0.255				10.82	0.08	0	5.0	
Orphenadrine citrate	0.13	0.037	0.13	0.074	0.13	0.144			0.12	0.204	0.10	0.285				.67		trace*	2.3	
Oxophenarsine HCl																				
Oxymetazoline HCl	0.22	0.063	0.22	0.124	0.20	0.232			0.19	0.335						4.92	0.18	86	5.7	
Oxyquinoline sulfate	0.24	0.068	0.21	0.113	0.16	0.182			0.14	0.236	0.11	0.315								
d-Pantothenyl alcohol	0.20	0.053	0.18	0.100	0.17	0.193			0.17	0.283	0.16	0.468				5.60	0.16	92	6.8	
Papaverine HCl			0.10	0.061 ^e																
Paraldehyde	0.25	0.071	0.25	0.142	0.25	0.288			0.25	0.430						3.65	0.25	97	5.3	
Pargyline HCl	0.30	0.083	0.29	0.165	0.29	0.327			0.28	0.491						3.18	0.28	91	3.8	
Penicillin G, potassium			0.18	0.102 ^e					0.17	0.29	0.16	0.46				5.48	0.16	0	6.2	
Penicillin G, procaine				0.06 ^d																
Penicillin G, sodium			0.18	0.100 ^e					0.16	0.28	0.16	0.46				5.90		18	5.2	
Pentazocine lactate	0.15	0.042	0.15	0.085	0.15	0.169			0.15	0.253	0.15	0.42				4.07		0	9.9	
Pentobarbital sodium				0.145 ^e												5.95		55*	3.4	
Pentolinium tartrate				0.09 ^d																
Phenacaine HCl				0.09																
Pheniramine maleate				0.09 ^d																
Phenobarbital sodium			0.24	0.135 ^e					0.23	0.40										
Phenol	0.35	0.20														3.95	0.23	0	9.2	
Phentolamine mesylate	0.18	0.052	0.17	0.096	0.16	0.173			0.14	0.244	0.13	0.364				2.8	0.32	0*	5.6	
Phenylephrine HCl			0.32	0.184 ^e					0.30							8.23	0.11	83	3.5	
Phenylephrine tartrate																3.0	0.30	0	4.5	
Phenylethyl alcohol	0.25	0.070	0.25	0.141	0.25	0.283										5.90		58*	5.4	
Phenylpropanolamine HCl			0.38	0.219 ^e														95	5.3	
Physostigmine salicylate			0.16	0.090 ^e																
Physostigmine sulfate				0.074 ^e																
Pilocarpine HCl			0.24	0.138 ^e					0.22	0.38						4.08	0.22	89	4.0	
Pilocarpine nitrate			0.23	0.132 ^e					0.20	0.35						4.84	0.20	88	3.9	
Piperocaine HCl				0.12 ^d												5.22		65	5.7	
Polyethylene glycol 300	0.12	0.034	0.12	0.069	0.12	0.141			0.12	0.216	0.13	0.378				6.73	0.13	53	3.8	

	0.5 %			1 %			2 %			3 %			5 %			Iso-Osmotic Concentration ^a				
	E	D		E	D		E	D		E	D		E	D		%	E	D	H	pH
Polyethylene glycol 400	0.08	0.022		0.08	0.047		0.09	0.098		0.09	0.153		0.09	0.272		8.50	0.11	0.52	0	4.4
Polyethylene glycol 1500	0.06	0.015		0.06	0.036		0.07	0.078		0.07	0.120		0.07	0.215		10.00	0.09	0.52	4	4.1
Polyethylene glycol 1540	0.02	0.005		0.02	0.012		0.02	0.028		0.03	0.047		0.03	0.094						
Polyethylene glycol 4000	0.02	0.004		0.02	0.008		0.02	0.02		0.020	0.033		0.02	0.067						
Polymyxin B sulfate				0.09	0.052			0.06		0.06	0.10		0.04	0.12						
Polysorbate 80	0.02	0.005		0.02	0.010		0.02	0.02		0.020	0.032		0.02	0.055						
Polyvinyl alcohol (99% hydrol)	0.02	0.004		0.02	0.008		0.02	0.020		0.02	0.035		0.03	0.075						
Polyvinylpyrrolidone	0.01	0.003		0.010	0.006		0.01	0.01		0.01	0.017		0.01	0.035						
Potassium acetate	0.59	0.172		0.59	0.342											1.53	0.59	0.52	0	7.6
Potassium chlorate																1.88			0	6.9
Potassium chloride				0.76	0.439 ^e											1.19	0.76	0.52	0	5.9
Potassium iodide				0.34	0.196 ^e											2.59	0.34	0.52	0	7.0
Potassium nitrate				0.56	0.324 ^e											1.62	0.56	0	5.9	
Potassium phosphate				0.46	0.27											2.08	0.43	0.52	0	8.4
Potassium phosphate, monobasic				0.44	0.25											2.18	0.41	0.52	0	4.4
Potassium sulfate				0.44												2.11	0.43		0	6.6
Pralidoxime chloride	0.32	0.092		0.32	0.183		0.32	0.364								2.87	0.32	0.52	0	4.6
Prilocaine HCl	0.22	0.062		0.22	0.125		0.22	0.250		0.22	0.375					4.18	0.22	0.52	45	4.6
Procainamide HCl				0.22	0.13					0.19	0.33		0.17	0.49						
Procaine HCl				0.21	0.122 ^e					0.19	0.33		0.18			5.05	0.18	0.52	91	5.6
Prochlorperazine edisylate	0.08	0.020		0.06	0.033		0.05	0.048		0.03	0.056		0.02	0.065						
Promazine HCl	0.18	0.050		0.13	0.077		0.09	0.102		0.07	0.112		0.05	0.137						
Proparacaine HCl	0.16	0.044		0.15	0.086		0.15	0.169		0.14	0.247		0.13	0.380		7.46	0.12	0.52		
Propiomazine HCl	0.18	0.050		0.15	0.084		0.12	0.133		0.10	0.165		0.08	0.215						
Propoxycaine HCl																6.40			16	5.3
Propylene glycol																2.00			100	5.5
Pyrazinazine HCl	0.22	0.065		0.17	0.095		0.11	0.123		0.08	0.140		0.06	0.17						
Pyridostigmine bromide	0.22	0.062		0.22	0.125		0.22	0.250		0.22	0.377					4.13	0.22	0.52	0	7.2
Pyridoxine HCl																3.05			31*	3.2
Quinacrine methanesulfonate					0.06 ^e															
Quinine bisulfate				0.09	0.05					0.09	0.16									
Quinine dihydrochloride				0.23	0.130 ^e					0.19	0.33		0.18			5.07	0.18	0.52	trace*	2.5
Quinine hydrochloride				0.14	0.077 ^e					0.11	0.19									
Quinine and urea HCl				0.23	0.13					0.21	0.36					4.50	0.20	0.52	64	2.9
Resorcinol		0.161 ^e														3.30			96	5.0

	0.5 %		1 %		2 %		3 %		5 %		Iso-Osmotic Concentration ^a					
	E	D	E	D	E	D	E	D	E	D	%	E	D	H	pH	
Sodium methicillin	0.18	0.050	0.18	0.099	0.17	0.192	0.16	0.281	0.15	0.445	6.00	0.15	0.52	0	5.8	
Sodium nafcillin	0.14	0.039	0.14	0.078	0.14	0.158	0.13	0.219	0.10	0.285						
Sodium nitrate			0.68								1.36	0.66		0	6.0	
Sodium nitrite			0.84	0.480 ^e							1.08	0.83		0*	8.5	
Sodium oxacillin	0.18	0.050	0.17	0.095	0.16	0.177	0.15	0.257	0.14	0.408	6.64	0.14	0.52	0	6.0	
Sodium phenylbutazone	0.19	0.054	0.18	0.104	0.17	0.202	0.17	0.298	0.17	0.488	5.34	0.17	0.52			
Sodium phosphate			0.29	0.168			0.27	0.47			3.33	0.27	0.52	0	9.2	
Sodium phosphate, dibasic (2 H ₂ O)			0.42	0.24							2.23	0.40	0.52	0	9.2	
Sodium phosphate, dibasic (12 H ₂ O)			0.22				0.21				4.45	0.20		0	9.2	
Sodium propionate			0.61	0.35							1.47	0.61	0.52	0	7.8	
Sodium salicylate			0.36	0.210 ^e							2.53	0.36	0.52	0	6.7	
Sodium succinate	0.32	0.092	0.32	0.184	0.31	0.361					2.90	0.31	0.52	0	8.5	
Sodium sulfate, anhydrous			0.58	0.34							1.61	0.56	0.52	0	6.2	
Sodium sulfite, exsiccated			0.65	0.38							1.45			0	9.6	
Sodium sulfobromophthalein	0.07	0.019	0.06	0.034	0.05	0.060	0.05	0.084	0.04	0.123						
Sodium tartrate	0.33	0.098	0.33	0.193	0.33	0.385					2.72	0.33	0.52	0	7.3	
Sodium thiosulfate			0.31	0.181 ^e							2.98	0.30	0.52	0	7.4	
Sodium warfarin	0.18	0.049	0.17	0.095	0.16	0.181	0.15	0.264	0.15	0.430	6.10	0.15	0.52	0	8.1	
Sorbitol (A H ₂ O)											5.48			0	5.9	
Sparteine sulfate	0.10	0.03	0.10	0.056	0.10	0.111	0.10	0.167	0.10	0.277	9.46	0.10	0.52	19*	3.5	
Spectinomycin HCl	0.16	0.045	0.16	0.092	0.16	0.185	0.16	0.280	0.16	0.460	5.66	0.16	0.52	3	4.4	
Streptomycin HCl			0.17	0.10 ^e			0.16	0.16								
Streptomycin sulfate			0.07	0.036 ^e			0.06	0.10	0.06	0.17						
Sucrose			0.08	0.047 ^e			0.09	0.16	0.09	0.26						
Sulfacetamide sodium			0.23	0.132 ^e			0.23	0.40			9.25	0.10	0.52	0	6.4	
Sulfadiazine sodium			0.24	0.14			0.24	0.38			3.85	0.23	0.52	0	8.7	
Sulfamerazine sodium			0.23	0.13			0.21	0.36			4.24	0.21	0.52	0	9.5	
Sulfapyridine sodium			0.23	0.13			0.21	0.36			4.53	0.20	0.52	0	9.8	
Sulfathiazole sodium			0.22	0.13			0.20	0.35			4.55	0.20	0.52	5	10.4	
Tartaric acid				0.143 ^e							4.82	0.19	0.52	0	9.9	
Tetracaine HCl	0.18		0.18	0.109 ^e			0.15	0.26	0.12	0.35	3.90			75*	1.7	
Tetracycline HCl	0.14		0.14	0.081 ^e		0.10										
Tetrahydrozoline HCl											4.10			60*	6.7	
Theophylline				0.02*												

	0.5 %			1 %			2 %			3 %			5 %			Iso-Osmotic Concentration ^a				
	E	D	E	E	D	E	E	D	E	E	D	E	D	E	%	E	D	H	pH	
Theophylline sodium glycinate															2.94		0	8.9		
Thiamine HCl					0.139 ^c		0.08	0.089	0.07	0.119		0.05	0.153		4.24		87*	3.0		
Thiethylperazine maleate	0.10	0.030	0.09		0.050										3.50		74	10.3		
Thiopental sodium					0.155 ^c															
Thiopropazate diHCl	0.20	0.053	0.16	0.16	0.090		0.12	0.137	0.10	0.170		0.08	0.222							
Thioridazine HCl	0.06	0.015	0.05	0.05	0.025		0.04	0.042	0.03	0.055		0.03	0.075							
Thiotepa	0.16	0.045	0.16	0.16	0.090		0.16	0.182	0.16	0.278		0.16	0.460		5.67	0.16	0.52	10*	8.2	
Tridihexethyl chloride	0.16	0.047	0.16	0.16	0.096		0.16	0.191	0.16	0.28		0.16	0.463		5.62	0.16	0.52	97	5.4	
Triethanolamine	0.20	0.058	0.21	0.21	0.121		0.22	0.252	0.22	0.383					4.05	0.22	0.52	100	10.7	
Trifluoperazine 2HCl	0.18	0.052	0.18	0.18	0.100		0.13	0.144												
Triflupromazine HCl	0.10	0.031	0.09	0.09	0.051		0.05	0.061	0.04	0.073		0.03	0.092							
Trimeprazine tartrate	0.10	0.023	0.06	0.06	0.035		0.04	0.045	0.03	0.052		0.02	0.061							
Trimethadone	0.23	0.069	0.23	0.23	0.133		0.22	0.257	0.22	0.378					4.22	0.21	0.52	100	6.0	
Trimethobenzamide HCl	0.12	0.033	0.10	0.10	0.062		0.10	0.108	0.09	0.153		0.08	0.232							
Tripelennamine HCl					0.13 ^d										5.50	0.26	0.52	100	6.3	
Tromethamine	0.26	0.074	0.26	0.26	0.15		0.26	0.30	0.26	0.45					3.45	0.26	0.52	0	10.2	
Tropicamide	0.10	0.03	0.09	0.09	0.050															
Trypan blue	0.26	0.075	0.26	0.26	0.150															
Trypsinamide					0.11 ^e															
Tubocurarine chloride					0.076 ^e															
Urea			0.59		0.34										1.63	0.55	0.52	100	6.6	
Urethan					0.18 ^b										2.93		100	6.3		
Uridine	0.12	0.035	0.12	0.12	0.069		0.12	0.138	0.12	0.208		0.12	0.333		8.18	0.11	0.52	0*	6.1	
Valethamate bromide	0.16	0.044	0.15	0.15	0.085		0.15	0.168	0.14	0.238		0.11	0.324							
Vancomycin sulfate	0.06	0.015	0.05	0.05	0.028		0.04	0.049	0.04	0.066		0.04	0.098							
Viomycin sulfate			0.08		0.05				0.07	0.12		0.07	0.20							
Xylometazoline HCl	0.22	0.065	0.21	0.21	0.121		0.20	0.232	0.20	0.342					4.68	0.19	0.52	88	5.0	
Zinc phenolsulfonate															5.40		0*		5.4	
Zinc sulfate			0.15		0.086 ^e				0.13	0.23		0.12	0.35		7.65	0.12	0.52			

^a The unmarked values were taken from Hammarlund *et al.*²⁶⁻²⁹ and Sapp *et al.*³⁰

^b Adapted from Lund *et al.*²⁵

^c Adapted from *British Pharmaceutical Codex*.³¹

^d Obtained from several sources.

^e, sodium chloride equivalents; *d*, freezing-point depression, °C; H, hemolysis, %, at the concentration that is iso-osmotic with 0.9% NaCl, based on freezing-point determination or equivalent test; pH, approximate pH of solution studied for hemolytic action; *, change in appearance of erythrocytes and/or solution³⁰⁻³²; † t, pH determined after addition of blood.

Note: See also Budavari S, ed. *Merck Index*, 11th edn, Rahway, NJ: Merck, 1988: MISC 79-103.

APPENDIX B: Isotonic Solution V — Values^{33, a,b}

Drug (0.3g)	Water Needed for Isotonicity(mL)	Drug (0.3g)	Water Needed for Isotonicity(mL)	Drug (0.3g)	Water Needed for Isotonicity(mL)
Alcohol	21.7	Epinephrine hydrochloride	9.7	Silver nitrate	11.0
Ammonium chloride	37.3	Ethylmorphine hydrochloride	5.3	Silver protein, mild	5.7
Amobarbital sodium	8.3	Fluorescein sodium	10.3	Sodium acetate	15.3
Amphetamine phosphate	11.3	Glycerin	11.7	Sodium bicarbonate	21.7
Amphetamine sulfate	7.3	Holocaine hydrochloride	6.7	Sodium biphosphate, anhydrous	15.3
Antipyrine	5.7	Homatropine hydrobromide	5.7		
Apomorphine hydrochloride	4.7	Homatropine methylbromide	6.3	Sodium biphosphate	13.3
Ascorbic acid	6.0	Hyoscyamine sulfate	4.7	Sodium bisulfite	20.3
Atropine methylbromide	4.7	Neomycin sulfate	3.7	Sodium borate	14.0
Atropine sulfate	4.3	Oxytetracycline hydrochloride	4.3	Sodium iodide	13.0
Bacitracin	1.7	Penicillin G, potassium	6.0	Sodium metabisulfite	22.3
Barbital sodium	10.0	Penicillin G, sodium	6.0	Sodium nitrate	22.7
Bismuth potassium tartrate	3.0	Pentobarbital sodium	8.3	Sodium phosphate	9.7
Boric acid	16.7	Phenobarbital sodium	8.0	Sodium propionate	20.3
Butacaine sulfate	6.7	Physostigmine salicylate	5.3	Sodium sulfite, exsiccated	21.7
Caffeine and sodium benzoate	8.7	Pilocarpine hydrochloride	8.0	Sodium thiosulfate	10.3
Calcium chloride	17.0	Pilocarpine nitrate	7.7	Streptomycin sulfate	2.3
Calcium chloride (6 H ₂ O)	11.7	Piperocaine hydrochloride	7.0	Sulfacetamide sodium	7.7
Chlorobutanol (hydrated)	8.0	Polymyxin B sulfate	3.0	Sulfadiazine sodium	8.0
Chlortetracycline sulfate	4.3	Potassium chloride	25.3	Sulfamerazine sodium	7.7
Cocaine hydrochloride	5.3	Potassium nitrate	18.7	Sulfapyridine sodium	7.7
Cupric sulfate	6.0	Potassium phosphate, monobasic	14.7	Sulfathiazole sodium	7.3
Dextrose, anhydrous	6.0	Procainamide hydrochloride	7.3	Tetracaine hydrochloride	6.0
Dibucaine hydrochloride	4.3	Procaine hydrochloride	7.0	Tetraeycline hydrochloride	4.7
Dihydrostreptomycin sulfate	2.0	Scopolamine hydrobromide	4.0	Viomycin sulfate	2.7
Ephedrine hydrochloride	10.0			Zinc chloride	20.3

Drug (0.3 g)	Water Needed for Isotonicity(mL)	Drug (0.3 g)	Water Needed for Isotonicity(mL)	Drug (0.3 g)	Water Needed for Isotonicity(mL)
Ephedrine sulfate	7.7	Scopolamine methylnitrate	5.3	Zinc sulfate	5.0
Epinephrine bitartrate	6.0	Secobarbital sodium	8.0		

^a This table of *Isotonic Solution Values* shows volumes in mL of water to be added to 300mg of the specified drug in sterile water to produce an isotonic solution. The addition of an isotonic vehicle (commonly referred to as diluting solution) to make 30 mL yields a 1% solution. Solutions prepared as directed above are isotonic with 0.9% sodium chloride solution but may not be isotonic with blood (see Appendix A for hemolysis data).

^b The v values for drugs that do not appear in Appendix B but are listed in Appendix A can be calculated from the sodium chloride equivalent for 1% drug.

Example—Calculate the V value for anileridine HCl (Appendix A defines *E* = 0.19).

$$\frac{100 \text{ mL Soln}}{0.9 \text{ NaCl}} \times \frac{0.19 \text{ g NaCl}}{1 \text{ g drug}} \times 0.3 \text{ g drug} = 6.33 \text{ mL Soln}$$

for dilute solution

$$6.33 \text{ mL soln} \equiv 6.33 \text{ mL water} \therefore V = 6.33 \text{ mL water/0.3 g drug}$$

Chapter 17

Chemical Kinetics

Rodney J. Wigent, PhD

CHEMICAL REACTIONS 301

REACTION MECHANISM 301

REACTION RATE 302

FIRST-ORDER REACTIONS 303

SECOND-ORDER REACTIONS 304

THIRD-ORDER REACTIONS 305

PSEUDO-ORDER REACTIONS 305

MORE COMPLEX REACTIONS 306

EFFECTS ON REACTION RATE 307

CATALYSIS 308

OTHER EFFECTS 311

DRUG STABILIZATION 314

This chapter is intended as a general introduction to Chemical Kinetics. A comprehensive review of experimental approaches and interpretation of data can be found in several texts, such as the books by House, Epenson, and Houston, and the compilation of information relative to kinetic studies on pharmaceuticals by Garrett.¹

CHEMICAL REACTIONS

A chemical reaction is said to have occurred when one or more molecules undergoes a process in which electrons are either lost or gained by an atom or molecule, or the electrons go through a transition process in which they are shared differently between atoms which result in the breaking or forming of bonds. These result in the formation of one or more new chemical species with different chemical structures and properties than that of the original starting material. Such chemical processes usually result in changes in energy and/or degree of randomness in going from the initial to the final state which is the basis for thermodynamics. Thermodynamic parameters, such as ΔG , ΔE , ΔH , and ΔS , are state functions that only depend on the initial and final states of a chemical process—usually defined as the state of the reactants and products at certain temperatures, pressures and physical states—and are independent of the pathway taken to get to the final state from the initial state. Spontaneous reactions are said to occur when the reactants proceed to products when the energy contained within the products is less than the energy contained in the reactants (i.e., $\Delta G < 0$) which usually result in the release in heat (i.e., $\Delta H < 0$) and/or an increase in randomness of the chemical system (i.e., $\Delta S > 0$). However, it has been observed that in some cases, whereas some reactions are thermodynamically favorable to proceed spontaneously, the conversion of reactants to products may proceed at very slow rates over relatively long periods of time. This serves to point out that thermodynamics is essentially a snap shot of the differences between the properties of the product and reactant states of a chemical reaction which provides no indication of the chemical processes or the time that it took to go from reactant to product state.

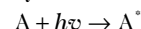
Chemical kinetics is the discipline that is concerned with the mechanism by which a chemical process gets to its final state from its initial state and the rate in which this reaction proceeds. Therefore, chemical kinetics involves the study of rate of chemical change and the way in which this rate is influenced by the conditions of the concentration of reactants, products, and other chemical species that may be present, and by factors such as solvent, pressure, and temperature. From these studies, one or more mechanisms involving a series of elementary processes may be postulated to explain how the reactants are converted to products during a chemical process.

Applied to pharmaceuticals, chemical kinetic information permits a rational approach to the stabilization of drug products, and prediction of shelf life and optimum storage conditions.

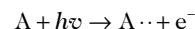
REACTION MECHANISM

There are two basic ways in which chemical reactions can occur. One method involves the absorption of electromagnetic radiation which excites certain electrons within an atom or molecule to a higher energy state or even the ejection of the electron creating a free radical which is a very reactive chemical species.

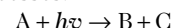
This may be indicated by the elementary processes such as:



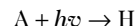
and



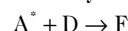
where $h\nu$ represents a quantum of electromagnetic radiation absorbed by chemical species A, A^* represents the higher energy, excited state of chemical species A, $A\cdot$ represents the free radical of species A, and e^- represents an electron. If the excited electrons were originally involved in a bond between atoms, then the bond may be weakened or broken which may cause the molecule to dissociate into smaller molecules or possibly cause the molecule to internally rearrange its bonds as shown by the elementary processes:



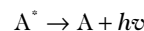
or



where species B and C are the products of the disintegration of species A, and H is a resulting product caused by the rearrangement of species A upon absorption of a quantum of electromagnetic energy. If an excited molecule collides with another atom/molecule under the right conditions, a new bond may form creating a new molecule as shown by the elementary process:

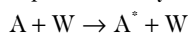


where D represent a reactant species that collides and reacts with the excited species A^* to produce the product species F. At other times, the excited electrons may revert back to a lower energy level or even the original energy level through a release of electromagnetic radiation as shown by the elementary process:

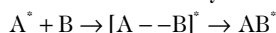


The other common method in which a reaction can occur is when a transfer of energy occurs when an atom or molecule collides with another atom, molecule or even the wall of a container. For example, an atom or molecule with a certain amount of kinetic and potential energy collides with a wall of a container that is at a higher temperature than that of the atom or molecule. The atom/molecule absorbs some of this thermal energy resulting in an increase in its kinetic energy, which causes it to increase its velocity, or, in the case of a molecule, may absorb enough energy to disrupt bonds within the molecule. A similar process can occur when there is a collision between

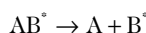
atoms/molecules where there may be a transfer of energy. In such cases where the atoms/molecules collide and bounce off one another, one atom/molecule may increase its kinetic or potential energy at the cost of energy of the other atom/molecule. Such processes can be represented by:



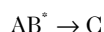
where W represents either a molecule or the wall of a container in which chemical species A has collided creating a higher energy, activated species A^* . It should be noted that W in such cases has lost some of its energy upon such collisions. The higher energy atom/molecule may then go on to collide with other atoms/molecules. If there is sufficient energy between the colliding molecules/atoms and other conditions, such as intermolecular orientation, are right, intermolecular interactions may be sufficient to cause the colliding entities to "stick" together creating an intermediate species as indicated by the elementary process:



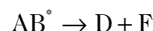
where $[A-B]^*$ and AB^* represent intermediate or transition state species. During this process, either a transfer of energy can occur before the entities separate creating an excited species, certain electrons of the two entities can form a bond creating a new chemical species, or the transition state entity can break apart to form a new chemical species as shown by the elementary processes:



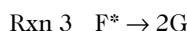
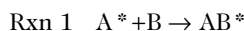
or



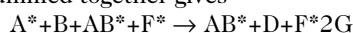
or



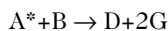
Thus, when a chemical reaction does occur, it may occur via a mechanism in a single reaction step, known as an elementary process, or by a series of two or more elementary processes which, when totaled, will give the net chemical reaction. For example:



which when summed together gives

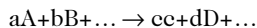


However, because the transition state AB^* and the activated molecule F^* appear in both reactant and product sides of the reaction in the same stoichiometric quantities, they can be treated as intermediates and removed from the reaction to give the net reaction:



REACTION RATE

The rate of a reaction is the velocity with which a reactant or reactants undergoes a chemical change. Experimentally, the rate of a reaction must be determined by directly or indirectly following the change in the concentration of the reactants or products as a function of time. When there is more than one reactant, such changes need to be normalized according to the stoichiometry of the reaction. For a reaction of the type



where the uppercase letters represent chemical species and the lowercase letters represent stoichiometric coefficients, the rate in which reactants go to products can be determined by following the rate of the disappearance of the reactants as a function of time.

$$\text{Rate} = -\frac{d[A]}{adt} = -\frac{d[B]}{bdt} \quad (1)$$

The brackets denote concentration (usually molar concentration unless otherwise indicated) and d represents the derivative

function. The negative sign signifies that the concentration of the reactants is decreasing, as the rate must always be positive as long as the reaction is progressing from reactants to products.

The rate at which a reaction proceeds for the reaction type shown above also can be determined by following the appearance of the products as a function of time:

$$\text{Rate} = +\frac{d[C]}{cdt} = +\frac{d[D]}{ddt} \quad (2)$$

where the positive signs indicate that the concentrations of the products are increasing. Note that these two expressions for rate are only for the type of reaction where the reactants go irreversibly to products, without going through any intermediates.

If $[A]_0$, $[B]_0$, $[C]_0$, and $[D]_0$ represent the initial concentration (i.e., $t = 0$) of each of the reactants and products, at some time t (i.e., $t = t$), the concentration of A decreases by aX (i.e., $[A]_t = [A]_0 - aX$) and the concentration of B decreases by bX (i.e., $[B]_t = [B]_0 - bX$). Similarly, the concentrations of the products C and D increase by cX and dX , respectively (i.e., $[C]_t = [C]_0 + cX$ and $[D]_t = [D]_0 + dX$) after some time t . Thus, upon normalization, the rate expressed in Equations 1 and 2 can be shown to reduce to Equation 3.

$$\text{Rate} = +\frac{dX}{dt} \quad (3)$$

The *law of mass action* relates these experimentally determined rates to the concentration of all of the reacting species. This law states that, at a given temperature, the rate of the reaction is at each instant proportional to the product of the concentration of each of the reacting species raised to a power equal to the number of molecules of each of these species participating in the process. Accordingly, the law of mass action applied to the above reaction gives the following rate equation,

$$\text{Rate} = k[A]^n[B]^m \quad (4)$$

where the proportionality constant k (referred to as the *specific rate constant* or as the *rate constant*) should be independent of the concentrations of all chemical species at a given temperature. The exponents, n and m , are known as the *order of the reaction* with respect to the components A and B, respectively; their sum represents the overall order of the reaction.

Equation 4 can be applied to each individual elementary process which makes up a reaction. In such cases, the exponents of each of the reacting species are equal to their stoichiometric coefficients for the balanced equation representing the processes. These exponents indicate the number of molecules/atoms of each reacting species that participate in a simultaneous collision to form the reactants with the rate of the process controlled by the specific rate constant, k , during that specific elementary process. It is interesting to note that statistically collisions involving more than two bodies are extremely rare and collisions involving more than three bodies are nearly impossible; the specific rate constants for these rare processes will be very small and the reactions very slow.

It is important to note that for a *net chemical reaction*, which is the sum of all of the elementary processes, there is no requirement that the order of the reaction with respect to a chemical species be identical to its stoichiometric coefficient indicated in the net chemical equation. This is because chemical species that appeared as both reactants and products in the various elementary processes have been removed to create a balanced net equation. Further, it may be necessary for some elementary processes to occur more than once in order to arrive at net balanced equation. Additionally, some elementary processes may be reversible (i.e., products revert to reactants) during an overall chemical reaction. Therefore, unlike for elementary processes in which the species which enter into Equation 4, can only be reactants, Equation 4 for a net chemical reaction can include products as well as reactants. However, a proper rate equation for a net chemical equation should not contain any chemical species that exists as an intermediate during a chemical reaction.

It should be noted that unless the stoichiometric coefficient of the reactant or product that is being followed to determine the rate of the reaction is *unity* (one), the rate of the reaction is not equivalent to the change in the concentration of the chemical species with respect to time. For the case where there is only one chemical reactant, which has a stoichiometric coefficient that is greater than one, authors of articles and textbooks on kinetics often base the reaction rate only on the disappearance of the reactant without accounting for the stoichiometry. When this occurs, the resulting rate constant will be greater than the true rate constant by a factor equal to the stoichiometric coefficient. Thus, care must be taken to determine how the rates of reactions were determined when comparing rate constants of a reaction.

FIRST-ORDER REACTIONS

When the rate of a reaction is proportional to the first power of the concentration of a reactant, the rate equation is given by

$$\frac{dX}{dt} = k[A]_t = k([A]_0 - aX) \quad (5)$$

where a represents the stoichiometric coefficient for reactant A. Rearrangement of Equation 5 gives

$$\int \frac{dX}{[A]_0 - aX} = k \int dt \quad (6)$$

When Equation 6 is integrated over the limits of $t = 0$ (at which $X = 0$) to $t = t$ (at which $X = X$), the following first-order integrated rate equation is obtained:

$$[A]_t = [A]_0 e^{-akt} \quad (7)$$

Figure 17-1 shows a typical plot where reactant A exponentially decays to products according to Equation 7. The rate of the reaction - that is - the negative value of the tangent of this curve at any time, decreases with time as the concentration of the reactant decreases. Equation 7 can be linearized by rearrangement to give Equation 8.

$$\ln[A]_t = -akt + \ln[A]_0 \quad (8)$$

Equation 8 suggests that a plot of the natural logarithm of the concentration of the reactant as a function of time should give a linear plot with a slope equal to $-ak$ and a y-intercept equal to the natural logarithm of the initial concentration of the reactant (Fig 17-2). Commonly, a plot of the common logarithm of the concentration versus time is found in the literature for first-order reactions. In this case, according to Equation 9, the slope of this line would be equal to $-ak/2.303$, and the y-intercept would be equal to the common logarithm of the initial concentration of the reactant.

$$\log[A]_t = -\frac{akt}{2.303} + \log[A]_0 \quad (9)$$

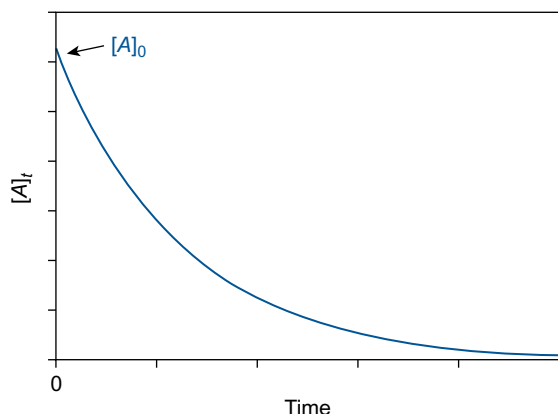


Figure 17-1. Plot of concentration of A versus time for a first-order reaction.

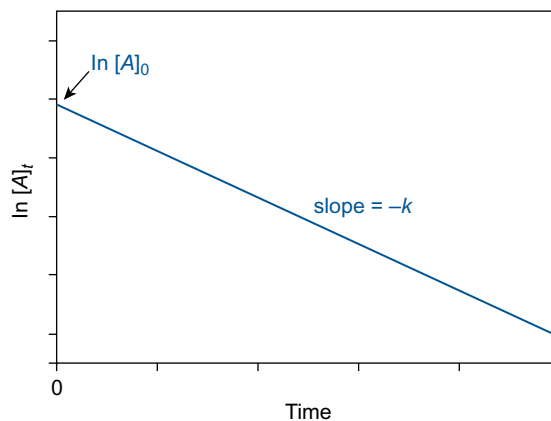


Figure 17-2. Plot of natural logarithm of the concentration of A versus time for a first-order reaction.

In such a case, the rate constant, k , for a first-order reaction has a unit of reciprocal time (e.g., s^{-1}).

Sometimes it may be necessary to determine the rate constant k from only two concentrations of the reactant, $[A]_1$ and $[A]_2$, obtained at two different times, t_1 and t_2 , in which case Equation 10 may be used.

$$k = \frac{1}{a(t_2 - t_1)} \ln \frac{[A]_1}{[A]_2} \quad (10)$$

Another useful method for determining k is the fractional-life method, of which the half-life method is the most common. The *half-life method* involves measuring the time ($t = t_{1/2}$) that it takes for half of the initial concentration of the reactant to undergo reaction: $[A]_t = [A]_0/2$. Substituting these values into Equation 7 and rearranging to solve for k yields

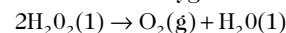
$$k = \frac{\ln 2}{at_{1/2}} \quad (11)$$

It is apparent from Equation 11 that the half-life period for first-order reactions is constant and independent of the amount of reactant present. Thus, half of the initial concentration of the reactant undergoes a reaction during the first half-life period, leaving 50% of the original concentration unreacted. During the second half-life period, which is identical in time to the first half-life period for a first-order reaction, half of the reactant remaining after the first half-life period reacts, leaving 25% of the initial concentration of the reactant unreacted. Similarly, after the third half-life period, 12.5% of the initial reactant would remain unreacted. After 10 first-order, half-life periods, only 0.098% of the original reactant remains unreacted. For precise studies, the rate of disappearance of a reactant should be followed over at least two or three half-life periods.

In some drug stability studies, it is necessary to determine the time that it takes for the loss of 10% of the drug, leaving 90% of the original drug concentration; that is, $[A]_t = 0.90[A]_0$ at $t = t_{0.90}$. This time can be determined with the knowledge of the rate constant and the stoichiometric coefficient a by substituting these expressions into Equation 7 and rearranging to yield

$$t_{0.90} = \frac{\ln 0.90}{ak} \quad (12)$$

A common example of a chemical reaction that follows first order chemical kinetics is the decomposition of liquid hydrogen peroxide to produce molecular oxygen and liquid water:



It should be noted that the “ a ” term, representing the stoichiometric coefficient for hydrogen peroxide is 2 and has to be used in all of the above kinetic equations. Because this is a first-order process, the rate of decomposition of hydrogen peroxide is directly proportional to the hydrogen peroxide concentration.

First-order rate processes are not restricted to chemical reactions. The passive diffusion of drugs across biological membranes and processes of drug absorption, distribution, metabolism, and excretion often can be shown to occur at rates proportional to the concentration of a drug, and thus can be described as first-order rate processes. The rate of growth and the rate of killing or inactivation of microorganisms by heat or chemical agents usually follow first-order kinetic processes. Radioactive decay always follows first-order kinetics.

SECOND-ORDER REACTIONS

There are two forms of second-order reactions. For the first case, it is assumed that the rate of reaction is proportional to the concentration of reactant A raised to the power of 2 - that is, the reaction is second order with respect to A, in which case the rate equation takes the form

$$\frac{dX}{dt} = [A]_t^2 = ([A]_0 - \alpha X)^2 \quad (13)$$

where α represents the stoichiometric coefficient of the reactant in the net equation. Equation 13 can be rearranged to give

$$\int \frac{dX}{([A]_0 - \alpha X)^2} = k \int dt \quad (14)$$

When Equation 14 is integrated over the limits of $t = 0$ (at which $X = 0$) to $t = t$ (at which $X = X$), the following second-order integrated rate equation is obtained.

$$\frac{1}{[A]_t} = \alpha kt + \frac{1}{[A]_0} \quad (15)$$

It should be noted that because the stoichiometry was taken into account in this derivation, the stoichiometric coefficient, α , has been incorporated into Equation 15. If the rate of reaction was determined solely on the disappearance of reactant A without considering the stoichiometry (i.e., α is assumed to be equal to 1), then the rate constant for this reaction would be twice as large as the true rate constant. The reader should be aware of this situation as this occurs quite frequently in the literature.

The decomposition of hydrogen iodide is a second-order reaction; in the gaseous state, hydrogen iodide forms hydrogen gas and molecular iodine according to the reaction



The integrated rate expression for this reaction follows the form given by Equation 15. With the value of " α " equal to 2.

Equation 15 suggests that, for a second-order reaction, if the reciprocal of the concentration of reactant A is plotted as a function of time, the slope of the line is equal to the rate constant $2k$, and the y-intercept is the reciprocal of the initial

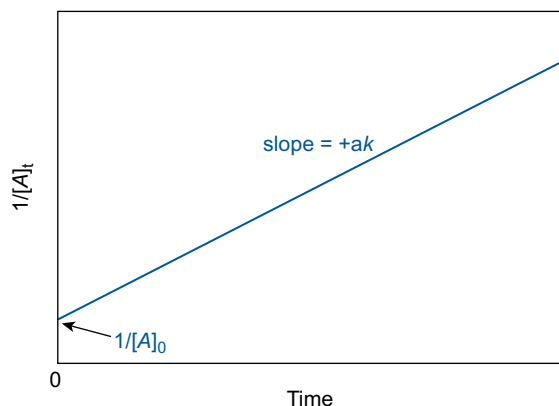


Figure 17-3. Plot of the reciprocal of the concentration of A versus time for a second-order reaction.

concentration of A (Figure 17-3). Rearranging Equation 15 and solving for k yields

$$k = \frac{1}{2t} \left(\frac{[A]_0 - [A]_t}{[A]_0[A]_t} \right) \quad (16)$$

The rate constant for second-order reactions has units of reciprocal concentration and seconds (e.g., $\text{M}^{-1}\text{s}^{-1}$).

The second type of a second-order reaction occurs if the rate of the reaction is proportional to the product of the concentrations of two reactants, each raised to the power of 1, that is, first order with respect to both reactants. Equation 17 shows the rate equation for such a reaction.

$$\frac{dX}{dt} = [A]_t[B]_t = ([A]_0 - \alpha X)([B]_0 - bX) \quad (17)$$

The stoichiometric coefficients of the reactants A and B are represented by α and b . Equation 17 can be arranged to give the following:

$$\int \frac{dX}{([A]_0 - \alpha X)([B]_0 - bX)} = k \int dt \quad (18)$$

When Equation 18 is integrated over the limits of $t = 0$ (at which $X = 0$) to $t = t$ (at which $X = X$), the following second-order integrated rate equation is obtained:

$$\ln \frac{[A]_t}{[B]_t} = (b[A]_0 - a[B]_0)kt + \ln \frac{[A]_0}{[B]_0} \quad (19)$$

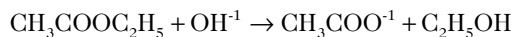
This suggests that if the left side of Equation 19 is plotted against time, the slope of the line would be equal to $(b[A]_0 - a[B]_0)k$ and the y-intercept is equal to the natural logarithm of the ratio of the initial concentrations of reactants A and B. Equation 19 does not apply if the initial concentrations of the two reactants consist of a stoichiometric mixture (i.e., $b[A]_0$ equals $a[B]_0$) in which case, Equation 18 reduces to a form similar to Equation 14 as shown in Equation 14b.

$$\frac{b}{a} \int \frac{dX}{([A]_0 - \alpha X)^2} = k \int dt \quad (14b)$$

for which the integrated rate equation for this system reduces to Equation 15b.

$$\frac{1}{[A]_t} = \frac{a^2}{b} kt + \frac{1}{[A]_0} \quad (15b)$$

An example of a second-order reaction in which two reactants are involved is the saponification of an ester, such as ethyl acetate, in alkaline solution:



The course of this reaction may be followed by determining, by titration at specified times, the concentration of hydroxide ions remaining unreacted during the course of the reaction. This information and the initial concentrations of the ethyl acetate and hydroxide can be used to determine the rate constant in Equation 19.

Fractional-life methods can be applied readily to second-order reactions for the case when the order of the reaction with respect to one reactant is 2, or for the case when the initial concentrations of each of two reactants are equal when the order with respect to each reactant is 1. For example, the half life of a second-order reaction is given by Equation 20.

$$t_{1/2} = \frac{1}{k[A]_0} \quad (20)$$

Unlike the half-life period for a first-order reaction, the half-life period for a second-order reaction is not constant, but rather is proportional to the reciprocal of the initial concentration of reactant. This means that the half-life period increases as a second-order reaction proceeds with time; thus, it takes twice as long to deplete a second-order reactant from 50 to 25% as it did to deplete the reactant from 100 to 50%.

THIRD-ORDER REACTIONS

Except for in a condensed phase system, true third-order reactions are rare, as they require a simultaneous three-body collision of chemical species. There are a number of ways in which third-order reaction can occur—from a combination of three different chemical entities, for which the order of the reaction with respect to each of these is 1, to the simplest case in which three identical substances react, for which the order of the reaction with respect to that species is three. For the latter case, assuming the stoichiometric coefficient of the single reacting entity A is 3, then the rearranged rate equation is given by Equation 21.

$$\int \frac{dX}{([A]_0 - 3X)^3} = k \int dt \quad (21)$$

Upon integration of Equation 21 over the limits of $t = 0$ (at which $X = 0$) to $t = t$ (at which $X = X$), the following third-order integrated rate equation is obtained.

$$\frac{1}{[A]_t^2} = 6kt + \frac{1}{[A]_0^2} \quad (22)$$

Again, it should be noted, that if the stoichiometry was not taken into account and the rate was only determined by the rate of disappearance of reactant A, then Equation 22 would have 2 for the coefficient of kt instead of 6 and the value for the rate constant would be three times the value of the rate constant in Equation 22.

The equation for the half-life period for the case of Equation 22 is given by

$$t_{1/2} = \frac{1}{2k[A]_0^2} \quad (23)$$

Another type of a third-order reaction occurs when the rate of the reaction is proportional to the product of the concentrations of two reactants, one raised to the power of 1 and the other raised to the power of 2; it is first order with respect to one reactant and second order with respect to the other reactant. Equation 24 shows the rate equation for such a reaction.

$$\frac{dX}{dt} = [A]_t^2[B]_t = ([A]_0 - \alpha X)^2([B]_0 - bX) \quad (24)$$

If the stoichiometric coefficients, α and b , are both equal to 1, then Equation 24 can be rearranged and integrated over the limits of $t = 0$ (at which $X = 0$) to $t = t$ (at which $X = X$). The rate constant from this resulting equation is given by

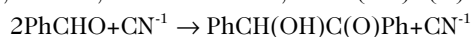
$$k = \frac{1}{t} \frac{1}{([B]_0 - [A]_0)} \frac{[A]_0 - [A]_t}{[A]_0[A]_t} \frac{1}{([B]_0 - [A]_0)^2} \ln \frac{[A]_t[B]_0}{[B]_t[A]_0} \quad (25)$$

However, if the stoichiometric coefficients are $\alpha = 2$ and $b = 1$, then when Equation 24 is rearranged and integrated over the limits of $t = 0$ (at which $X = 0$) to $t = t$ (at which $X = X$), the rate constant is determined by Equation 26.

$$k = \frac{1}{t} \frac{1}{(2[B]_0 - [A]_0)} \frac{[A]_0 - [A]_t}{[A]_0[A]_t} \frac{1}{(2[B]_0 - [A]_0)^2} \ln \frac{[A]_t[B]_0}{[B]_t[A]_0} \quad (26)$$

The rate constant for third-order reactions has units of reciprocal of the square of concentration per second (e.g., $M^{-2}s^{-1}$).

One well known example of a third order reaction that follows Equation 24 is the cyanide catalyzed dimerization of benzaldehyde, $PhCHO$, to form benzoin, $PhCH(OH)C(O)Ph$:



Whereas this reaction appears to be third order, the reaction mechanism is proposed to occur through several elementary processes in which the net reaction appears to be second order with respect to the benzaldehyde and first order with respect to the cyanide catalyst. It should be noted that the individual elementary processes are often simple first and/or second order reactions, which when combined to give a net reaction, gives the appearance that the net reaction is a higher order reaction. Care must be taken not to interpret reactions, such as this

benzoin condensation, that requires an unlikely elementary process in which all three reactants come together in the rate limiting step.

Because of the rigors of the mathematics, when a third-order reaction is suspected, experimental conditions are often chosen so as to simplify the calculations. For example, for the third-order reaction in which the stoichiometric coefficients of the two reacting species are $a = 2$ and $b = 1$, such as that which led to the development of Equation 26, if the experimental conditions are set such that $[A]_0 = 2[B]_0$, it will lead to a much simpler integrated rate equation.

PSEUDO-ORDER REACTIONS

For some reactions, the rate of the reaction may be independent of the concentration of one or more of the reacting species over a wide range of concentrations. This may occur under these conditions:

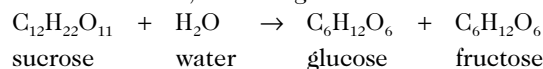
1. One or more of the reactants enters into the rate equation in great excess compared to the others.
2. One of the reactants is a catalyst.
3. One or more of the reactants is constantly replenished during the course of a reaction.

If this happens, the constant concentration term(s) in the rate equation is combined with the rate constant to give an *apparent rate constant*. For example, if the concentration of A in Equation 4 remains constant, then Equation 4 can be rewritten as

$$\text{Rate} = (k[A]^n)[B]^m = k_{\text{app}}[B]^m \quad (27)$$

where the apparent rate constant, k_{app} (sometimes referred to as the *pseudo-order rate constant*) now depends on the concentration of A raised to its power, n . Unfortunately, no information about n , the order of reaction with respect to A, can be determined from a single experiment. Rather, to gain an understanding of n , multiple experiments must be performed where the concentration of A is varied. A plot of the natural logarithm of k_{app} versus the natural logarithm of the concentration of A will give a slope that is equal to n .

In 1850, Wilhelmy performed the first quantitative kinetics study by following the rate of hydrolysis (inversion) of sucrose to glucose and fructose, according to the reaction:



Wilhelmy found that this reaction followed the rate equation:

$$-\frac{d[C_{12}H_{22}O_{11}]}{dt} = k_{\text{app}}[C_{12}H_{22}O_{11}] \quad (28)$$

which, upon rearrangement and integration, gives Equation 29.

$$\ln[C_{12}H_{22}O_{11}] = -k_{\text{app}}t + \ln[C_{12}H_{22}O_{11}]_0 \quad (29)$$

This reaction is now known to be a second-order reaction, as it is first order with respect to both sucrose and water. As for most typical aqueous solutions, the molar concentration of water (approximately 55.5 moles of water per liter) greatly exceeds the concentration of the solute sucrose. Therefore, even at moderate concentrations of sucrose, there is only a minor change in the molar concentration of water and the concentration of the solvent is practically constant over the course of the reaction. This allows the concentration of water to be incorporated into the apparent rate constant and the reaction appears to be first order.

As another example, if component A reacts in aqueous solution to go to product B, according to the first-order rate equation given by Equation 5 and the stoichiometric coefficient a is 1 (unity), then the concentration of A as a function of time should follow the exponential form of the integrated rate equation given by Equation 7. However, if this reaction occurs in a saturated solution of A (i.e., $[A]_{\text{sat}}$) in the presence of excess solid A, and if the rate of converting solid A to aqueous A is

greater than the rate of reaction in solution, then the rate of disappearance of A is given by:

$$\frac{dX}{dt} = k[A]_{\text{sat}} = k_{\text{app}} \quad (30)$$

If Equation 30 is rearranged and integrated between the limits of $t = 0$ (at which $X = 0$) to $t = t$ (at which $X = X$) and defining $[B]_t = X$, then the following zero-order rate equation is obtained,

$$[B]_t = k_{\text{app}} t \quad (31)$$

which shows that as long as the solution remains saturated with A, the formation of B will occur at a constant rate. As an example, if a compound for which decomposition in solution is first order is present in excess of its maximum solubility (a suspension), the concentration of the reactant in solution will be invariant so long as there is excess solid reactant present. The kinetics of such a system would then follow Equation 30.

First- and second-order reactions are by far the most common types of rate processes encountered regarding drug stability. If a reaction is of higher order than first order, it often is convenient to adjust experimental conditions so that the concentrations of all but one of the reactants remain constant throughout the experiment. If, for example, the concentration of hydroxide ion in the saponification of an ester is in great excess of the concentration of ester, or if a buffer system is employed to control hydroxide-ion concentration, then the concentration of hydroxide ion essentially is invariant throughout the course of the experiment. The observed rate of the reaction, therefore, depends only on the changing concentration of the ester, and the reaction is said to be *pseudo first-order*. The apparent first-order rate constant, k_{app} , thus obtained is $k[\text{OH}^{-1}]$ and, of course, is different for each hydroxide-ion concentration. The actual rate constant, k , can be obtained easily by dividing the experimentally determined apparent first-order rate constant, $k[\text{OH}^{-1}]$, by the concentration of hydroxide ion maintained throughout the study.

In the study of complex reactions, it is often desirable to use this approach of maintaining the concentration of all but one of the reactants constant to facilitate determining the dependency of the reaction rate on each of the reactants in turn.

MORE COMPLEX REACTIONS

Many chemical reactions do not follow the simple reaction kinetics listed above, but rather they often consist of two or more elementary processes that may lead to more complicated rate equations. For example, comparison of experimental measurements of the rate of the disappearance of the reactants and the appearance of the products may indicate that the reactants must be forming one or more intermediates before proceeding to form the products. Often chemical reactions proceed reversibly to form products before an equilibrium is established. There are many cases where the reactants simultaneously proceed through different mechanisms to form two or more products. These situations can lead to negative or non-integer orders of reactions with respect to reactants and products within the rate equation. Quite often, a series of experiments must be performed in which certain conditions are controlled in order to establish the order of the reaction of individual species involved in the chemical reaction before an overall rate equation can be established. The next several sections will look at some of the more common complex reactions.

REVERSIBLE REACTIONS

Many reactions are known to be reversible where the reactants go to form products but the products will reversibly revert back to reactants. The simplest example of this is in the case where reactant A follows a first-order kinetic process with a forward rate constant, k_f , to produce product B.



However, product B then follows a first-order rate process with a reverse rate constant, k_r , to reform reactant A.



Because, during the course of this reaction, reactant A is being simultaneously depleted and formed, the rate at which reactant A disappears is related to the forward and reverse rates according to Equations 32 and 33:

$$-\frac{d[A]_t}{dt} = \frac{d[A]_{\text{forward}}}{dt} - \frac{d[A]_{\text{reverse}}}{dt} \quad (32)$$

or

$$-\frac{d[A]_t}{dt} = k_f[A]_t - k_r[B]_t \quad (33)$$

If the initial concentration of B is zero, then at time $t = 0$ (initially) the rate equation is given solely by the forward rate equation. As the reaction proceeds, the reverse rate equation begins to contribute more and more substantially to the overall rate equation. Finally, a point will be reached at which the rate of the forward reaction is equal to the rate of reverse reaction and the overall rate is equal to 0. This is defined as a *dynamic equilibrium* and the concentration equilibrium constant, K_c , given by Equation 34, is equal to the ratio of the forward and reverse rate constants, where

$$K_c = \frac{[B]_{\text{eq}}}{[A]_{\text{eq}}} = \frac{k_f}{k_r} \quad (34)$$

where $[B]_{\text{eq}}$ and $[A]_{\text{eq}}$ are the equilibrium concentrations of the product and reactant, respectively.

The rate equation expressed in Equation 33 can be rewritten to give

$$\frac{dX}{dt} = k_f([A]_0 - X) - k_r([B]_0 - X) \quad (35)$$

Rearrangement and integration of Equation 35 between the limits of $t = 0$ (at which $X = 0$) to $t = t$ (at which $X = X$), the following expressions for the concentrations of A and B as a function of time are obtained:

$$[A]_t = \frac{(k_f[A]_0 - k_r[B]_0) \exp[-(k_f + k_r)t] + k_r([A]_0 + [B]_0)}{(k_f + k_r)} \quad (36)$$

and

$$[B]_t = \frac{k_f([A]_0 + [B]_0) - (k_f[A]_0 - k_r[B]_0) \exp[-(k_f + k_r)t]}{(k_f + k_r)} \quad (37)$$

SIMULTANEOUS REACTIONS

Another very common reaction is when the reaction of one or more reactants lead to the formation of multiple products through different mechanistic pathways, each with characteristic rates:



and



For the case in which both reaction pathways are first order, then the rate of disappearance of reactant A is then given by Equation 38.

$$-\frac{d[A]_t}{dt} = k_1[A]_t + k_2[A]_t = (k_1 + k_2)[A]_t \quad (38)$$

Rearrangement and integration of Equation 38 gives:

$$[A]_t = [A]_0 \exp[-(k_1 + k_2)t] \quad (39)$$

Because the rate of formation of product B is given by

$$\frac{d[B]_t}{dt} = k_1[A]_t \quad (40)$$

then assuming that the initial concentration of B is 0, rearranging and integrating, and substituting Equation 39 into Equation

40 yields the following expression for the concentration of B as a function of time:

$$[B]_t = \frac{k_1[A]_0}{(k_1 + k_2)}(1 - \exp[-(k_1 + k_2)t]) \quad (41)$$

Using similar arguments, the concentration of C as a function of time is given by Equation 42.

$$[C]_t = \frac{k_2[A]_0}{(k_1 + k_2)}(1 - \exp[-(k_1 + k_2)t]) \quad (42)$$

It is of particular interest to note that if Equation 41 is divided by Equation 42, the ratio of the concentration of the products at any time is given by the ratio of the rate constants.

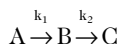
$$\frac{[B]_t}{[C]_t} = \frac{k_1}{k_2} \quad (43)$$

An example of this type of simultaneous reaction is the reaction of phenol with nitric acid to form both ortho- and para-nitrophenol through two simultaneous first-order reaction pathways. The relative concentrations of these two products are found to be given by Equation 43.

It is clear that if a kinetic experiment was performed without any *a priori* knowledge that the reaction is a simultaneous reaction, there is a danger that only the disappearance of a reactant or the appearance of only one of the products may lead to a faulty conclusion of the reaction mechanism. Care must be taken to attempt to identify and account for all of the chemical species in a chemical reaction to ensure that a proper rate mechanism is obtained.

CONSECUTIVE REACTIONS

One of the more common complex reactions is when a reactant decays through a series of consecutive reactions, forming one or more intermediates before forming a product. A simple case of a consecutive reaction is when reactant A proceeds through a first-order process to intermediate B which then decays to product C through another first-order process.



For cases such as this, it is often convenient to consider the situation in which the initial concentrations of B and C are 0 and the sum of the concentrations of A, B, and C at any time is equal to the initial concentration of the reactant A. In this case, the rate of disappearance of A is given by Equation 44 and the rate of appearance of product C is given by Equation 45.

$$-\frac{d[A]_t}{dt} = k_1[A]_t \quad (44)$$

and

$$+\frac{d[C]_t}{dt} = k_2[B]_t \quad (45)$$

The derivative of the concentration of the intermediate B with respect to time consists of the rate of formation of B from the product A and the disappearance of B as it proceeds to product C, as shown by

$$\frac{d[B]_t}{dt} = k_1[A]_t - k_2[B]_t \quad (46)$$

Upon integration and rearrangement of Equation 44, the concentration of reactant A as a function of time can be expressed by

$$[A]_t = [A]_0 \exp(-k_1 t) \quad (47)$$

It should be noted that Equations 45 and 46 are not considered to be valid rate equations because, by convention, the concentration of an intermediate may not appear in a final rate equation. Therefore, an expression for the concentration of B as a function of time in terms of only the reactant or product must be developed. Substituting Equation 47 into Equation 46, rearranging and integrating yields the following expression for the concentration of B as a function of time.

$$[B]_t = \frac{k_1[A]_0}{(k_2 - k_1)}(\exp[-k_1 t] - \exp[-k_2 t]) \quad (48)$$

Equation 48 can be substituted into Equations 45 and 46 to give appropriate rate expressions. Then Equation 46 can be rearranged and integrated to give an expression for the concentration of C as a function of time.

$$[C]_t = [A]_0 \left(1 - \frac{k_2 \exp[-k_1 t]}{(k_2 - k_1)} \right) + \frac{k_1 \exp[-k_2 t]}{(k_2 - k_1)} \quad (49)$$

EFFECTS ON REACTION RATE

TEMPERATURE

The application of heat to increase the rate of a chemical reaction is a common laboratory procedure. The rate of most solvolytic reactions of pharmaceuticals is increased roughly 2- to 3-fold by a 10°C increase near room temperature. In 1889, Arrhenius noted that the variation with temperature of the rate constant of chemical reactions could be expressed by

$$k = A \exp[-E_a / RT] \quad (50)$$

where, according to collision theory, E_a is the Arrhenius activation energy (i.e., the difference between the average energy of reactive molecules and the minimum energy required for reactants to proceed to products); the quantity $\exp[-E_a/RT]$ is the Boltzmann factor, which represents the fraction of molecules having energies greater than or equal to E_a ; the pre-exponential term A is a constant called the frequency factor; R is the gas constant (8.314 joule/mol-K or 1.987 cal/mol-K); and T is the absolute temperature. The Arrhenius equation can be expressed in a linear form according to Equation 51.

$$\ln k = -\frac{E_a}{R} \frac{1}{T} + \ln A \quad (51)$$

Equations 50 and 51 are valid so long as the reaction mechanism does not change over the temperature range studied. A plot of the natural logarithm of the rate constant versus the reciprocal of the absolute temperature in which the rate constants are determined gives a negative slope that is equivalent to $-E_a/R$ (Figure 17-4). If a nonlinear plot is obtained, a thermally induced change in the reaction mechanism has probably occurred.

Differentiation of Equation 51 with respect to temperature, and then integrating between the limits of k_2 and k_1 at temperatures between T_2 and T_1 yields

$$\ln \frac{k_2}{k_1} = \frac{E_a}{R} \left(\frac{T_2 - T_1}{T_1 T_2} \right) \quad (52)$$

This equation allows E_a to be calculated for a reaction when the rate constants are known at two temperatures, or the rate constant at one temperature to be calculated if E_a and the rate constant at another temperature are known.

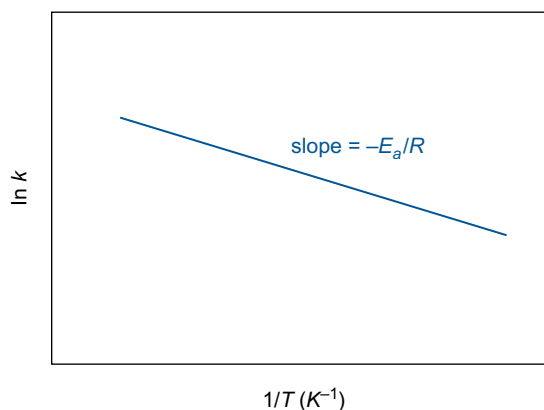
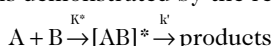


Figure 17-4. Variation of the rate constant with reciprocal absolute temperature, illustrating the Arrhenius equation.

Most solvolytic reactions of pharmaceuticals exhibit activation energies in the range of 8 to 20 kcal/mol. Using Equation 51 and the appropriate activation energy, one readily can calculate that a reaction having an activation energy of 8 kcal/mol would show an approximately 1.5-fold increase in k for a temperature increase from 25° to 35°; a reaction having an activation energy of 20 kcal/mol would show a 3-fold increase in k for a similar temperature increase.

When two molecules undergo chemical interaction, it is reasonable to suppose that they first must collide and then, if conditions are right, undergo a rearrangement of certain electrons to form the bonds characteristic of the new molecules. However, not all collisions can cause a chemical change, or else chemical reactions would occur with great rapidity because collision frequencies are very high. Whereas molecules or atoms must first collide if a reaction is to occur, the colliding molecules may not have an energy greater than or equal to the activation energy sufficient to overcome the mutual repulsion of the interacting molecules and enable them to approach close enough to each other to effect certain bond ruptures and/or establish new bonds characteristic of the products. The greater this energy requirement will result in a smaller proportion of colliding molecules that will have energies sufficient to cause a reaction, which, in turn, will lead to slower reaction rates. In the Arrhenius equation, A is a factor related to frequency of collisions, and the term $\exp[-E_a/RT]$ expresses the probability that the colliding molecules will have sufficient energies sufficient to cause a reaction at temperature T . The concept of energy of activation, in relationship to the energy of the reactants and of the products, is illustrated in Figure 17-5.

Eyring, in his transition state theory, proposed that reactants must proceed through an activated complex before proceeding to products. This is demonstrated by the reaction



where the reactants are considered to be in a rapid equilibrium with the activated complex or transition state, represented by $[AB]^*$, which then decays to products by a first-order process, according to the rate equation

$$\text{Rate} = k'[AB]^* \quad (53)$$

However, as Equation 53 contains the concentration of the activated complex, an intermediate, it is not a valid rate equation and an expression in terms which include only the reactants or products must be substituted for this expression. Because the activated complex is in equilibrium with the reactants, the concentration of the activated complex can be given by

$$[AB]^* = K^*[A][B] \quad (54)$$

where K^* is the equilibrium constant. Substituting Equation 54 into Equation 53 yields

$$\text{Rate} = K^*k'[A][B] = k[A][B] \quad (55)$$

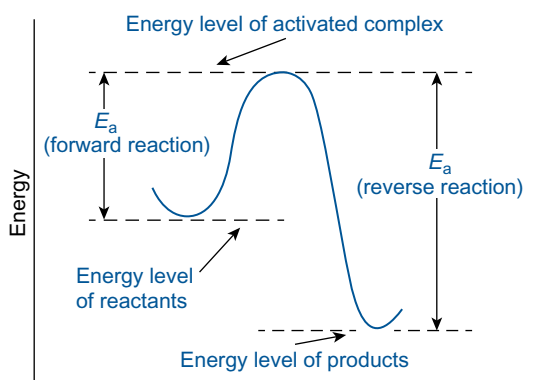


Figure 17-5. Relation between activation energy and energy levels of reactants, products and activated complex.

where k is equal to K^*k' and Equation 55 is a proper rate equation. Eyring was able to show that the rate constant, k' , of any reaction is given by the expression

$$k' = \frac{RT}{N_a h} K^* \quad (56)$$

where R is equal to 8.314 ergs/mol-K, N_a is Avogadro's number, and h is Planck's constant, which is equal to 6.625×10^{-27} erg-sec. K^* can be related to the thermodynamic parameters ΔG^* , ΔH^* , and ΔS^* through the equation

$$K^* = \exp(-\Delta G^*/RT) = \exp(T\Delta S^* - \Delta H^*)/RT \quad (57)$$

If Equation 57 is substituted into Equation 56, after it has been divided by the absolute temperature, the following linear equation is obtained.

$$\ln \frac{k'}{T} = -\frac{\Delta H^*}{R} \frac{1}{T} + \frac{\Delta S^*}{R} + \ln \frac{R}{N_a h} \quad (58)$$

Thus, the thermodynamics of the formation of the activated complex can be determined from a plot of the natural logarithm of the ratio of the rate constant to absolute temperature versus the reciprocal absolute temperature.

CATALYSIS

In catalytic reactions, a molecule, called a catalyst, interacts with a reactant in a series of elementary processes in such a fashion as to lower the activation energy barrier (i.e., E_a in Figure 17-5) of an uncatalyzed reaction. This change in mechanism causes the catalyzed reaction to run faster without changing the relative energy levels of either the reactants or products. During a catalytic reaction, the catalyst reacts with a reactant to form an intermediate that undergoes additional reaction(s) to form the product and the original catalytic molecule. Thus, while the catalyst is both consumed and produced in several elementary processes, there is no net change in the concentration of the catalyst during a catalyzed reaction but there is usually a substantial change in the rate in which the reaction occurs.

Whereas there are several types of catalysis, such as homogeneous and heterogeneous catalysis, and autocatalysis, only a few general examples will be discussed here. Additional information about catalysis can be found in some of the references cited in the attached bibliography.

SPECIFIC ACID AND SPECIFIC BASE CATALYSIS

The terms *specific acid catalysis* and *specific base catalysis* refer to catalysis by the hydronium or hydrogen ion, and by the hydroxide ion, respectively. For example, if the rate of hydrolysis of an ester, such as ethyl acetate, is studied at a constant pH in a strongly buffered solution, the rate of disappearance of the intact ester will be an apparent first-order reaction. If the reaction is studied in solutions buffered at several different pH values in a sufficiently acid pH region, a different apparent first-order rate constant will be observed for each pH value. The observed rate actually depends on the concentration of both the ester and hydrogen ion and, therefore, is a second-order reaction that appears to be a pseudo first-order reaction at the constant hydrogen-ion concentration in the buffer. Therefore, the observed first-order rate constant, k_{obs} , is proportional to the hydrogen ion concentration of the buffer system as shown by Equation 59.

$$k_{\text{obs}} = k_{\text{acid}}[H^+] \quad (59)$$

Taking the logarithm of Equation 59 yields

$$\log k_{\text{obs}} = \log k_{\text{acid}} + \log[H^+] \quad (60)$$

this upon applying the definition of pH yields

$$\log k_{\text{obs}} = \log k_{\text{acid}} - \text{pH} \quad (61)$$

Equation 61 suggests that a plot of $\log k_{\text{obs}}$ versus pH will be linear with a slope of -1 with a y-intercept equal to the term $\log k_{\text{acid}}$.

Similarly, if the same hydrolysis reaction is studied in buffered solutions at several pH values in a sufficiently alkaline region of the pH scale, the observed apparent first-order rate constants will be found to vary with hydroxide-ion concentration:

$$k_{\text{obs}} = k_{\text{base}}[\text{OH}^-] \quad (62)$$

and

$$\log k_{\text{obs}} = \log k_{\text{base}} + \log[\text{OH}^-] \quad (63)$$

However, the hydroxide ion concentration is related to the hydrogen ion concentration through the ionization constant of water, K_w , and Equation 61 becomes

$$\log k_{\text{obs}} = \log k_{\text{base}} + \log K_w + \text{pH} \quad (64)$$

Therefore, a plot of $\log k_{\text{obs}}$ versus pH in a heavily buffered alkaline solution would yield a straight line with a slope of +1 and a y-intercept equal to the term $\log k_{\text{base}} + \log K_w$.

Because of the equilibrium that exists between hydroxide and hydronium ions in aqueous solutions, each of these ions exists at all values of pH and the observed rate constant is actually given by the sum of Equations 59 and 62.

$$k_{\text{obs}} = k_{\text{acid}}[\text{H}^+] + k_{\text{base}}[\text{OH}^-] \quad (65)$$

The complete logarithm k_{obs} versus pH profile would be similar to that illustrated in Figure 17-6 for the hydrogen ion and hydroxide ion (specific acid and specific base) catalyzed hydrolysis of the ester atropine.² At relatively low values of pH the acid-catalyzed hydrolysis predominates; at relatively high values of pH the base-catalyzed hydrolysis predominates. The pH at which the minimum rate of hydrolysis is observed is a function of the relative magnitude of the specific rate constants k_{acid} and k_{base} . In the atropine example, the minimum rate of hydrolysis is at pH of 3.7, which indicates that $k_{\text{base}} > k_{\text{acid}}$. If k_{base} equals k_{acid} , then, at 25°, the expected minimum rate of the reaction would be expected to occur at pH 7.

A reaction may be catalyzed not only by hydrogen ion and hydroxide ion, but also by other Brønsted acids or bases such as the solvent water. This is referred to as *general acid/base catalysis*. In this case, the observed rate constant is given by

$$k_{\text{obs}} = k_{\text{water}} + k_{\text{acid}}[\text{H}^+] + k_{\text{base}}[\text{OH}^-] \quad (66)$$

where k_{water} is a pseudo-order rate constant that has the concentration of water, which is in large excess, incorporated into it. Figure 17-7 shows how a plot of the logarithm k_{obs} versus pH might appear in such a case. The flat region, where the rate of reaction apparently is not pH dependent, is the region where

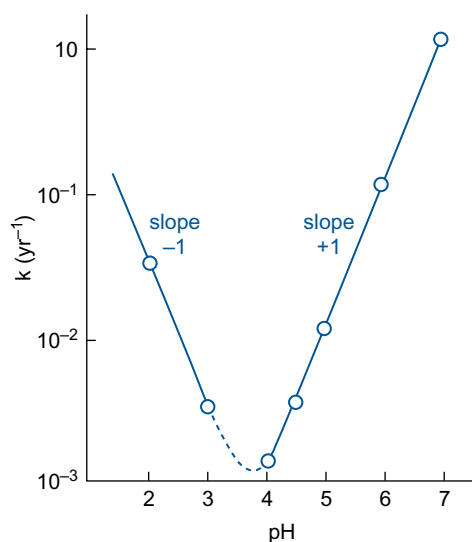


Figure 17-6. Apparent first-order rate of hydrolysis of atropine as a function of pH at 30°C. The reaction is an illustration of specific hydrogen and hydroxide-ion catalysis. (Courtesy Kondritzer AA, Zvirblis P. Stability of atropine in aqueous solution. *J APhA Sci Ed* 1957; 46: 531–5).

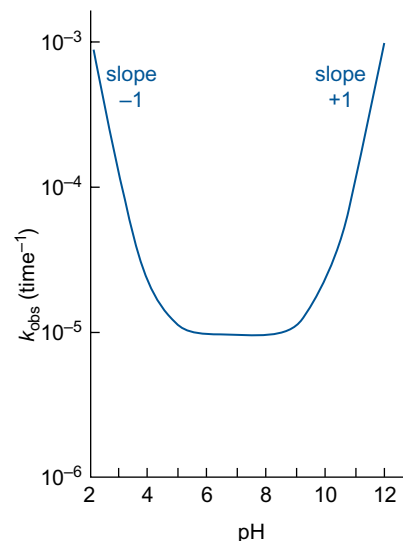


Figure 17-7. Apparent first-order rate of decomposition as a function of pH for a hypothetical case where $k_{\text{H}^+} = k_{\text{OH}^-} = 0.1$, $k_{\text{H}_2\text{O}} = 1 \times 10^{-5}$. The uncatalyzed reaction predominates in the pH region 5 to 9.

the solvent is much more important as a catalyst than either the hydrogen or hydroxide ions.

In some cases, compounds may be weak acids or weak bases in solution and will be in equilibrium with its conjugate acid or base form with the relative concentration of each species of the compound dependent on the pH of the solution and the magnitude of the dissociation constants of the weak acid, K_a , and/or the weak base, K_b . The decomposition of each of these forms of the compound may be catalyzed by hydronium ions, hydroxide ions and water in aqueous solution but with different rate constants. For example, a plot of k_{obs} versus pH for the hydrolysis of the weakly basic drug procaine is similar to that seen in Figure 17-6, which indicates that procaine undergoes both acid and base hydrolysis³. However, there are notable aberrations in this plot at high pH values (*i.e.*, low pOH values) when compared to Figure 17-6. It has been proposed that the kinetics of this system when the pH is greater than 7.0 can be represented by¹

$$-d[\text{Pr}]_{\text{total}}/dt = k_1[\text{OH}^-][\text{Pr}] + k_2[\text{OH}^-][\text{PrH}^+] \quad (67)$$

which neglects the acid catalyzed hydrolysis of procaine in this pH range. *Pr* represents the nonionized procaine molecule, *PrH*⁺ represents the protonated form of procaine, and the total concentration of procaine, $[\text{Pr}]_{\text{total}}$, is the sum of the concentrations of ionized and protonated forms. The concentration of each species can be related to the total procaine concentration, $[\text{Pr}]_{\text{total}}$, by the relationships

$$[\text{Pr}] = \frac{[\text{OH}^-]}{K_b + [\text{OH}^-]} [\text{Pr}]_{\text{total}} \quad (68)$$

and

$$[\text{PrH}^+] = \frac{K_b}{K_b + [\text{OH}^-]} [\text{Pr}]_{\text{total}} \quad (69)$$

where K_b is the classical dissociation constant for the weak base procaine. The complete rate expression for procaine hydrolysis in this pH range is given by Equation 70.

$$-\frac{d[\text{Pr}]_{\text{total}}}{dt} = \left[\frac{k_1[\text{OH}^-]^2 + k_2[\text{OH}^-]K_b}{K_b + [\text{OH}^-]} \right] d[\text{Pr}]_{\text{total}} = k_{\text{obs}}d[\text{Pr}]_{\text{total}} \quad (70)$$

Thus, the observed rate constant, k_{obs} , of procaine hydrolysis will depend on the concentration of the hydroxide concentration as illustrated graphically in Figure 17-8,⁹ by a plot of k_{obs} versus pOH for the pH region 7 to 13. (Note: the values for k_{obs} are plotted versus pOH rather than pH to better illustrate the relationship of k_{obs} and hydroxide concentration as indicated in equation 70.) At high concentrations of hydroxide ion (*i.e.*,

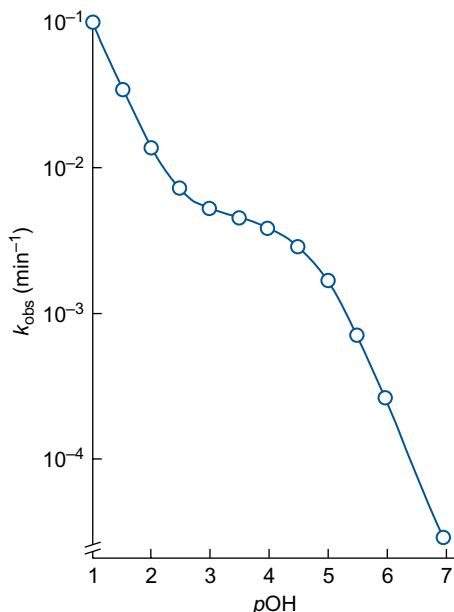


Figure 17-8. Apparent first-order rate of hydrolysis of procaine as a function of hydroxide-ion concentration at 40°C. (From Higuchi T, Lachman L. Inhibition of hydrolysis of esters in solution by formation of complexes. I. Stabilization of benzocaine with caffeine. *J APhA Sci Ed* 1955; 44: 521-6.)

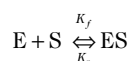
pOH values from 1 to 2.5), the hydroxide ion concentration is much greater than the value of K_b , and the denominator in equation 70 is predominated by the hydroxide concentration causing k_{obs} to decrease with decrease in hydroxide ion concentration (i.e., increase in pOH) via a first-order equation. Above pOH of 4.5, the hydroxide concentration is much smaller than the value of K_b , making the denominator constant with increasing values of pOH, (i.e., decreasing values of $[OH^-]$). This causes k_{obs} to decrease with a decrease in hydroxide ion concentration (i.e., increase in pOH) via a second order equation as seen in the numerator of Equation 70. Over the range of pOH values of 2.5 to 4.5, the values of the hydroxide concentration and K_b are similar and k_{obs} follows the more complex hydroxide ion concentration dependence shown in the bracketed term in equation 70 which shows a slower rate of decrease in k_{obs} with decrease in hydroxide concentration relative to that exhibited at higher values of hydroxide concentration.

GENERAL ACID OR BASE CATALYSIS

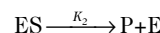
Acid or base catalysis is not restricted to the effect of hydrogen or hydroxide ion. Undissociated acids and bases often can be demonstrated to produce a catalytic effect and, in some instances, metal ions and various anions can serve as catalysts. Mutarotation of glucose in acetate buffer is catalyzed by hydrogen ion, hydroxide ion, acetate ion, and undissociated acetic acid. Also, the rate of barbiturate hydrolysis in ammonia buffers is increased by increasing buffer concentration at constant pH as a result of catalysis by NH_3 . Hydrolysis of the amide function of chloramphenicol exhibits, in addition to solvent and specific acid—base catalysis, general acid—base catalysis in phosphate and citrate buffers. General acid—base catalysis is to be anticipated if there is evidence of a significant solvent catalysis, as illustrated in the pH-rate profile of Figure 17-7.

ENZYME CATALYSIS

In biological systems, catalytic molecules, called enzymes (E), reversibly bind to a substrate (S) to form an intermediate (ES) which then decomposes to give a product (P) and the original enzyme.



and



The rate of this reaction, v , will be given by $d[P]/dt = k_2[ES]$. However, this is an improper rate equation as ES is an intermediate. Michaelis and Menten used a steady state approximation (i.e., at sometime during the reaction the time rate of change of the intermediate will be zero) to calculate the concentration of ES.

$$[ES] = \frac{k_f[E]_0[S]_0}{k_f[S]_0 + k_r + k_2} \quad (71)$$

Upon substitution of Equation 71 into the rate equation, the initial rate, v_0 , (i.e., The rate at time equal to zero) is given by Equation 72.

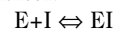
$$v_0 = \frac{k_f k_2 [E]_0}{k_f + \frac{k_r + k_2}{[S]_0}} = \frac{v_m}{1 + \frac{K_m}{[S]_0}} = \frac{v_m [S]_0}{K_m + [S]_0} \quad (72)$$

where K_m , the Michaelis-Menten constant, is equal to the term $(k_r + k_2)/k_f$ and v_m , the maximum initial velocity of the reaction, is equal to the term $k_2[E]_0$. The rate constant k_2 is often referred to as the turnover number which represents the number of molecules of product, P, created per second per mole of enzyme.

Equation 71 does not lend itself well to analysis as plots of v_0 versus $[S]_0$ only asymptotically approaches the maximum velocity, v_m . Rearrangement of Equation 72 into Equation 73, known as the Lineweaver-Burk Equation, lends itself more readily to analysis as shown in Figure 17-9.

$$\frac{1}{v_0} = \frac{1}{v_m} + \frac{K_m}{v_m [S]_0} \quad (73)$$

An important area of study in enzyme kinetics is enzyme inhibition. Although there are several cases in which an inhibitor, I, can inhibit an enzyme-catalyzed reaction, only two basic mechanisms will be discussed here. Competitive inhibition occurs when the inhibitor competes with the substrate, S, for the active binding site on the enzyme and blocks the catalytic action of the enzyme. In such a case, the formation of the enzyme-inhibitor complex, EI, is assumed to be in rapid equilibrium with the enzyme and inhibitor.



Equation 74 shows the resulting Lineweaver-Burk Equation for the case of a competitive inhibitor.

$$\frac{1}{v_0} = \frac{1}{v_m} + \left[1 + \frac{[I]}{K_i} \right] \frac{K_m}{v_m [S]_0} \quad (74)$$

K_i is the dissociation constant for the enzyme-inhibitor complex and is equal to the term $[E][I]/[EI]$. Figure 17-10 shows a typical

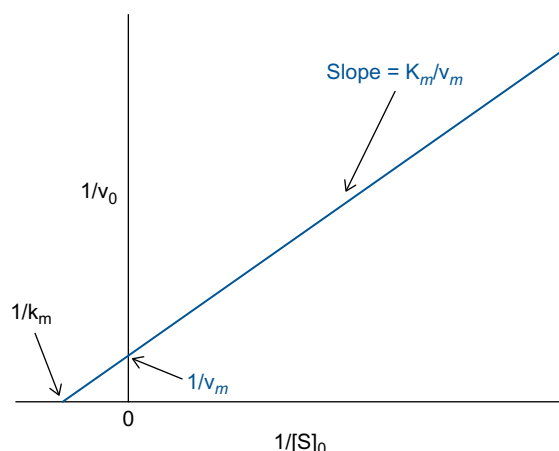


Figure 17-9. Lineweaver-Burk plot of modified Michaelis-Menten equation (Equation 73).

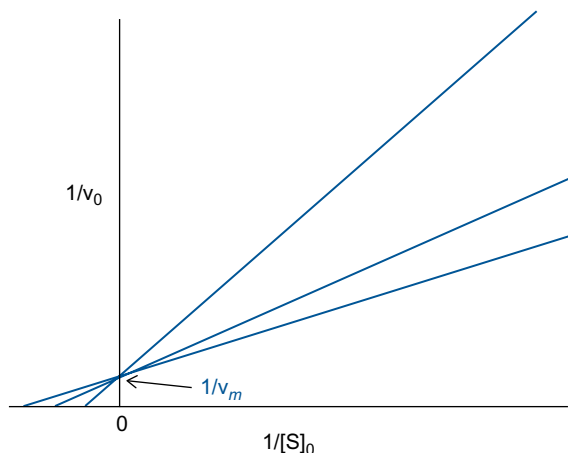


Figure 17-10. Lineweaver-Burk plot for competitive enzyme inhibition (Equation 74).

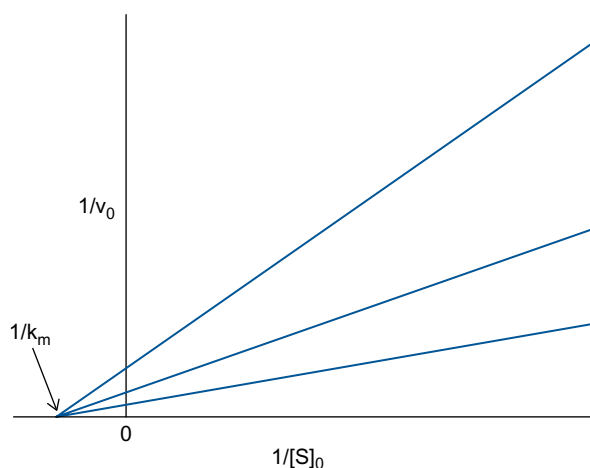
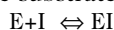


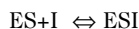
Figure 17-11. Lineweaver-Burk plot for noncompetitive enzyme inhibition (Equation 75).

Lineweaver-Burk graph for competitive inhibition. Each plot represents a different concentration of the competitive inhibitor. Note that all three plots intersect at the same point on the y-axis indicating that the reactions all have the same maximum velocity. However, because they do not intersect at the same point on the x-axis then they have different Michaelis-Menten constants, K_m .

Another type of enzyme inhibition is noncompetitive inhibition. In this case, the inhibitor does not bind to the active site of the enzyme but rather binds to another part of the enzyme which modifies the normal enzyme binding site changing the ability to form the enzyme-substrate complex, ES.



and



In this case, the equilibrium constant, K_i , is given by both $[E][I]/[EI]$ and $[ES][I]/[ESI]$. The resulting Lineweaver-Burk equation is given by Equation 75.

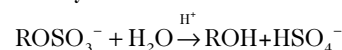
$$\frac{1}{v_0} = \left[\frac{1}{v_m} + \frac{K_m}{v_m[S]_0} \right] \left[1 + \frac{[I]}{K_i} \right] \quad (75)$$

Figure 17-11 shows a typical Lineweaver-Burk graph for noncompetitive inhibition. Each plot represents a different concentration of the competitive inhibitor. Note that all three plots intersect at the same point on the x-axis indicating that the reactions all have the same Michaelis-Menten constant, K_m . However, because the plots do not intersect at the same point on the y-axis, then the reactions have different maximum velocities.

OTHER EFFECTS

IONIC STRENGTH

The effects of increasing concentrations of electrolytes on the reaction rate can be largely predicted by consideration of the influence of ionic strength of the solution on interionic attraction through solvation effects. In general, the Debye-Hückel equation may be used to demonstrate that at a constant temperature, increased ionic strength of the solution would be expected to decrease the rate of reaction between oppositely charged ions, and increase the rate of reaction between similarly charged ions when compared to lower ionic strength solutions. The rationale for this is at higher ionic strengths the repulsion affects of like charged reactants and the attractive effects of unlike charged reactants are mediated through shielding by the increased ionic strength solution. Thus, the hydrogen-ion catalyzed hydrolysis of sulfate esters is inhibited by increasing electrolyte concentration.



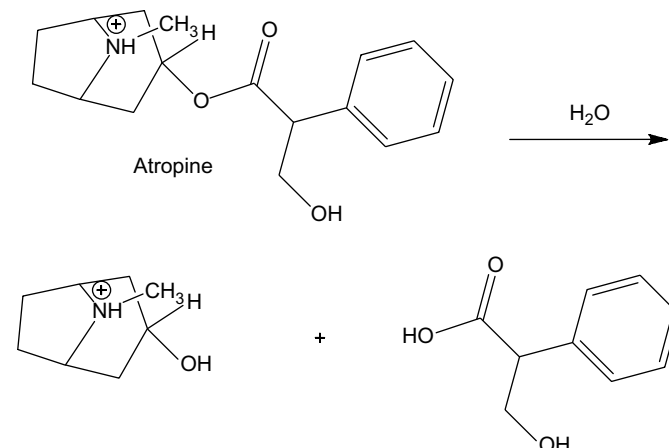
Reactions between ions and molecules with dipole moments, and reactions between neutral molecules generally are less sensitive to ionic strength effects than are reactions between ionic compounds causing little change in the reaction rate with increased ionic strength of the solution. However, reactions that result in formation of oppositely charged ions as products may exhibit a considerable increase in the reaction rate with increasing ionic strength.

DIELECTRIC CONSTANT OF SOLVENT

Reactions involving ions of opposite charge are accelerated by solvents with low dielectric constants. For example, the rate of hydrogen ion-catalyzed hydrolysis of sulfate esters is much greater in low dielectric constant solvents, such as methylene chloride, than in water. Reactions between similarly charged species are favored by high dielectric constant solvents. Reactions between neutral molecules, which produce a highly polar transition state, such as the reaction of triethylamine with ethyl iodide to produce a quaternary ammonium salt, also will be enhanced by high dielectric constant solvents.

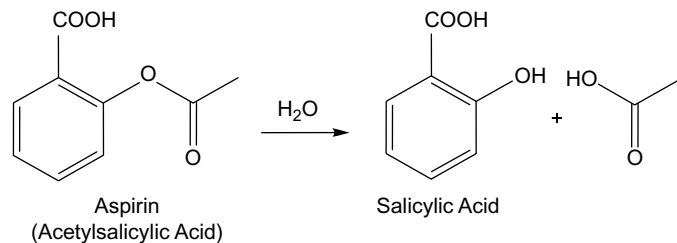
HYDROLYSIS (SOLVOLYSIS)

Hydrolysis of esters, such as procaine, aspirin, or atropine, represents one of the more common types of drug instability. *Ester hydrolysis* is either hydrogen- or hydroxide-ion catalyzed, although the catalysis that is important from the viewpoint of drug-product stability depends upon the specific compound and the pH of the solution. Amides generally are more stable than esters but are subject to catalysis by hydrogen and hydroxide ions, and often by general acids and bases. Some examples of the kinds of functional groups subject to hydrolytic cleavage and species shown to be catalysts for the reactions are presented below.



Hydrolysis of the ester function of atropine is typical of ester hydrolysis in that only catalysis by the hydrogen or hydroxide ions is important. Figure 17-6 illustrates a pH-rate profile which might be considered typical for such a reaction. Below pH 3, the principal reaction is hydrogen-ion catalyzed hydrolysis of the protonated form of atropine. Above pH 5, the principal reaction is hydroxide ion catalyzed hydrolysis of the same species. The maximum stability at 30° is at pH 3.7.

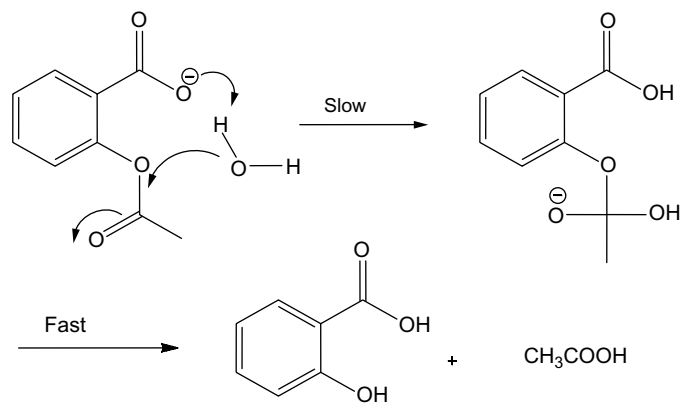
Hydrolytic cleavage of aspirin to salicylic acid and acetic acid was studied by Edwards.⁵



Edwards obtained the interesting pH-rate profile reproduced in Figure 17-9. The unusual pH-rate profile obtained for aspirin was attributed to a reaction of the form

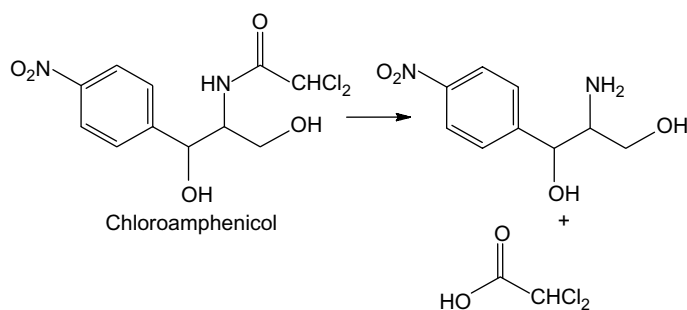
$$-\frac{d[\text{Aspirin}]_{\text{total}}}{dt} = k_1[H^+][HA] + k_2[H^+][A^-] + k_3[OH^-][A^-] + k_0[A^-] \quad (76)$$

where *HA* represents undissociated aspirin and *A⁻* represents aspirin anion. The pH-independent anion hydrolysis indicated for the pH region 5 to 9 has been attributed to intramolecular catalysis by orthocarboxylate anion rather than to general acid-base catalysis by water. It is principally this intramolecular catalysis that is responsible for the high instability of aqueous solutions of aspirin in the pharmaceutically useful pH range. Fersht and Kirby⁶ represented the intramolecular carboxylate ion reaction as a general base catalysis of attack by a water molecule.



For nucleophiles such as ethanol, the terminal hydroxyl of polyethylene glycol (PEG) and the lysine ε-amino function in serum albumin also can participate in this reaction in the same manner as water. Thus, from aspirin in ethanol solution, ethyl acetate appears as a product; in polyethylene glycol, a polyethylene glycol acetate is formed; and in a solution containing serum albumin (both *in vitro* and *in vivo*) aspirin produces an acetylated serum albumin. Whitworth et al.⁷ reasoned that an aspirin solution, prepared in a PEG solvent containing no free hydroxyl groups, would provide an aspirin solution of improved stability. They used acetylated PEG 400 as a solvent for aspirin and demonstrated that in such a solvent, less than 1% aspirin loss occurred after 30 days at 45°C.

Chloramphenicol decomposition below pH 7 proceeds primarily through hydrolytic cleavage of the amide function.

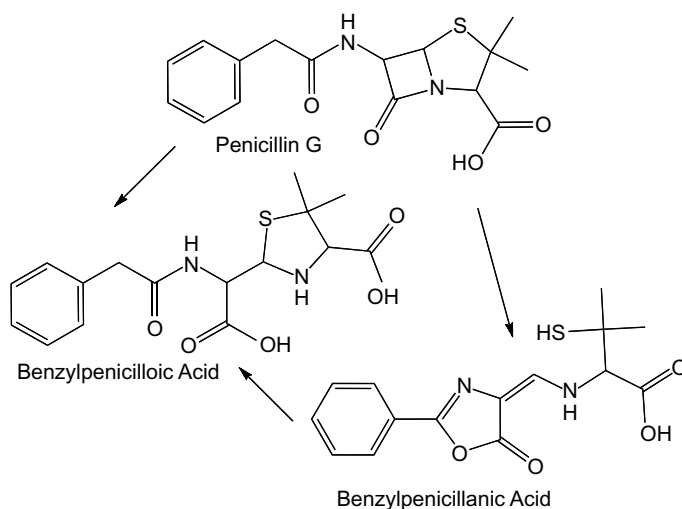


In the presence of a buffer, the reaction may be represented as

$$-\frac{d[\text{chloramp}]}{dt} = (k_0 + k_1[H^+] + k_2[OH^-] + k_{HB}[HB] + k_{B^-}[B^-])[\text{chloramp}] \quad (77)$$

where [Chloramp] is the concentration of chloramphenicol, and [HB] and [B⁻] are the concentrations of the weak acid and its conjugate base of the buffer species, respectively. In addition to hydrogen and hydroxide-ion catalysis, there is an uncatalyzed (or water) reaction, and there may be general acid—base catalysis, represented above by the buffer species HB and B⁻. In general, the rate of hydroxide-ion-catalyzed hydrolysis of amides is greater than rate of hydronium-ion-catalyzed hydrolysis.

Amides generally are much more stable than esters. Penicillins and cephalosporins are important exceptions to this rule because the amide bond is part of a strained four-membered ring (i.e., β-lactam). The decomposition of these compounds in aqueous solution is catalyzed by hydrogen ion, solvent, hydroxide ion, sugars, and many buffer species. Maximum stability occurs at about pH 7, but β-lactam antibiotics are too unstable to be formulated as solutions. For example, a buffered aqueous solution of penicillin G under refrigeration has a useful life of only about 1 week. Formation of the penicillanic acid by water-catalyzed rearrangement in acidic and neutral solutions is thought to be the first step in the degradation process.⁷



Barbiturate hydrolysis involves hydroxide-ion attack on both the undissociated acid, HP, and the ionized species, P⁻. Hydrogen ion catalyzed hydrolysis is not observed in the pH range of interest in pharmaceutical products. Therefore, the rate equation would take the form of Equation 78.

$$-\frac{d[\text{Barb}]}{dt} = k_1[OH^-][HP] + k_2[OH^-][P^-] \quad (78)$$

Hydrolysis of the amide (peptide) bond also occurs in protein and peptide drugs. This can occur by cleavage of the primary peptide linkage (R-NH-CO-R) between adjacent amino acids

in the peptide chain. Hydrolysis of the free side-chain amide groups of asparagine and glutamine (deamidation) is another degradation pathway for proteins. Insulin and recombinant human growth hormone undergo deamidation in solution. Further discussion of drugs resulting from biotechnology can be found in Chapter 34.

RACEMIZATION

Many drugs are chiral and racemization is a common mechanism of degradation resulting in loss of biological activity. In proteins, a mixture of the D and L enantiomers is formed by base-catalyzed reaction of the natural L configuration. Acid-catalyzed racemization of epinephrine or base-catalyzed racemization of pilocarpine results in a loss of pharmacological activity.

OXIDATION

Compounds such as phenols, aromatic amines, aldehydes, ethers, and unsaturated aliphatic compounds are subject to oxidation upon exposure to air or oxidizing chemicals. Epinephrine, ascorbic acid, phenothiazines, and vitamin A are examples of important pharmaceutical products that are oxidized readily. Proteins can undergo oxidative degradation by oxidation of methionine, a thioether, to its corresponding sulf-oxide. Oxidation of the carbon—carbon double bonds in unsaturated fatty acids (e.g., oleic acid) results in the fats and oils tasting rancid.

Of particular concern are oxidations that occur when solutions are exposed to atmospheric oxygen. Such reactions, termed *autoxidation* or *self-oxidation*, are complex reactions that proceed via a free-radical mechanism. A free radical is a highly unstable (highly reactive) species containing an unpaired electron. Autoxidation reactions are autocatalytic in that free-radical reactions generate additional free radicals, causing a chain reaction.

A technique used to protect pharmaceuticals susceptible to autoxidation is to include in the formulation agents that will react readily with free radicals, but that will terminate the chain propagation either by forming relatively stable, resonance-stabilized free radicals or by forming products that do not include additional free radicals.

PHOTOCHEMICAL DECOMPOSITION

Numerous dyes and drugs are subject to photochemical decomposition. Light-catalyzed oxidations and reductions of photoexcited species are common and are often mechanistically complex reactions involving free-radical intermediates. Pharmaceuticals such as riboflavin, nifedipine, and the phenothiazines are examples of common drugs that are extremely light sensitive.

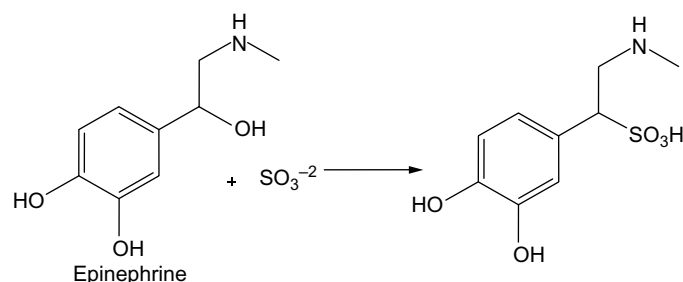
INTERACTION BETWEEN COMPONENTS

Because drugs are often combined in solution with buffers, antioxidants, flavoring agents, antimicrobial preservatives, and other drugs, potential interaction between the components of a formulation must be considered in pharmaceutical formulation development. Some obvious interactions, such as the possibility of the reaction of a drug having a primary amino function with an aldehyde, such as vanillin, to produce a Schiff base can be predicted; however, a number of interesting, less well-recognized reactions have been encountered.

In addition to buffer species acting as general acid—base catalysts, as previously indicated, some buffer species undergo specific interactions with drug molecules to form new chemical compounds. The formation of amides in aqueous solution from amines, such as benzocaine, and buffers, such as citric acid, has been observed.

The aromatic function of procaine reacts with glucose to form procaine *N*-glycoside, also, phenylethylamine reacts with dehydroacetic acid to form a Schiff base-type compound. Catechols have been shown to catalyze penicillin hydrolysis.

It has been demonstrated that bisulfite, an agent commonly employed to protect epinephrine against oxidative decomposition, is capable of inducing epinephrine degradation through attack on the chiral side chain.



Although a solution of folic acid alone is stable to light, a combination of riboflavin and folic acid showed a rapid loss of folic acid through formation of a coupled oxidation—reduction system in which riboflavin was photoreduced, with folic acid being used as a reducing substrate and being itself irreversibly oxidized. In the dark and in the presence of oxygen, the riboflavin was regenerated, and, when the solution was again irradiated, the cycle was repeated with further destruction of folic acid. In this case, the riboflavin acts as a photosensitizer in this reaction and would cause the decomposition not only of folic acid, but also of ascorbic acid or any other easily oxidized substrate.

The presence of micellar surfactants and certain high-molecular-weight polymers commonly employed in pharmaceuticals also have been shown to lead to decreased drug stability in some cases. Both nonionic and anionic surfactants, as well as polymers, such as polyvinylpyrrolidone, accelerate the photodecomposition of riboflavin in aqueous solution. Nonionic surfactants also are capable of increasing the rate of hydrolysis of sulfate esters which may be incorporated in or on the micellar surface.

PHYSICAL INSTABILITY

The introduction of an increasing number of drugs derived from developments in biotechnology necessitates greater awareness of instability occurring as a result of loss of drug activity through structural changes unrelated to disruption of covalent

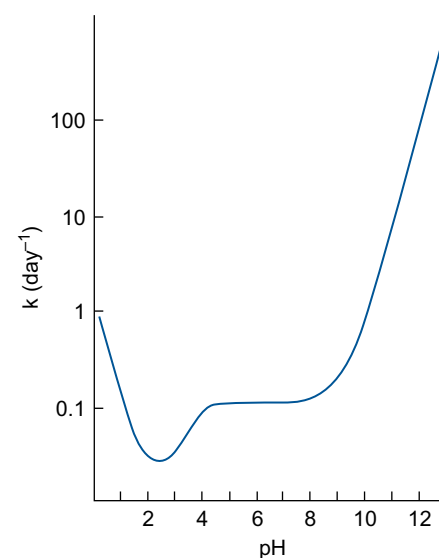


Figure 17-12. Apparent first-order rate of hydrolysis of aspirin as a function of pH. (From Edwards LJ. The hydrolysis of aspirin. A determination of the thermodynamic dissociation constant and a study of the reaction kinetics by ultra-violet spectrophotometry. *Trans Faraday Soc.* 1950; 46: 723–35. Reproduced by permission of the Royal Society of Chemistry.)

bonds. Protein-based drugs may lose activity as a result of a change in conformation (secondary, tertiary, and quaternary structures) that is independent of chemical modification. Conformational changes, which may alter protein drug activity, include denaturation (unfolding), aggregation, surface adsorption, and precipitation. Treatments with potential for inducing such changes include temperature changes, pH extremes, and agitation, foaming resulting from shaking and conditions which may expose drugs, such as proteins, to manufacturing or administering processes that cause shear. Detection of instability of a physical nature generally requires one or more biological assays, or physical assay methods that are sensitive to the critical conformational changes.

DRUG STABILIZATION

Some drug decomposition reactions, such as photolytic and oxidative reactions, are relatively easy to avoid by protecting the components from light (photodecomposition) or exclusion of oxygen and by use of chain-terminating reagents or free-radical scavengers to minimize free-radical-mediated reactions. Solvolysis reactions, however, cannot be stopped by such procedures, but several techniques may be employed to retard reactions sufficiently to permit the formulation of a suitable drug product. The following approaches may be useful in attempts to retard solvolytic reactions.

SELECTION OF OPTIMUM pH, BUFFER, AND SOLVENT

Consideration of the mechanism of the reaction and the way in which the reaction rate is influenced by pH, buffer species, and solvent permits the selection of the optimum conditions for drug stability. Often, however, ideal conditions for maximum stability may be unacceptable from the viewpoint of pharmaceutically acceptable formulation or therapeutic efficacy; thus, it may be necessary to prepare a formulation with conditions less than optimum for drug stability. If a suitable compromise between conditions for maximum stability and conditions for a pharmaceutically acceptable formulation cannot be achieved, techniques, such as those described below, may be useful in retarding solvolysis reactions.

SPECIFIC COMPLEXING AGENTS

The technique of stabilization by forming complexes in solution was introduced by Higuchi and Lachman⁹ who demonstrated that the rate of hydrolysis of the ester function of benzocaine was retarded significantly in the presence of caffeine, a reagent with which the benzocaine formed a soluble complex. It was demonstrated further that, in these systems, the complexed drug did not hydrolyze at all and that the observed rate of hydrolysis could be ascribed to the concentration of the free or uncomplexed drug that was in equilibrium with the drug complex.

There are many recognized examples in which drug stabilization occurs by disrupting the kinetic mechanisms of degradation. Boric acid chelation of the catechol function of epinephrine stabilizes epinephrine against attack by bisulfite and sulfite. The complex of povidone (polyvinylpyrrolidone) and iodine was used for many years as a topical antiseptic because of its higher iodine concentration, slow release of iodine from the complex, lower toxicity, and its ability to stabilize and protect the iodine from degradation before application.

SURFACTANTS

It has been demonstrated that the incorporation of benzocaine into surfactant micelles could retard significantly the rate of ester hydrolysis. Nonionic and anionic surfactants retarded the hydroxide-ion-catalyzed hydrolysis, but cationic surfactants somewhat increased the rate of hydroxide-ion-catalyzed hydrolysis. Similar observations have been reported for a number of drugs that are sufficiently lipophilic to be solubilized by surfactant micelles.

SUSPENSIONS

If the solubility of a labile drug is reduced and the drug is prepared in a suspension form, the rate at which the drug degrades will be related only to the concentration of dissolved drug rather than to the total concentration of drug in the product. Thus it has been demonstrated that penicillin G procaine suspensions degraded at a rate proportional to the low concentration of penicillin in solution. Because the penicillin in solution was in equilibrium with excess solid penicillin G procaine, the penicillin concentration in solution was constant and the observed order of reaction was apparently zero order.

REFRIGERATION

Storage below room temperature usually will retard solvolytic reactions. Storage in the frozen state generally is an effective means of retarding degradative reactions. Several antibiotics are sold as frozen solutions in flexible plastic bags. An exception is sodium ampicillin dissolved in 5% dextrose solution, which showed approximately 10% decomposition after 4 h of storage at 5°C and more than 13% loss after storage for the same period in the frozen state at -20°C.

STABILITY TESTING OF PHARMACEUTICAL PRODUCTS

If a product is to be marketed, it must be stable over relatively long storage times at room temperature or at the actual temperature at which it will be shipped and stored prior to its ultimate use. Thus, the rate of degradation may have to be studied over an undesirably long period of time in order to determine the product's stability under normal storage conditions.

To avoid this undesirable delay in evaluating possible formulations, the manufacturer attempts to predict stability under conditions of room temperature or actual storage conditions by using data for the rate of decomposition obtained at several elevated temperatures. This is accomplished using an Arrhenius plot to predict, from high-temperature data, the rate of product breakdown to be expected at actual lower temperature storage conditions. See Chapter 4.

Prediction based on data obtained at elevated temperatures generally is satisfactory for solution dosage forms. Success is more uncertain when non-homogeneous products are involved. Suspensions of drugs may not provide linear Arrhenius plots because often there is the possibility that the solid phase, which exists at elevated temperature, may not be the same solid phase that exists at room temperature. Such differences in the solubility of the several solid phases may invalidate the usual Arrhenius plots due to a change in mechanism as the phase changes with increase in temperature. These difficulties should be anticipated when polymorphic crystal forms or several different solvates are known to exist for a specific solute. Also, when solid dosage forms (e.g., tablets) are subjected to high temperatures, changes in the quantity of moisture in the product may greatly influence the stability of the product.

Arrhenius plots also suffer limitations when applied to reactions that have relatively low activation energies and, therefore, are not accelerated greatly by an increase in temperature. Where usually it is desirable to determine drug stability by analyzing samples for the amount of intact drug remaining—in instances where there is very little drug decomposition and particularly when it is not convenient to accelerate the reaction by increasing temperature—it sometimes is advantageous to determine initial reaction rates from the determination of the amount of reaction product formed.

Using modern methods of analysis, such as high-performance liquid chromatography (HPLC), it is often possible to measure the rate of formation of a degradation product. By using this technique, very small amounts of degradation (less than 1% loss of parent compound) can be detected, resulting in a more sensitive indication of product stability than can be obtained by analyzing potency.

Because manufacturers are interested primarily in the time required to produce just a few-percent breakdown in their product, it is not uncommon to employ terminology such as $t_{0.90}$ or $t_{0.95}$, which is the time required for the drug to decompose to 90 or 95%, respectively, of original potency.

An Arrhenius-type plot, analogous to that illustrated in Figure 17-4, can be obtained by plotting the natural logarithm of the specific rate constants versus the reciprocal of absolute temperature. The rate constant at the temperature of interest can be determined and applied to the appropriate kinetic model to determine the time required for the product to decrease in potency to 90% of original potency.

REFERENCES

1. Garrett ER. In: Bean HS, *et al.* eds. *Advances in Pharmaceutical Sciences*, vol. 2. New York: Academic Press, 1967: Chapter 2.
2. Kondritzer AA, Zvirblis P. Stability of atropine in aqueous solution. *J APhA Sci Ed* 1957; 46:531–535.
3. Connors KA *et al.* *Chemical Stability of Pharmaceuticals: A Handbook for Pharmacists*, 2nd ed., John Wiley & Sons, New York, 1986: 696.
4. Higuchi T *et al.* The kinetics of the hydrolysis of procaine. *J APhA Sci Ed* 1950; 39:405–410.
5. Edwards LJ. The hydrolysis of aspirin. A determination of the thermodynamic dissociation constant and a study of the reaction kinetics by ultra-violet spectrophotometry. *Trans Faraday Soc* 1950; 46:723–735.
6. Fersht AR, Kirby AJ. The hydrolysis of aspirin. Intramolecular general base catalysis of ester hydrolysis. *J Am Chem Soc* 1967; 89:4857–4863.
7. Whitworth CA *et al.* Stability of aspirin in liquid and semisolid bases. I. Substituted and nonsubstituted polyethylene glycols. *J Pharm Sci* 1973; 62:1184–1185.
8. Yamana T *et al.* Physicochemical properties of beta-lactam antibacterials: deuterium solvent isotope effect on penicillin G degradation rate. *J Pharm Sci* 1977; 66:861–866.
9. Higuchi T, Lachman L. Inhibition of hydrolysis of esters in solution by formation of complexes. I. Stabilization of benzocaine with caffeine. *J APhA Sci Ed* 1955; 44:521–526.

BIBLIOGRAPHY

- Carstensen JT and Rhodes CT. *Drug Stability, Principles and Practices*. New York: Dekker, 2000.
- Connors KA *et al.* *Chemical Stability of Pharmaceuticals*, 2nd edn. New York: Wiley, 1986.
- Espenson JH. *Chemical Kinetics and Reaction Mechanisms*, 2nd edn. New York: McGraw-Hill, Inc., 1995.
- Fung HL. In: Banker GS, Rhodes CT, eds. *Modern Pharmaceutics*, 2nd ed. New York: Dekker, 1990: Chapter 6.
- House JE. *Principles of Chemical Kinetics*. London: Elsevier, 2007.
- Houston, PL. *Chemical Kinetics and Reaction Dynamics*, Boston: McGraw-Hill, Inc., 2001.
- Lachman L *et al.*, eds. *The Theory and Practice of Industrial Pharmacy*, 3rd ed. Philadelphia: Lea & Febiger, 1986: Chapter 26.

Chapter 18

Complex Formation

Thorsteinn Loftsson, MSPharm, MSc, PhD and Marcus E. Brewster, PhD

COMPLEX BONDING 317

METAL-ION COORDINATION COMPLEXES 318

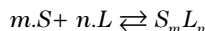
MOLECULAR COMPLEXES 319

COMPLEX STABILITY 320

CYCLODEXTRINS 323

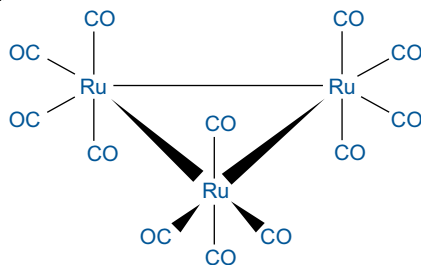
COMPLEXES IN PHARMACY 325

In chemistry and chemical processes the word *complex* usually refers to molecules or molecular assemblies formed by combination of *substrates*, *S*, and *ligands*, *L*. Most often complex (S_mL_n) formation is a reversible process:



where *m* substrate molecules, associate with *n* ligand molecules to form a complex of *m:n* stoichiometry. In this context, *complex formation*, *complexation*, *binding*, *association*, and *chelation* are often synonymous. The substrate and ligand are kept together by relatively strong coordinate covalent bonds or by weak non-covalent forces such as hydrogen bonding, van der Waals forces, electrostatic interactions, dipole forces, charge transfer, release of conformational strain, or hydrophobic interactions. The complex formation changes the physicochemical properties of its constituents, both of the substrate and the ligand, including their aqueous solubility, molar absorptivity, NMR chemical shifts, adsorption to solid surfaces, partitioning behavior, conductivity, chemical reactivity and/or pKa values. By studying such properties, for example of the substrate as a function of the ligand concentration, the complex can be identified and quantitatively described. Furthermore, the methods of chemical kinetics and thermodynamics can be applied to describe the formation and dissociation of a complex. Although most frequently, substrate and ligand molecules are associated by weak chemical forces, there are complexes where bonds are quite strong and formation of some metal complexes are virtually irreversible. Complexes are usually broadly classified into two groups based on the type of S-L bonding involved, namely *coordination complexes* and *molecular complexes*.

Coordination complexes consist of ionic substrates, most frequently transition metal ions, with bases or, in other words, products of Lewis acid-base reactions where the metal ion (an acid) accepts a pair of electrons from the ligand (the base) to form a coordinate covalent bond. Examples of such complexes are $[Ag(NH_3)_2]^+$, $[Co(NH_3)_6]^{3+}$ and $[Fe(CN)_6]^{4-}$. Other common types of coordination complexes are organometallic complexes that are complexes formed between organic groups and metal atoms such as vitamin B12 (a porphyrin containing a cobalt atom), and cluster complexes where the central metal ion consists of a three-dimensional cell of several directly bonded metal ions such as triruthenium dodecacarbonyl ($Ru_3(CO)_{12}$):



A **coordination complex** is called a *chelate* if the same substrate (metal ion) binds with two or more sites on a ligand.

Molecular complexes consist of non-covalently bound substrates and ligands. These include complexes of relatively small substrates and ligands such as drug-cyclodextrin complexes,

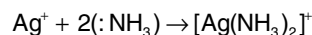
complexes between small substrates and a large ligand such as drug-protein complexes (e.g., plasma protein binding) and complexes between large substrates and a small ligand such as some protein-polyalcohol complexes. Molecular complexes also include molecular dimers, ion-pairs, intramolecular interactions (such as base-base interactions in the DNA helix) and clathrate compounds (cage compounds) where a cage-like lattice of one type of molecules (e.g., hydroquinone molecule) entraps a second type of molecules (e.g., methanol molecule). Pharmaceutical co-crystals can also be considered a type of molecular complex with components subjected to hydrogen bonding and other forces in the crystal lattice. Examples of these types of molecular complexes include isoniazid and 4-aminosalicylic acid where the carboxylic acid function can be shown to interact with the pyridine nitrogen.

The following is a more detailed description of these complexes with several examples of pharmaceutical interest.

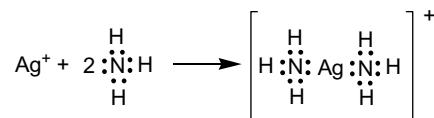
COMPLEX BONDING

Chemical bonding and bonding forces within and between molecules is covered in Chapter 11. The following relates to forces that participate in complex formation.^{1,2}

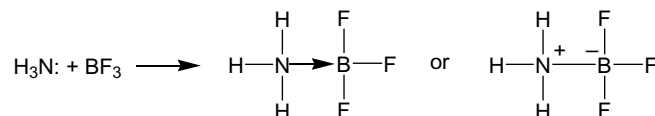
A **coordinate covalent bond** is a covalent bond where both electrons come from the same atom unlike a simple covalent bond which is formed between two atoms with each atom supplying one electron. For example, when silver ions (Ag^+) interact with ammonia (NH_3) to form the silver-ammonia coordinate complex ($[Ag(NH_3)_2]^+$) both of the electrons forming the covalent bond come from ammonia. However, each of the three H-N covalent bonds in ammonia is formed by one electron from a hydrogen atom and one electron from the ammonia atom.



or



Coordinate covalent bonds are sometimes designated by arrows indicating that both electrons are coming from, for example, the nitrogen or by plus and minus signs indicating that the nitrogen end of the bond has become positive because the electron pair has moved away from the nitrogen towards the boron which has thus become negative:



Several different theories have been developed in order to describe coordinate covalent bonding such as *molecular orbital theory* (also called *ligand field theory*) that applies quantum mechanical description of coordinate bonds, *electrostatic theory* where the substrate ions are treated as spherical charges

Table 18-1. Some examples of the Hard-Soft Classification of Lewis Acids and Bases

	Acids	Bases
Hard	H ⁺ , Li ⁺ , Na ⁺ , K ⁺ , Mg ²⁺ , Ca ²⁺ , Mn ²⁺ , Al ³⁺ , Zn ²⁺ , AlCl ₃ , CO ₂	H ₂ O, OH ⁻ , F ⁻ , Cl ⁻ , PO ₄ ³⁻ , SO ₄ ²⁻ , ClO ₄ ⁻ , NO ₃ ⁻ , NH ₃ , CH ₃ COO ⁻
Soft	Cu ⁺ , Ag ⁺ , Au ⁺ , Hg ₂ ²⁺ , Hg ₂ ²⁺ , Pd ²⁺ , Pt ²⁺ ; I ₂ , Br ₂	I ⁻ , SCN ⁻ , CN ⁻ , CO, C ₆ H ₆

and the ligands as dipoles, and *valence bond theory*, a quantum mechanical theory wherein the ligand donates a pair of electrons to a vacant orbital associated with the ion.

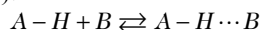
Another concept that is used to describe coordination complexes is the concept of *hard and soft acids and bases* (HSAB). A *hard acid* is defined as a small Lewis acid (i.e. electron-pair acceptor) atom with high positive charge density and low polarizability whereas a *soft acid* is large and polarizable. A *hard base* has high electronegativity and low polarizability whereas a *soft base* is polarizable (Table 18-1). Polarizability is a measure of the ease with which the electron cloud can be deformed under the influence of a charge field. In general, hard acids bind preferably with hard bases, and soft acids bind preferably with soft bases.

Non-covalent intermolecular forces (i.e., between molecules) participating in a complex formation are, in general, relatively weak forces compared to intramolecular forces (i.e., within a molecule) such as covalent bonds. The following types of non-covalent forces are known to participate in complex formations:

Electrostatic interactions are the consequence of classical attraction and repulsion effects between charges such as charge-charge, charge-dipole, and dipole-dipole interactions.

Dispersion force is a quantum mechanical effect where synchronization of the electronic motion between two molecules results in momentary dipole moments and consequently attraction between the molecules. Dispersion force is also called the *van der Waals force*. However, some authors use the term *van der Waals force* for all non-covalent forces.

Hydrogen bonding is an interaction involving formation of hydrogen bond (*H-bond*) between a proton donor (*HA*) and a proton acceptor (*B*):



The *A-H* bond has a covalent character but the *H ... B* bond is predominantly electrostatic. The strength of the hydrogen bond is controlled in part by the acid strength of *HA* and the base strength of *B*, but it is also affected by the solvent.

Charge-transfer interaction is a consequence of electron transfer from an electron donor molecule to an electron acceptor molecule. Frequently, charge-transfer complexes involve electron transfer between metal atoms and ligands. However, charge-transfer interactions are also known in other inorganic as well as organic compounds. A well-known charge-transfer complex is the dark blue or purple iodine/starch complex.

Hydrophobic interaction results from the tendency of liquid water to exclude non-polar molecules. The water molecules form a “cage” around two or more non-polar molecules keeping them together in a kind of a complex or loosely associated molecules. Hydrophobic interactions are commonly observed during inclusion complex formations where a non-polar moiety of larger molecule enters a somewhat lipophilic cavity.

Release of conformational strain is known to participate in formation of cyclodextrin inclusion complexes where the cyclodextrin molecule frequently undergoes significant conformation changes upon complex formation. However, it has been argued that relief of conformational strain is not a driving force for cyclodextrin complex formation.²

Release of high-energy water molecules from, for example, a cyclodextrin cavity does participate in inclusion complex formation. In aqueous solutions the somewhat lipophilic cyclo-

dextrin cavity is occupied by high-energy water molecules (i.e., water not capable of fulfilling all of its hydrogen-bonding requirements) that are expelled during inclusion complex formation. However, based on some thermodynamic observations it has been argued that the exclusion of water molecules from the cavity is not a driving force of cyclodextrin complex formation.² This is based on the fact that although the cavity-bound water molecules are of higher energy (i.e., are enthalpy rich) they have more conformational freedom (i.e., form fewer hydrogen bonds). Consequently, although expulsion of water molecules from the cavity is accompanied by a negative enthalpy change, the free energy change of the overall process is not necessarily negative.

METAL-ION COORDINATION COMPLEXES

Metal-ion coordination complexes, sometimes simply called metal complexes, consist of a central metal ion bonded to one or more ligands that are electron-pair donors such as a nitrogenous base (e.g., ammonia), an ion (e.g., chloride ion), or an aromatic compound (e.g., ferrocene). The number of bonds between the metal ion and the ligand or ligands is called the *coordination number* of the complex. Metal ions can have more than one coordination number. The maximum number is defined by the size, charge and electronic structure of the metal ion. Coordination numbers are normally between 2 and 9 with the most common coordination numbers being 4 and 6. For example, the anticancer drug cisplatin has a coordination number of 4 (Fig. 18-1).

The *trans* stereoisomer of cisplatin is transplatin (Fig. 18-1). Both of these geometrical isomers have square planar structure, that is, the metal substrate and the four ligand groups all lie in the same plane. Most metal-ion coordination complexes of coordination number 4 have a square planar structure but some are tetrahedral where the central metal ion is in the center of a tetrahedron with each of the four ligands located in the four corners. Most complexes with coordination number 6 are octahedral, that is, the bonds lie along the x, y, and z axes of a coordinate system with the metal ion at the origin. Example of such a complex are the *cis* and *trans* isomers of dichlorotetraamminecobalt(III) chloride (Figure 18-2).

Metal ions that are found within cells form coordination complexes with small molecules such as porphyrins that are themselves bound to proteins. The function of such metalloproteins can be the transport of oxygen or nitric oxide, or can be enzymatic in nature. For example, the heme unit (Figure 18-3) is a coordination complex of iron and a porphyrin that is responsible for binding oxygen in hemoglobin and myoglobin.

A chelate (from Greek word “chelè” meaning “claw”) is formed when a ligand uses more than one donor atom to bind a single metal atom. Such ligands are called chelating agents, chelants, or chelators. Chelates tend to be more stable than comparable complexes containing only one binding site and are used in drug formulations to, for example, bind metal ions that catalyze drug oxidation, thus increasing the shelf-life of the drug product. Examples of such chelating agents include citric acid,

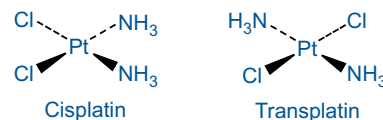


Figure 18-1. Cisplatin and transplatin.

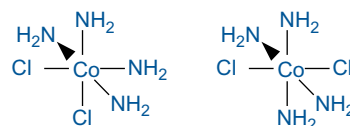


Figure 18-2. Equivalent ways to draw an octahedral complex.

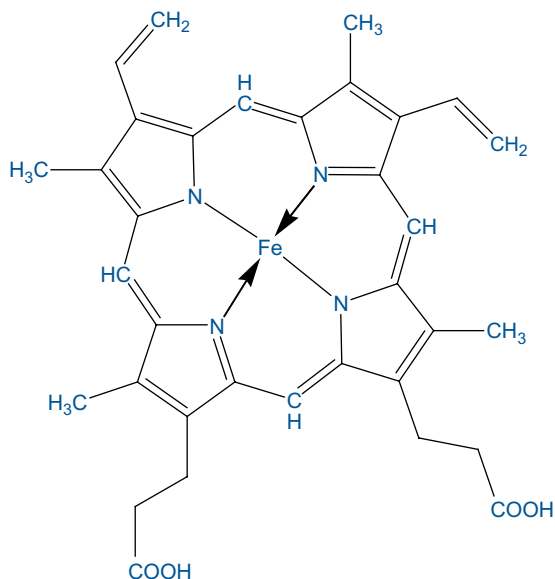


Figure 18-3. Heme, a coordination complex of iron and a porphyrin. Two of the bonds are coordination covalent bonds indicated by arrows.

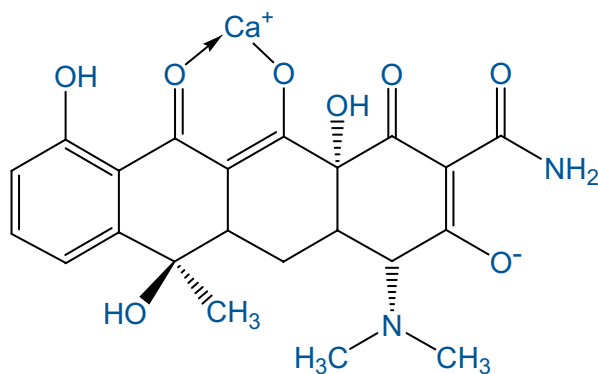


Figure 18-4. Calcium (Ca^{2+}) tetracycline complex (chelate).

tartaric acid and EDTA (ethylenediamine tetraacetic acid). Some drugs can also be chelating agents and bind ions. For instance tetracycline forms hydrophilic chelates with ions such as calcium (Ca^{2+}), iron (Fe^{3+} , Fe^{2+}), aluminium (Al^{3+}), and magnesium (Mg^{2+}), i.e. complexes that possess poor oral bioavailability (Fig. 18-4).³ Milk and milk products, mineral supplements and antacids containing polyvalent cations ingested simultaneously with tetracycline antibiotics can reduce their oral bioavailability by as much as 90%.⁴ Nalidixic acid, ciprofloxacin and other quinolones do also form chelates with polyvalent ions that can reduce their oral bioavailability. The intensity of the effect depends both on the nature of the drug and the cation, as well as on the doses used. The drugs that bind metal ions should be taken either two to three hours after or before ingestion of cation containing products such as dairy products, antacids and mineral supplements.

β -Lactam antibiotics form chelates with metal ions such as Cu^{2+} (Fig. 18-5). The β -lactam ring is much more susceptible to specific base hydrolysis (i.e., toward OH^- attack) when complexed than when uncomplexed. Thus, formation of such chelates can significantly decrease the shelf-life of β -lactam antibiotics.

MOLECULAR COMPLEXES

Molecular complexes can be classified according to 1) the bonding or interaction between substrate and ligand (e.g.,

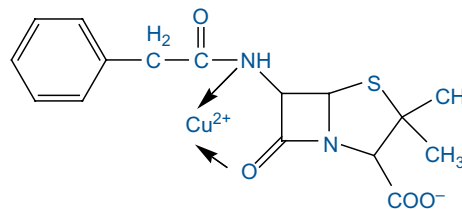


Figure 18-5. Copper (Cu^{2+}) benzylpenicillin complex (chelate).

electrostatic interaction, charge-transfer, hydrogen bonding and hydrophobic interaction), 2) type of substrate and ligand forming the complex (e.g., small molecule—small molecule, small molecule—macromolecule, enzyme—substrate, drug—receptor and antigen—antibody) or 3) type of structure formed (e.g., self-assembled aggregate, micelle, clathrate, and inclusion complex). Molecular complexes consist of one or more substrates and ligands that are, in general, held together by relatively weak, non-covalent forces. In aqueous solutions, free molecules are most often in dynamic equilibrium with molecules bound within the complex but tightly bound complexes are not unknown. Much of the pioneering work on molecular complexes was published by Takeru Higuchi and coworkers.⁵

Drug-exciptent interactions are quite common in pharmaceutical formulations and frequently such interactions are the result of formation of molecular complexes. Formation of such complexes will affect the physicochemical and biological properties of the drug bound to the excipient such as its aqueous solubility, release from the drug formulation and bioavailability.

Non-ionic water-soluble polymers, such as polyvinyl pyrrolidone (PVP) and hydroxypropyl methylcellulose (HPMC), are commonly used to enhance viscosity and to form hydrogels. In aqueous solutions, PVP forms coils and is able to bind drug molecules via non-covalent binding, and PVP is known to form complexes with various drugs.⁶ PVP binds iodine in the presence of iodide ions (povidone-iodine) and as such is used as antibacterial agent for treatment and prevention of skin or wound infections (Fig. 18-6). Both PVP and HPMC can enhance aqueous solubility of drugs through the formation of water-soluble complexes and have been used as solubility and dissolution enhancers.^{7,8}

Ionic water-soluble polymers, such as carboxymethylcellulose (CMC), are able to bind drugs. For example, in aqueous solutions CMC forms complexes with some basic drugs such as atenolol, diphenhydramine, lidocaine and propranolol.⁹ Gum Arabic and alginates, both of which are anionic polysaccharides, are known to interact with drugs affecting their aqueous solubility and chemical stability. Chitosan (a linear polysaccharide composed of randomly distributed glucosamine) is positively charged at acidic pH and forms nano-sized complexes with negatively charged DNA and RNA. Chitosan has been used in gene delivery.¹⁰ Alginates form drug complexes and have also been used to modify drug delivery.¹¹

Drug-drug interactions in the form of complexes are also well known. Thus, salicylates form complexes with benzocaine, and the anticancer drugs paclitaxel, doxorubicin, and etoposide have been shown to form dimers and trimers in aqueous solutions as well as etoposide-paclitaxel complexes.¹² Cocaine and morphine form a binary complex in aqueous solutions that

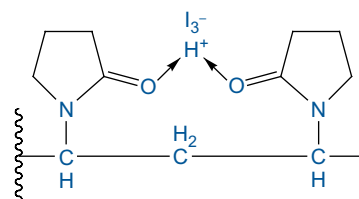


Figure 18-6. The iodine—PVP complex.

can affect their pharmacokinetics and biological activity.¹³ β -Lactam antibiotics form complexes with aminoglycosides,¹⁴ heparin forms complexes with drugs such as polymyxin B and streptomycin, and caffeine and nicotinamide form water-soluble complexes with a number of drugs and vitamins.^{5,15,16}

Drug-food interactions can also be due to drug complexation.¹⁷ Dietary fibers reduce oral drug absorption by drug complexation.¹⁸ These material can include psyllium, guar gums, and other carbohydrates such as cyclodextrins. Cholestyramine and related bile acid sequestrates interact with bile acids to reduce their ability to be reabsorbed and thereby accelerating cholesterol elimination. These are strong ion exchange resins which interact with substrates through electrostatic forces to form the desired complexes.

INCLUSION COMPLEXES

Inclusion complexes are molecular complexes characterized by entrapment of the substrate (guest) in a cage (host) consisting of one or more ligand molecules. Cyclodextrin complexes are examples of guest-host complexes where a single ligand molecule entraps one or more guest molecules. Some other relatively large molecules, such as, water-soluble dextrans, crown ethers and calixarenes, are also capable of forming this type of monomolecular inclusion complex.^{19,20} In contrast several small molecules, like hydroquinone, bile acids, thiourea, and urea assemble to form channels (tunnels), clathrates (cages), molecular sandwiches or other types of hydrophilic nanostructures around lipophilic molecules. For example, hydroquinone molecules form a cage like structure around small molecules like methanol through hydrogen bonding of the hydroquinone molecules whereas urea can form channels around straight-chain hydrocarbons and 13-*cis*-retinoic acid.^{21,22}

COMPLEX STABILITY

STOICHIOMETRY AND BINDING CONSTANTS

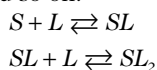
The two most important characteristics of complexes are their stoichiometry and the numerical values of their stability constants. If m substrate molecules (S) associate with n ligand molecules (L) to form a complex (S_mL_n) the following overall equilibrium is attained:



where $K_{m,n}$ is the stability constant (also known as binding constant, formation constant or association constant) of the substrate-ligand complex. The stability constant can be written as:

$$K_{m,n} = \frac{[S_mL_n]}{[S]^m \cdot [L]^n} \quad (2)$$

where the brackets denote molar concentrations. In general, higher order complexes are formed in a stepwise fashion where a 1:1 complex is formed in the first step and 1:2 (or 2:1) complex in the next step and so on:



Consequently the stability constants can be written as:

$$K_{1:1} = \frac{[SL]}{[S] \cdot [L]} \quad (3)$$

$$K_{1:2} = \frac{[SL_2]}{[SL] \cdot [L]} \quad (4)$$

The **reciprocal values** (e.g., $1/K_{1:1}$) are known as dissociation constants or instability constants. The physicochemical properties of free substrate molecules are somewhat different from those of the compound bound to the ligand molecules. Likewise, the physicochemical properties of free ligand molecules are different from those of the ligand in the complex. In general, any

methodology that can be used to observe these changes in additive physicochemical properties may be utilized to determine the stoichiometry of the complexes formed and the numerical values of their stability constants. These include changes in solubility, changes in chemical reactivity, changes in UV/VIS absorbance, changes in fluorescence, NMR chemical shifts, changes in substrate retention (e.g., in liquid chromatography), changes in pKa values, potentiometric measurements and effects on substrate permeability through artificial membranes. Furthermore, because complexation will influence the physicochemical properties of the complexation media, methods that monitor these media changes can also be applied to study the complexation. These include measurements of conductivity changes of an aqueous medium, determinations of freezing point depression, viscosity measurements and solution calorimetric titrations. The stoichiometry of the substrate/ligand complexes and the numerical values of their stability constants should be independent of the methodology applied and sometimes that is the case. In some cases, method-dependent differences in the stability constants can be discerned. These inconsistencies can be explained by coexistence of several different types of substrate-ligand complexes in the complexation media and concentration dependent changes in their relative contribution to the overall complexation effect.

The van't Hoff equation is used to explain the effects of temperature on equilibrium constants such as stability constants of complexes:

$$\ln K = \text{Intercept} - \frac{\Delta H^0}{R} \cdot \frac{1}{T} \quad (5)$$

where ΔH^0 is the standard enthalpy change, R is the gas constant and T is the temperature in Kelvin. A plot of T^{-1} versus $\ln K$ will give a straight line with slope equal to $-(\Delta H/R)$. The standard free energy change (ΔG) for equilibrium processes is calculated from Eq. 6:

$$\Delta G^0 = -R \cdot T \cdot \ln K \quad (6)$$

with the standard free entropy change (ΔS^0) derived from the following relationship:

$$\Delta G^0 = \Delta H^0 - T \cdot \Delta S^0 \quad (7)$$

DETERMINATION OF BINDING CONSTANTS

Various analytical methods can be applied to obtain the stoichiometry and numerical values of a given substrate-ligand complex. The following are descriptions of just a few common methods that are applied during drug formulation studies. However, it must be emphasized that most of these methods were developed under ideal conditions (e.g., in dilute aqueous solutions) or conditions that rarely can be found in pharmaceutical formulations. Thus, the determined values should be characterized as observed and not exact values (i.e., observed stoichiometry, observed stability constant etc.).

The solubility method

The most common application of complexes in pharmaceutical formulations is their usage as solubilizers of poorly soluble drugs. In such cases, the stoichiometry of the substrate-ligand complexes and the numerical values of their stability constants are frequently obtained from phase-solubility diagrams, i.e., plots of substrate solubility vs. ligand concentration. The phase-solubility technique was developed by Higuchi and Connors and it is based on research related to how complexes of different ligands such as caffeine, polyvinylpyrrolidone, and some aromatic acids affects the aqueous solubility of substrates.^{5,23,24} Experimentally, an excess of a poorly water-soluble substrate is introduced into several vials to which a constant volume of an aqueous vehicle containing successively higher concentrations of ligand are added. The vials are shaken or otherwise agitated at constant temperature until equilibrium is established. The suspensions are then filtered and the total concentration of dissolved substrate ($[S]_T$) determined based on an appropriate

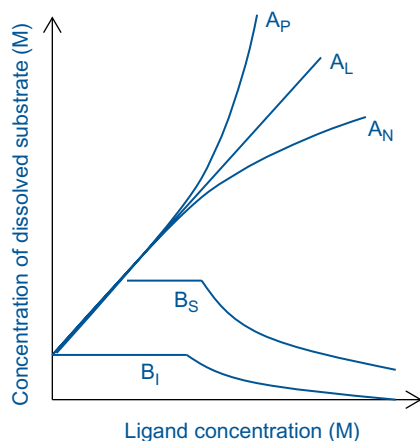


Figure 18-7. Schematic presentation of A- and B-type phase-solubility diagrams with applicable subtypes (A_P , A_L , A_N , B_S , and B_I).

analytical technique. The phase-solubility profile is then constructed by assessing the effect of the ligand concentration on the apparent solubility of the substrate. The phase-solubility method does not give insight as to how the complexes are formed or as to the forces involved. To study these types of effects, only methodologies that monitor physicochemical properties of the included substrate molecule or the ligand molecule can be applied, such as changes in the UV-visible or NMR spectra. Also, due to the coexistence of several different types of substrate–ligand complexes, the stability constants obtained from the phase-solubility profiles are observed constants. Based on the shape of the generated phase-solubility relationships, several types of behaviors can be identified.²⁵ Phase-solubility diagrams fall into two major types: A and B (Fig. 18-7).

A-TYPE PROFILES

In A systems, the apparent solubility of the substrate increase as a function of ligand concentration. Three subtypes have been defined: A_L profiles indicate a linear increase in solubility as a function of ligand concentration, A_P systems indicate an isotherm wherein the curve deviates in a positive direction from linearity (i.e., the ligand solubilization is proportionally more effective at higher concentrations) and A_N relationships indicate a negative deviation from linearity (i.e., the ligand solubilization is proportionally less effective at higher concentrations). Taken as a whole, these isotherms indicate that water-soluble complexes are being formed with solubilities higher than that of the uncomplexed substrate. A_L -type relationships are first order with respect to ligand and first or higher order with respect to the substrate, that is, SL , S_2L , S_3L , and so on. If the slope of the A_L isotherm is greater than unity, higher order complexes are assumed to be involved in the solubilization. Although a slope of less than one does not exclude the occurrence of higher order complexes, a one-to-one complex is often assumed in the absence of other information. A_P systems suggest the formation of higher order complexes with respect to the ligand at higher ligand concentrations, for example, SL_2 , SL_3 , and so on. The stoichiometry of the formed complexes has historically been implied by the extent of curvature of the phase-solubility profile. Thus, an isotherm best fit to a quadratic function suggest the formation of a one-to-two (SL_2) complex, one best fit to a cubic function suggests a one-to-three complex (SL_3), and so forth. A_N profiles have several explanations including bulk changes imparted to the solvent by the ligand at various concentrations, that is, the ligand is acting as a chaotrope or kosmotrope or is altering the bulk properties of the media by changing its viscosity, surface tension, or conductivity and/or self-association of the ligand at high concentrations.

Equilibrium constants can be derived from A-type phase-solubility profiles in a number of ways. As discussed previously, the equilibrium constant for a complexation of interest is given by Equation 1. The intrinsic substrate solubility is given as S_0 and a formed complex is represented by SL :

$$[S] = S_0 \quad (8)$$

$$[S]_T = S_0 + m \cdot [S_m L_n] \quad (9)$$

$$[L]_T = [L] + n \cdot [S_m L_n] \quad (10)$$

where $[S]$ is the concentration of free substrate (i.e., dissolved substrate that is not bound to ligand), $[S]_T$ is the total concentration of dissolved substrate (i.e., both free and bound to a water-soluble complex), $[L]_T$ is the total concentration of ligand in solution and $[L]$ is the concentration of free ligand (i.e., dissolved ligand molecules that are not forming complex with the substrate). The values for $[S_m L_n]$, $[S]$ and $[L]$ can be derived as:

$$[S_m L_n] = \frac{[S]_T - S_0}{m} \quad (11)$$

$$[L] = [L]_T - n \cdot [S_m L_n] \quad (12)$$

where S_0 is the solubility of the substrate in the absence of the L . For equilibria that are first order with respect to the ligand ($n = 1$), the following equation can be obtained:

$$[S]_T = \frac{m \cdot K \cdot (S_0)^m \cdot [L]_T}{1 + K \cdot (S_0)^m} + S_0 \quad (13)$$

A plot of $[S]_T$ versus $[L]_T$ for the formation of $S_m L$ complex should, therefore, give a straight line with the y-intercept representing S_0 and the slope defined as:

$$\text{Slope} = \frac{m \cdot K \cdot (S_0)^m}{1 + K \cdot (S_0)^m} \quad (14)$$

Therefore, if m is known, the K can be calculated, meaning that for one-to-one complexation, that is, $m = 1$, the following equation will apply:

$$K_{1:1} = \frac{\text{Slope}}{S_0 \cdot (1 - \text{Slope})} \quad (15)$$

It should be noted that in the circumstance where a series of complexes of the form $SL + S_2L + \dots + S_m L$ are present, an A_L -type profile would still be observed and simple phase-solubility analysis would not be able to distinguish between the simple 1:1 (one-to-one) and higher order complexes. If the slope of the isotherm is greater than unity, higher order complexes are indicated and equation 13 can be used with substitution of various trial values of m to suggest the stoichiometry and magnitude for the equilibrium constant.

Frequently, the y-intercept is not equal to S_0 and this can cause considerable error in the value of K . A more accurate method for determination of the solubilizing effect of ligands is to determine their complexation efficiency (CE), i.e., the concentration ratio between ligand in a complex and free ligand. CE is calculated from the slope of the phase-solubility diagrams, it is independent of both S_0 and the intercept, and it is more reliable when the influences of different pharmaceutical excipients on the solubilization are being investigated. For 1:1 SL complexes the complexation efficiency (CE) can be calculated from the slope of the phase-solubility diagram:

$$CE = \frac{[SL]}{[L]} = S_0 \cdot K_{1:1} = \frac{\text{Slope}}{(1 - \text{Slope})} \quad (16)$$

And the substrate:ligand molar ratio can be calculated from the CE :

$$S : L \text{ molar ratio} = 1 : \frac{(CE + 1)}{CE} \quad (17)$$

For A_P -type profiles, the equilibrium constants can also be calculated. Equations 3 and 4 apply for a system in which one

substrate molecules forms a complex with two ligand molecules and the following mass balance equations apply:

$$[S]_T = S_0 + [SL] + [SL_2] \quad (18)$$

$$[L]_T = [L] + [SL] + 2 \cdot [SL_2] \quad (19)$$

Equations 3, 4, 18, and 19 can be combined and converted into the following quadratic relationship:

$$[S]_T = S_0 + K_{1:1} \cdot S_0 \cdot [L] + K_{1:1} \cdot K_{1:2} \cdot S_0 \cdot [L]^2 \quad (20)$$

indicating that a plot of $[S]_T$ versus $[L]_T$ (assuming that $[L] \approx [L]_T$) fitted to the quadratic relationship will allow for the estimation of $K_{1:1}$ and $K_{1:2}$. Note that at low L concentrations, $[L]_T$ is sometimes used as an estimate for $[L]$, assuming low CE, meaning that a plot of the ligand concentration versus substrate solubilized can be used to estimate the K values. A linear form of this equation can also be derived:

$$\frac{[S]_T - S_0}{[L]} = K_{1:1} \cdot S_0 + K_{1:1} \cdot K_{1:2} \cdot S_0 \cdot [L] \quad (21)$$

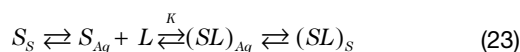
Theoretically, higher order complexation can be further examined with higher order curve fitting meaning that the same formalism can be used when $K_{1:3}, K_{1:4}, \dots, K_{1:m}$ values are present (the cubic equation is given as an example):

$$[S]_T = S_0 + K_{1:1} \cdot S_0 \cdot [L] + K_{1:1} \cdot K_{1:2} \cdot S_0 \cdot [L]^2 + K_{1:1} \cdot K_{1:2} \cdot K_{1:3} \cdot S_0 [L]^3 \quad (22)$$

It should be emphasized that frequently the K -values obtained are apparent values that do not describe the actual process on a molecular level. Coexistence of several different types of ligand complexes and the nonideality of substrate saturated solutions complicate exact determination of the K -values from phase-solubility diagrams. In addition substrate molecules are known to self-associate in aqueous solution and to interact with other pharmaceutical excipients and because these equilibria will reduce the availability of the substrate for complex formation, this may lead to errors in the K -value determinations. Although correlation is often found between phase-solubility diagrams and the stoichiometry of substrate-ligand complexes determined by other means such as NMR, some discrepancies can be found in the literature.²⁶

The origin of A_N -type phase-solubility profiles is uncertain. One possibility is ligand enforced structural changes of the aqueous complexation medium. Another possibility is self-association of ligand molecules at higher concentrations.

B-type profiles indicate formation of complexes with limited water solubility. Two subclasses have been described including B_s and B_l systems (Fig. 18-7). B_s -type isotherms can be interpreted in the following manner. As the ligand concentration increases, a soluble complex is formed which increases the total solubility of the substrate. At a particular point in this solubilization process, the maximum solubility of the substrate is achieved which is the sum of S_0 plus any substrate solubilized in the form of the ligand complex (SL). Additional ligand generates additional complex which precipitates but so long as solid substrate remains, dissolution and complexation can occur to maintain the value of $[S]_T$. During this plateau phase, the following equilibrium is assumed to occur:



where S_s is the solid substrate, S_{Aq} is the dissolved substrate, K is the stability constant of the complex, $(SL)_{Aq}$ is the dissolved complex, and $(SL)_s$ is the solid complex. At some point, all of the solid substrate will have been consumed in the above described process and further addition of the ligand results in the formation of additional insoluble inclusion complex which precipitates and further depletes the total substrate concentration, $[S]_T$. Finally, the solubility observed in the systems is associated with the solubility of the precipitated complex. If the same complex which forms in the ascending portion the phase-solubility profile precipitates in the plateau phase, the increase

in the substrate concentration from S_0 to the plateau should be equal to the intrinsic solubility of the complex.

The kinetic method

The second most common application of complexes in pharmacy is to influence drug kinetics, usually to decrease drug degradation in pharmaceutical formulations. Complexes are always under kinetic equilibrium in aqueous solutions, i.e., the molecules forming a complex are in equilibrium with free molecules in the solution, where the rate of formation and dissociation of the complex is diffusion controlled. In kinetic studies the stability constant ($K_{m:n}$ in Equation 1) can be determined from the stabilizing or destabilizing effects of a ligand on a given substrate. If a ligand has for example stabilizing effect on a substrate molecule in aqueous solution, then the rate of disappearance of the substrate will decrease when the substrate-to-ligand concentration ratio is decreased (i.e., at increasing ligand concentration). In the following, we assume that the substrate degradation is first order and that a 1:1 SL complex is being formed. The first-order rate constant for degradation of the free substrate (k_f) is determined in the aqueous complexation media when no ligand is present. The first-order rate constant for the degradation of the substrate within the ligand complex (k_c) and $K_{1:1}$ can then be determined from the degradation profile (Fig. 18-8) and the observed first-order rate constant (k_{obs}) for the rate of disappearance of the substrate:

$$k_{obs} = k_f \cdot f_f + k_c \cdot f_c \quad (24)$$

where f_f is the fraction of free substrate and f_c is the fraction of substrate in complex. If we assume that only a 1:1 substrate-ligand complex (SL) is being formed and that the total substrate concentration ($[S]_T$) is the sum of the concentration of free substrate ($[S]$) and the concentration of the complex ($[SL]$) the following equations are obtained:

$$[S]_T = [S] + [SL] \quad (25)$$

$$[L]_T = [L] + [SL] \quad (26)$$

$$K_{1:1} = \frac{[SL]}{[S] \cdot [L]} \quad (27)$$

$$f_f = \frac{[S]}{[S] + [SL]} = \frac{1}{1 + K_{1:1} \cdot [L]} \quad (28)$$

$$f_c = 1 - f_f = \frac{K_{1:1} \cdot [L]}{1 + K_{1:1} \cdot [L]} \quad (29)$$

$$k_{obs} = \frac{k_f + k_c \cdot K_{1:1} \cdot [L]}{1 + K_{1:1} \cdot [L]} \quad (30)$$

$$-\frac{d[S]_T}{dt} = k_{obs} \cdot [S]_T = \left(\frac{k_f + k_c \cdot K_{1:1} \cdot [L]}{1 + K_{1:1} \cdot [L]} \right) \cdot [S]_T \quad (31)$$

If the total ligand concentration is much greater than the total substrate concentration ($[L]_T \geq 10 \cdot [S]_T$) then it can be assumed that $[L] \approx [L]_T$:

$$k_{obs} = \left(\frac{k_f + k_c \cdot K_{1:1} \cdot [L]}{1 + K_{1:1} \cdot [L]} \right) \quad (32)$$

Equation 32 can then be rearranged into several different formats including that of the Lineweaver-Burk plot where $(k_f - k_{obs})^{-1}$ versus $([L]_T)^{-1}$ will give a straight line in which k_c can be obtained from the intercept and $K_{1:1}$ from the slope:

$$\frac{1}{k_f - k_{obs}} = \frac{1}{K_{1:1} \cdot (k_f \cdot k_c)} \cdot \frac{1}{[L]_T} + \frac{1}{k_f - k_c} \quad (33)$$

Alternatively, k_c and $K_{1:1}$ can be obtained by simple non-linear fitting of k_{obs} according to equation 32.

UV/vis spectroscopic method

It is also possible to determine the value of $K_{1:1}$ by observing spectrophotometric or spectroscopic changes of the substrate

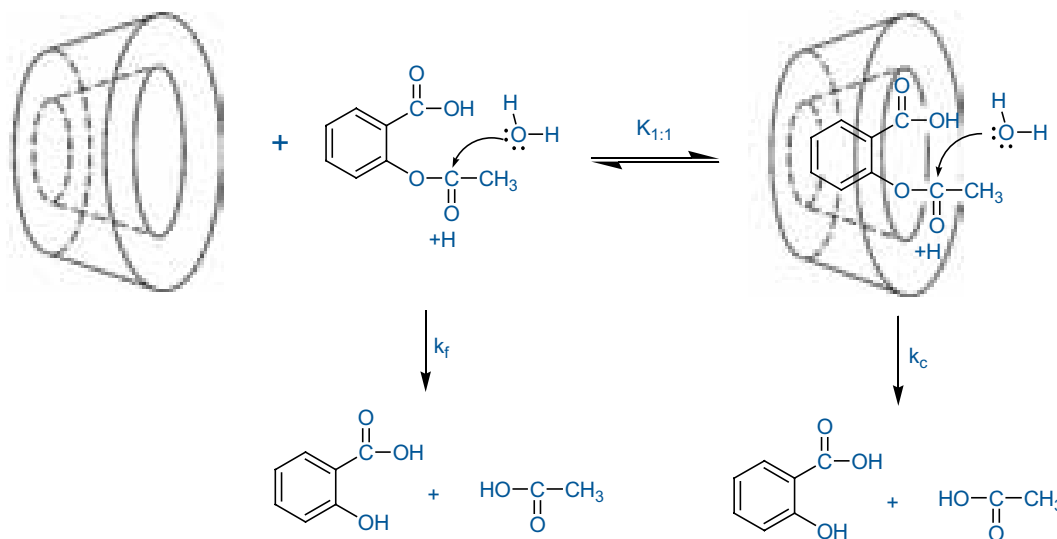


Figure 18-8. Under acidic conditions unionized aspirin forms 1:1 complex with β -cyclodextrin and its derivatives. When cyclodextrin is added to the aqueous aspirin solution the lipophilic aromatic ring will be taken into the cyclodextrin cavity where the ester linkage is more sterically hindered against nucleophilic attack by water molecules than outside it (i.e., $k_f > k_c$).

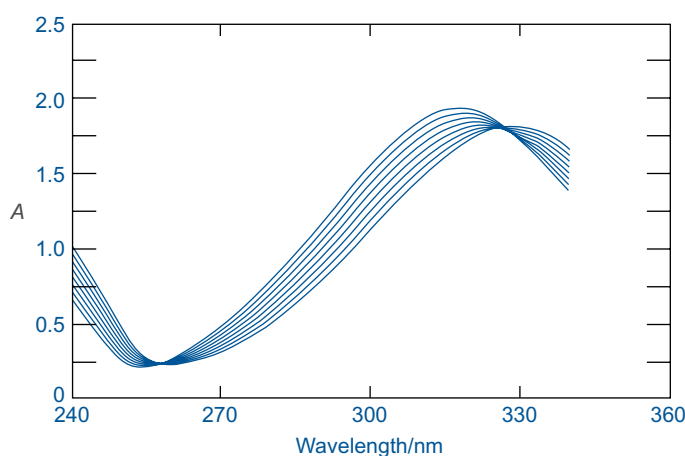


Figure 18-9. Ultraviolet absorption spectrum of *p*-nitrophenol in the presence of varying concentrations of α -cyclodextrin. The *p*-nitrophenol concentration is $1.99 \cdot 10^{-4}$ M, and the cyclodextrin concentration ranges from zero (topmost spectrum) to 0.01 M.

upon binding. Figure 18-9 shows how the ultraviolet spectrum of *p*-nitrophenol changes upon complexation with α -cyclodextrin. The spectrometric changes which occur are similar to those encountered when a substrate is dissolved in solvents of decreased polarity (e.g., water $\rightarrow$ dioxane). In the case of cyclodextrin guest-host complexes, the chromophore of the guest (substrate) is transferred from a polar aqueous environment to a less polar environment within the cyclodextrin cavity (ligand). The presence of well-defined isosbestic points is consistent with the assumption of 1:1 stoichiometry.

A wavelength showing significant changes in absorption is selected and the changes in absorption titrated assuming that Beer's law is obeyed:

$$A_0 = \epsilon_s \cdot b \cdot [S]_T \quad (34)$$

where A_0 is the absorbance of a substrate solution when no ligand is present, ϵ_s is the molar absorptivity of the substrate, b is the path length and $[S]_T$ is the total substrate concentration. In the presence of a ligand the absorbance is:

$$A_L = \epsilon_s \cdot b \cdot [S] + \epsilon_L \cdot b \cdot [L] + \epsilon_{11} \cdot b \cdot [SL] \quad (35)$$

where ϵ_L is the absorptivity of the ligand and ϵ_{11} is the absorptivity of the complex. Combining equation 35 with the mass

balances $[S]_T = [S] + [SL]$ and $[L]_T = [L] + [SL]$ (i.e., Equations 25 and 26) gives:

$$A_L = \epsilon_s \cdot b \cdot [S]_T + \epsilon_L \cdot b \cdot [L]_T + \Delta\epsilon_{11} \cdot b \cdot [SL] \quad (36)$$

where $\Delta\epsilon_{11} = \epsilon_{11} - \epsilon_S - \epsilon_L$. If the solution absorbance is measured against reference solution containing same concentration of ligand ($[L]_T$) but no substrate ($[S]_T = 0$) the measured absorbance will be:

$$A = \epsilon_s \cdot b \cdot [S]_T + \Delta\epsilon_{11} \cdot b \cdot [SL] \quad (37)$$

Combining Equations 3 and 37 gives the Benesi-Hildebrand equation:

$$\frac{b}{\Delta A} = \frac{1}{K_{1:1} \cdot [S]_T \cdot \Delta\epsilon_{11}} \cdot \frac{1}{[L]_T} + \frac{1}{[S]_T \cdot \Delta\epsilon_{11}} \quad (38)$$

where ΔA is the difference in absorbance in the presence and absence of ligand. Most frequently the $[S]_T$ is kept constant whereas $[L]_T$ is varied. Then a plot of $b/\Delta A$ versus $1/[L]_T$ should give a straight line for a 1:1 complex with the ratio intercept/slope yielding $K_{1:1}$.

NMR spectrometry

Changes in the NMR spectra of the substrate and/or the ligand (i.e., chemical shifts, coupling constants, nuclear Overhauser effects, and spin-spin and spin-lattice relaxation times) can be used to probe the solution geometry of complexes as well as to give kinetic information on their association and dissociation. If the changes in chemical shift of, for example, the substrate molecule is titrated then a modified Benesi-Hildebrand Equation 38 can be used:

$$\frac{1}{\Delta\delta} = \frac{1}{K_{1:1} \cdot \Delta\delta_{\max}} \cdot \frac{1}{[L]_T} + \frac{1}{[S]_T \cdot \Delta\delta_{\max}}$$

where $\Delta\delta$ is the change in chemical shift at particular ligand concentration and $\Delta\delta_{\max}$ is the limiting change in chemical shift at infinite ligand concentration (i.e., when all the substrate molecules in the solutions are bound to the ligand). A plot of $1/\Delta\delta$ versus $1/[L]_T$ should give a straight line for a 1:1 complex with the ratio intercept/slope yielding $K_{1:1}$.

CYCLODEXTRINS

Cyclodextrins are cyclic oligosaccharides consisting of (α -1,4)-linked D-glucopyranose units, with a hydrophilic outer surface and a lipophilic central cavity. The natural α -, β -, and γ -cyclodextrins consist of 6, 7, and 8 glucopyranose units, re-

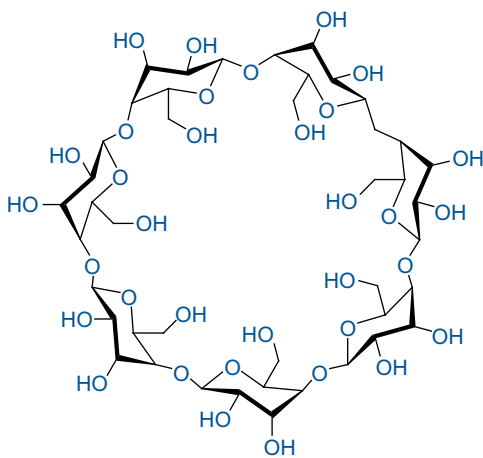


Figure 18-10. The structure of β -cyclodextrin.

spectively (Fig. 18-10 and Table 18-2).^{25,27,28} The diameters of the central cavities are about 5, 6, and 7 Å for α -, β -, and γ -cyclodextrin, respectively. Larger cyclodextrins containing more than eight glucopyranose units do exist but are of limited pharmaceutical interest. Although the natural cyclodextrins and their complexes are hydrophilic, their aqueous solubility can be rather limited. This is thought to be due to relative strong binding of the cyclodextrin molecules in the crystal state. Random substitution of the hydroxy groups can result in dramatic improvements in their aqueous solubility. Cyclodextrin derivatives of pharmaceutical interest include the hydroxypropyl derivatives of β - and γ -cyclodextrin, the randomly methylated β -cyclodextrin and sulfobutylether β -cyclodextrin (Table 18-2). The molar degree of substitution (MS) is defined as the average number of substituents that have reacted with one glucopyranose repeat unit. In an aqueous environment, cyclodextrins form inclusion complexes with many lipophilic molecules through a process in which water molecules located inside the central cavity are replaced by either a whole molecule, or more frequently by some lipophilic structure of the molecule (Fig. 18-8). Cyclodextrin complexation of a drug molecule changes the physicochemical properties of the drug, such as its aqueous solubility, chemical stability and ability to permeate biological membranes. Because the cyclodextrin molecule is hydrophilic on the outside the complex formation usually increases the water-solubility of lipophilic water-insoluble drugs. Once included in the cyclodextrin cavity, the drug molecules may be dissociated from the cyclodextrin molecules through complex dilution or competitive binding by some other suitable molecule (e.g., lipids) or, if the complex is located in close approximation to a lipophilic biological membrane (e.g., mucosa), the drug may be transferred to the matrix for which it has the highest affinity. Importantly, because no covalent bonds are formed or broken during the guest-host complex formation, the complexes are

in dynamic equilibrium with free drug and cyclodextrin molecules. In aqueous solutions the rates for formation and dissociation of drug-cyclodextrin complexes are very close to the diffusion-controlled limits and the complexes are continuously being formed and dissociated.²⁹ For cyclodextrin complexes the value of $K_{1:1}$ (Equations 1 and 15) is frequently between 10^1 and 10^3 M⁻¹ and $K_{1:1}$ greater than 5×10^3 M⁻¹ is rarely observed. The effects of cyclodextrins on drug solubility, bioavailability, chemical stability and delivery through biological membranes have been investigated by a number of research groups.^{28,30}

Cyclodextrins enhance drug delivery through biological membranes by increasing drug permeation through the unstirred water layer that is located adjacent to the membrane surface or, in other words, by increasing the availability of dissolved drug molecules juxtaposed to the membrane surface. Cyclodextrins only enhance drug permeation when a water layer is present at the membrane exterior. Such water layers can consist of mucus or an aqueous vehicle such as o/w creams or hydrogels. Cyclodextrins do not enhance drug permeation from vehicles that do not form an unstirred water layer, such as lipophilic ointments and w/o creams. The effect also depends on the physicochemical properties of the drug. Better enhancement is obtained for lipophilic drugs that are poorly soluble in water that form water-soluble complexes with cyclodextrins with stability constants ($K_{1:1}$) that are between about 50 and 5000 M⁻¹.³⁰

In aqueous solutions free drug molecules are in equilibrium with drug molecules bound in a complex and, thus, the release of drug molecules from cyclodextrin-containing vehicle will depend on the drug/cyclodextrin molar ratio and the $K_{m:n}$ value (Eq. 1). Furthermore, pharmaceutical excipients will affect the $K_{m:n}$ value. Thus it is of utmost importance to optimize the drug vehicle with regard to the amount of cyclodextrin. Too much or too little cyclodextrin will result in less than optimal drug bioavailability.

Recently it has been discovered that cyclodextrin molecules and their complexes self-associate in aqueous solutions to form nanoparticles.³¹ The general observation is that the aggregate formation increases with increasing cyclodextrin concentration. The anomalously low solubility of β -cyclodextrin is explained by the intensity of nanoparticle formation (i.e., aggregation). The same explanation is valid for the observed peculiarity of aqueous γ -cyclodextrin solutions, which are known to become spontaneously turbid at concentrations of about 1% (w/v) or above. The fraction of molecules participating in nanoparticle formation is often very low. For example, the mass contribution of the nanoparticles in aqueous 12 mM α -cyclodextrin solution does not exceed 0.8%, that of β -cyclodextrin is only 0.0011% in 10 mM in β -cyclodextrin solution, and that of γ -cyclodextrin is only 0.02% in 12 mM γ -cyclodextrin solution.³²⁻³⁴ However, formation of cyclodextrin complexes enhances the cyclodextrin aggregation.³⁵ In some cases the mass contribution of the nanoparticles in aqueous complexation media can be well above 50%. Another interesting feature of cyclodextrin nanoparticles is their shape, such as disks, rods and fibers. Thus, cyclodextrins

Table 18-2. The natural cyclodextrins and some of their derivatives.

Cyclodextrin	Synonyms	MS	MW* (Da)	Solubility (mg/ml)
α -Cyclodextrin	alfadex		972.8	130
β -Cyclodextrin	betadex		1135	18.4
2-Hydroxypropyl- β -cyclodextrin	hydroxypropyl betadex	0.65	1400	> 600
Sulfobutylether β -cyclodextrin sodium salt		0.9	2163	> 500
Methylated β -cyclodextrin		1.8	1312	> 600
		0.57	1191	200
γ -Cyclodextrin	gammadex		1297	249
2-Hydroxypropyl- γ -cyclodextrin	hydroxypropylgammadex	0.6	1576	> 600

*The molecular weights (MW) of the cyclodextrin derivatives will depend on their molar degree of substitution (MS), i.e., the number of substituents per glucopyranose repeat unit.

and cyclodextrin complexes form supramolecular complexes at elevated cyclodextrin concentrations.

Due to their favorable toxicological profile cyclodextrins are frequently preferred to organic solvents during *in vitro/in vivo* evaluation of new chemical entities.

Worldwide there are between 30 and 40 different cyclodextrin-containing drug products on the market and few examples are shown in Table 18-3. In most cases cyclodextrins are used as solubilizers, either to enhance dissolution and oral bioavailability of poorly-soluble drugs in solid dosage forms or to replace organic solvents in parenteral dosage forms. Cyclodextrins are also used to increase both chemical and physical stability of drugs, including both peptide and protein drugs, to reduce drug-drug and drug-excipient interactions and to convert liquids to solid powders. Cyclodextrins can also have some adverse effects. For example, excess cyclodextrin can hamper drug absorption from the gastro-intestinal tract and permeation of drugs through skin and other biological membranes. In fact, cyclodextrins have been used to prevent permeation of topically applied sunscreen agents into skin and to reduce absorption of fat from the gastrointestinal tract.

COMPLEXES IN PHARMACY

APPLICATION TO DRUG DELIVERY

Complexation may affect physicochemical and biopharmaceutical properties of drugs, such as aqueous solubility, chemical stability, dissolution rate, partition coefficient, permeability, rate of absorption, bioavailability, biological activity, volatility and physical state (e.g., converting liquid drug into solid drug complex). Pharmaceutical formulators sometimes apply

complexation agents in their formulation design. For example, cyclodextrins and their derivatives are commonly applied functional pharmaceutical excipients with products containing these materials associated with numerous therapeutic categories and administration routes (Table 18-3). These materials are present in formulations intended for parenteral (intravenous, intramuscular, intracavernosal), oral, sublingual, buccal, nasal, ophthalmic, and dermal use. Parenteral use of cyclodextrins include drug products intended to deliver prostaglandins with examples including prostaglandin E1 formulated with α -cyclodextrin (α CD) (alprostadil alfadex). This formulation is used in the treatment of various vascular complications including Buerger's disease. More recently, the material has been shown to be of benefit in male erectile dysfunction when administered intracavernosally. Studies have found that it is useful in patients who do not respond to oral sildenafil (Viagra®, Pfizer) treatment. Although β -cyclodextrin (β CD) is contraindicated parenterally, its derivatives can be safely administered using the oral route. 2-Hydroxypropyl- β -cyclodextrin (HP β CD) is available in several intravenous products including in the Sporanox® IV solution (itraconazole) product. Itraconazole is a triazole-type drug which exerts its effect by inhibiting fungal P450 and inhibiting the biosynthesis of ergosterol, an essential component of the fungal membrane. The compound is noteworthy in that it was the first approved orally bioavailable agent with significant clinical activity against both candidiasis and *Aspergillus spp.*, the two most common human fungal pathogens. Intravenous use of itraconazole is indicated for empiric therapy of idiopathic fever as well as for blastomycosis (pulmonary and extra-pulmonary), histomycosis (pulmonary and disseminated, non-meningeal) and aspergillosis (pulmonary and non-pulmonary). Cyclodextrins were enabling in this

Table 18-3. Examples of cyclodextrin-containing drug products.

Drug/Cyclodextrin	Therapeutic usage	Formulation	Trade Name
<i>α-Cyclodextrin (αCD)</i>			
Alprostadil	To treat erectile dysfunction	IV solution	Caverject Dual
<i>β-Cyclodextrin (βCD)</i>			
Cetirizine	Antihistamine drug	Chewing tablets	Cetirizin
Dexamethasone	Anti-inflammatory steroid	Ointment, tablets	Glymesason
Ethinylestradiol and drospirenone	Birth control	Tablets	Yaz
Iodine	Throat disinfection	Solution	Mena-Gargle
Nicotine	Nicotine replacement product	Sublingual tablets	Nicorette
Nimesulide	Non-steroidal anti-inflammatory drug	Tablets	Nimedex
Omeprazole	To treat gastroesophageal reflux	Tablets	Omebeta
Piroxicam	Non-steroidal anti-inflammatory drug	Tablets, suppository	Brexin
Tiaprofenic acid	Non-steroidal anti-inflammatory drug	Tablets	Surgamyl
<i>2-Hydroxypropyl-β-cyclodextrin (HPβCD)</i>			
Indomethacin	Non-steroidal anti-inflammatory drug	Eye drop solution	Indocid
Itraconazole	Antifungal agent	Oral and i.v. solutions	Sporanox
Mitomycin	Anticancer agent	IV infusion	MitoExtra
<i>Sulfobutylether β-cyclodextrin sodium salt (SBEβCD)</i>			
Aripiprazole	Antipsychotic drug	IM solution	Abilify
Maropitant	Anti-emetic drug (motion sickness in dogs)	Parenteral solution	Cerenia
Voriconazole	Antifungal agent	IV solution	Vfend
Ziprasidone mesylate	Antipsychotic drug	IM solution	Geodon
<i>Randomly methylated β-cyclodextrin (RMβCD)</i>			
Cloramphenicol	Antibacterial agent	Eye drop solution	Clorocil
<i>γ-Cyclodextrin (γCD)</i>			
Tc-99 Teboroxime	Diagnostic aid, cardiac imaging	IV solution	CardioTec
<i>2-Hydroxypropyl-γ-cyclodextrin (HPγCD)</i>			
Diclofenac sodium salt	Non-steroidal anti-inflammatory drug	Eye drop solution	Voltaren Ophtha
Tc-99 Teboroxime	Diagnostic aid, cardiac imaging	IV solution	CardioTec

product due to their solubilizing effect on itraconazole which has an estimated aqueous solubility at neutral pH of about 1 ng/mL. The formulation contains 40% w/v HPβCD and increases the solubility of itraconazole to 10 mg/mL (or approximately 100,000-fold). Based on i.v. doses of itraconazole between 200 and 400 mg and the formulation containing 10 mg/mL itraconazole in a 40% HPβCD solution, the i.v. dose of HPβCD is between 8 and 16 g/day. Another widely used β-cyclodextrin derivative is the sulfobutylether β-cyclodextrin sodium salt (SBEβCD). This anionic excipient is found in a number of parenteral products including the intravenous formulation for voriconazole (Vfend, Pfizer) as well as intramuscular preparations for ziprasidone (Geodon, Zeldox, Pfizer) and aripiprazole (Abilify, BMS). In the case of voriconazole, the cyclodextrin solubilizes the antifungal such that its solubility increases from 0.2 mg/mL at pH 3 in the absence of cyclodextrin to 10 mg/mL using 15% w/v SBEβCD. Based on standard injection doses of the API, 3.5 to 7 g of the cyclodextrin are administered. 2-Hydroxypropyl-γ-cyclodextrin (HPγCD) is also available in an approved intravenous product.

Oral use of cyclodextrins includes products containing αCD, βCD, and HPβCD. An oral formulation has been developed and marketed for itraconazole. This dosage form provides for increased and more consistent oral bioavailability of itraconazole relative to other solid formulation in various subpopulations with a fraction absorbed of 85% and oral bioavailability of 55%. Based on oral doses of 200 mg itraconazole, the dose of HPβCD is 8 g/day. This formulation also allowed the effective treatment of esophageal candidiasis. Solid oral dosage forms containing βCD include tablets containing ethinylestradiol and drospirenone (Yaz®, Bayer), piroxicam (Brexin®, Chiesi) and prostaglandin E2 (Prostarmon E®, Ono). Cyclodextrins are also used to enhance drug absorption from suppositories as well as to increase the bioavailability of drug in eye drops and nasal sprays.^{28,30}

Complexation does not always improve the physicochemical and biopharmaceutical properties of drugs and, as previously mentioned, some *incompatibilities* (i.e., drug-drug and drug-excipient interactions) as well as drug-food interactions are the results of unwanted drug complexation.

Other complexing agents commonly used in drug formulations are, for example, EDTA and citric acid for complexation of metal ions. Metal ions catalyze various drug degradation pathways such as oxidation. However, the complex bound metal ions are inactive. Polyols, such as sorbitol, are added as stabilizers to lyophilized protein drugs. The polyols form complexes with proteins and prevent their self-aggregation.

COMPLEXES IN PHARMACEUTICAL ANALYSIS

Complexation is an essential aspect of many types of bioanalysis. Determination of metal ions is frequently based on metal ion coordination complexes. *Complexometric* (chelometric) titration is often based on EDTA complexation of metal ions. Spectrophotometric (i.e., colorimetric) determinations of metal ions are sometimes based on complex formations. Recently, complexes have been applied in *chiral chromatography*. Cyclodextrins have found numerous applications in this regard in part due to their ability to interact with compounds by complexation as well as their molecular recognition abilities. These properties have allowed cyclodextrin to aid in chiral analysis in gas and high performance liquid chromatography as well as capillary electrophoresis and NMR assessments. Chromatographically, cyclodextrins can be added to the mobile phase or bound to the stationary phase with both approaches being widely exploited. Cyclodextrins are chiral sugars and can be useful adjuncts in circular dichroism measurements due to their ability to induce chirality in non-chiral substrates. Furthermore, cyclodextrins can increase fluorescence quantum yield by shielding the excited singlet state in the central cavity thus protecting it from quenching and related effects. *Phosphorescence* can likewise be allowed in solution by providing

pseudo-ordered stated within the cyclodextrin cavity in which the triplet state can relax.

DRUG-PROTEIN BINDING

After a drug has been absorbed into plasma, or after it has been injected into the blood circulation, the drug molecules are carried by the blood circulation to the target tissue within the body where they bind to receptors. The drug molecules are also carried to the eliminating organs such as the liver and kidneys. Blood plasma and the various body tissues contain proteins that are able to form complexes with drugs. Formation of such complexes is reversible and is normally referred to as drug-protein binding. Irreversible drug-protein binding is usually a consequence of covalent binding of drug molecules to proteins, not complexation, and may account for certain types of adverse effects such as allergy, carcinogenicity, teratogenicity or tissue toxicity. Drug-protein binding does affect the ability of drug molecules to permeate biological membranes and their ability to interact with enzymes and receptors. Only unbound drug molecules permeate membrane barriers and interact with drug receptors or undergo metabolism and glomerular filtration, and, thus, drug-protein binding will affect the drug pharmacokinetics.

Plasma protein binding is the result of complex formation between drug molecules and plasma proteins, such as human serum albumin (HSA), lipoprotein, glycoprotein, and α-, β-, and γ-globulin, of which HSA is the most important and most studied. HSA consists of 585 amino acid residues having calculated molecular weight of 66.47 kDa. Normal plasma concentration of HSA is between 35 and 55 mg/ml but it varies with age, exercise, stress and disease. One of the functions of HSA is to act as a transporter for poorly soluble drugs. However, HSA is very non-specific complexing agent that besides drugs binds water, various cations, fatty acids and thyroid hormones. Also plasma α₁-acid-glycoprotein (α-AGP) has been shown to bind a variety of drugs. Other plasma proteins play a smaller role in drug-protein binding.^{36,37} In general, weak bonds such as hydrophobic bonds, van der Waals dispersion forces, hydrogen bonds, and ionic interactions are involved in the protein binding of drugs. Each protein molecule may possess several binding sites (i.e. bind several drug molecules) and, thus Equation 1 is not precisely applicable but it can be used to explain the basis of binding of drug (*D*) to a protein (*P*) to form a drug-protein complex (*PD*):



If one drug molecule is bound to one protein molecule then (see Equations 28 and 29 for comparison):

$$r = \frac{\text{concentration of bound drug}}{\text{total concentration of protein}} = \frac{[PD]}{[PD] + [P]} \quad (41)$$

$$= \frac{[PD]}{[P]_T} = \frac{K \cdot [D]}{1 + K \cdot [D]}$$

where *K* is the binding constant or the equilibrium constant of the complex formation and [*D*] is the concentration of free drug. However, because each protein has *n* number of independent binding sites we get:

$$r = \frac{n \cdot K \cdot [D]}{1 + K \cdot [D]} \quad (42)$$

Then the large protein molecule may contain more than one type of binding sites:

$$r = \frac{n_1 \cdot K_1 \cdot [D]}{1 + K_1 \cdot [D]} + \frac{n_2 \cdot K_2 \cdot [D]}{1 + K_2 \cdot [D]} + \dots \quad (43)$$

where *n*₁ is the number of binding sites of type 1 with binding constant *K*₁, *n*₂ is the number of binding sites of type 2 with binding constant *K*₂, and so on. If only one type of binding site exists then Equation 42 can be converted to:

$$\frac{1}{r} = \frac{1}{n \cdot K} \cdot \frac{1}{[D]} + \frac{1}{n} \quad (44)$$

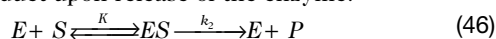
Plotting $1/r$ against $1/[D]$ will give a straight line where the value of K is obtained from the slope and the number of binding sites (n) from the intercept. The concentration of free drug can be determined by dialysis. The fraction of bound drug (β) is expressed as:

$$\beta = \frac{\text{concentration of bound drug}}{\text{total concentration of drug}} = \frac{[PD]}{[PD] + [D]} \quad (45)$$

$$= \frac{n \cdot K \cdot [P]_T}{1 + K \cdot [D] + n \cdot K \cdot [P]_T}$$

The plasma protein binding is non-linear and, thus, the fraction of drug bound is dependent on the concentration of both drug and protein. However, at very low free drug concentrations ($[D]$) β becomes essentially independent of drug concentration but β increases with increasing protein concentration ($[P]_T$). Drugs with high K value may saturate the protein resulting in a decrease in β with increasing drug concentration. Drugs with relatively high β values are susceptible to drug-drug interactions due to competitive drug protein binding.

Enzymes are proteins that catalyze reactions and like all proteins, enzymes are long, linear chains of amino acids that fold to produce a three-dimensional structure. Enzymes (E) form an intermediate complex (ES) with the substrate which is then converted to product upon release of the enzyme:



K is the binding constant or the equilibrium constant of the complex formation and k_2 is the rate constant for the catalyzed reaction. Enzymes can be highly specific catalyzing only certain class of chemical reactions or even only one particular reaction. Enzyme catalysis is covered in Chapter 17.

Drug receptors (R) are most often, but not always, proteins that interact with the drug molecules (D) to form a drug-receptor complex (DR) (see Chapter 38):



The complex formation changes the receptor that consequently leads to a biological activity and pharmacological response and/or adverse toxic effects. The forces involved in drug-receptor binding are the same as responsible for drug-protein binding such as ionic interactions, hydrogen bonds van der Waals dispersion forces and hydrophobic bonds. Formation of irreversible covalent bonds between drug and receptor leads to long-lasting effects.

COMPLEXES IN THERAPEUTICS

Complexes are very common in biological systems. As mentioned above drugs interact with proteins, including enzymes and receptors, through complex formations. For example, biological activity of some antimicrobial and antineoplastic agents is based on complex formation with DNA base-pairs. Molecular complexes in biological systems include DNA base-pairing and folding of proteins. Charge-transfer interactions play an important role in some membrane-transport processes. Metal ion coordination complexes are important parts of many biologically active compounds.³⁸ Examples of such compounds are hemoglobin (iron), cytochrome (iron), carboxypeptidase A (zinc), carbonic anhydrase (zinc), superoxide dismutase (zinc and copper), vitamin B₁₂ (cobalt), chlorophyll (magnesium), and urease (nickel).

ACKNOWLEDGMENTS

Kenneth A. Connors, PhD is acknowledged for his contribution to previous editions of this work.

REFERENCES

- Müller-Dethlefs K, Hobza P. Noncovalent interactions: a challenge for experiment and theory. *Chem Rev* 2000; 100:143–167.
- Liu L, Guo Q-X. The driving forces in the inclusion complexation of cyclodextrins. *J Incl Phenom Macroc Chem* 2002; 42:1–14.

- Palm GJ *et al.* Specific binding of divalent metal ions to tetracycline and to the Tet repressor/tetracycline complex. *J Biol Inorg Chem* 2008; 13:1097–1110.
- Leyden JJ. Absorption of minocycline hydrochloride and tetracycline hydrochloride. Effect of food, milk, and iron. *J Am Acad Dermatol* 1985; 12:308–312.
- Riley CM *et al.* eds. *Takeru Higuchi, a Memorial Tribute. Volume 3 - Equilibria and Thermodynamics*. Lawrence: Allen Press, 1991.
- Plaizier-Vercammen JA, de Nève RE. Interaction of povidone with aromatic compounds II: Evaluation of ionic strength, buffer concentration, temperature, and pH by factorial analysis. *J Pharm Sci* 1981; 70:1252–1256.
- Rácz I. *Drug Formulations*. Budapest: John Wiley and Sons, 1989.
- Loftsson T *et al.* The effect of water-soluble polymers on aqueous solubility of drugs. *Int J Pharm* 1996; 127:293–296.
- Ramírez Rigo MV *et al.* A linear free energy relationship treatment of the affinity between carboxymethylcellulose and basic drugs. *Mol Pharm* 2004; 1:383–386.
- Mao S *et al.* Chitosan-based formulations for delivery of DNA and siRNA. *Adv Drug Deliv Rev* 2010; 62:12–27.
- Cafaggi S *et al.* Preparation and evaluation of nanoparticles made of chitosan or N-trimethyl chitosan and a cisplatin-alginate complex. *J Control Rel* 2007; 121:110–123.
- Lorenz SA *et al.* Using solution phase hydrogen/deuterium (H/D) exchange to determine the origin of non-covalent complexes observed by electrospray ionization mass spectrometry: in solution or in vacuo? *J Am Soc Mass Spectrom* 2001; 12:795–804.
- Garrido JMPJ *et al.* Spectroscopic and electrochemical studies of cocaine-opioid interactions. *Anal Bioanal Chem* 2007; 388:1799–1808.
- Greenslade MD *et al.* Interactions between β -lactam and aminoglycoside antibiotics. *Int J Pharm* 1985; 26:133–144.
- Lim L-Y, Go M-L. Caffeine and nicotinamide enhances the aqueous solubility of the antimalarial agent halofantrine. *Eur J Pharm Sci* 2000; 10:17–28.
- Evstigneev MP *et al.* Effect of a mixture of caffeine and nicotinamide on the solubility of vitamin (B2) in aqueous solution. *Eur J Pharm Sci* 2006; 28:59–66.
- Singh BN. Effects of food on clinical pharmacokinetics. *Clin Pharmacokinet* 1999; 37:213–255.
- Dongowski G *et al.* Interactions between food components and drugs. Part 4: Influence of pectins and bile salts on propranolol absorption. *Int J Pharm* 1996; 144:233–239.
- Kralj M *et al.* Biomedical potentials of crown ethers: prospective antitumor agents. *Chem Med Chem* 2008; 3:1478–1492.
- Rodik RV *et al.* Calixarenes in bio-medical researches. *Curr Med Chem* 2009; 16:1630–1655.
- Brown ME, Hollingsworth MD. Stress-induced domain reorientation in urea inclusion compounds. *Nature* 1995; 376:323–327.
- Thakral S, Madan AK. Urea co-inclusion compounds of 13 cis-retinoic acid for simultaneous improvement of dissolution profile, photostability and safe handling characteristics. *J Pharm Pharmacol* 2008; 60:823–832.
- Higuchi T, Connors KA. Phase-solubility techniques. *Adv Anal Chem Instrum* 1965; 4:117–212.
- Repta AJ. Alteration of apparent solubility through complexation. In: Yalkowski SH, ed. *Techniques of solubilization of drugs*. New York: Marcel Dekker, 1985: 135–157.
- Brewster ME, Loftsson T. Cyclodextrins as pharmaceutical solubilizers. *Adv Drug Deliv Rev* 2007; 59:645–666.
- Loftsson T *et al.* Self-association of cyclodextrins and cyclodextrin complexes. *J Pharm Sci* 2004; 93:1091–1099.
- Loftsson T, Duchêne D. Cyclodextrins and their pharmaceutical applications. *Int J Pharm* 2007; 329:1–11.
- Loftsson T, Brewster ME. Pharmaceutical applications of cyclodextrins: basic science and product development. *J Pharm Pharmacol* 2010; 62:1607–1621.
- Stella VJ *et al.* Mechanisms of drug release from cyclodextrin complexes. *Adv Drug Deliv Rev* 1999; 36:3–16.

30. Loftsson T, Brewster ME. Pharmaceutical applications of cyclodextrins: effects on drug permeation through biological membranes. *J Pharm Pharmacol* 2011; 63:1119–1135.
31. Messner M *et al.* Self-assembled cyclodextrin aggregates and nanoparticles. *Int J Pharm* 2010; 387:199–208.
32. Bonini M *et al.* Self-assembly of beta-cyclodextrin in water. Part 1: Cryo-TEM and dynamic and static light scattering. *Langmuir* 2006; 22:1478–1484.
33. Wu A *et al.* Micrometer-sized rodlike structure formed by the secondary assembly of cyclodextrin nanotube. *J Coll Interf Sci* 2006; 302:87–94.
34. He W *et al.* Cyclodextrin-based aggregates and characterization by microscopy. *Micron* 2008; 39:495–516.
35. Messner M *et al.* Self-assembly of cyclodextrins: the effect of the guest molecule. *Int J Pharm* 2011; 408:235–247.
36. Bertucci C, Domenici E. Reversible and covalent binding of drugs to human serum albumin: methodological approaches and physiological relevance. *Curr Med Chem* 2002; 9:1463–1481.
37. Vuignier K *et al.* Drug-protein binding: a critical review of analytical tools. *Anal Bioanal Chem* 2010; 398:53–66.
38. Frezza M *et al.* Novel metals and metal complexes as platforms for cancer therapy. *Curr Pharm Des* 2010; 16:1813–1825.

Chapter 19

Interfacial Phenomena

Paul M. Bummer, PhD and Yvonne Perrie, PhD

INTRODUCTION	329
INTERFACIAL FORCES AND ENERGETICS	329
ADHESIONAL AND COHESIONAL FORCES	331
WETTING PHENOMENA	332
CAPILLARITY	333

PRESSURE DIFFERENCES ACROSS CURVED SURFACES	334
---	-----

ADSORPTION	334
------------	-----

SURFACE-ACTIVE AGENTS	339
-----------------------	-----

INTRODUCTION

Very often it is desirable or necessary in the development of pharmaceutical dosage forms to produce multiphasic dispersions by mixing together two or more ingredients that are not mutually miscible and capable of forming homogeneous solutions. Examples of such dispersions include:

- Suspensions (solid in liquid)
- Emulsions (liquid in liquid)
- Foams (vapor in liquids)

Because these systems are not homogeneous and thermodynamically stable, over time they will show some tendency to separate on standing. By doing so, this produces the minimum possible surface area of contact between phases. Thus, suspended particles agglomerate and sediment, emulsified droplets cream and coalesce, and the bubbles dispersed in foams collapse to produce unstable and nonuniform dosage forms. One way to prevent or slow down this natural tendency for further phase separation is to add materials that can accumulate at the interface to provide some type of energy barrier to aggregation and coalescence. Such materials are said to exhibit *surface activity* or to act as *surface-active agents*.

In this chapter the fundamental physical chemical properties of molecules situated at interfaces will be discussed so that the reader can gain a better understanding of how problems involving interfaces can be resolved in designing pharmaceutical dosage forms by the use of surface-active agents.

INTERFACIAL FORCES AND ENERGETICS

In the bulk portion of each phase, molecules are attracted to each other equally in all directions, such that no resultant forces are acting on any one molecule. The strength of these forces determines whether a substance exists as a vapor, liquid, or solid at a particular temperature and pressure.

At the interface between phases, however, molecules are acted upon unequally because they are in contact with other molecules exhibiting different forces of attraction. For example, the primary intermolecular forces in water are due to hydrogen bonds, whereas those responsible for intermolecular bonding in hydrocarbon liquids, such as mineral oil, are due to London dispersion forces.

Thus, molecules situated at the interface experience interaction forces dissimilar to those experienced in each bulk phase. In liquid systems such unbalanced forces can be satisfied by spontaneous movement of molecules from the interface into the bulk phase. This leaves fewer molecules per unit area at the interface (greater intermolecular distance) and reduces the actual contact area between dissimilar molecules.

Any attempt to reverse this process by increasing the area of contact between phases—that is, bringing more molecules into

the interface—causes the interface to resist expansion and behave as though it is under a tension everywhere in a tangential direction. The force of this tension per unit length of interface generally is called the *interfacial tension*, except when dealing with the air–liquid interface, where the terms *surface* and *surface tension* are used.

To illustrate the presence of a tension in the interface, consider an experiment in which a circular metal frame with a looped piece of thread loosely tied to it is dipped into a liquid. When the frame is removed and exposed to the air, a film of liquid will be stretched entirely across the circular frame, as when one uses such a frame to blow soap bubbles. Under these conditions (Figure 19-1A) the thread will remain collapsed. If a heated needle is used to puncture and remove the liquid film from within the loop (Figure 19-1B), the loop will stretch spontaneously into a circular shape.

The result of this experiment demonstrates the spontaneous reduction of interfacial contact between air and the liquid remaining; indeed, it illustrates that a tension causing the loop to remain extended exists parallel to the interface. The circular shape of the loop indicates that the tension in the plane of the interface exists at right angles or normal to every part of the looped thread. The total force on the entire loop divided by the circumference of the circle, therefore, represents the tension per unit distance of surface, or the surface tension.

Just as work is required to extend a spring under tension, work should be required to reverse the process seen in Figure 19-1A and B, thus bringing more molecules to the interface. This may be seen quantitatively by considering an experiment in which tension and work may be measured directly. Assume that we have a rectangular wire with one movable side (Figure 19-2). Assume further that by dipping this wire into a liquid, a film of liquid will form within the frame when it is removed and exposed to the air. As seen earlier in Figure 19-1, when it comes in contact with air, the liquid surface will tend to contract with a force F as molecules leave the surface for the bulk. To keep the movable side in equilibrium, an equal force must be applied to oppose this tension in the surface. The surface tension γ of

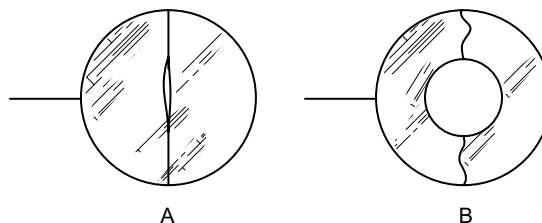


Figure 19-1. A circular wire frame with a loop of thread loosely tied to it: (A) a liquid film on the wire frame with a loop in it; (B) the film inside the loop is broken. (From Semat H. *Fundamentals of Physics*, 3rd edn. New York: Holt Rinehart Winston, 1957.)

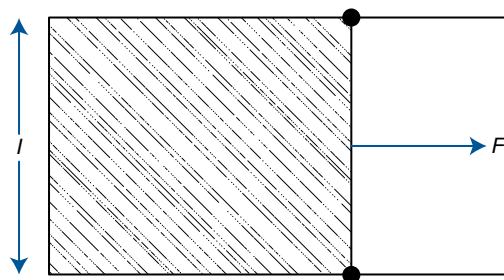


Figure 19-2. A movable wire frame containing a film of liquid being expanded with a force F .

the liquid may be defined as $F/2l$, where $2l$ is the distance of surface over which F is operating. The factor 2 arises out of considering two surfaces, top and bottom. Upon expansion of the surface by a very small distance, Δx , the work done (W) is

$$W = F\Delta x \quad (1)$$

and, therefore,

$$W = \gamma 2l\Delta x \quad (2)$$

Since

$$\Delta A = 2l\Delta x \quad (3)$$

where ΔA is the change in area due to the expansion of the surface, it may be concluded that

$$W = \gamma\Delta A \quad (4)$$

Thus, the work required to create a unit area of surface, known as the *surface free energy/unit area*, is equivalent to the surface tension of a liquid system—the greater the area of interfacial contact between phases, the greater the free-energy increase for the total system. Because a prime requisite for equilibrium is that the free energy of a system be at a minimum, it is not surprising to observe that phases in contact tend to reduce area of contact spontaneously.

Liquids, being mobile, may assume spherical shapes (smallest interfacial area for a given volume), as when ejected from an orifice into air or when dispersed into another immiscible liquid. If a large number of drops are formed, further reduction in area can occur by having the drops coalesce, as when a foam collapses or when the liquid phases making up an emulsion separate.

In the centigrade-gram-second (cgs) system, surface tension is expressed in units of dynes per centimeter (dyne/cm), while surface free energy is expressed in erg/cm². Because an erg is a dyne-cm, both sets of units are equivalent. In the SI (international units) system, surface tension is expressed in mN/m and surface free energy in mJ/m².

Values for the surface tension of a variety of liquids are given in Table 19-1, and interfacial tension values for various liquids against water are given in Table 19-2. Other combinations of immiscible phases could be given, but most heterogeneous systems encountered in pharmacy usually contain water. Values for these tensions are expressed for a particular temperature. Because an increased temperature increases the thermal energy of molecules, the work required to bring molecules to the interface should be less, and thus the surface and interfacial tension will be reduced. For example, the surface tension of water is 76.5 dynes/cm at 0°C and 63.5 dynes/cm at 75°C.

As would be expected from the discussion so far, the relative values for surface tension should reflect the nature of intermolecular forces present, hence the relatively large values for mercury (metallic bonds) and water (hydrogen bonds) and the lower values for benzene, chloroform, carbon tetrachloride, and the n alkanes.

Benzene, with π electrons, exhibits a higher surface tension than the alkanes of comparable molecular weight, but increasing the molecular weight of the alkanes (and hence intermolecular

Table 19-1. Surface Tension of Various Liquids at 20°C

Substance	Surface Tension (dyne/cm)
Mercury	476
Water	72.8
Glycerin	63.4
Oleic acid	32.5
Benzene	28.9
Chloroform	27.1
Carbon tetrachloride	26.8
1-Octanol	26.5
Hexadecane	27.4
Dodecane	25.4
Decane	23.9
Octane	21.8
Heptane	19.7
Hexane	18.0
Perfluoroheptane	11.0
Nitrogen (at 75° K)	9.4

Table 19-2. Interfacial Tension of Various Liquids Against Water at 20°C

Substance	Interfacial Tension (dyne/cm)
Decane	52.3
Octane	51.7
Hexane	50.8
Carbon tetrachloride	45.0
Chloroform	32.8
Benzene	35.0
Mercury	428
Oleic acid	15.6
1-Octanol	8.51

attraction) increases their surface tension closer to that of benzene. The lower values for the more nonpolar substances, perfluoroheptane and liquid nitrogen, demonstrate this point even more strongly.

Values of interfacial tension should reflect the differences in chemical structure of the two phases involved—the greater the tendency to interact, the less the interfacial tension. The 20-dyne/cm difference between air–water tension and that at the octane–water interface reflects the small, but significant, interaction between octane molecules and water molecules at the interface. This is seen also in Table 19-2 by comparing the values for octane and octanol, oleic acid and the alkanes, or chloroform and carbon tetrachloride. In each case the presence of chemical groups capable of hydrogen bonding with water markedly reduces the interfacial tension, presumably by satisfying the unbalanced forces at the interface. These observations strongly suggest that molecules at an interface arrange themselves, or orient, so as to minimize differences between bulk phases.

That this phenomenon occurs even at the air–liquid interface is seen when one notes the relatively low surface-tension values of very different chemical structures, such as the n alkanes, octanol, oleic acid, benzene, and chloroform. Presumably, in each case the similar nonpolar groups are oriented toward the air, with any polar groups oriented away toward the bulk phase. This tendency for molecules to orient at an interface is a basic factor in interfacial phenomena and will be discussed more fully in succeeding sections.

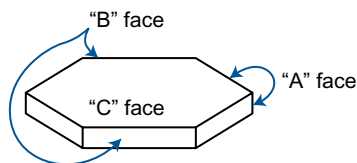


Figure 19.3. Adipic acid crystal showing various faces. (From Michaels AS. Modification of growth rate and habit of adipic acid crystals with surfactants. *J Phys Chem* 1961; 65: 1730.)

Solid substances, such as metals, metal oxides, silicates, and salts, all containing polar groups exposed at their surface, may be classified as *high-energy solids*, whereas nonpolar solids, such as carbon, sulfur, glyceryl tristearate, polyethylene, and polytetrafluoroethylene (Teflon), may be classified as *low-energy solids*. It is of interest to measure the surface free energy of solids; however, the lack of mobility of molecules at the surface of solids prevents the observation and direct measurement of a surface tension. It is possible to measure the work required to create new solid surface by cleaving a crystal and measuring the work involved. However, this work not only represents free energy due to exposed groups but also takes into account the mechanical energy associated with crystal fracture (i.e., plastic and elastic deformation and strain energies due to crystal structure and imperfections in that structure).

Also contributing to the complexity of a solid surface is the heterogeneous behavior as a result of the exposure of different crystal faces, each having a different surface free energy/unit area. For example, adipic acid, $\text{HOOC}(\text{CH}_2)_4\text{COOH}$, crystallizes from water as thin, hexagonal plates with three different faces, as shown in Figure 19-3. Each unit cell of such a crystal contains adipic acid molecules, oriented such that the hexagonal planes (faces) contain exposed carboxyl groups, while the sides and edges (A and B faces) represent the side view of the carboxyl and alkyl groups and, thus, are quite nonpolar. Indeed, interactions involving these different faces reflect the differing surface free energies.²

Other complexities of solid surfaces include roughness and porosity.³ Even in the absence of chemical contamination, such as that occurring during recrystallization, surface energy changes in a solid can be induced by unit operations such as milling, resulting in an altered pattern of drug dissolution.^{4,5} In view of all these potential complications that are difficult to quantify, surface free energy values for solids, when reported, should be regarded as average values, often dependent on the method used and not necessarily the same for other samples of the same substance.

Table 19-3 lists some average values of γ_{sv} for a variety of solids, ranging in polarity from Teflon to copper, obtained by various indirect techniques.

Table 19-3. Values of γ_{sv} for Solids of Varying Polarity

Solid	γ_{sv} (dyne/cm)
Teflon	19.0
Paraffin	25.5
Polyethylene	37.6
Polymethyl methacrylate	45.4
Nylon	50.8
Indomethacin	61.8
Griseofulvin	62.2
Hydrocortisone	68.7
Sodium Chloride	155
Copper	1300

ADHESIONAL AND COHESIONAL FORCES

Of prime importance to those dealing with heterogeneous systems is the question of how two phases will behave when brought in contact with each other. It is well known, for instance, that some liquids, when placed in contact with other liquid or solid surfaces, will remain retracted in the form of a drop (known as a *lens*), while other liquids may exhibit a tendency to spread and cover the surface of this liquid or solid.

Based upon concepts developed to this point, it is apparent that the individual phases will exhibit a tendency to minimize the area of contact with other phases, thus leading to phase separation. On the other hand, the tendency for interaction between molecules at the new interface will offset this to some extent and give rise to the spontaneous spreading of one substance over the other.

In essence, therefore, phase affinity is increased as the forces of attraction between different phases (*adhesional forces*) become greater than the forces of attraction between molecules of the same phase (*cohesional forces*). If these adhesional forces become great enough, miscibility will occur and the interface will disappear. The present discussion is concerned only with systems of limited phase affinity, where an interface still exists.

A convenient approach used to express these forces quantitatively is work of adhesion and work of cohesion. The *work of adhesion*, W_a , is defined as the free energy/cm² required to separate two phases at their boundary and is equal but opposite in sign to the free energy/cm² released when the interface is formed. In an analogous manner, the *work of cohesion* for a pure substance, W_c , is the work/cm² required to produce two new surfaces, as when separating different phases, but now both surfaces contain the same molecules. This is equal and opposite in sign to the free energy/cm² released when the same two pure liquid surfaces are brought together and eliminated.

By convention, when the work of adhesion between two substances, A and B, exceeds the work of cohesion for one substance (e.g., B), spontaneous spreading of B over the surface of A should occur with a net loss of free energy equal to the difference between W_a and W_c . If W_c exceeds W_a , no spontaneous spreading of B over A can occur. The difference between W_a and W_c is known as the *spreading coefficient*, S . Only when S is positive will spreading occur.

The values for W_a and W_c (and hence S) may be expressed in terms of surface and interfacial tensions, when one considers that upon separation of two phases, A and B, γ_{AB} ergs of interfacial free energy/cm² (interfacial tension) are lost but that γ_A and γ_B erg/cm² of energy (surface tensions of A and B) are gained; upon separation of bulk-phase molecules in an analogous manner, $2\gamma_A$ or $2\gamma_B$ erg/cm² will be gained. Thus,

$$W_a = \gamma_A + \gamma_B - \gamma_{AB} \quad (5)$$

and

$$W_c = 2\gamma_A \text{ or } 2\gamma_B \quad (6)$$

for B spreading on the surface of A. Therefore,

$$S_B = \gamma_A + \gamma_B - \gamma_{AB} - 2\gamma_B \quad (7)$$

or

$$S_B = \gamma_A - (\gamma_B + \gamma_{AB}) \quad (8)$$

Using Equation 8 and the values of surface and interfacial tension given in Tables 19-1 and 19-2, the spreading coefficient can be calculated for three representative substances—decane, benzene, and oleic acid—on water at 20°C.

$$\text{Decane: } S = 72.8 - (23.9 + 52.3) = -3.4$$

$$\text{Benzene: } S = 72.8 - (28.9 + 35.0) = 8.9$$

$$\text{Oleic Acid: } S = 72.8 - (32.5 + 15.6) = 24.7$$

As expected, relatively nonpolar substances such as decane exhibit negative values of spreading coefficient, whereas the more-polar materials yield positive values—the greater the polarity of the molecule, the more positive the value of S .

The importance of the cohesive energy of the spreading liquid may be noted also by comparing the spreading coefficients for hexane on water and water on hexane.

$$S_{H/W} = 72.8 - (18.0 + 50.8) = 10.0$$

$$S_{W/H} = 18.0 - (72.8 + 50.8) = -105.6$$

Here, despite the fact that both liquids are the same, the high cohesion and air-liquid tension of water prevents spreading on the low-energy hexane surface, while the very low value for hexane allows spreading on the water surface. This also is seen when comparing the positive spreading coefficient of hexane to the negative value for decane on water.

To see whether spreading does or does not occur, a powder such as talc or charcoal can be sprinkled over the surface of water such that it floats; then, a drop of each liquid is placed on this surface. As predicted, decane will remain as an intact drop, while hexane, benzene, and oleic acid will spread out, as shown by the rapid movement of solid particles away from the point where the liquid drop was placed originally.

An apparent contradiction to these observations may be noted for hexane, benzene, and oleic acid when more of each substance is added: lenses now appear to form even though initial spreading occurred. Thus, in effect, a substance does not appear to spread over itself.

It is now established that the spreading substance forms a monomolecular film that creates a new surface that has a lower surface free energy than pure water. This arises because of the apparent orientation of the molecules in such a film so that their most hydrophobic portion is oriented toward the spreading phase. It is the lack of affinity between this exposed portion of the spread molecules and the polar portion of the remaining molecules that prevents further spreading. This may be seen by calculating a final spreading coefficient where the new surface tension of water plus monomolecular film is used. For example, the presence of benzene reduces the surface tension of water to 62.2 dyne/cm so that the final spreading coefficient is

$$S = 62.2 - (28.9 + 35.0) = -1.7$$

The lack of spreading exhibited by oleic acid should be reflected in an even more negative final spreading coefficient, for the very polar carboxyl groups should have very little affinity for the exposed alkyl chain of the oleic acid film. Spreading so as to form a second layer with polar groups exposed to the air also would seem very unlikely, thus leading to the formation of a lens.

WETTING PHENOMENA

In the experiment described above, it was shown that talc or charcoal sprinkled onto the surface of water floats, despite the fact that their densities are much greater than that of water. In order for immersion of the solid to occur, the liquid must displace air and spread over the surface of the solid; when liquids

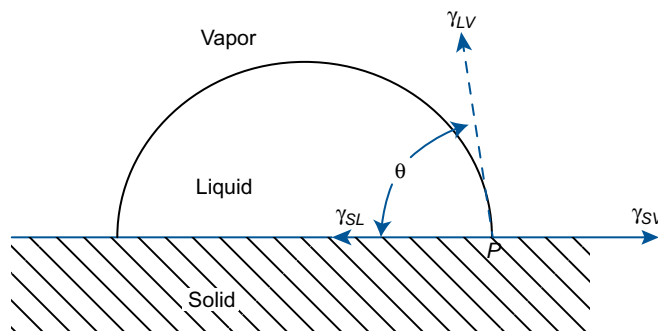


Figure 19.4. Forces acting on a nonwetting liquid drop exhibiting a contact angle of. (From Zisman WA. Relation of the equilibrium contact angle to liquid and solid constitution. *Adv Chem Ser* 1964; 43: 1.)

cannot spread over a solid surface spontaneously, and, therefore, S , the spreading coefficient, is negative, we say that the solid is not wetted.

An important parameter reflecting the degree of wetting is the angle made by the liquid with the solid surface at the point of contact (Figure 19-4). By convention, when wetting is complete, the contact angle is 0° ; in nonwetting situations it theoretically can increase to a value of 180° , where a spherical droplet makes contact with solid at only one point.

To express contact angle in terms of solid-liquid-air equilibria, one can balance forces parallel to the solid surface at the point of contact between all three phases (see Figure 19-4), as expressed in

$$\gamma_{SV} = \gamma_{SL} + \gamma_{LV} \cos \theta \quad (9)$$

where γ_{SV} , γ_{SL} , and γ_{LV} represent the surface free energy/unit area of the solid-air, solid-liquid and liquid-air interfaces, respectively. Although difficult to use quantitatively because of uncertainties with γ_{SV} and γ_{SL} measurements, conceptually the equation, known as the Young equation, is useful because it shows that the loss of free energy due to elimination of the air-solid interface by wetting is offset by the increased solid-liquid and liquid-air area of contact as the drop spreads out.

The $\gamma_{LV} \cos \theta$ term arises as the horizontal vectorial component of the force acts along the surface of the drop, as represented by γ_{LV} . Factors tending to reduce γ_{LV} and γ_{SL} , therefore, will favor wetting, while the greater the value of γ_{SV} , the greater the chance for wetting to occur. This is seen in Table 19-4 for the wetting of a low-energy surface, paraffin (hydrocarbon), and a higher energy surface, nylon (polyhexamethylene adipamide). Here, the lower the surface tension of a liquid, the smaller the contact angle on a given solid, and the more polar the solid, the smaller the contact angle with the same liquid.

With Equation 9 in mind and looking at Figure 19-5, it is now possible to understand how the forces acting at the

Table 19-4. Contact Angle on Paraffin and Nylon for Various Liquids of Differing Surface Tension

Substance	Surface Tension (dyne/cm)	Contact Angle ($^\circ$)	
		Paraffin	Nylon
Water	72.8	105	70
Glycerin	63.4	96	60
Formamide	58.2	91	50
Methylene iodide	50.8	66	41
Bromonaphthalene	44.6	47	16
Tert-butyl naphthalene	33.7	38	spreads
Benzene	28.9	24	spreads
Dodecane	25.4	17	spreads
Decane	23.9	7	spreads
Nonane	22.9	spreads	spreads

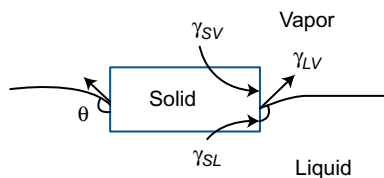


Figure 19.5. Forces acting on a nonwetable solid at the air+liquid+solid interface: contact angle θ greater than 90° .

solid-liquid-air interface can cause a dense, nonwetted solid to float if γ_{SL} and γ_{LV} are large enough, relative to γ_{SV} .

The significance of reducing γ_{LV} was first developed empirically by Zisman⁶ when he plotted $\cos \theta$ versus the surface tension of a series of liquids and found that a linear relationship, dependent on the solid, often was obtained. When such plots are extrapolated to $\cos$ equal to 1, or 0° contact angle, a value of surface tension required to just cause complete wetting is obtained. Doing this for a number of solids, it was shown that this surface tension (known as the critical surface tension, γ_c) parallels expected solid surface energy γ_{SV} —the lower γ_c , the more nonpolar the surface.

Table 19-5 indicates some of these γ_c values for different surface groups, indicating such a trend. Thus, water with a surface tension of about 72 dyne/cm will not wet polyethylene ($\gamma_c = 31$ dyne/cm), but heptane, with a surface tension of about 20 dyne/cm, will. Likewise, Teflon (polytetrafluoroethylene) ($\gamma_c = 19$) is not wetted by heptane but is wetted by perfluoroheptane with a surface tension of 11 dyne/cm.

One complication associated with the wetting of high-energy surfaces is the lack of wetting after the initial formation of a monomolecular film, caused by the spreading substance. As in the case of oleic acid spreading on the surface of water, the remaining liquid retracts because of the low-energy surface produced by the oriented film. This phenomenon, often called *autophobic behavior*, is an important factor in many systems of pharmaceutical interest because many solids, expected to be wetted easily by water, may be rendered hydrophobic if other molecules dissolved in the water can form these monomolecular films at the solid surface.

CAPILLARITY

Because water shows a strong tendency to spread out over a polar surface such as clean glass (contact angle equal to 0°),

Table 19-5. Critical Surface Tensions of Various Polymeric Solids

Polymeric Solid	γ_c (dyne/cm AT 20°C)
Polymethacrylic ester of π' -octanol	10.6
Polyhexafluoropropylene	16.2
Polytetrafluoroethylene	19
Polytrifluoroethylene	22
Poly(vinylidene fluoride)	25
Poly(vinyl fluoride)	28
Polyethylene	31
Polytrifluorochloroethylene	31
Polystyrene	33
Poly(vinyl alcohol)	37
Poly(methyl methacrylate)	39
Poly(vinyl chloride)	39
Poly(vinylidene chloride)	40
Poly(ethylene terephthalate)	43
Poly(hexamethylene adipamide)	46

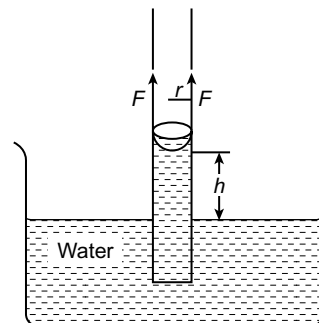


Figure 19.6. Capillary rise for a liquid exhibiting 0° contact angle. (From Semat H. *Fundamentals of Physics*, 3rd edn. New York: Holt Rinehart Winston, 1957.)

one would expect to observe a meniscus forming when water is contained in a glass vessel such as a pipet or buret. This behavior is accentuated dramatically if a fine-bore capillary tube is placed into the liquid (Figure 19-6). Not only will the wetting of the glass produce a more highly curved meniscus, but the level of the liquid in the tube will be appreciably higher than the level of the water in the beaker.

The spontaneous movement of a liquid into a capillary or narrow tube due to surface forces is defined as *capillarity* and is responsible for a number of important processes involving the penetration of liquids into porous solids. In contrast to water in contact with glass, if the same capillary is placed into mercury (contact angle on glass: 130°), not only will the meniscus be inverted (Figure 19-7), but the level of the mercury in the capillary will be lower than in the beaker. In this case one does not expect mercury or other *nonwetting* liquids to penetrate pores easily unless external forces are applied.

To examine more closely the factors giving rise to the phenomenon of capillarity, consider the case of a liquid that rises to a height, h , above the bulk liquid in a capillary having a radius r . As shown in Figure 19-6, if the contact angle of water on glass is 0 , a force F will act upward and vertically along the circle of liquid-glass contact. Based upon the definition of surface tension, this force will be equal to the surface tension γ multiplied by the circumference of the circle, $2\pi r$. Thus,

$$F = \gamma 2\pi r \quad (10)$$

This force upward must support the column of water, and because the mass m of the column is equal to the density d multiplied by the volume of the column, $\pi r^2 h$, the force W opposing the movement upward will be

$$W = mg = \pi r^2 dgh \quad (11)$$

where g is the gravity constant.

Equating the two forces at equilibrium gives

$$\pi r^2 dgh = \gamma 2\pi r \quad (12)$$

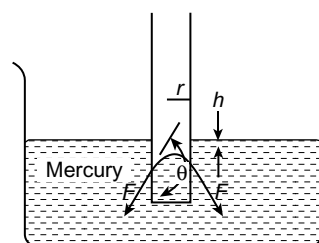


Figure 19.7. Capillary fall for a liquid exhibiting a contact angle, θ , that is greater than 90° . (From Semat H. *Fundamentals of Physics*, 3rd edn. New York: Holt Reinhart Winston, 1957.)

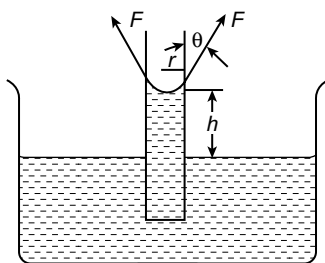


Figure 19.8. Capillary rise for a liquid exhibiting a contact angle, θ , that is greater than 0° but less than 90° . (From Semat H. *Fundamentals of Physics*, 3rd edn. New York: Holt Rinehart Winston, 1957.)

so that

$$h = \frac{2\gamma}{rdg} \quad (13)$$

Thus, the greater the surface tension and the finer the capillary radius, the greater the rise of liquid in the capillary.

If the contact angle of liquid is not 0 (Figure 19-8), the same relationship may be developed, except the vertical component of F , which opposes the weight of the column, is $F \cos \theta$ and, therefore

$$h = \frac{2\gamma \cos \theta}{rdg} \quad (14)$$

This indicates the very important fact that if θ is less than 90° but greater than 0° , the value of h will decrease with increasing contact angle until at 90° ($\cos \theta = 0$), $h = 0$. Above 90° , values of h will be negative, as indicated in Figure 19-7 for mercury. Thus, based on these equations it may be concluded that capillarity will occur spontaneously in a cylindrical pore even if the contact angle is greater than 0° , but it will not occur at all if the contact angle becomes 90° or more. In solids with irregularly shaped pores, the relationships between parameters in Equation 14 will be the same, but they will be more difficult to quantitate because of nonuniform changes in pore radius throughout the porous structure.

PRESSURE DIFFERENCES ACROSS CURVED SURFACES

From the preceding discussion of capillarity, another important concept follows. In order for the liquid in a capillary to rise spontaneously, it must develop a higher pressure than the lower level of the liquid in the beaker. However, because the system is open to the atmosphere, both surfaces are in equilibrium with the atmospheric pressure. To be raised above the level of liquid in the beaker and produce a hydrostatic pressure equal to hgd , the pressure just below the liquid meniscus, in the capillary, P_1 , must be less than that just below the flat liquid surface, P_0 , by hgd , and therefore,

$$P_0 - P_1 = hgd \quad (15)$$

Because, according to Equation 14,

$$h = \frac{2\gamma \cos \theta}{rdg}$$

then

$$P_0 - P_1 = \frac{2\gamma \cos \theta}{r} \quad (16)$$

For a contact angle of 0° , where the radius of the capillary is the radius of the hemisphere making up the meniscus,

$$P_0 - P_1 = \frac{2\gamma}{r} \quad (17)$$

The consequences of this relationship (known as the Laplace equation) are important for any curved surface when r becomes very small and is γ relatively significant. For example, a spherical

droplet of air formed in a bulk liquid, and having a radius r , will have a greater pressure on the inner concave surface than on the convex side, as expressed in Equation 17. Direct measurement of the pressure difference ($P_0 - P_1$) for an air bubble of known radius allows the determination of the surface tension of either a pure liquid or a solution of surface active substance. Both static (constant radius) and dynamic (radius changing in a cyclic fashion as a function of time) measurements have been employed. The latter treatment, known as the pulsating bubble method, has been very useful in the study of some of the biophysical properties and associated disease states of pulmonary surfactant, a mixture of surface active materials lining the small airways of the mammalian lung.⁷ One of the less appreciated advantages of this method for measuring surface tension is the need for only a very small sample size, typically on the order of 50 μL .

Another direct consequence of what Equation 17 expresses is the fact that very small droplets of liquid, having highly curved surfaces, will exhibit a higher vapor pressure, VP , than are observed over a flat surface of the same liquid at VP' . Equation 18, called the *Kelvin equation*, expresses the ratio of VP/VP' to droplet radius r , and surface tension γ :

$$\log \frac{P}{P'} = \frac{2\gamma M}{2.303RT\rho r} \quad (18)$$

where M is the molecular weight, R is the gas constant in erg/mol/degree , T is temperature, and ρ is the density in g/cm^3 . Values for the ratio of vapor pressures are given in Table 19-6 for water droplets of varying size. Such ratios indicate why it is possible for very fine water droplets in clouds to remain uncondensed, despite their close proximity to one another.

This same behavior may be seen when measuring the solubility of very fine solid particles, for both vapor pressure and solubility are measures of the escaping tendency of molecules from a surface. Indeed, the equilibrium solubility of extremely small particles has been shown to be greater than the usual value noted for coarser particles; the greater the surface energy and smaller the particles, the greater this effect.

ADSORPTION

VAPOR ADSORPTION ON SOLID SURFACES

It was suggested earlier that a high surface or interfacial free energy may exist at a solid surface if the unbalanced forces at the surface and the area of exposed groups are quite great.

Substances such as metals, metal oxides, silicates, and salts—all containing exposed polar groups—may be classified as high-energy or hydrophilic solids; nonpolar solids, such as carbon, sulfur, polyethylene, or Teflon (polytetrafluoroethylene), may be classified as low-energy or hydrophobic solids (see Table 19-3). Whereas liquids satisfy their unbalanced surface forces by changes in shape, pure solids (which exhibit negligible surface mobility) must rely on reaction with molecules, either in the

Table 19-6. Ratio of Observed Vapor Pressure (P) to Expected Vapor Pressure (P') of Water at 25°C With Varying Droplet Size

P/P'	Droplet Size (μm)
1.001	1
1.01	0.1
1.1	0.01
2.0	0.005
3.0	0.001
4.2	0.00065
5.2	0.00060

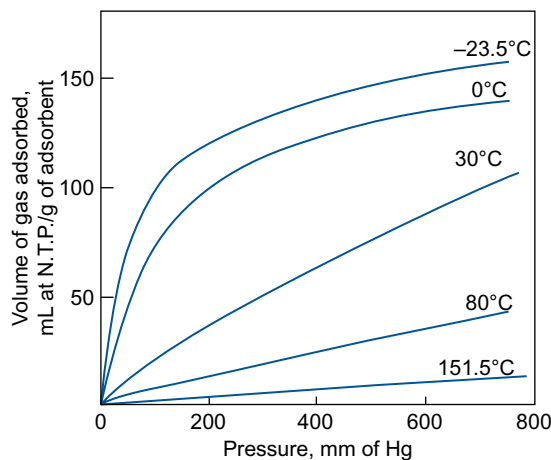


Figure 19.9. Adsorption isotherms for ammonia on charcoal. (From Titoff Z. Die Adsorption von Gasen durch Kohle. *Z Phys Chem Leipzig* 1910; 74: 641–678.)

vapor state or in a solution that comes in contact with the solid surface to accomplish this.

Vapor adsorption is the simplest model demonstrating how solids reduce their surface free energy in this manner. Depending on the chemical nature of the adsorbent (solid) and the adsorbate (vapor), the strength of interaction between the two species may vary from strong specific chemical bonding to interactions produced by the weaker, more nonspecific London dispersion forces. Ordinarily, these latter forces are those responsible for the condensation of relatively nonpolar substances, such as N_2 , O_2 , CO_2 , or hydrocarbons.

When chemical reaction occurs, the process is called *chemisorption*; when dispersion forces predominate, the term *physisorption* is used. Physisorption occurs at temperatures approaching the liquefaction temperature of the vapor; for chemisorption, temperatures depend on the particular reaction involved. Water-vapor adsorption to various polar solids can occur at room temperature through hydrogen-bonding, with binding energies intermediate to physisorption and chemisorption.

To study the adsorption of vapors onto solid surfaces, one must measure the amount of gas adsorbed/unit area or unit mass of solid, at different pressures of gas. Because such studies usually are conducted at constant temperature, plots of volume

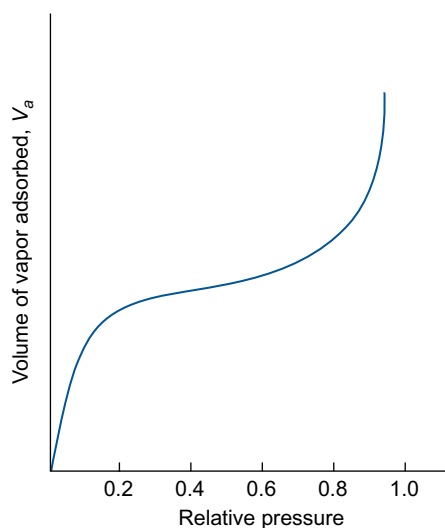


Figure 19.10. Typical plot for multilayer physical adsorption of a vapor on a solid surface.

adsorbed versus pressure are known as *adsorption isotherms*. If the physical or chemical adsorption process is monomolecular, the adsorption isotherm should appear similar to those shown in Figure 19-9. Adsorption significantly increases with increasing pressure, followed by a leveling off, which is due either to a saturation of available specific chemical groups, as in chemisorption, or to the entire available surface being covered by physically adsorbed molecules. Adsorption reduction with increasing temperature occurs because the adsorption process is exothermic. In the case of physical adsorption at low temperatures after adsorption levels off, often a marked increase in adsorption occurs, presumably due to multilayered adsorption. In this case vapor molecules essentially condense upon themselves as the liquefaction pressure of the vapor is approached. Figure 19-10 illustrates one type of isotherm generally seen with multilayered physisorption.

To have a quantitative understanding of the adsorption process and to be able to compare different systems, two factors must be evaluated. It is important to know the capacity of the solid or the maximum amount of adsorption under a given set of conditions and the affinity of a given substance for the solid surface—how readily does it adsorb for a given amount of pressure? In effect, the second term is the equilibrium constant for the process. For many systems vapor-adsorption data may fit a very general, but somewhat empirical, equation, the Freundlich equation:

$$V_a = kp^n \quad (19)$$

where V_a is the volume of gas adsorbed, p is the gas pressure, and k and n are constants, reflecting adsorption affinity and capacity.

A significant theoretical improvement along these lines was the theory of *monomolecular adsorption*, proposed by Langmuir. He postulated that for adsorption to occur, a solid must contain uniform adsorption sites, each capable of holding a single gas molecule. Molecules colliding with the surface may bounce off elastically, or they may remain in contact for a period of time. It is this contact over a period of time that Langmuir termed *adsorption*.

Two major assumptions were made in deriving the adsorption equation:

Only those molecules striking an empty site can be adsorbed; hence, only monomolecular adsorption occurs.

The forces of interaction between adsorbed molecules are negligible and, therefore, the probability of a molecule adsorbing onto or desorbing from any site is independent of the surrounding sites.

With these assumptions and applying the kinetic theory of gases, it can be shown that

$$V_a = (V_m k' p) / (1 + k' p) \quad (20)$$

where V_m is the volume of gas covering all of the adsorption sites with a single layer of molecules, and k' is a constant that reflects the affinity of the gas for the solid.

A test of fit to this equation can be made by expressing it in linear form.

$$\frac{p}{V_a} = \frac{1}{V_m k'} + \frac{p}{V_m} \quad (21)$$

The value of k' is, in effect, the equilibrium constant and may be used to compare affinities of different substances for the solid surface. The value of V_m is valuable because it indicates the maximum number of sites available for adsorption. In the case of physisorption, the maximum number of sites is actually the total surface area of the solid; therefore, the value of V_m can be used to estimate surface area if the volume and area/molecule of vapor are known.

Since physisorption most often involves some multilayered adsorption, an equation based on the Langmuir equation, the B.E.T. equation, normally is used to determine V_m and solid surface areas. Equation 22 is the B.E.T. equation:

$$V_a = \frac{V_m c p}{(p_0 - p) \left[1 + (C - 1)(p/p_0) \right]} \quad (22)$$

where c is a constant and p_0 is the vapor pressure of the adsorbing substance.⁹ Experimentally, the most widely used vapor for this purpose is nitrogen, which adsorbs nonspecifically on most solids near its boiling point at 195°C and appears to occupy about 16 Å²/molecule on a solid surface.

ADSORPTION FROM SOLUTION

By far one of the most important aspects of interfacial phenomena encountered in pharmaceutical systems is the tendency for substances dissolved in a liquid to adsorb to various interfaces. Adsorption from solution is generally more complex than that from the vapor state because of the influence of the solvent and any other solutes dissolved in the solvent. Although such adsorption generally is limited to one or two molecular layers at most, the presence of other molecules often makes the interpretation of adsorption mechanisms much more difficult than for chemisorption or physisorption of a vapor. Because monomolecular adsorption from solution is so widespread at all interfaces, we will first discuss the nature of monomolecular films and then return to a discussion of adsorption from solution.

INSOLUBLE MONOMOLECULAR FILMS

It was suggested above that molecules exhibiting a tendency to spread out at an interface might be expected to orient so as to reduce the interfacial free energy produced by the presence of the interface. Direct evidence for molecular orientation has been obtained from studies dealing with the spreading on water of insoluble polar substances containing long hydrocarbon chains, such as fatty acids.

In the late 19th century Pockels and Rayleigh showed that a very small amount of olive or castor oil, when placed on the surface of water, spreads out, as discussed above. If the amount of material was less than could physically cover the entire surface, only a slight reduction in the surface tension of water was noted. However, if the surface was compressed between barriers, as shown in Figure 19-11, the surface tension was reduced considerably.

Devaux extended the use of this technique by dissolving small amounts of solid in volatile solvents and dropping the solution onto a water surface. After assisting the water-insoluble molecules to spread, the solvent evaporated, leaving a surface film containing a known amount of solute.

Compression and measurement of surface tension indicated that a maximum reduction of surface was reached when the number of molecules/unit area was reduced to a value corresponding to complete coverage of the surface. This suggested that a monomolecular film forms and that surface tension is reduced upon compression because contact between air and water is reduced by the presence of the film molecules. Beyond the point of closest packing, the film apparently collapses very much as a layer of corks floating on water would be disrupted when laterally compressed beyond the point of initial physical contact.

Using a refined quantitative technique based on these studies, Langmuir¹¹ spread films of pure fatty acids, alcohols, and

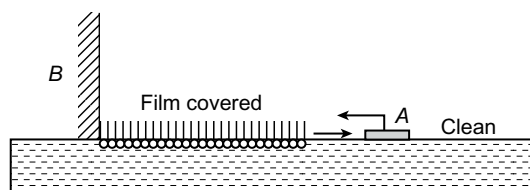


Figure 19.11. Insoluble monomolecular film compressed between a fixed barrier B and a movable barrier A . (Osipow L.I. *Surface Chemistry: Theory and Applications*. New York: Reinhold, 1962.)

esters on the surface of water. Comparing a series of saturated fatty acids, differing only in chain length, he found that the area/molecule at collapse was independent of chain length, corresponding to the cross-sectional area of a molecule oriented in a vertical position (see Figure 19-11). He further concluded that this molecular orientation involved association of the polar carboxyl group with the water phase and the nonpolar alkyl chain out toward the vapor phase.

In addition to the evidence for molecular orientation, Langmuir's work with surface films revealed that each substance exhibits film properties which reflect the interactions between molecules in the surface film. This is seen best by plotting the difference in surface tension of the clean surface γ_0 and that of the surface covered with the film γ , versus the area/molecule A produced by film compression (total area/the number of molecules). The difference in surface tension is called the surface pressure, π , and thus

$$\pi = \gamma_0 - \gamma \quad (23)$$

Figure 19-12 depicts such a plot for a typical fatty acid monomolecular film. At areas greater than 50 Å²/molecule, the molecules are far apart and do not cover enough surface to reduce the surface tension of the clean surface to any extent, and thus the lack of appreciable surface pressure. Because the molecules in the film are quite free to move laterally in the surface, they are said to be in a two-dimensional *gaseous* or *vapor* state.

As the intermolecular distance is reduced upon compression, the surface pressure rises because the air–water surface is being covered to a greater extent. The rate of change in γ with A , however, will depend on the extent of interaction between film molecules—the greater the rate of change, the more “condensed” the state of the film.

In Figure 19-12, from 50 to 30 Å²/molecule, the curve shows a steady increase in π , representative of a two-dimensional “liquid” film, where the molecules become more restricted in their freedom of movement because of interactions. Below 30 Å²/molecule, the increase in π occurs over a narrow range of A , characteristic of closest packing and a two-dimensional “solid” film.

Any factor tending to increase polarity or bulkiness of the molecule—such as increased charge, number of polar groups, reduction in chain length, or the introduction of aromatic rings, side chains, and double bonds—should reduce molecular interactions. On the other hand, the longer the alkyl chain and the less bulky the polar group, the closer the molecules can approach and the stronger the extent of interaction in the film.

SOLUBLE FILMS AND ADSORPTION FROM SOLUTION

If a fatty acid exhibits highly gaseous film behavior on an aqueous surface, a relatively small change in π with A over a

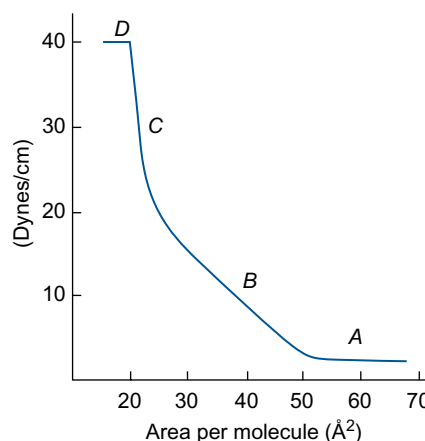


Figure 19.12. A surface pressure–area curve for an insoluble monomolecular film: Region A, gaseous film; Region B, liquid film; Region C, solid film; Region D, film collapse.

considerable range of compression should be expected. Indeed, for short-chain compounds, such as lauric acid (12 carbons) or decanoic acid, not only is the change small with decreasing A , but at a point just before the expected closest packing area, the surface pressure becomes constant without any collapse.

If lauric acid is converted to the laurate ion, or if a shorter chain acid such as octanoic acid is used, spreading on water and compression of the surface produces no increase in π . These results illustrate that the more polar the molecule (hence, the more *gaseous* the film), the higher the area/molecule where a constant surface pressure occurs. This behavior may be explained by assuming that polar molecules form monomolecular films when spread on water, but that, upon compression, they are caused to enter the aqueous bulk solution rather than to remain as an intact, insoluble film. The constant surface pressure with increased compression arises because a constant number of molecules/unit area remain at the surface in equilibrium with dissolved molecules. The extent of such behavior will be greater for substances exhibiting weaker intermolecular interaction and greater water solubility.

Starting from the other direction, it can be shown that short-chain acids and alcohols (when dissolved in water) reduce the surface tension of water, thus producing a surface pressure, just as with insoluble films (see Equation 23). That dissolved molecules are accumulating at the interface in the form of a monomolecular film is suggested from the similarity in behavior to systems where lightly soluble molecules are spread on the surface. For example, compressing the surface of a solution containing "surface-active" molecules has no effect on the initial surface pressure, whereas increasing bulk-solution concentration tends to increase surface pressure, presumably by shifting the equilibrium between surface and bulk molecules.

At this point one may ask, why should water-soluble molecules leave an aqueous phase and accumulate or *adsorb* at an air-solution interface? Because any process will occur spontaneously if it results in a net loss in free energy, such must be the case for the process of adsorption. A number of factors will produce such a favorable change in free energy:

- The presence of the oriented monomolecular film reduces the surface free energy of the air-water interface.
- The hydrophobic group on the molecule is in a lower state of energy at the interface, where it no longer is as surrounded by water molecules, than when it is in the bulk-solution phase.
- Increased interaction between film molecules also will contribute to this process.

A further reduction in free energy occurs upon adsorption because of the gain in entropy associated with a change in water structure. Water molecules in the presence of dissolved alkyl chains are more highly organized or *ice-like* than they are as a pure bulk phase; hence, the entropy of such structured water is lower than that of bulk water.

The process of adsorption requires that the ice-like structure *melt* as the chains go to the interface, and thus an increase in the entropy of water occurs. The adsorption of molecules dissolved in oil can occur, but it is not influenced by water structure changes, and hence, only the first factors mentioned are important here.

It is very rare that significant adsorption can occur at the hydrocarbon-air interface because little loss in free energy comes about by bringing hydrocarbon chains with polar groups attached to this interface. On the other hand, at oil-water interfaces, the polar portions of the molecule can interact with water at the interface, leading to significant adsorption.

Thus, whereas water-soluble fatty acid salts are adsorbed from water to air-water and oil-water interfaces, their undissociated counterparts, the free fatty acids, which are water insoluble, form insoluble films at the air-water interface, are not adsorbed from oil solution to an oil-air interface but show

significant adsorption at the oil-water interface when dissolved in oil.

From this discussion it is possible also to conclude that adsorption from aqueous solution requires a lower solute concentration to obtain the same level of adsorption if the hydrophobic chain length is increased or if the polar portion of the molecule is less hydrophilic. On the other hand, adsorption from nonpolar solvents is favored when the solute is quite polar.

Because soluble or adsorbed films cannot be compressed, there is no simple, direct way to estimate the number of molecules/unit area coming to the surface under a given set of conditions. For relatively simple systems it is possible to estimate this value by application of the Gibbs equation, which relates surface concentration to the surface-tension change produced at different solute activities. The derivation of this equation is beyond the scope of this discussion, but it arises from a classical thermodynamic treatment of the change in free energy when molecules concentrate at the boundary between two phases. The equation may be expressed as

$$\Gamma = -\frac{a}{RT} \frac{d\gamma}{da} \quad (24)$$

where Γ is the moles of solute adsorbed/unit area, R is the gas constant, T is the absolute temperature, and $d\gamma$ is the change in surface tension, with a change in solute activity, da , at activity a .

For dilute solutions of nonelectrolytes, or for electrolytes when the Debye-Hückel equation for activity coefficient is applicable, the value of a may be replaced by solute concentration, c . Because the term dc/c is equal to $d \ln c$, the Gibbs equation is often written as

$$\Gamma = -\frac{1}{RT} \frac{d\gamma}{d \ln c} \quad (25)$$

In this way the slope of a plot of γ versus $\ln c$ multiplied by $1/RT$ should give Γ at a particular value of c .

Figure 19-13 depicts typical plots for a series of water-soluble surface-active agents differing only in the alkyl chain length. A greater reduction of surface tension occurs at lower concentrations for longer chain-length compounds. In addition, there are greater slopes with increasing concentration, indicating more adsorption (Equation 25), and an abrupt leveling of surface tension at higher concentrations. This latter behavior reflects the self-association of surface-active agent to form micelles, which exhibit no further tendency to reduce surface tension. The topic of micelles will be discussed later in Chapter 20.

If one plots the values of surface concentration, Γ versus concentration c , for substances adsorbing to the vapor-liquid and liquid-liquid interfaces, using data such as those given in Figure 19-13, one generally obtains an adsorption isotherm shaped like those in Figure 19-9 for vapor adsorption. Indeed, it can be

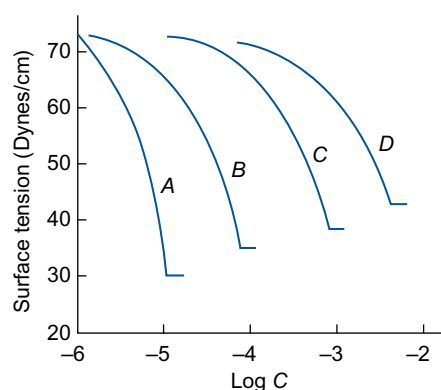


Figure 19.13. The effect of increasing chain length on the surface activity of a surfactant at the air-aqueous solution interface (each curve differs from the preceding or succeeding by two methylene groups, with A, the longest chain, and D, the shortest).

shown that the Langmuir equation (Equation 20) can be fitted to such data when written in the form

$$\Gamma = \frac{\Gamma_{\max} k'c}{1 + k'c} \quad (26)$$

where $\Gamma_{\max}$ is the maximum surface concentration attained with increasing concentration and k' is related to k in Equation 20. Combining Equations 24 and 26 leads to a widely used relationship between surface-tension change π (see Equation 23), and solute concentration c , known as the Szyszkowski equation.

$$\Pi = \Gamma_{\max} RT \ln(1 + k'c) \quad (27)$$

MIXED FILMS

It would seem reasonable to expect that the properties of a surface film could be varied greatly if a mixture of surface-active agents were in the film. As an example, consider that a mixture of short- and long-chain fatty acids would be expected to show a degree of *condensation* varying from the gaseous state, when the short-chain substance is used in high amount, to a highly condensed state, when the longer chain substance predominates. Thus, each component in such a case would operate independently by bringing a proportional amount of film behavior to the system.

More often, the ingredients of a surface film do not behave independently but rather interact to produce a new surface film. An obvious example would be the combination of organic amines and acids, which are charged oppositely and would be expected to interact strongly. In addition to such polar-group interactions, chain-chain interactions strongly favor mixed condensed films. An important example of such a case occurs when a long-chain alcohol is introduced along with an ionized long-chain substance. Together the molecules form a highly condensed film, despite the presence of a high number of like charges. Presumably this occurs, as seen in Figure 19-14, by arranging the molecules so that ionic groups alternate with alcohol groups; however, if chain-chain interactions are not strong, the ionic species often will be displaced by the more nonpolar un-ionized species and will "desorb" into the bulk solution.

On the other hand, sometimes the more soluble surface-active agent produces surface pressures in excess of the collapse pressure of the insoluble film and displaces it from the surface. This is an important concept because it is the underlying principle behind cell lysis by surface-active agents and some drugs, and behind the important process of detergency.

ADSORPTION FROM SOLUTION ON TO SOLID SURFACES

Adsorption to solid surfaces from solution may occur if the dissolved molecules and the solid surface have chemical groups capable of interacting. Nonspecific adsorption also will occur if the solute is surface active and if the surface area of the solid is high. This latter case would be the same as occurs at the vapor-liquid and liquid-liquid interfaces. As with adsorption to liquid interfaces, adsorption to solid surfaces from solution generally leads to a monomolecular layer, often described by the Langmuir equation in the form:

$$x/M = [(x/M)mk^*c]/(1 + k^*c) \quad (28)$$

where x is the amount of adsorbed solute, M is the total weight of solid, x/M is the amount of solute adsorbed per unit weight

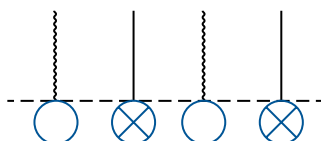


Figure 19.14. A mixed monomolecular film; a long-chain ion; a long-chain non-ionic compound.

of solid at concentration c , k^* is a constant, and $(x/M)m$ is the amount of solute per unit weight covering the surface with a complete monolayer. However, as Giles¹² has pointed out, the variety of combinations of solutes and solids, and hence the variety of possible mechanisms of adsorption, can lead to a number of more complex isotherms. In particular, adsorption of surfactants and polymers, of great importance in a number of pharmaceutical systems, still is not understood well on a fundamental level and may in some situations even be multilayered.

Adsorption from solution may be measured by separating solid and solution and either estimating the amount of adsorbate adhering to the solid or the loss in concentration of adsorbate from solution. In view of the possibility of solvent adsorption, the latter approach really only gives an apparent adsorption. For example, if solvent adsorption is great enough, it is possible to end up with an increased concentration of solute after contact with the solid; here, the term *negative adsorption* is used.

Solvent not only influences adsorption by competing for the surface, but as discussed in connection with adsorption at liquid surfaces, the solvent will determine the escaping tendency of a solute; for example, the more polar the molecule, the less the adsorption that occurs from water. This is seen in Figure 19-15, where adsorption of various fatty acids from water onto charcoal increases with increasing alkyl chain length or nonpolarity. It is difficult to predict these effects, but, in general, the more chemically unlike the solute and solvent and the more alike the solid surface groups and solute, the greater the extent of adsorption. Another factor that must be kept in mind is that charged solid surfaces, such as polyelectrolytes, will strongly adsorb oppositely charged solutes. This is similar to the strong specific binding seen in gas chemisorption, and it is characterized by significant monolayer adsorption at very low concentrations of solute. See Figure 19-16 for an example of such adsorption.

Adsorption onto activated charcoal has been shown to be extremely useful in the emergency treatment of acute overdosage of a variety of drugs taken by the oral route.¹⁵ Overall effectiveness of commercially available activated-charcoal suspensions as an antidote in oral poisonings appears to be directly related to the total charcoal surface area.¹⁶ Drug adsorption to charcoal tends to follow the Langmuir model as well as the Freundlich model. In addition, a drug that is un-ionized at gastric pH will

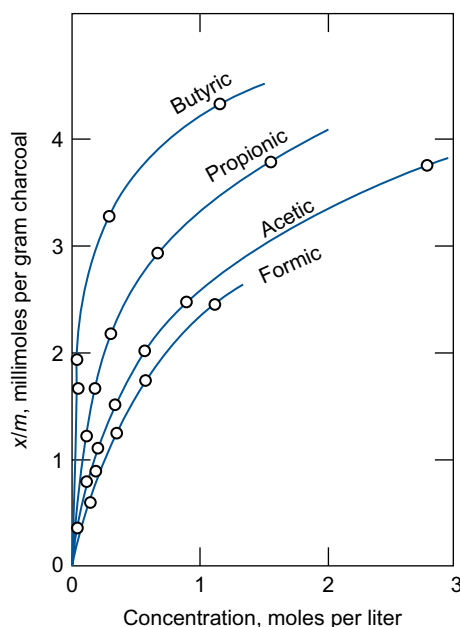


Figure 19.15. The relation between adsorption and molecular weight of fatty acids. (Data from Weiser HB. *A Textbook of Colloid Chemistry*. New York: Elsevier, 1949.)

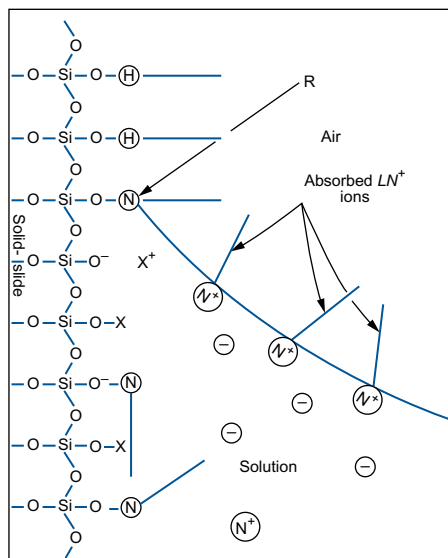


Figure 19.16. The adsorption of a cationic surfactant, LN, onto a negatively charged silica or glass surface, exposing a hydrophobic surface as the solid is exposed to air. (From TerMinassianSaraga L. *Chemisorption and dewetting of glass and silica. Adv Chem Ser* 1964; 43: 232–49.)

adsorb to charcoal to a greater extent than will the ionized form of the drug, probably because of less repulsive interactions in the adsorbed state of neutral molecules. Great care must be exercised in the formulation of activated-charcoal suspensions because pharmaceutical adjuvants employed in suspensions have the potential to adsorb to the charcoal and block sites for drug adsorption.

SURFACE-ACTIVE AGENTS

Throughout the discussion so far, examples of surface-active agents (surfactants) have been restricted primarily to fatty acids and their salts. It has been shown that both a hydrophobic portion (alkyl chain) and a hydrophilic portion (carboxyl and carboxylate groups) are required for their surface activity, the relative degree of polarity determining the tendency to accumulate at interfaces. It now becomes important to look at some of the specific types of surfactants available and to see what structural features are required for different pharmaceutical applications.

The classification of surfactants is quite arbitrary, but one based on chemical structure appears best as a means of introducing the topic. It is generally convenient to categorize surfactants according to their polar portions because the nonpolar portion usually is made up of alkyl or aryl groups. The major polar groups found in most surfactants may be divided as follows: anionic, cationic, amphoteric, and nonionic. As shall be seen, the last group is the largest and most widely used for pharmaceutical systems, so that it will be emphasized in the discussion that follows.

TYPES

Anionic Agents

The most commonly used anionic surfactants are those containing carboxylate, sulfonate, and sulfate ions. Those containing carboxylate ions are known as soaps and generally are prepared by the saponification of natural fatty acid glycerides in alkaline solution. The most common cations associated with soaps are sodium, potassium, ammonium, and triethanolamine; the chain length of the fatty acids ranges from 12 to 18.

The extent of solubility in water is influenced greatly by the length of the alkyl chain and the presence of double bonds. For

example, sodium stearate is quite insoluble in water at room temperature, whereas sodium oleate under the same conditions is quite water soluble.

Multivalent ions, such as calcium and magnesium, produce marked water insolubility, even at lower alkyl chain lengths; thus, soaps are not useful in hard water that is high in content of these ions. Soaps, being salts of weak acids, are subject also to hydrolysis and the formation of free acid plus hydroxide ion, particularly when in more concentrated solution.

To offset some of the disadvantages of soaps, a number of long alkyl chain sulfonates, as well as alkyl aryl sulfonates such as sodium dodecylbenzene sulfonate, may be used; the sulfonate ion is less subject to hydrolysis and precipitation in the presence of multivalent ions. A popular group of sulfonates, widely used in pharmaceutical systems, are the dialkyl sodium sulfosuccinates, particularly sodium bis (2-ethylhexyl)sulfosuccinate, best known as Aerosol OT or docusate sodium. This compound is unique in that it is soluble both in oil and water and, hence, forms micelles in both phases. It reduces surface and interfacial tension to low values and acts as an excellent wetting agent in many types of solid dosage forms (Table 19-7).

A number of alkyl sulfates are available as surfactants, but by far the most popular member of this group is sodium lauryl sulfate, which is used widely as an emulsifier and solubilizer in pharmaceutical systems. Unlike the sulfonates, *sulfates* are susceptible to pH-dependent hydrolysis, leading to the formation of the long-chain alcohol.

Cationic Agents

A number of long-chain cations, such as amine salts and quaternary ammonium salts, often are used as surface-active agents when dissolved in water; however, their use in pharmaceutical preparations is limited to that of antimicrobial preservation rather than as surfactants. This arises because the cations adsorb so readily at cell membrane structures in a nonspecific manner, leading to cell lysis (e.g., hemolysis), as do anionics to a lesser extent. It is in this way that they act to destroy bacteria and fungi.

Since anionic and nonionic agents are not as effective as preservatives, one must conclude that the positive charge of these compounds is important; however, the extent of surface activity has been shown to determine the amount of material needed for a given amount of preservation. Quaternary ammonium salts are preferable to free amine salts because they are not subject to effect by pH in any way; however, the presence of organic anions, such as dyes and natural polyelectrolytes, is an important source of incompatibility, and such a combination should be avoided.

Amphoteric Agents

The major groups of molecules falling into the amphoteric category are those containing carboxylate or phosphate groups

Table 19-7. Effect of Aerosol OT Concentration on the Surface Tension of Water and The Contact Angle of Water With Magnesium Stearate

Concentration (M $\times 10^6$)	γ_{SV}	$\theta(^{\circ})$
1.0	60.1	120
3.0	49.8	113
5.0	45.1	104
8.0	40.6	89
10.0	38.6	80
12.0	37.9	71
15.0	35.0	63
20.0	32.4	54
25.0	29.5	50

such as the anion, and amino or qua-ternary ammonium groups such as the cation. The former group is represented by various polypeptides, proteins, and the alkyl betaines; the latter group consists of natural phospholipids, such as the lecithins and cephalins. In general, long-chain amphoteries, which exist in solution in zwitterionic form, are more surface active than are ionic surfactants having the same hydrophobic group because, in effect, the oppositely charged ions are neutralized. However, when compared to nonionics, they appear somewhere between ionic and nonionic.

Proteins

Considering the rapidly growing importance of proteins as therapeutic agents, the unique surface characteristics of these biological macromolecules deserve some special attention. Therapeutic proteins have been shown to be extremely surface active, and they adsorb to clinically important surfaces, such as glass bottles and syringes, sterile filters, and plastic IV bags and administration sets; the result is treatment failures. In general, proteins can adsorb to a whole variety of surfaces, both hydrophobic and hydrophilic. From the standpoint of the surface, protein adsorption appears to be maximized when the electrical charge of the surface is opposite that of the protein or when the surface is extremely hydrophobic. From the standpoint of the protein, the extent of adsorption depends on the molecular weight, the number of hydrophobic side chains, and the relative distribution of cationic and anionic side chains. The effect of ionic strength is usually to enhance adsorption by shielding adjacent proteins from repulsive electrical interactions. Adsorption is also maximized when the pH of the protein solution is equal to the pI (isoelectric point) of the molecule, again due to minimized electrical repulsion.

When different proteins compete for adsorption sites on a single surface, the effect of molecular weight becomes most striking. Early in the adsorption process the protein with the smaller molecular weight, which can diffuse to the surface more rapidly, initially occupies the interface. After some time, it is found that the larger molecular weight protein has displaced the smaller protein since the larger molecule has more possible interaction points with the surface and, thus, greater total energy of interaction.

The most important consequence of therapeutic protein adsorption is the loss of bioactivity, the reasons for which include loss of therapeutic agent by irreversible adsorption to the surface, possible structural changes in the protein induced by the interface, and surface-associated aggregation and precipitation of the protein. Each of these consequences is related to the structure adopted by the protein in the interfacial region. The native three-dimensional structure of a protein in solution is the result of a complex balance between attractive and repulsive forces. Surface can easily disrupt the balance of forces in proteins residing in the interfacial region and cause the molecule to undergo a change, unfolding from the native to the extended configuration. Because it is unlikely that the extended configuration will refold back to the native state upon release from the interface, the protein is considered to be denatured. Like other polymers, the unfolding of the protein at the interface is thought to minimize the contact of apolar amino acid side chains with water.

In addition, electrical interactions, both within the protein and between the protein and the surface, strongly modulate the configuration assumed at the interface. Motion of the interface, such as comes about during shaking of a solution, appears to accelerate the surface-associated denaturation. Some proteins appear to be rather vulnerable to surface-induced structural alterations, whereas others are very resistant. Algorithms for predicting those proteins most vulnerable to the structure-damaging effects of interfaces are not yet available. Empirical observations suggest that those proteins easily denatured in solution by elevated temperatures may also be most sensitive to interfacial denaturation.

The best defense against untoward effects on the structure of proteins induced by surfaces appears to be prevention of adsorption. Research in the field of biomaterials has shown that surfaces that are highly hydrophilic are less likely to serve as sites for protein adsorption. Steric hindrance of adsorption by bonding hydrophilic polymers, such as polyethylene oxide, to a surface also appears to be successful in minimizing adsorption. Formulations of proteins intended for parenteral administration frequently contain synthetic surfactants to preserve bioactivity. The specific molecular mechanism of protection is not understood and can involve specific blocking of adsorption to the interface or enhanced removal from the interface before protein unfolding can occur. In support of the former mechanism is the observation that surfactants most successful at protecting proteins from interfacial denaturation contain long polyethylene oxide chains capable of blocking access of the protein to the surface.

Phospholipids

All lecithins contain the L-glycerophosphocholine skeleton esterified to two long-chain fatty acids (often oleic, palmitic, stearic, and linoleic). Typically, for pharmaceutical use, lecithins are derived from egg yolk or soybean. Although possessing a polar zwitterionic *head* group, the twin hydrocarbon *tails* result in a surfactant with very low water solubility in the monomer state. With the exception of the skin, phospholipids make up a vast majority of the lipid component of cell membranes throughout the body. As a result, the biocompatibility of lecithin is high, accounting for the increasing popularity of use in formulations intended for oral, topical, and intravenous use. Egg yolk lecithins are used extensively as the main emulsifying agent in the fat emulsions intended for intravenous use.

The ability of the lecithins to form a tough but flexible film between the oil and water phases is responsible for the excellent physical stability shown in the IV fat emulsions. In aqueous media phospholipids are capable of assembling into concentric bilayer structures, known as liposomes. The therapeutic advantage of such a lipid assembly for drug delivery depends upon the encapsulation of the active ingredient either within the interior aqueous environment or within the hydrophobic region of the bilayer. Deposition of the liposome within the body appears to be dependent upon a number of factors, including the composition of the phospholipids employed in the bilayer and the diameter of the liposome.

The unique surface properties of phospholipids are critical to the function of the pulmonary system. Pulmonary surfactant is a mixture of phospholipids and other associated molecules secreted by type II pneumocytes. In the absence of pulmonary surfactant (as in a neonate born prematurely), the high surface energy of the pulmonary alveoli and airways can be diminished only by physical collapse of these structures and resulting elimination of the air-water interface. As a consequence of airway collapse, the lung fails to act as an organ of gas exchange. Pulmonary surfactant maintains the morphology and function of the alveoli and airways by markedly decreasing surface energy through decreasing the surface tension of the air-water interface.

The most prevalent component of pulmonary surfactant, dipalmitoylphosphatidylcholine (DPPC), is uniquely responsible for forming the very rigid surface film necessary to reduce the surface tension of the interface to a value near 0. Such an extreme reduction in surface tension is most critical during the process of exhalation of the lung, where the air-water interfacial area is decreasing. Although DPPC does form the rigid film, in the absence of additives, it is unable to respread over an expanding interface typical of a lung during the inhalation phase. An anionic phospholipid, phosphatidylglycerol, in conjunction with a surfactant-associated protein, SP-C, appears to aid the respreading of DPPC and to maintain mechanical stability of the interface. A truly remarkable feature is that pulmonary surfactant is able to carry out the cycle of reducing surface tension

to near 0 during exhalation and then re-expanding over the interface during inhalation at whatever rate is necessary by the respiratory pattern.

Commercially available pulmonary surfactant replacement preparations contain DPPC as the primary ingredient. Agents that aid in the respreading of DPPC may differ, depending upon the source of the surface-active material.

Nonionic Agents

The major class of compounds used in pharmaceutical systems are the nonionic surfactants, for their advantages with respect to compatibility, stability, and potential toxicity are quite significant. It is convenient to divide these compounds into those that are relatively water insoluble and those that are quite water soluble. The major types of compounds making up this first group are the long-chain fatty acids and their water-insoluble derivatives. These include

- Fatty alcohols, such as lauryl, cetyl (16 carbons), and stearyl alcohols
- Glyceryl esters, such as the naturally occurring mono-, di-, and triglycerides
- Fatty acid esters of fatty alcohols and other alcohols, such as propylene glycol, polyethylene glycol, sorbitan, sucrose, and cholesterol. Included also in this general class of nonionic water-insoluble compounds are the free steroidal alcohols, such as cholesterol.

To increase the water solubility of these compounds and to form the second group of nonionic agents, polyoxyethylene groups are added through an ether linkage with one of their alcohol groups. The list of derivatives available is much too long to cover completely, but a few general categories will be given.

The most widely used compounds are the polyoxyethylene sorbitan fatty acid esters, found in pharmaceutical formulations that are to be used both internally and externally. Closely related compounds include polyoxyethylene glyceryl and steroidal esters, as well as the comparable polyoxypropylene esters. It is also possible to have a direct ether linkage with the hydrophobic group, as with a polyoxyethylene-stearyl ether or a polyoxyethylene-alkyl phenol. These ethers offer advantages because, unlike the esters, they are quite resistant to acidic or alkaline hydrolysis.

Besides the classification of surfactants according to their polar portion, it is useful to have a method that categorizes them in a manner that reflects their interfacial activity and their ability to function as wetting agents, emulsifiers, and solubilizers. Variation in the relative polarity or nonpolarity of a surfactant significantly influences its interfacial behavior, so some measure of polarity or nonpolarity should be useful as a means of classification.

One such approach assigns a hydrophile-lipophile balance (HLB) number for each surfactant; although the method was developed by a commercial supplier of one group of surfactants, it has received widespread application.

The HLB value, as originally conceived for nonionic surfactants, is merely the percentage weight of the hydrophilic group divided by 5 in order to reduce the range of values. On a molar basis, therefore, a 100% hydrophilic molecule (polyethylene glycol) would have a value of 20. Thus, an increase in polyoxyethylene chain length increases polarity and, hence, the HLB value; at constant polar chain length, an increase in alkyl chain length or number of fatty acid groups decreases polarity and the HLB value. One immediate advantage of this system is that, to a first approximation, one can compare any chemical type of surfactant to another type when both polar and nonpolar groups are different.

Values of HLB for nonionics are calculable on the basis of the proportion of polyoxyethylene chain present; however, to determine values for other types of surfactants, it is necessary to compare physical chemical properties reflecting polarity with those surfactants having known HLB values.

Relationships between HLB and phenomena, such as water solubility, interfacial tension, and dielectric constant have been used. Those surfactants exhibiting values greater than 20 (e.g., sodium lauryl sulfate) demonstrate hydrophilic behavior in excess of the polyoxyethylene groups alone. Refer to Chapter 21 for further information.

ACKNOWLEDGMENT: The author is grateful to Professor George Zografi for his continuing mentorship and support.

REFERENCES

1. Semat H. *Fundamentals of Physics*, 3rd edn. New York: Holt Rinehart Winston, 1957.
2. Michaels AS *et al.* Modification of growth rate and habit of adipic acid crystals with surfactants. *J Phys Chem* 1961; 65(10): 1730-1737
3. Ring TA. Kinetic effects on particle morphology and size distribution during batch precipitation. *Powder Tech* 1991; 65: 195-205
4. Elamin AA *et al.* Increased metastable solubility of milled griseofulvin, depending on the formation of a disordered surface structure. *Int J Pharmaceut* 1994; 111: 159-170.
5. Dirksen JARing TA Fundamental of crystallization: kinetic effects on particle size distributions and morphology. *Chem Eng Sci* 1991; 46: 2389-2427
6. Zisman WA. Relation of the equilibrium contact angle to liquid and solid constitution. *Adv Chem Ser* 1964; 43: 1-51
7. Putz G *et al.* Comparison of captive and pulsating bubble surfactometers with use of lung surfactants. *J Appl Physiol* 1994; 76: 1425-1431.
8. Titoff ZA. Die Adsorption von Gasen durch Kohle. *Z Phys Chem* 1910; 74(64): 641-678
9. Brittain HG. *Physical Characterization of Pharmaceutical Solids*. New York: Dekker, 1995.
10. Osipow LI. *Surface Chemistry: Theory and Applications*. New York: Reinhold, 1962.
11. Langmuir I. The Constitution and fundamental properties of solids and liquids. II. Liquids. *J Am Chem Soc* 1917; 39: 1848-1906
12. Giles CH. In: EH Lucassen-Reynders, ed. *Anionic Surfactants*. New York: Dekker, 1981; Chap 4.
13. Weiser HB. *A Textbook of Colloid Chemistry*. New York: Elsevier, 1949.
14. TerMinassianSaraga L. Chemisorption and dewetting of glass and silica. *Adv Chem Ser* 1964; 43: 232-249
15. Cooney DO. *Activated Charcoal in Medical Applications*. New York: Dekker, 1995.
16. Modi NB *et al.* Phenobarbital removable characteristics of three brands of activated charcoals: a system analysis approach. *Pharm Res* 1994; 11: 318-323.

BIBLIOGRAPHY

- Adamson AW. *Physical Chemistry of Surfaces*, 5th edn. New York: Wiley Interscience, 1990.
- David JT, Rideal EK. *Interfacial Phenomena*, 2nd edn. New York: Academic Press, 1963.
- Hiemenz PC. *Principles of Colloid and Surface Chemistry*, 2nd edn. New York: Dekker, 1986.
- MacRitchie F. *Chemistry at Interfaces*. San Diego: Academic Press, 1990.
- Shaw DJ. *Introduction to Colloid and Surface Chemistry*, 4th edn. London: Butterworths, 1992.

Chapter 20

Colloidal Dispersions

Bill J. Bowman, RPh, PhD; Clyde M. Ofner III, PhD; Hans Schott, PhD and Yvonne Perrie, PhD

DEFINITIONS AND CLASSIFICATIONS 343

PROPERTIES OF COLLOIDAL DISPERSIONS 345

LYOPHILIC DISPERSIONS 355

LYOPHOBIC DISPERSIONS 356

ASSOCIATION COLLOIDS 361

PHARMACEUTICAL APPLICATIONS OF COLLOIDAL SYSTEMS 365

The British chemist Thomas Graham applied the term “colloid” (derived from the Greek word for glue) about 1850 to polypeptides such as albumin and gelatin, to polysaccharides such as acacia, starch, and dextrin, and to inorganic compounds such as gelatinous metal hydroxides and Prussian blue (ferric ferrocyanide). Contemporary colloid and surface chemistry deals with an unusually wide variety of industrial and biological systems.

Within the field of pharmacy, the colloidal systems we commonly come across include some protein and polymer solutions, micellar systems, liquid crystals and emulsions, suspensions, aerosols, foams and other drug delivery systems that fall within the colloidal size range. In some instances, what was formally referred to as colloidal systems have been replaced by the term nanotechnology (structures that have one or more dimension between approximately 1 and 100 nm).

DEFINITIONS AND CLASSIFICATIONS

COLLOIDAL SYSTEMS AND INTERFACES

Except for high-molecular-weight polymers, most soluble substances can be prepared as either low-molecular-weight solutions, colloidal dispersions, or coarse suspensions, depending upon the choice of dispersion medium and dispersion technique. Colloidal dispersions consist of at least two phases: one or more dispersed or internal phases, and a continuous or external phase called the *dispersion medium* or *vehicle*. Colloidal dispersions are distinguished from solutions and coarse dispersions by the particle size of the dispersed phase, not its composition. Whilst somewhat arbitrary, colloidal dispersions can be characterized as containing particles in the size range of between approximately 1 nm and 1 micrometer, however a smaller size range of up to 500 nm is also quoted. Thus, blood, cell membranes, micelles, thinner nerve fibers, milk, rubber latex, fog, and beer foam are colloidal systems. Some materials, such as emulsions and suspensions of most organic drugs, are coarser than true colloidal systems but exhibit similar behavior. Even though serum albumin, acacia, and povidone form true or molecular solutions in water, the size of the individual solute molecules places such solutions in the colloidal range (particle size >1 nm).

In general, there are several features that distinguish colloidal dispersions from coarse suspensions and emulsions. Colloidal particles are usually too small for visibility in a light microscope, because at least one of their dimensions measures 1 micrometer or less. They are, however, often visible in an ultramicroscope, and almost always in an electron microscope. Conversely, coarse suspended particles are usually visible to the naked eye and always in a light microscope. In addition, colloidal particles, as opposed to coarse particles, pass through ordinary filter paper but are retained by dialysis or ultrafiltration membranes. Also, unlike coarse particles, colloidal particles diffuse very slowly and undergo little or no sedimentation or creaming. Brownian motion (see Diffusion and Sedimentation below) maintains the dispersion of the colloidal internal phases.

An appreciable fraction of atoms, ions, or molecules of colloidal particles are located in the boundary layer between the particle and the dispersion medium. The boundary layer between a particle and air is commonly referred to as a *surface*, whereas the boundary layer between a particle and a liquid or solid is commonly referred to as an *interface*. The ions or molecules within the particle and within the medium are surrounded on all sides by similar ions or molecules and have balanced force fields; however, the ions or molecules at surfaces or interfaces are subjected to unbalanced forces of attraction. Consequently, a surface free-energy component is added to the total free energy of colloidal particles and becomes important as the particles become smaller and a greater fraction of their atoms, ions, or molecules are located in the surface or interfacial region. As a result, the solubility of very fine solid particles and the vapor pressure of very small liquid droplets are greater than the corresponding values for coarse particles and drops of the same materials.

Specific Surface Area

Decreasing particle size increases the surface-to-volume ratio, expressed as the *specific surface area* (A_{sp}). Specific surface area can be expressed as the area (A , cm^2) per unit volume (V , cm^3) or per unit mass (M , grams). For a sphere, $A = 4\pi r^2$ and $V = 4/3\pi r^3$, then A_{sp} is:

$$A_{sp} = \frac{A}{V} = \frac{4\pi}{4/3\pi r^3} = \frac{3}{r} \frac{\text{cm}^2}{\text{cm}^3} = \frac{3}{r} \text{cm}^{-1} \quad (1)$$

For density (d) of the material expressed as g/cm^3 , the specific surface area is:

$$A_{sp} = \frac{A}{M} = \frac{A}{Vd} = \frac{4\pi r^2}{4/3\pi r^3 d} = \frac{3}{rd} \frac{\text{cm}^2}{\text{g}} \quad (2)$$

Table 20-1 illustrates the effect of comminution on the specific surface area of a material initially consisting of one sphere

Table 20-1. Effect of Comminution on the Specific Surface Area of a Volume of $4/3 \text{ cm}^3$, Divided into Uniform Spheres of Radius R

Number of Spheres	R	$A_{sp} \text{ cm}^2/\text{cm}^3$
1	1 cm	3
10^3	0.1 cm > 1 mm	3×10
10^6	0.1 mm	3×10^2
10^9	0.01 mm	3×10^3
10^{12}	1 0m	3×10^4
10^{15}	0.1 micrometer	3×10^5
10^{18}	0.01 micrometer	3×10^6
10^{21}	1 nm	3×10^7
10^{23}	0.1 nm	3×10^8

Shaded region corresponds to colloidal particle-size range.

having a 1-cm radius. The specific surface area increases as the material is broken into a larger number of smaller and smaller spheres. Activated charcoal and kaolin are solid adsorbents having specific surface areas of about $6 \times 10^6 \text{ cm}^2/\text{g}$ and $10^4 \text{ cm}^2/\text{g}$, respectively. One gram of activated charcoal has a surface area equal to approximately 700 square meters because of its extensive porosity and internal voids.

PHYSICAL STATES OF DISPERSED AND CONTINUOUS PHASES

A useful classification of colloidal systems is based upon the state of matter of the dispersed phase and the dispersion medium (i.e., whether they are solid, liquid, or gaseous).^{1,2} Common examples and various combinations are shown in Table 20-2. The terms *sols* and *gels* are often applied to colloidal dispersions of a solid in a liquid or gaseous medium. Sols tend to have a lower viscosity and are fluid. If the solid particles form bridged structures possessing some mechanical strength, the system is then called a gel. Prefixes typically designate the dispersion medium. For example, hydrosol (or hydrogel), alcisol (or alcogel), and aerosol (or aerogel) designate water, alcohol, and air, respectively.

INTERACTION BETWEEN DISPERSED PHASES AND DISPERSION MEDIUMS

Ostwald originated another useful classification of colloidal dispersions based on the affinity or interaction between the dispersed phase and the dispersion medium.² This classification refers mostly to solid-in-liquid dispersions. Colloidal dispersions are divided into the two broad categories, *lyophilic* and *lyophobic*. Some soluble, low-molecular-weight substances have molecules with both tendencies and associate in solution, forming a third category called *association colloids*.

Lyophilic Dispersions

The system is said to be *lyophilic* (solvent-loving) if there is considerable attraction between the dispersed phase and the liquid vehicle (i.e., extensive solvation). The system is said to be *hydrophilic* if the dispersion medium is water. Due to the presence of high concentrations of hydrophilic groups, solids such as bentonite, starch, gelatin, acacia, and povidone swell, disperse, or dissolve spontaneously in water to the greatest degree possible without breaking covalent bonds. Hydrophilic colloids often contain ionized groups that dissociate into highly

hydrated ions (e.g., carboxylate, sulfonate, and alkylammonium ions) and/or organic functional groups that bind water through hydrogen bonding (e.g., hydroxyl, carbonyl, amino, and imino groups).

Hydrophilic colloidal dispersions can be further subdivided as:

- *True solutions*: water-soluble polymers (e.g., acacia and povidone)
- *Gelled solutions, gels, or jellies*: polymers present at sufficiently high concentrations and/or at temperatures where their water solubility is low, such as relatively concentrated solutions of gelatin and starch (which set to gels upon cooling) and methylcellulose (which gels upon heating)
- *Particulate dispersions*: solids that do not form molecular solutions but remain as discrete though minute particles (e.g., bentonite and microcrystalline cellulose)

Lipophilic or *oleophilic* substances have a strong affinity for oils. Oils are nonpolar liquids consisting mainly of hydrocarbons having few polar groups and low dielectric constants. Examples include mineral oil, benzene, carbon tetrachloride, vegetable oils (e.g., cottonseed or peanut oil), and essential oils (e.g., lemon or peppermint oil). Oleophilic colloidal dispersions include polymers such as polystyrene and unvulcanized or gum rubber dissolved in benzene, magnesium, or aluminum stearate dissolved or dispersed in cottonseed oil, and activated charcoal, which forms sols or particulate dispersions in all oils.

Lyophobic Dispersions

The dispersion is said to be *lyophobic* (solvent-hating) when there is little attraction between the dispersed phase and the dispersion medium. *Hydrophobic* dispersions consist of particles that are only hydrated slightly or not at all, because water molecules prefer to interact with one another instead of solvating the particles. Therefore, such particles do not disperse or dissolve spontaneously in water. Examples of materials that form hydrophobic dispersions include organic compounds consisting largely of hydrocarbon portions with few, if any, hydrophilic functional groups (e.g., cholesterol and other steroids); some nonionized inorganic substances (e.g., sulfur); and oleophilic materials such as polystyrene or gum rubber, organic lipophilic drugs, paraffin wax, magnesium stearate, and cottonseed or soybean oils. Materials such as sulfur, silver chloride, and gold

Table 20-2. Classification of Colloidal Dispersions According to State of Matter

Disperse Phase	Dispersion Medium (Vehicle)		
	Solid	Liquid	Gas
Solid	Zinc Oxide Paste USP, toothpaste (dicalcium phosphate or calcium carbonate with aqueous sodium carboxymethylcellulose binder), and pigmented plastics (titanium dioxide in polyethylene)	Sols: Bentonite Magma NF, Trisulfapyrimidines Oral Suspension USP, Alumina and Magnesia Oral Suspension USP, Tetracycline Oral Suspension USP, Betamethasone Valerate Lotion USP, and Ophthalmic Suspension USP	Solid aerosols: Epinephrine Bitartrate Inhalation Aerosol USP, Isoproterenol Sulfate Inhalation Aerosol USP, smoke, and dust
Liquid	Absorption bases (aqueous medium in Hydrophilic petrolatum USP), emulsion bases (oil in Hydrophilic Ointment USP, Lanolin USP), and butter	Emulsions: Mineral Oil Emulsion USP, Benzyl Benzoate Lotion USP, and soybean oil in water for parenteral nutrition, milk, and mayonnaise	Liquid aerosols: Metaproterenol Sulfate Inhalation Aerosol USP, Povidone Iodine Topical Aerosol USP, mist, and fog
Gas	Solid foams (foamed plastics and rubbers) and pumice	Foams, carbonated beverages, and effervescent salts in water.	No colloidal dispersions

form hydrophobic dispersions without being lipophilic. There is no sharp dividing line between hydrophilic and hydrophobic dispersions. For example, gelatinous hydroxides of polyvalent metals (e.g., aluminum and magnesium hydroxide) and clays (e.g., bentonite and kaolin) possess some characteristics of both.² Common *lipophobic* dispersions include water-in-oil (W/O) emulsions, which are essentially lyophobic dispersions in lipophilic vehicles.

Association Colloids

Organic compounds that contain large hydrophobic moieties on the same molecule with strongly hydrophilic groups are said to be *amphiphilic*. The individual molecules are generally too small to be in the colloidal size range, but they tend to associate into larger aggregates when dissolved in water or oil. These compounds are designated *association colloids*, because their aggregates are large enough to qualify as colloidal particles. Examples include surfactant molecules that associate into micelles above their critical micelle concentration (CMC) and phospholipids that associate into cellular membranes and liposomes, which have been used for drug delivery.

PROPERTIES OF COLLOIDAL DISPERSIONS

PARTICLE SHAPE

Particle shape depends upon the chemical and physical nature of the dispersed phase and the method employed to prepare the dispersion (preparation methods are described in later sections of this chapter). Primary particles exist in a wide variety of shapes, and their aggregation produces an even wider variety of shapes and structures. Preparation methods such as mechanical comminution and precipitation generally produce randomly shaped particles unless the precipitating solids possess pronounced crystallization habits or the solids being ground possess strongly developed cleavage planes. For example, micronized particles of sulfonamides and other organic powders, as well as precipitated aluminum hydroxide gels, typically have irregular random shapes. An exception is bismuth subnitrate; hydrolyzing bismuth nitrate solutions with sodium carbonate precipitates lath-shaped particles. In addition, precipitated silver chloride particles show their cubic nature under the electron microscope. Lamellar or plate-like solids often preserve their lamellar shape during mechanical comminution, because milling and micronization break up the stacks of thin plates, in addition to fragmenting plates in the lateral dimensions. In these materials, the molecular cohesion between layers is much weaker than the cohesion within layers. Examples of such materials include graphite, mica, and kaolin (Figure 20-1). In a like manner, macroscopic asbestos and cellulose fibers consist of bundles of microscopic and submicroscopic fibrils that have very small diameters. Mechanical comminution splits these bundles into their component fibrils as well as cutting them shorter. Figure 20-2 shows the individual, needle- or rod-shaped cellulose crystallites formed after breaking up the aggregated bundles of *microcrystalline cellulose*. These crystallites average 0.3 micrometers in length and 0.02 micrometers in width, which places them in the colloidal size range. Microcrystalline cellulose is a fibrous thickening agent and tablet additive made by the controlled hydrolysis of cellulose.

Except in the special cases of clay and cellulose just mentioned, regular-shaped particles are typically produced by condensation rather than disintegration methods. For example, *colloidal silicon dioxide* is a white powder consisting of submicroscopic spherical particles of rather uniform size (i.e., narrow particle size distribution). It is manufactured by high-temperature, vapor-phase hydrolysis of silicon tetrachloride in an oxyhydrogen flame (i.e., a flame produced by burning hydrogen in a stream of oxygen). It is commonly referred to as fumed or pyrogenic silica because of this manufacturing process. Different grades are produced by different reaction conditions.

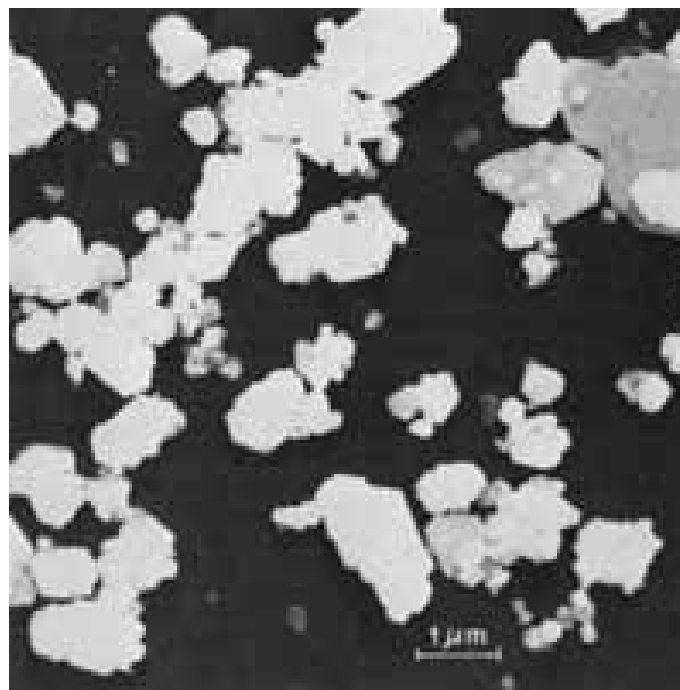


Figure 20-1. Transmission electron micrograph of a well-crystallized, fine-particle kaolin. Note hexagonal shape of the clay platelets. (Courtesy of John L. Brown, Engineering Experiment Station, Georgia Institute of Technology.)

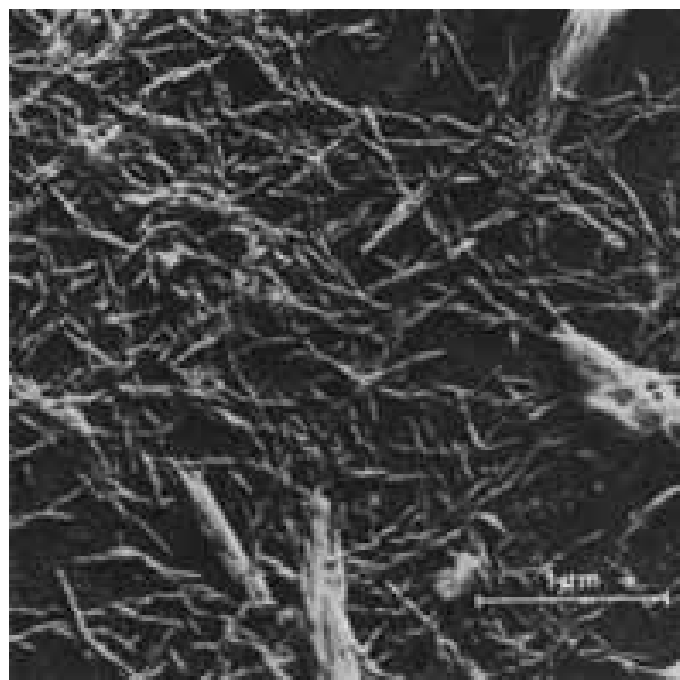


Figure 20-2. Transmission electron micrograph of Avicel RC-591 thickening-grade microcrystalline cellulose. The needles are individual cellulose crystallites; some are aggregated into bundles. (Courtesy of FMC Corporation; Avicel is a registered trademark of FMC Corporation.)

Figure 20-3 shows the relatively large, single spherical particles of colloidal silicon dioxide. Their average diameter of 50 nm corresponds to the comparatively small specific surface area of 50 m²/g. Smaller spherical particles have a larger specific surface area. For example, the grade with the smallest average diameter, 5 nm, has a specific surface area of 380 m²/g. The finer

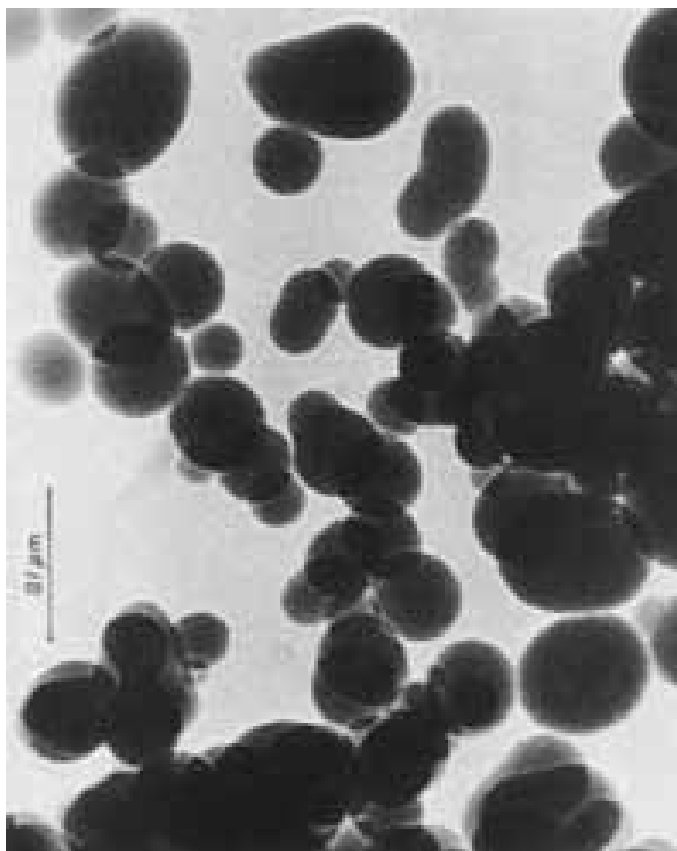


Figure 20-3. Transmission electron micrograph of Aerosil OX 50, ground and dusted on. The spheres are translucent to the electron beam, causing overlapping portions to be darker as a result of an increased thickness. The suffix 50 indicates the specific surface area in m²/g. (Courtesy of Degussa AG, Hanau, Germany; Aerosil is a registered trademark of Degussa.)

grade particles tend to sinter or grow together into chain-like aggregates resembling pearl necklaces during the manufacturing process (Figure 20-4). *Latexes* of polymers, such as latex-based paints, are aqueous dispersions prepared by emulsion polymerization. Their particles are spherical, because polymerization of the solubilized liquid monomers takes place inside spherical surfactant micelles. Some *clays* grow as plate-like particles possessing straight edges and hexagonal angles (e.g., bentonite and kaolin) (Figure 20-1). Other clays have lath-shaped (nontronite) or rod-shaped particles (attapulgite).³

Emulsification produces spherical droplets to minimize the oil-water interfacial area. Cooling an emulsion below the melting point of the dispersed phase freezes the dispersed particles in a spherical shape. For instance, paraffin may be emulsified in 80°C water and then cooled to room temperature to produce a hydrosol containing spherical particles. Sols of viruses and globular proteins, which are hydrophilic, contain compact particles possessing definite geometric shapes. For example, the poliomyelitis virus is spherical, the tobacco mosaic virus is rod-shaped, and the serum albumins and globulins are prolate ellipsoids of revolution (football-shaped).

DIFFUSION AND SEDIMENTATION

The molecules of a gas or liquid are engaged in a perpetual and random thermal motion causing collisions with one another and with the container wall billions of times per second. Each collision changes the direction and the velocity of these molecules. The continuous motion of molecules of a dispersion medium randomly buffets any dissolved molecules and suspended colloidal particles. The random bombardment imparts an erratic movement called *brownian motion* to solutes and particles.

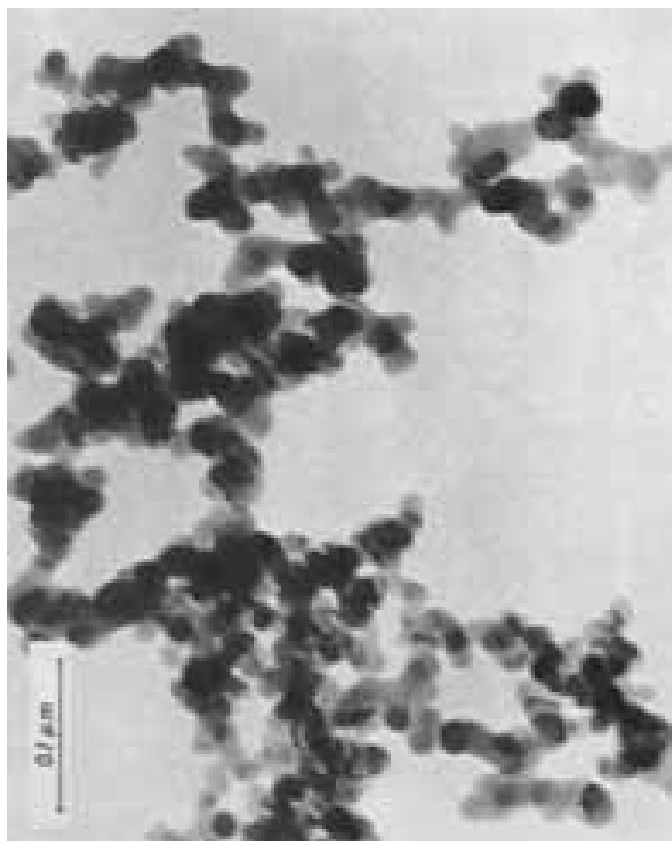


Figure 20-4. Transmission electron micrograph of Aerosil 130, ground and dusted on. The spheres are fused together into chain-like aggregates. The suffix 130 indicates the specific surface area in m²/g. (Courtesy of Degussa AG, Hanau, Germany; Aerosil is a registered trademark of Degussa.)

This phenomenon is named after the botanist Robert Brown, who first observed it under the microscope in an aqueous pollen suspension. The brownian motion of colloidal particles magnifies the random movements of molecules in the liquid or gaseous suspending medium and represents a three-dimensional random walk.

Suspended colloidal particles and solute molecules undergo both rotational and translational brownian movements. For translational motion, Einstein derived the equation:

$$\bar{x} = \sqrt{2Dt} \quad (3)$$

where x is the mean displacement in the x -direction in time t , and D is the *diffusion coefficient*. Einstein also showed that for spherical particles of radius r , under conditions valid for Stokes' law and Einstein's law of viscosity:

$$D = \frac{RT}{6\pi\eta rN} \quad (4)$$

where R is the gas constant, T is the absolute temperature, N is Avogadro's number, and η is the viscosity of the suspending medium.

A common measure of the mobility of a dissolved molecule or suspended particle in a liquid medium is the diffusion coefficient. At room temperature, using units of cm²/s, the value of sucrose in water is 4.7×10^{-6} and the value of serum albumin in water is 6.1×10^{-7} . Diffusion is a slow process on a macroscopic scale. Using the value of 1×10^{-7} cm²/s, brownian motion causes a particle to move in one direction an average distance of 1 cm in 58 days, 1 mm in 14 hours, or 1 micrometer in 0.05 seconds. As seen in the above equation, smaller molecules diffuse faster

in a given medium. The radius of a sucrose molecule is smaller than that of a serum albumin molecule; the calculated values are 0.44 nm and 3.5 nm, respectively (assuming a spherical shape). Steroids have only slightly higher molecular weights than sucrose; however, their diffusion coefficients in petrolatum-based absorption bases are in the range of 10^{-10} to 10^{-8} cm²/s. These much smaller diffusion coefficients are caused by the much higher vehicle viscosity. Passive diffusion (driven by a concentration gradient and carried out through brownian motion) is important in the release of drugs from topical preparations and in the gastrointestinal absorption of drugs.

Brownian motion and convection currents maintain dissolved molecules and small colloidal particles in suspension indefinitely. This is true for all intrinsically stable systems when dissolution or dispersion occurs spontaneously and the corresponding free energy change is negative (see equation 5 below). In meta-stable or *diuturna* (i.e., durable, lasting) dispersions, brownian motion prevents sedimentation and may extend their life for years.

As particle size or r increases, brownian motion decreases as seen by the x proportionality to $r^{-1/2}$. Larger particles have a greater tendency than smaller particles of the same material to settle to the bottom of the dispersion, provided the densities of the dispersed phase, dP , and the liquid vehicle, dL , are sufficiently different (sedimentation, when $dP > dL$). On the other hand, larger particles will rise to the top of the dispersion when $dP < dL$. This is known as creaming. The Stokes equation reflects the rate of *sedimentation/creaming*; it is expressed as:

$$h = \frac{2(d_p - d_L)r^2gt}{9\eta} \quad (5)$$

where h is the height (or distance) that a spherical particle moves in time t , and g is the acceleration of gravity. The equation illustrates that this rate is proportional to r^2 . Consequently, as brownian motion diminishes with increasing particle size, the tendency of particles to sediment or cream is increased. At a critical radius, the distance, h , that a particle settles/creams equals the mean displacement, x , due to brownian motion over the same time interval, t , and therefore, the two are equal.⁴ Intravenous vegetable oil emulsions have little tendency to cream, because their mean droplet size, approximately 0.2 micrometers, is smaller than the critical radius. In most pharmaceutical suspensions, sedimentation prevails.

LIGHT SCATTERING

The optical properties of a medium are determined by its refractive index. Light will pass through the medium undeflected when the refractive index is uniform throughout. However, when there are discrete variations in the refractive index from the presence of particles or caused by small-scale density fluctuations, part of the passing light will be scattered in all directions. When a narrow beam of sunlight is admitted through a small hole into a darkened room, bright flashing points reveal the presence of the minute dust particles suspended in the air. A beam of light striking a particle polarizes the atoms and molecules of that particle and induces dipoles, which act as secondary sources and re-emit weak light of the same wavelength as the incident light. This phenomenon is called *light scattering*. The scattered radiation propagates in all directions away from the particle. In a bright room, the light scattered by the dust particles is too weak to be noticeable.

Colloidal particles suspended in a liquid also scatter light. When an intense, narrowly defined beam of light is passed through a suspension, its path becomes visible because of the light scattered by the particles in the beam. This *Tyndall beam* is characteristic of colloidal systems and becomes most visible when viewed against a dark background in a direction perpendicular to the incident beam. The magnitude of the turbidity or opalescence depends upon the nature, size, and concentration of the dispersed particles. For example, when clear mineral oil

is dispersed in an equal volume of a clear, aqueous surfactant solution, the resultant emulsion is milky white and opaque as a result of light scattering. However, microemulsions containing emulsified droplets that are only about 40 nm in diameter (i.e., much smaller than the wavelength of visible light) are transparent and clear to the naked eye.

The concentration of inorganic and organic colloidal dispersions and of bacterial suspensions can be measured by their Tyndall effect or turbidity. Turbidity, τ , is defined by an equation analogous to Beer's law for the absorption of light,^{1,2} namely:

$$\tau = \frac{1}{l} \ln \frac{I_0}{I_t} \quad (6)$$

where I_0 and I_t are the intensities of the incident and transmitted light beams, and l is the length of the dispersion through which the light passes. The concentration of dispersed particles can be measured in two ways using turbidity. In *turbidimetry* a spectrophotometer or photoelectric colorimeter is used to measure the intensity of the light transmitted in the incident direction. If the dispersion is less turbid, the intensity of light scattered at 90 degrees to the incident beam is measured with a *nephelometer*. Both methods require careful standardization, using suspensions that contain known amounts of particles similar to those being studied. The turbidity of hydrophilic colloidal systems such as aqueous solutions of gums, proteins, and other polymers is far weaker than that of lyophobic dispersions. These solutions appear clear to the naked eye; however, their turbidity can be measured with a photoelectric cell/photomultiplier tube and used to determine the molecular weight of the solute.

The theory of light scattering was developed in detail by Lord Rayleigh. For white, nonabsorbing nonconductors or dielectrics like sulfur and insoluble organic compounds, the equation obtained for spherical particles whose radius is small compared to the wavelength of light (λ) is^{1,2}:

$$I_s = I_0 \frac{4\pi^2 n_0^2 (n_1 - n_0)^2}{\lambda^4 d^2 c} (1 + \cos^2 \theta) \quad (7)$$

I_0 is the intensity of the unpolarized incident light; I_s is the intensity of light scattered in a direction making an angle, θ , with the incident beam and measured at a distance, d . The scattered light is largely polarized. The concentration, c , is expressed as the number of particles per unit volume. The refractive indices, n_1 and n_0 , refer to the dispersion and the dispersion medium, respectively. Because the intensity of scattered light is inversely proportional to the fourth power of the wavelength, blue light ($\lambda \approx 450$ nm) is scattered much more strongly than red light ($\lambda \approx 650$ nm). Colloidal dispersions of colorless particles appear blue when the incident white light is viewed in scattered light (i.e., in lateral directions such as 90 degrees to the incident beam). Loss of the blue rays due to preferential scattering leaves the transmitted light yellow or red. Preferential scattering of blue radiation sideways accounts for the blue color of the sky, sea, cigarette smoke, and diluted milk, and for the yellow-red color of the rising and setting sun viewed head-on.

The particles in pharmaceutical suspensions, emulsions, and lotions are generally larger than the wavelength of light, λ . When the particle size exceeds $\lambda/20$, destructive interference between the light scattered by different portions of the same particle lowers the intensity of the scattered light and changes its angular dependence. Rayleigh's theory was extended to large and strongly absorbing and conducting particles by German physicist Gustav Mie and to nonspherical particles by German physicist Richard Gans.^{1,2} It is possible to determine the average particle size and even the particle size distribution of colloidal dispersions and coarser suspensions by means of turbidity measurements using appropriate precautions in experimental techniques and interpretations.

Dynamic Light Scattering

Light scattered by a moving particle undergoes a Doppler shift; its frequency increases slightly when the particle moves toward the photodetector and decreases slightly when it moves away. This shift is so small that it can only be detected by very intense, strictly monochromatic laser light. Because they are engaged in random brownian motion, a set of colloidal particles scatters light with a broadened frequency. Smaller particles diffuse faster than larger ones and therefore produce greater Doppler broadening. If the particles are spherical and monodisperse, and their concentration is so dilute that they neither attract nor repel one another, the frequency broadening can be used to estimate the particle diffusion coefficient, D . As noted above, the diffusion coefficient is inversely proportional to the particle radius. The measured radius is actually the hydrodynamic radius (r_H), which comprises the particle plus its attached water of hydration. The technique is called *dynamic or quasi-elastic light scattering*. The technique is also called *photon correlation spectroscopy* (PCS), because it counts and correlates the number of scattered photons over very short time intervals. For polydisperse spherical colloidal particles, it estimates the particle size distribution.^{2,5} Particles that are asymmetric rather than spherical and/or extensively hydrated have a larger r_H and hence smaller D value than unsolvated spherical particles with the same dry volume. It is not possible to separate the effect of hydration upon r_H and D from the effect of asymmetry by PCS alone; either hydration or particle shape must be determined by other means.^{1,2,5}

In a related technique that uses a *fiber-optic Doppler anemometer*, a laser beam is carried into the interior of a colloidal dispersion via a fiber-optic cable. Particles in the small volume of dispersion around the immersed tip scatter light with the Doppler frequency shift back into the same fiber to the detector. This method is suitable for concentrated dispersions that are opaque to the laser beam and would have to be diluted extensively for conventional dynamic light scattering measurements.⁶

VISCOSITY

Most lyophobic dispersions have viscosities only slightly greater than that of the liquid vehicle. This holds true even at comparatively high volume fractions of the disperse phase unless the particles form continuous network aggregates throughout the vehicle, in which case yield stresses are observed. By contrast, the apparent viscosities of lyophilic dispersions, especially of polymer solutions, are several orders of magnitude greater than the viscosity of the solvent or vehicle even at concentrations of only a few percent solids. Lyophilic dispersions are also generally much more pseudoplastic or shear-thinning than lyophobic dispersions.

GEL FORMATION

The flexible chains of dissolved polymers interpenetrate and entangle because of the constant brownian motion of their segments. The chains constantly writhe and change their conformations. Each chain is encased in a sheath of solvent molecules that solvate its functional groups. For example, water molecules are hydrogen-bonded to the hydroxyl groups of polyvinyl alcohol, hydroxyl groups, and ether linkages of polysaccharides, ether linkages of polyethylene oxide or PEG, amide groups of polypeptides and povidone, and carboxylate groups of anionic polyelectrolytes. This envelope of water of hydration prevents chain segments that are in close proximity from touching and attracting one another through interchain hydrogen bonds and van der Waals forces as they do in the solid state. The free solvent between the chains' solvation sheaths acts as a lubricant allowing the solvated chains to slip past one another when the solution flows. However, any factor that lowers the hydration of dissolved macromolecules will reduce or thin out the sheath of hydration separating adjacent chains. When hydration is low,

contiguous chains tend to attract one another through secondary valence forces such as hydrogen bonds and van der Waals forces, which establish weak and reversible crosslinks between the chains at their points of contact or entanglement, thus bringing about phase separation or precipitation. Hydrophobic bonding is an important contribution to the interchain attraction between polypeptide chains even in solution.

Most water-soluble polymers have a higher solubility in hot water than in cold water and tend to precipitate upon cooling because the sheaths of hydration surrounding adjacent chains become too sparse to prevent interchain attraction. Cooling dilute solutions tends to separate them into a solvent phase and a viscous liquid phase that contains practically the entire amount of polymer but still a large excess of solvent. This process is called *simple coacervation*, and the polymer-rich liquid phase is referred to as a *coacervate*.⁷ If the polymer solution is concentrated enough and/or the temperature is low enough, cooling causes the formation of a continuous network of precipitating chains attached to one another through weak crosslinks that consist of interchain hydrogen bonds and van der Waals forces at the points of mutual contact. Segments of regularly sequenced polymer chains will associate laterally into crystalline bundles or crystallites. However, irregular chain structures, such as those found in random copolymers, randomly substituted cellulose ethers and esters, and highly branched polymers like acacia, will prevent crystallization during precipitation from solution. In these cases chain entanglements provide the sole temporary crosslinks. A network of associated polymer chains immobilizes the solvent, which may result in the separation of gelatinous precipitates or highly swollen flocs, and in the case of more concentrated polymer solutions, may even cause the solution to set to a gel.

The most important factors causing phase separation, precipitation, and gelation of polymer solutions are the chemical nature of the polymer and the solvent, temperature, polymer concentration, and polymer molecular weight. Lower temperatures, higher concentrations, and higher molecular weights promote *gelation* and produce stronger gels. For a typical *gelatin*, 10 percent solutions acquire yield values and begin to gel at about 25°C, 20 percent solutions gel at about 30°C, and 30 percent solutions gel at about 32°C. The gelation is reversible, and the gels liquefy when heated above these temperatures. Regardless of concentration, gelation is rarely observed above 34°C, and therefore, gelatin solutions do not gel at body temperature (i.e., 37°C). Agar and pectic acid solutions set to gels at only a few percent of solids. Unlike most water-soluble polymers, methylcellulose, hydroxypropyl cellulose, and polyethylene oxide are more soluble in cold water than in hot water. Therefore, their solutions tend to gel upon heating (i.e., *thermal gelation*) instead of cooling.

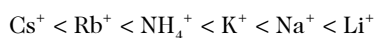
When dissolving powdered polymers in water, temporary gel formation often slows dissolution considerably. As water diffuses into loose clumps of the powder, their exterior frequently turns to a cohesive gel of solvated particles encasing the remaining dry powder. Such globs of gel dissolve very slowly because of their high viscosity and the low diffusion coefficients of the polymers. For large-scale dissolution, it is helpful to disperse the polymer powder in water at temperatures where the solubility of the polymer is lowest before it can agglomerate into lumps of gel. Most polymer powders, such as sodium carboxymethylcellulose, are dispersed with high shear in *cold* water before the particles can hydrate and swell into sticky gel grains that agglomerate into lumps. Once the powder is well dispersed, the mixture is heated with moderate shear to about 60°C for the quickest dissolution. Because methylcellulose hydrates more slowly in hot water, the powder is dispersed with high shear in one-fifth to one-third of the required amount of water heated to 80°C to 90°C. Once the powder is finely dispersed, the remaining amount of water is added cold, or even as ice, and moderate stirring causes prompt dissolution. For maximum clarity, fullest hydration, and highest viscosity, the solution should be cooled to 0°C to 10°C

for about an hour. Alternatively, the polymer powder may be prewetted with a water-miscible organic solvent (e.g., ethyl alcohol or propylene glycol) that does not swell the polymer. The solvent should be added in a proportion of three to five parts of solvent to one part of polymer. If other nonpolymeric, powdered adjuvants are to be incorporated into the solution, they are dry-blended with the polymer powder and should comprise only one-quarter or less of the blend for the best results.

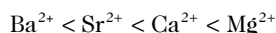
Large increases in the concentration of polymer solutions may lead to precipitation and gelation. One way of effectively increasing the concentration of aqueous polymer solutions is to add inorganic salts. The salts will bind part of the water in the solution so as to become hydrated. Competition for water of hydration dehydrates the polymer molecules and precipitates them. This phenomenon is called **salting out** and may cause the polymer to separate as a concentrated, viscous liquid solution, a simple coacervate, or a solid gel. Because of its high solubility in water, ammonium sulfate is often used to precipitate and separate proteins from dilute solutions. However, salting out is reversible and subsequent addition of water redissolves the precipitated polymers and liquefies their gels.

Hofmeister or Lyotropic Series

The effectiveness of electrolytes to cause salting out depends upon their extent of hydration. The *Hofmeister or lyotropic series* arranges ions in order of increasing hydration and increasing effectiveness in salting out hydrophilic colloids. The series for monovalent and divalent cations are:

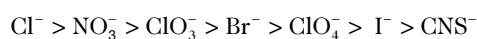
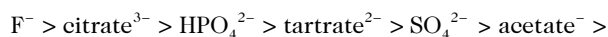


and



The Hofmeister series governs many colloidal phenomena, including the effect of salts upon the temperature of gelation, the swelling of aqueous gels, and the viscosity of hydrosols, and the permeability of membranes toward salts. The series is observed in many phenomena involving small atoms or ions and true solutions, including the ionization potential and electronegativity of metals; the heats of hydration of cations; the size of hydrated cations; the viscosity, surface tension, and infrared spectra of salt solutions; and the solubility of gases in salt solutions. This series also arranges cations in order of increasing ease in displacement from cation-exchange resins based on the smaller hydrated species size (e.g., K^+ displaces Na^+ and Li^+). Adsorption in the Stern layer of particles (see below) also illustrates the series. The lithium ion is more extensively hydrated, and therefore, Li^+ (aq), including the hydration shell, is larger than Cs^+ (aq). Because of its smaller size, the hydrated cesium ion can approach a negative particle's surface more closely than the hydrated lithium ion. Moreover, because of its greater electron cloud, the Cs^+ ion is more polarizable than the Li^+ ion. Therefore, the Cs^+ ion is more strongly adsorbed in the Stern layer.

For anions, in order of decreasing effectiveness in salting out, the lyotropic series is



Iodides and thiocyanates, and to a lesser extent bromides and nitrates, actually tend to increase the solubility of polymers in water (i.e., salt them in).^{1,2} These large polarizable anions reduce the extent of hydrogen bonding among water molecules and thereby make more of the hydrogen-bonding capacity of water available to the solute. Most salts, except for nitrates, bromides, perchlorates, iodides, and thiocyanates, raise the temperature of precipitation or gelation of most hydrophilic

colloidal solutions. Exceptions among hydrophilic colloids are methylcellulose, hydroxypropyl cellulose, and polyethylene oxide, whose gelation temperatures or gel melting points are lowered by salting out.

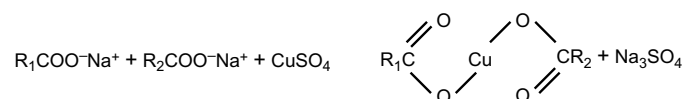
Most hydrophilic sols require electrolyte concentrations of 1 M or higher to induce precipitation or gelation. In addition, hydrophilic colloids disperse or dissolve spontaneously in water, and their sols are intrinsically stable. Therefore, the polymer can be redissolved by removing the coagulating salt through dialysis or by adding more water. Whenever hydrophilic colloidal dispersions undergo irreversible precipitation or gelation, chemical reactions are involved. Neither dilution with water, heating, nor attempts to remove the gelling or precipitating agent by washing or dialysis will liquefy these gels.

Most of the hydrophilic and water-soluble polymers mentioned previously are only slightly soluble or insoluble in alcohol. Addition of alcohol to their aqueous solutions may cause precipitation or gelation, because it lowers the dielectric constant of the medium and tends to dehydrate the hydrophilic solute. Alcohol also lowers the concentrations at which electrolytes salt out hydrophilic colloids. Therefore, alcohol is often referred to as a nonsolvent or precipitant. However, the addition of alcohol to an aqueous polymer solution may cause coacervation (i.e., the separation of a concentrated viscous liquid phase) rather than precipitation or gel formation. Sucrose also competes for water of hydration with hydrophilic colloids and may cause phase separation. However, most hydrophilic sols tolerate substantially higher concentrations of sucrose than of electrolytes or alcohol. Lower viscosity grades of a given polymer are usually more resistant to the effects of electrolytes, alcohol, and sucrose than grades having higher viscosities and molecular weights.

The gelation temperature or gel point of gelatin is highest at its isoelectric point, where the attachment of adjacent chains through ionic bonds between carboxylate ions and alkylammonium, guanidinium, or imidazolium groups is most extensive. Because carboxyl groups are not ionized in strongly acidic media such as gastric juices, interchain ionic bonds are practically nonexistent in this environment and interchain attraction is limited to hydrogen bonds and van der Waals forces. Therefore, the combination of an acidic pH that is considerably below the isoelectric point and a temperature of 37°C completely prevents the gelation of gelatin solutions. Conversely, if a polymer owes its solubility to the ionization of these weakly acid groups, reducing the pH of its solution below 3 may lead to precipitation or gelation. This is observed with carboxylated polymers such as many gums, sodium carboxymethylcellulose, and carbomer. Adjusting the media to higher pH values returns the carboxyl groups to their ionized state and reverses the gelation or precipitation. However, gelation temperatures typically depend more upon temperature and concentration than pH.^{8,9}

Hydrogen carboxymethylcellulose swells and disperses but does not dissolve in water. Only the sodium, potassium, ammonium, and triethanolammonium salts of carboxylated polymers are well soluble in water. In the case of carboxymethylcellulose, salts with heavy-metal cations (e.g., silver, copper, mercury, lead) and trivalent cations (e.g., aluminum, chromic, ferric) are practically insoluble. Salts with divalent cations, especially of the alkaline earth metals, have borderline solubilities. Generally, higher degrees of substitution tend to increase the tolerance of carboxymethylcellulose toward salts.

When inorganic salts of heavy or trivalent cations are mixed with alkali metal salts of carboxylated polymers in solution, precipitation or gelation occurs due to metathesis. For instance, if a soluble copper salt is added to a solution of sodium carboxymethylcellulose, the double decomposition can be schematically written as:



R_1 and R_2 represent two carboxymethylcellulose chains, which are crosslinked by a chelated copper ion. Dissociation of the cupric carboxylate complex is negligible.

ELECTRIC PROPERTIES

Origin of Electric Charges

Particles can acquire charges from several sources. In *proteins*, one end group of the polypeptide chain and any aspartic and glutamic acid units contribute carboxylic acid groups, which are ionized into negatively charged carboxylate ions in neutral to alkaline media. The other chain end group and any lysine units contribute amino groups, while arginine units contribute guanidine groups, and histidine units contribute imidazole groups. The nitrogen atoms of these groups become protonated in neutral to acid media. These covalently attached anions and cations confer a negative and positive charge to the molecule, respectively. Therefore, proteins may be referred to as *polyelectrolytes* (polymeric electrolytes or salts). However, they are not the only organic polymers that contain ionic groups, and thus, many substances can be considered to be polyelectrolytes. For example, natural polysaccharides of vegetable origin such as acacia, tragacanth, alginic acid, and pectin contain carboxylic acid groups, which are ionized in neutral to alkaline media. Agar and carrageenan, as well as the animal polysaccharides heparin and chondroitin sulfate, contain sulfate groups, which are strongly acidic and ionize even in acid media. Cellulosic polyelectrolytes include *sodium carboxymethylcellulose*, while synthetic carboxylated polymers include *carbomer*, a copolymer of acrylic acid.

Counterions are required for electroneutrality of the ionizing groups on polyelectrolytes. Counterions dissociate from ionogenic functional groups and can be replaced by other ions of like charge. For example, in neutral and alkaline media, Na^+ , K^+ , Ca^{2+} , and Mg^{2+} are among the counterions neutralizing the negative charges of the carboxylate groups, and if hydrochloric acid was used to make the medium acidic and to supply the protons, Cl^- is present to neutralize any of the cationic groups mentioned previously. These counterions are not an integral part of the protein particle but are located in its immediate vicinity. Alternatively, at a specific intermediate pH value (4.5–7 for most proteins), the carboxylate anions and the alkylammonium, guanidinium, and imidazolium cations on the same molecule neutralize each other exactly. There is no need for counterions, because the ionized functional groups are in exact balance. At this pH value, called the *isoelectric point*, the protein particle or molecule is neutral; its electrical charge is neither negative nor positive, but zero.²

Most inorganic particulate compounds are also charged. *Aluminum hydroxide*, $\text{Al}(\text{OH})_3$, can be dissolved by acids and alkalis to form aluminum ions, Al^{3+} , and aluminate ions, $\text{Al}(\text{OH})_4^-$, respectively. In neutral or weakly acid media (i.e., acid concentrations too low to cause dissolution), an aluminum hydroxide particle has some positive charges attributed to Al^{3+} valences that have not been completely neutralized. Figure 20-5A represents a portion of the surface of an aluminum hydroxide particle that has one such positive charge neutralized by a Cl^- counterion.

In weakly alkaline media (i.e., base concentrations too low to transform the aluminum hydroxide particles completely into aluminate and dissolve them), the aluminum hydroxide particles bear some negative charges due to the presence of a few aluminate groups. Figure 20-5B represents a portion of the surface of an aluminum hydroxide particle that has one such negative group neutralized by a Na^+ counterion. At a pH of 8.5 to 9.1 there are neither $\text{Al}(\text{OH})_2^+$ nor $\text{Al}(\text{OH})_4^-$ ions on the particle surface but only neutral $\text{Al}(\text{OH})_3$ molecules.^{10,11} Therefore, the particles have no charge and do not need counterions for charge neutralization. This pH is considered to be the isoelectric point. In the case of inorganic particulate compounds such as aluminum hydroxide, it also is called the *zero point of charge*.

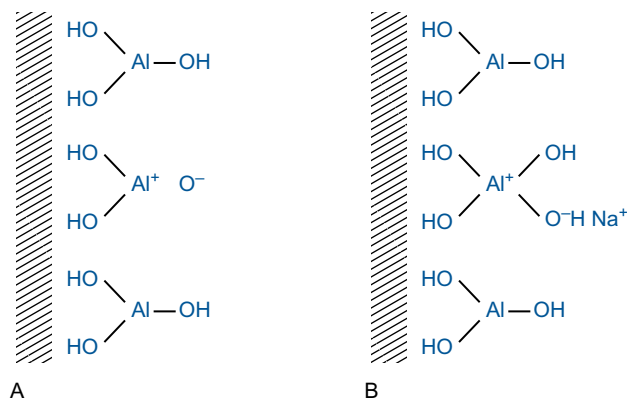
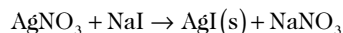


Figure 20-5. Schematic representation of a portion of the surface of an aluminum in a neutral or weakly acid media (A) or a weakly alkaline media (B).

Bentonite clay is a lamellar aluminum silicate. Each lattice layer consists of a sheet of hydrated alumina sandwiched between two silica sheets. Isomorphous replacement of Al^{3+} by Mg^{2+} or of Si^{4+} by Al^{3+} confers net negative charges to the thin clay lamellas in the form of cation-exchange sites resembling silicate ions built into the lattice. The counterions producing electroneutrality are usually Na^+ (in sodium bentonite) or Ca^{2+} (in calcium bentonite).

Silver iodide sols can be prepared by the reaction:



In the bulk of the silver iodide particles, there is a 1:1 stoichiometric ratio of Ag^+ ions to I^- ions. If the above reaction is carried out with an excess of silver nitrate, there will be more Ag^+ ions than I^- ions in the surface layer of the particles. The particles will then be positively charged, and the counterions surrounding them will be NO_3^- . If the reaction is carried out using an exact stoichiometric 1:1 ratio of silver nitrate to sodium iodide or with an excess of sodium iodide, the surface of the particles will contain more I^- ions than Ag^+ ions.^{1,2} The particles will then be negatively charged, and the counterions surrounding them will be Na^+ .

An additional mechanism through which particles acquire electric charges is by the adsorption of ions,^{1,2,12} including ionic surfactants. This is discussed in more detail in a later section (Stabilization of Colloidal Systems).

Electric Double Layers

As described previously, the surface layer of a silver iodide particle prepared using an excess of sodium iodide contains more I^- ions than Ag^+ ions; in contrast, the bulk of the particle contains the two ions in an equimolar proportion. The aqueous solution in which such particles are suspended contains relatively high concentrations of Na^+ and NO_3^- , a lower concentration of I^- , and traces of H^+ , OH^- , and Ag^+ . The negatively charged particle surface attracts positive ions from the solution and repels negative ions. Therefore, the solution immediately surrounding the particle surface contains a much higher concentration of Na^+ (counterions) and a much lower concentration of NO_3^- ions than in the bulk solution. A number of Na^+ ions equal to the number of excess I^- ions in the surface (i.e., the number of I^- ions in the surface layer minus the number of Ag^+ ions in the surface layer) and equivalent to the net negative surface charge of a particle are pulled toward its surface. These counterions tend to approach the particle surface as closely as their hydration spheres permit (Helmholtz double layer); however, thermal agitation of the water molecules tends to disperse them throughout the solution. Consequently, the layer of counterions surrounding the particle is spread out. The Na^+

ion concentration is highest in the immediate vicinity of the negative surface, where the ions form a compact layer called the Stern layer. The Na^+ ion concentration decreases with distance from the surface, throughout a diffuse layer called the Gouy-Chapman layer. Therefore, the sharply defined, negatively charged particle surface is surrounded by a cloud of Na^+ counterions required for electroneutrality. The combination of the two layers of oppositely charged ions constitutes an electric double layer, which is illustrated in the top part of Figure 20-6.

The electric potential of a plane is equal to the work required to bring a unit electric charge from infinity (in this case, from the bulk of the solution) to that plane against electrostatic forces. If the plane is the surface of a particle, the potential is called surface or ψ_0 potential, which measures the total potential of the double layer (Figure 20-5). This is the thermodynamic potential that operates in galvanic cells. Upon moving away from the particle surface toward the bulk solution in the direction of the horizontal axis, the potential drops rapidly across the Stern layer because the Na^+ ions in the immediate vicinity of the particle surface screen Na^+ ions that are farther removed in the diffuse part of the double layer from the effect of the negative surface charge. The decrease in potential across the Gouy-Chapman layer is more gradual. As the composition of the diffuse double layer gradually approaches that of the bulk liquid, where the anion concentration equals the cation concentration, the potential asymptotically approaches zero. In view of this

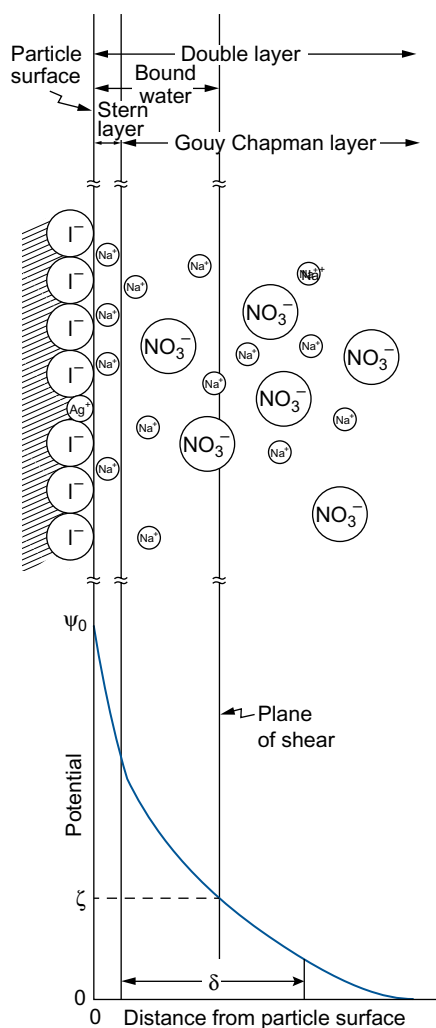


Figure 20-6. Electric double layer at the surface of a silver iodide particle (upper part) and the corresponding potentials (lower part). The distance from the particle surface, plotted on the horizontal axis, refers to both the upper and lower portions of the figure.

indefinite end point, the thickness of the diffuse double layer (δ) is arbitrarily defined as the distance over which it takes the potential at the boundary between the Stern and Gouy-Chapman layers to drop to 0.37 (equal to $1/e$) of its value (Figure 20-6).^{1,2,5,12} The thickness of double layers usually ranges from 1 to 100 nm and decreases as the concentration of electrolytes in solution increases. This occurs more rapidly for higher valence counterions. The value of δ is approximately equal to the reciprocal of the Debye-Hückel theory parameter (κ).

The electrokinetic or zeta potential (ζ -potential) has practical importance, because it can be measured experimentally. In aqueous dispersions, organic particles containing polar functional groups, and even relatively hydrophobic inorganic particles, are surrounded by a layer of water of hydration, which is associated with the particles through ion-dipole and dipole-dipole interactions. When a particle moves, this shell of water and all of the ions located inside it move along with the particle. Conversely, if water or an aqueous solution flows through a fixed bed of these solid particles, the hydration layer surrounding each particle remains attached to it. The electric potential at the plane of shear or slip separating the bound water from the free water is the ζ -potential. It does not include the Stern layer and includes only the part of the Gouy-Chapman layer that lies outside the hydration shell (Figure 20-6).

Stabilization by Electrostatic Repulsion

When two uncharged hydrophobic particles are in close proximity, they attract each other by van der Waals secondary valences, mainly London dispersion forces. For individual atoms and molecules, these forces decrease with the seventh power of the distance between them. In the case of two particles, every atom of one particle attracts every atom of the other particle. Because the attractive forces are nearly additive, they decay much less rapidly with interparticle distance, approximately with the second or third power of the distance between them. Therefore, whenever two particles approach each other closely, the attractive forces take over and cause them to adhere. Coagulation occurs as the primary particles aggregate into increasingly larger secondary particles or flocs. If the dispersion consists of two kinds of particles, one having positive and the other negative charges, the electrostatic attraction between such oppositely charged particles is superimposed on the attraction by van der Waals forces and coagulation is accelerated. If the dispersion contains only one kind of particle with the same surface charge and charge density (the most common case), then electrostatic repulsion tends to prevent the particles from approaching closely enough to come within the effective range of each other's van der Waals attractive forces. This stabilizes the dispersion against interparticle attachments or coagulation. The electrostatic repulsive energy has a range in the order of δ .

A quantitative theory of the interaction between lyophobic disperse particles was worked out independently by Derjaguin and Landau in the USSR and by Verwey and Overbeek in the Netherlands in the early 1940s.^{1,2,5} The so-called DLVO theory (after the surnames of the four authors just mentioned) predicts and explains many, but not all, experimental data; its refinement to account for discrepancies is ongoing. The DLVO theory is summarized in Figure 20-7, where curve WA represents the van der Waals attractive energy, which decreases approximately by the second power of the interparticle distance, and curve ER represents the electrostatic repulsive energy, which decreases exponentially with interparticle distance. Because of the combination of these two opposing effects, attraction predominates at small and large distances; whereas repulsion may predominate at intermediate distances. Negative energy values indicate attraction, and positive values indicate repulsion. The resultant curve DPBAS, obtained by algebraic addition of curves WA and ER, gives the total, net energy of interaction between two particles.

Interparticle attraction depends mainly upon the chemical nature and particle size of the dispersed material. Once these

have been selected, the attractive energy between particles is fixed and cannot be readily altered. Electrostatic repulsion depends upon the ψ_0 potential, or the density of the surface charge, and upon the thickness of the double layer, both of which govern the magnitude of the ζ -potential. Thus, dispersion stability correlates to some extent with this potential. The ζ -potential can be widely adjusted using additives, especially ionic surfactants, water-miscible solvents, and electrolytes. If the absolute value of the ζ -potential is small, the resultant potential energy is negative and van der Waals forces of attraction predominate over electrostatic repulsion at all interparticle distances. Such sols coagulate rapidly.

The two identical particles, whose interaction is depicted in Figure 20-7, have a large (positive or negative) ζ -potential, resulting in an appreciable positive or repulsive potential energy at intermediate distances. However, brownian motion, convection currents, sedimentation, or stirring of the dispersion will eventually put them on a collision course. As the two particles approach each other, the two counterion atmospheres begin to interpenetrate or overlap at point A, corresponding to the distance, d_A . This produces a net repulsive (positive) energy because of the work involved in distorting the diffuse double layers and pushing water molecules and counterions aside. If the particles continue to approach each other, the repulsion between their surface charges increases the net potential energy of interaction to its maximum positive value at B, where most of the intervening water and counterions have been displaced. If the height of the potential energy barrier B exceeds the kinetic energy of the approaching particles, they will not come any closer to each other than the distance d_B and will then move away from each other. A net positive potential energy of about $25 kT$ units usually suffices to keep them apart and renders the dispersion stable (k is the Boltzmann constant and T is absolute temperature). At $T = 298$ K, the required potential energy for stabilization corresponds to 1×10^{-12} erg or 1×10^{-5} J. The kinetic energy of a particle is of the order of kT .

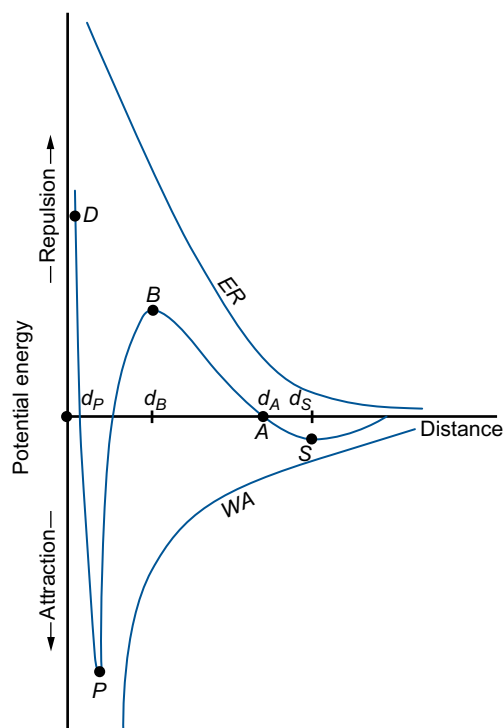


Figure 20-7. Curves representing the van der Waals energy of attraction (WA), the energy of electrostatic repulsion (ER), and the net energy of interaction (DPBAS) between two identical charged particles, as a function of the interparticle distance.

On the other hand, if the kinetic energy of the approaching particles exceeds the potential energy barrier B , the particles will continue to approach each other past d_B , where the van der Waals forces of attraction become increasingly more important compared to the electrostatic repulsion. Therefore, the net potential energy of particle interaction decreases to zero and then becomes negative. This now pulls the particles closer together. When the particles touch, at a distance, d_P , the net energy has acquired a large negative value of P . This deep minimum in potential energy corresponds to a very stable situation in which the particles adhere. Because it is unlikely that enough kinetic energy can be supplied to the particles or that their ζ -potential can be increased sufficiently to cause them to climb out of the potential energy well P , they are permanently attached to each other. When most or all of the primary particles agglomerate into secondary particles by this process, the sol coagulates. Any closer approach of the two particles than the touching distance d_P will cause a very rapid rise in potential energy along PD because the solid particles would interpenetrate each other and cause atomic orbitals to overlap (Born repulsion).

Coagulation of Hydrophobic Dispersions

The height of the potential energy barrier and the range over which the electrostatic repulsion is effective (or the thickness of the double layer) determine the stability of hydrophobic dispersions. Both factors are reduced by the addition of electrolytes. The transition between a coagulating and a stable sol is gradual and depends upon the time of observation. Therefore, standardized conditions must be used to classify a sol as either coagulated or coagulating, or stable (i.e., fully dispersed).

To determine the coagulating concentration of a given electrolyte for a given sol, a series of test tubes is filled with equal portions of the sol. Identical volumes of electrolyte solutions having increasing concentrations are added to the test tubes with vigorous stirring. After a rest period of a certain duration (e.g., 2 hours), the mixtures are agitated again. After an additional, shorter rest period (e.g., 30 minutes) they are inspected for signs of coagulation. The tubes are then classified into two groups – one showing no signs of coagulation and the other showing at least some signs such as visible flocs. Alternatively, they can be classified into one group showing complete coagulation and another showing none or incomplete coagulation, such as some deflocculated colloid left in the supernatant. In either case, the separation between the two classes is quite sharp. The intermediate agitation breaks the weakest interparticle bonds and brings smaller particles into contact with larger ones, thus sharpening the distinction between coagulation and stability. After repeating the experiment with a narrower range of electrolyte concentrations, the coagulation value (ccv) of the electrolyte (i.e., the lowest concentration at which the electrolyte coagulates the sol) is established with good reproducibility.^{1,2}

Typical ccv 's for a silver iodide sol prepared with an excess of iodide are listed in Table 20-3. The following conclusions can be drawn from the left half of Table 20-3:

1. The ccv does not depend upon the valence of the anion. For example, nitrate and sulfate salts of the same metal have nearly identical ccv 's.
2. The differences between the ccv 's of cations having the same valence are relatively minor. However, there is a slight but significant trend of decreasing ccv with increasing atomic number in both the alkali and the alkaline earth metal groups. Arranging these cations in order of decreasing ccv produces the Hofmeister or lyotropic series in decreasing size of hydrated species. For monovalent cations, the lyotropic series is described above. The atomic weight of cesium is 19 times greater than that of lithium, but the Cs^+ ion in aqueous solution is less hydrated and therefore smaller than the hydrated Li^+ ion. It is also more polarizable. Therefore, the Cs^+ ion can approach the surface of a negatively charged particle

suspended in water more closely and is more extensively adsorbed. This reduces its *ccv*, making cesium salts more effective coagulants than lithium salts for negatively charged hydrosols.

- The coagulation values depend primarily upon the valence of the counterions, decreasing by one to two orders of magnitude for each unit increase in the valence of the counterions (Schulze-Hardy rule). According to the DLVO theory, the coagulation values vary inversely with the sixth power of the valence of the counterions. For mono-, di-, and trivalent counterions, they should be in the ratio:

$$\frac{1}{1^6} : \frac{1}{2^6} : \frac{1}{3^6} \text{ or } 100 : 1.6 : 0.14$$

The mean *ccv*'s for the mono-, di-, and trivalent counterions in Table 20-3 are 141, 2.45, and 0.068 mmol/L, respectively. This results in a ratio of 100:1.7:0.05, which is in satisfactory agreement with the DLVO theory.

The following conclusions can be drawn from the right half of Table 20-3:

- The cations on the right side of Table 20-3 constitute obvious exceptions to the preceding conclusions. Ag^+ is a potential-determining counterion (i.e., those whose concentration determines the surface potential). When silver nitrate is added to the negatively charged silver iodide dispersion, some of its silver ions are incorporated into the negatively charged surface of the particles. This lowers the magnitude of the particle's surface charge by reducing the excess of I^- ions in the surface. Therefore, silver salts are exceptionally effective coagulating agents because they reduce the magnitude of both the ψ_0 -potential and the ζ -potential. Indifferent salts, which only reduce the ζ -potential, require much higher salt concentrations for comparable reductions in this potential. The other potential-determining ion of silver iodide is the I^- ion. Alkali iodides have *ccv*'s higher than 141 mmol/L because they supply iodide ions that enter the surface layer of the silver iodide particles and increase its excess of I^- ions over Ag^+ ions, thereby making the ψ_0 -potential more negative. Bromide and chloride ions act similarly but less effectively. The principal potential-determining ion for proteins is H^+ (and hence OH^-); those for aluminum hydroxide are OH^- (and hence H^+) and, Al^{3+} as well as Fe^{3+} and Cr^{3+} , which form mixed hydroxides with Al^{3+} .
- The cationic surfactant and the alkaloidal salts (which also behave as cationic surfactants) on the right side of Table 20-3 constitute the second exception to the Schulze-Hardy rule. Surface-active compounds contain both hydrophilic and hydrophobic moieties within the same molecule. The dual nature of these compounds causes them to accumulate at interfaces. Dodecylammonium and alkaloidal cations are able to displace inorganic monovalent cations from the Stern layer of a negatively charged silver iodide particle. This occurs because they are not only attracted to the particle by electrostatic forces but also by van der Waals forces between their hydrocarbon moieties (i.e., dodecyl chains in the case of the dodecylammonium ions) and the solid particle. Because they are strongly adsorbed from solution onto the particle surface and do not tend to dissociate from it, surface-active cations are very effective in reducing the negative ζ -potential of silver iodide particles. Therefore, they have lower *ccv*'s than purely inorganic cations with the same valence.
- Anionic surfactants, like those containing lauryl sulfate ions, also have a tendency to adsorb at solid-liquid interfaces. However, electrostatic repulsion between the negatively charged surface of the silver iodide particles, whose surface layer contains an excess of iodide ions,

Table 20-3. Coagulation Values for Negative Silver Iodide Sol

Electrolyte	<i>ccv</i> (mmol/L)	Electrolyte	<i>ccv</i> (mmol/L)
LiNO_3	165	AgNO_3	0.01
NaNO_3	140	$\frac{1}{2}(\text{C}_{12}\text{H}_{25}\text{NH}_3)_2\text{SO}_4$	0.07
$\frac{1}{2}(\text{Na}_2\text{SO}_4)$	141	Strychnine nitrate	1.7
KNO_3	136	$\frac{1}{2}(\text{Morphine sulfate})$	2.5
$\frac{1}{2}(\text{K}_2\text{SO}_4)$	138		
RbNO_3	126		
Mean	141		
$\text{Mg}(\text{NO}_3)_2$	2.60	Quinine sulfate	0.7
MgSO_4	2.57		
$\text{Ca}(\text{NO}_3)_2$	2.40		
$\text{Sr}(\text{NO}_3)_2$	2.38		
$\text{Ba}(\text{NO}_3)_2$	2.26		
$\text{Zn}(\text{NO}_3)_2$	2.50		
$\text{Pb}(\text{NO}_3)_2$	2.43		
Mean	2.45		
$\text{Al}(\text{NO}_3)_3$	0.067		
$\text{La}(\text{NO}_3)_3$	0.069		
$\text{Ce}(\text{NO}_3)_3$	0.069		
Mean	0.068		

ccv, Coagulation value.

(Data from Kruyt HR. *Colloid Science*, vols I and II. Houston: Elsevier, 1949 and 1952 and unpublished data.)

and the surface-active anions usually prevents adsorption from occurring below the CMC. If adsorption does occur, it increases the density of the negative charges in the particle surface, and therefore, raises the *ccv* of the anionic surfactant above the value corresponding to its valence.

The addition of water-miscible solvents such as alcohol, glycerin, propylene glycol, or PEGs to aqueous dispersions lowers the dielectric constant of the medium. This reduces the thickness of the double layer and therefore reduces the range over which electrostatic repulsion is effective, and lowers the size of the potential energy barrier. As a result, the addition of sufficient amounts of such solvents tends to coagulate aqueous dispersions. At lower concentrations, these solvents do not induce coagulation themselves but make the dispersions more sensitive to coagulation by added electrolytes (i.e., they lower the *ccv*).

Progressive addition of salts having counterions of high valence gradually reduces the ζ potential of colloidal particles to zero. Eventually, the sign of the ζ -potential may be inverted, and its magnitude may then increase in the opposite direction. The ψ_0 and ζ -potentials of aqueous sulfamerazine suspensions are negative above their isoelectric points, and those of bismuth subnitrate are positive. However, the addition of Al^{3+} to the former and of PO_4^{3-} to the latter in large enough amounts inverts the sign of their ζ -potentials, while their ψ_0 potentials remain unchanged. Surface-active ions of opposite charge may also produce such charge inversions.

The superposition of the van der Waals attractive energy, with its long-range effectiveness, and the electrostatic repulsive energy with its intermediate-range effectiveness, frequently produces a shallow minimum (designated *S* in Figure 20-7) in the resultant energy-distance curve at interparticle distances *dS*. If this minimum in potential energy is small compared to *kT*, brownian motion prevents aggregation. For large particles, such as those of many pharmaceutical suspensions, and for particles

that are large in one or two dimensions (e.g., rods and plates), this *secondary minimum* may be deep enough to trap them at distances dS from each other. This requires a depth of several kT units. Such fairly long-range and weak energies of attraction produce loose aggregates or flocs that can be redispersed by agitation or by reducing the concentration of flocculating electrolytes.^{1,2,12,13} This reversible aggregation process involving the secondary minimum is called *flocculation*. By contrast, aggregation in the deep primary minimum P is called *coagulation* and is irreversible.

Electrokinetic Phenomena

When a dc electric field is applied to a dispersion, the particles move toward the electrode having a charge opposite to that on their surface. The counterions located inside their hydration shell are dragged along while the counterions in the diffuse double layer outside the plane of slip, in the free or mobile solvent, move toward the other electrode. This phenomenon is called *electrophoresis*. If the charged surface is immobile, as is the case with a packed bed of particles or a tube filled with water, application of an electric field causes the counterions in the free water to move toward the opposite electrode, dragging solvent with them. This flow of liquid is called *electro-osmosis*, and the pressure produced by it is called *electro-osmotic pressure*. Conversely, if the liquid is made to flow past charged surfaces by applying hydrostatic pressure, displacement of the counterions in the free water produces a potential difference between the two ends of the tube or bed called *streaming potential*. These three phenomena depend upon the relative motion of the charged surface and the diffuse double layer outside the plane of slip surrounding the surface. Actually, most of the diffuse double layer lies within the free solvent and can therefore move along the surface.^{1,2,14} All three electrokinetic phenomena measure the same zeta potential, which is the potential at the plane of slip.

MICROELECTROPHORESIS—MEASURING ZETA POTENTIAL

The particles of pharmaceutical suspensions and emulsions, bacteria, erythrocytes and other isolated cells, latex particles, and many contaminant particles in pharmaceutical solutions are visible in a microscope. Therefore, their zeta potentials can be measured by *microelectrophoresis*. A potential difference, E , applied between two electrodes that are dipped into a dispersion and separated by a distance, d , produces the potential gradient or field strength, E/d , expressed as V/cm. The average velocity, v , of the particles in response to the applied potential difference is determined using the eyepiece micrometer of a microscope and a stopwatch, and used to calculate the zeta potential by the Smoluchowski equation:

$$\zeta = \left(\frac{4\pi\eta}{D} \right) \left(\frac{v}{E/d} \right) = \left(\frac{4\pi\eta}{D} \right) \mu \quad (8)$$

The electrophoretic mobility, μ , is equal to $v/(E/d)$ and is the velocity caused by a potential gradient of 1 V/cm. According to the Smoluchowski equation, particle size and shape do not affect the zeta potential. However, if the particle radius is smaller than or comparable to δ (in which case the particles cannot be detected in a microscope), a factor of 6 replaces the 4. The viscosity, η , and the dielectric constant, D , refer to the aqueous medium within the double layer and cannot be measured directly.¹⁵ By using the values for water at 25°C, expressing the velocity in micrometers/s and the electrophoretic mobility in (micrometers/s)/(volts/cm), and converting into the appropriate units, the Smoluchowski equation is reduced to $\zeta = 12.9 \mu$ with zeta potential given in millivolts (mV). Whilst the above method can be used, zeta potential is normally measured using systems based on laser scattering methods to measure the electrophoretic mobility of particles and a range of systems are available. When using these techniques it is important to

remember that zeta potential is a function of the particle surface and the dispersant. Given that the zeta potential of a particle depends largely on the ionic nature, ionic strength and pH of the suspending medium, any dilution of samples should take this into consideration and dilutions should be made with solutions of identical composition to the continuous phase of the sample, rather than water.

CAPILLARY ELECTROPHORESIS

Capillary electrophoresis (CE) is a widely used separation technique best suited for charged, water-soluble molecules having molecular weights ranging from those of amino acids and peptides to nucleic acids. CE has the following advantages: it provides fast and efficient separations, requires only minute amounts of sample, is applicable to a wide range of analytes, and (in contrast to high-pressure liquid chromatography, HPLC) employs aqueous media rather than organic solvents.

Electrophoresis is carried out in horizontal capillaries of fused silica (which is transparent to ultraviolet) with a bore of 20 to 100 micrometers and length of 20 to 100 cm. Both capillary ends are bent downward. Each end is immersed in a vial filled with a buffer solution that contains an electrode assembly. The electrodes are connected to an adjustable high-voltage dc power supply. As the dissolved analytes migrate past a detection window in the capillary, their concentrations are measured by ultraviolet absorption or by (often laser-induced) fluorescence. The silica capillaries are often coated externally with a very thin layer of polyimide to reduce their fragility. This plastic coating is burned off in the window area.

The capillary is filled with buffer solution, the sample is injected at one end, and a constant high potential difference E is applied, which produces a potential gradient or field strength in the range of 300 to 400 V/cm. If $E = 20,000$ V and the total capillary length $d_r = 57$ cm, then $E/d_r = 350$ V/cm. The velocity of migration, v , is the capillary length to the detector, dL , divided by the migration time, t , of the analyte from the capillary injection end to the detection window. If $dL = 50$ cm and $t = 10$ minutes = 600 seconds, then $v = 50/600 = 0.083$ cm/s. The electrophoretic mobility [$v/(E/d)$] is $(0.083 \text{ cm/s})/(350 \text{ V/cm}) = 2.4 \times 10^{-4} \text{ cm}^2/\text{V s}$.

In addition to electrophoresis, electro-osmosis may play an important role in CE. The isoelectric point of hydrated silica is approximately 1.8. The weakly acidic silanol groups become increasingly ionized with increasing pH. Conditioning the capillary with a NaOH solution and then with the buffer ensures that its wall is charged uniformly with partially ionized silanol groups. These negatively charged sites attract cationic counterions from the buffer to form an electric double layer. When an electric potential is applied, the cations in the diffuse part of the double layer beyond the plane of shear (Figure 20-6) migrate toward the negative electrode, entraining water of hydration. Because of the small bore of the capillaries compared to the diffuse double-layer thickness, the electro-osmotic flow (EOF) can be substantial. The EOF moves in a plug profile rather than in the customary parabolic profile of laminar flow. At high pH the EOF is strong, because the silanol groups are extensively ionized. Amphoteric peptides are negatively charged and try to migrate toward the positive electrode or anode. This motion is often overwhelmed by a strong and opposite EOF, which drags them toward the negative electrode or cathode. These analytes move from the injection point in the anode compartment toward the cathode. The order of arrival at the detection window is: cationic analytes (whose electrophoretic migration toward the cathode is superimposed and accelerated by the EOF), non-ionic analytes (which migrate exclusively by EOF), and anionic analytes (whose net migration velocity is that of the EOF minus their electrophoretic migration velocity toward the anode).

At low pH the EOF is small, and the peptides are positively charged. Their migration toward the cathode is then superimposed and accelerated by the EOF in the same direction. The direction of the EOF can be reversed by adding to the buffer a

cationic surfactant, such as cetyltrimethylammonium bromide. It will react and neutralize the silanol groups, and excess surfactant adsorbed on the silica will confer a positive charge to it. The Br^- counterions cause an EOF toward the anode. The EOF can be suppressed by coating the silica capillaries with a polymer or by using Teflon capillaries.

Unless the heat generated by the electric resistance of the buffer solution (Joule heating) is dissipated, it causes the temperature to increase with time and promotes temperature gradients across the capillary. This interferes with reproducibility and sharpness of the CE separations. The amount of heat generated is directly proportional to the square of the field strength (which is large) and to the conductivity of the buffer solution. While decreasing the voltage, using longer or smaller bore capillaries and/or more dilute buffer solution would reduce the rate of heat generation; it would also increase the separation times.

Short separation times reduce the band broadening due to analyte diffusion, improving the resolution. Therefore, the capillary is cooled with a thermostatted liquid.

A variation of CE is micellar electrokinetic (capillary) chromatography (MEKC). The analytes are solubilized in micelles of an ionic surfactant, such as sodium dodecyl sulfate (SDS), which is added to the buffer solution at a concentration well above its CMC of 8×10^{-3} M. MEKC is suited particularly to separate neutral analytes of limited water solubility, which are extensively partitioned into the micelles, but is also applicable to ionic analytes. Anionic analytes, being well soluble in water, either do not partition into the micelles or form mixed micelles with the anionic surfactant. Cationic analytes often form precipitates with anionic surfactants. Therefore, their separation by MEKC is best carried out with cationic surfactants.

The combination of a nonionic and an ionic surfactant produces mixed micelles. These have lower surface charge densities and larger sizes than the micelles of ionic surfactants, and hence they have lower electrophoretic mobilities. Thus, addition of a nonionic to an ionic surfactant narrows the migration time window, which is the difference between the migration times of the bulk solution in EOF and of the micelles, and often shortens the analysis time.

Analytes with molecular weights of 5000 or higher are not solubilized by micelles. Therefore, while they can be analyzed by CE, they cannot be analyzed by MEKC. Although anionic micelles tend to migrate toward the anode, a strong EOF in neutral or alkaline media will drag them toward the cathode, but with a velocity retarded by their own migration velocity. If a neutral analyte is partly solubilized in micelles and partly dissolved molecularly in the buffer solution, the latter portion has a shorter migration time, because its velocity is that of the EOF.

*Capillary gel electrophoresis*¹⁶ employs capillaries filled with a gel of cross-linked polyacrylamide, which suppresses EOF. Isotachopheresis and isoelectric focusing, described in previous editions of this text, are other modalities of CE.

LYOPHILIC DISPERSIONS

Lyophilic dispersions consist either of polymers dissolved in a solvent or of insoluble but extensively solvated particles dispersed in a liquid medium that has a high affinity or attraction for them. The free energy of dissolution or dispersion is $\Delta G_s = \Delta H_s - T\Delta S_s$, where ΔH_s and ΔS_s are the heat or enthalpy change, and the entropy change of dissolution or dispersion, respectively. For dissolution of polymers and dispersion of particulate solids to occur spontaneously, ΔG_s must be negative. Because both of these processes are exothermic (i.e., occur with the evolution of heat), their ΔH_s is negative. Because the number of available conformations of the polymer chains increases considerably upon dissolution, and the number of positions and orientations of the solid particles increases considerably upon their dispersion in the liquid, their ΔS_s is positive (i.e., there is an increase in randomness). The negative enthalpy change and the positive entropy change of dissolution/dispersion both

contribute to making ΔG_s negative. Therefore, both types of colloidal systems are formed spontaneously when powders of solid polymers and particulate solids are brought into contact with the liquid dispersion medium. They are thermodynamically stable and reversible (i.e., they are easily reconstituted after the dispersion medium has been removed). The van der Waals energies of attraction between dissolved macromolecules or between dispersed lyophilic solid particles are smaller than ΔG_s and are therefore insufficient to cause flocculation or coagulation of the dispersed phase. Furthermore, the solvation layers surrounding dissolved macromolecules and dispersed lyophilic particles form a physical barrier preventing their close approach.

Most liquid dispersed systems of pharmaceutical interest are aqueous. Therefore, most of the lyophilic colloidal systems discussed below consist of hydrophilic solids dissolved or dispersed in water. Most of the products mentioned below are official in the US Pharmacopeial Convention (USP) or National Formulary (NF), where more detailed descriptions can be found; they are also discussed in detail elsewhere in this text. Hydrophilic colloids can be divided into two classes (i.e., soluble and particulate materials). Solutions of water-soluble polymers molecularly dissolved in water can be classified as colloidal dispersions, because the individual molecules are in the colloidal particle size range: the diameter of a randomly coiled polymer chain commonly exceeds 10 nm. Particulate or corpuscular hydrophilic colloidal dispersions are formed by solids that swell and are peptized in water but whose primary particles do not dissolve or break down into individual molecules or ions.

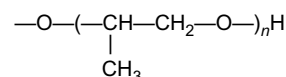
WATER-SOLUBLE POLYMERS

Most of the hydrophilic colloidal systems used to prepare pharmaceutical dosage forms are molecular solutions of water-soluble, high-molecular-weight polymers. These polymers are either linear or slightly branched but not cross-linked. Water-soluble polymers can be divided into three classes according to their origin:

Natural polymers include polysaccharides (acacia, agar, heparin sodium, pectin, sodium alginate, tragacanth, xanthan gum) and polypeptides (casein, gelatin, protamine sulfate). Of these, agar and gelatin are soluble only in hot water.

Cellulose derivatives are produced by chemically modifying cellulose obtained from wood pulp or cotton to produce soluble polymers. *Cellulose* is an insoluble, linear polymer of glucose units in the ring or pyranose form joined by β 1,4 glucosidic linkages. Each glucose unit (except for the two at the terminal chain ends) contains a primary hydroxyl group on C6 (the carbon atom at position 6) and two secondary hydroxyls on C2 and C3. Chemical modification of cellulose involves substitutions at these hydroxyl groups, with the primary hydroxyl group being the most reactive. The extent of such reactions is expressed as *degree of substitution* (DS), the number of substituted hydroxyl groups per glucose residue. The highest value for DS is 3. Fractional values are most common, because the DS is averaged over a multitude of glucose residues. A DS value of 0.6 indicates that some glucose repeat units are not substituted while others substituted once or even twice.

Some soluble cellulose derivatives are listed below. Their DS values correspond to their respective pharmaceutical grades; the groups shown replace the hydrogen atoms of the cellulosic hydroxyls. Official derivatives include *methylcellulose* (DS = 1.65–1.93); $(-\text{O}-\text{CH}_3)$ and *sodium carboxymethylcellulose* (DS = 0.60–1.00); $(-\text{O}-\text{CH}_2\text{COO}-\text{Na}^+)$. *Hydroxyethylcellulose* (DS $\approx$ 1.0); $(-\text{O}-(\text{CH}_2\text{CH}_2\text{O})_{\text{nl}})$ and *hydroxypropylcellulose* (DS $\approx$ 2.5);



are manufactured by adding ethylene oxide and propylene oxide, respectively, to alkali-treated cellulose. The value of n is about 2.0 for hydroxyethylcellulose and not much greater than

1.0 for hydroxypropylcellulose. *Hydroxypropylmethylcellulose* is prepared by reacting alkali-treated cellulose first with methyl chloride to introduce methoxy groups (DS = 1.1–1.8) and then with propylene oxide to introduce propylene glycol ether groups (DS = 0.1–0.3). In general, the introduction of hydroxypropyl groups into cellulose slightly reduces its water solubility while promoting its solubility in polar organic solvents such as short-chain alcohols, glycols, and some ethers.

The molecular weight of native cellulose is so high that soluble derivatives of approximately the same degree of polymerization would dissolve too slowly and their solutions would be excessively viscous even at concentrations of ≤ 1 percent. To overcome these difficulties, controlled degradation is used to break the cellulose chains into shorter segments. Commercial-grade cellulose derivatives, such as sodium carboxymethylcellulose, come in various molecular weights or viscosity grades as well as with various degrees of substitution.

Official cellulose derivatives that are insoluble in water but soluble in some organic solvents include *ethylcellulose* (DS = 2.2–2.7); $(-\text{OC}_2\text{H}_5)$; *cellulose acetate phthalate* (DS = 1.70 for acetyl and 0.77 for phthalyl); *hydroxypropylmethylcellulose phthalate*; and *polyvinyl acetate phthalate*. *Collodion*, a 4.0 percent (w/v) solution of pyroxylin (cellulose dinitrate) in a mixture of 75 percent (v/v) ether and 25 percent (v/v) ethyl alcohol, is also a cellulose-based, lyophilic colloidal system.

Water-soluble synthetic polymers consist mostly of high-molecular-weight PEGs, or *polyethylene oxides*, and vinyl derivatives such as *polyvinyl alcohol*, *povidone* or polyvinylpyrrolidone, and *carbomer* (Carbopol; Lubrizol Corp.), a copolymer of acrylic acid.

A second classification of hydrophilic polymers is based upon their charge. *Nonionic* or uncharged polymers include methylcellulose, hydroxyethyl and hydroxypropyl cellulose, ethylcellulose, pyroxylin, polyethylene oxide, polyvinyl alcohol, and povidone. *Anionic* or negatively charged *polyelectrolytes* include carboxylated polymers (e.g., acacia, alginic acid, pectin, tragacanth, xanthan gum, and carbomer) at pH values that result in the ionization of their carboxyl groups. Sodium alginate, sodium carboxymethylcellulose, and polypeptides (e.g., sodium caseinate) at pH values above their isoelectric points are also anionic. Sulfuric acid is a stronger acidic group that exists as a monoester in agar and heparin and as a monoamide in heparin. *Cationic* or positively charged *polyelectrolytes* are rare. Examples include chitin, a polysaccharide found in the shells of beetles and crustaceans, and polypeptides at pH values below their isoelectric points. Protamines such as protamine sulfate are strongly basic due to their high arginine content and have isoelectric points around pH 12.

PARTICULATE HYDROPHILIC DISPERSIONS

The dispersed phase of these sols consists of solids that swell in water and spontaneously break up into particles having colloidal dimensions. The dispersed particles have high specific surface areas and are extensively hydrated. *Bentonite NF* is a hydrated aluminum silicate that crystallizes in a layer structure with individual lamellas 0.94 nm thick. Their top and bottom surfaces consist of sheets of oxygen ions from silica and an occasional sodium ion neutralizing a silicate ion-exchange site. The clay particles contain stacks of these lamellae. Water penetrates between these lamellae to hydrate the oxygen ions and causes extensive swelling. The bentonite particles in bentonite magma consist of single lamellae or packets of a few lamellae with intercalated water. Their specific surface area amounts to several hundred square meters per gram. *Kaolin USP* is also a hydrated aluminum silicate having a layer structure. In kaolin, hydrated alumina lattice layers alternate with silica layers. Therefore, one of the two external surfaces of a kaolin plate consists of a sheet of oxygen ions from silica, whereas the other is a sheet of hydroxide ions from hydrated alumina. Both surfaces are well hydrated, but water cannot penetrate into the individual lattice layers. Therefore, the particles do not swell in

water or exfoliate into thin plates. As a result, kaolin plates dispersed in water are much thicker than those of bentonite, about 0.04 to 0.2 micrometers. *Magnesium Aluminum Silicate NF*, also known as Veegum® (R.T. Vanderbilt Co.), is a clay similar to bentonite but contains magnesium; it is white, whereas bentonite is gray. *Colloidal Activated Attapulgite USP* also consists of magnesium aluminum silicate. However, instead of having a lamellar habit like the other three clays, it crystallizes in the form of long needles approximately 20 nm wide.

The following additional hydrophilic particles can also produce colloidal dispersions in water. *Titanium dioxide* is a white pigment with excellent covering power. *Colloidal silicon dioxide* consists of roughly spherical particles that are covered with siloxane and silanol groups. *Microcrystalline cellulose* is hydrophilic because of the hydroxyl and ether groups on the surface of the cellulose crystals. Gelatinous precipitates of hydrophilic compounds such as *aluminum hydroxide gel*, *aluminum phosphate gel*, and *magnesium hydroxide* consist of coarse flocs produced by agglomeration of the colloidal particles formed in the initial stages of precipitation.

LYOPHOBIC DISPERSIONS

Lyophobic dispersions are intrinsically unstable and irreversible because of the lack of attraction between the dispersed and continuous phases. Unlike lyophilic dispersions, their large surface free energy is not lowered by solvation, their dispersion process does not take place spontaneously, and they are not easily reconstituted. For lyophobic dispersions, ΔG_s is positive because of a positive (endothermic) ΔH_s term, which makes the reverse process (agglomeration) the spontaneous one. Aqueous dispersions of hydrophobic solids or liquids can be prepared by physical means that supply an appropriate amount of energy to the system. However, they are unstable. The van der Waals attractive forces between the particles are stronger than the solvation forces that promote particle dispersal, and therefore, the particles tend to aggregate. Most of the discussion of lyophobic dispersions deals with hydrosols consisting of hydrophobic solids or liquids dispersed in an aqueous medium, because water is the most widely used vehicle. Such hydrosols consist of aqueous dispersions of insoluble organic and inorganic compounds, which usually have low degrees of hydration. Organic compounds that are preponderantly hydrocarbon in nature and possess few hydrophilic or polar groups are hydrophobic, and therefore, insoluble in water. Like all lyophobic dispersions, hydrophobic dispersions are intrinsically unstable. In their most thermodynamically stable state, the dispersed phase has coalesced into large crystals or drops, so that the specific surface area and surface free energy are minimized. Therefore, mechanical, chemical, or electrical energy must be supplied to break up the dispersed phase into smaller particles and overcome the resulting increase in surface free energy that occurs from the parallel increase in specific surface area.

Hydrophobic dispersions can be prepared by either dispersion methods (the reduction of coarse particles to colloidal dimensions through comminution or peptization) or condensation methods (the aggregation of small molecules or ions into particles having colloidal dimensions). Dispersion methods tend to produce sols that have wide particle size distributions. Conversely, condensation methods may produce essentially monodisperse sols provided specialized techniques are employed.

PREPARATION BY DISPERSION METHODS

The first method, *mechanical disintegration* of solids and liquids into smaller particles before or during dispersion within a fluid vehicle, is frequently carried out by the input of mechanical energy via shear or attrition. Equipment such as colloid and ball mills, micronizers and (for emulsions) homogenizers is described in Chapter 23 ("Powders"). Dry grinding with inert, water-soluble diluting agents also produces colloidal dispersions. For example, sulfur hydrosols may be prepared by

tritulating the powder with urea or lactose followed by shaking with water. Ultrasonic generators provide exceptionally high concentrations of energy. However, the successful dispersion of solids by means of ultrasonic waves can only be achieved with comparatively soft materials such as many organic compounds, sulfur, talcum, and graphite. In cases where fine emulsions are mandatory, such as soybean oil-in-water (O/W) emulsions for intravenous feeding, emulsification by ultrasound waves is the method of choice.

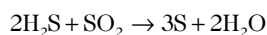
Peptization is a second dispersion method used to prepare colloidal dispersions. The term is defined as the breaking up of aggregates (or secondary particles) into smaller aggregates (or primary particles) that are within the colloidal size range. Primary particles are those particles that are not formed from smaller ones. Peptization is synonymous with *deflocculation*. It can be brought about by the removal of flocculating agents, usually electrolytes, or by the addition of deflocculating or peptizing agents, usually surfactants, water-soluble polymers, or ions that adsorb onto the particle surface.² When powdered activated charcoal is added to water with stirring, the aggregated grains cannot be completely broken up and the resulting suspension is gray and translucent. The addition of ≤ 0.1 percent sodium lauryl sulfate or octoxynol 9 deflocculates the grains into finely dispersed particles and results in a deep-black and opaque dispersion. Ferric or aluminum hydroxide that has been freshly precipitated with ammonia can be peptized with small amounts of acids that reduce the pH below the isoelectric points of the hydroxides. Even washing the gelatinous precipitate of $\text{Al}(\text{OH})_3$ with water tends to peptize it. Therefore, in quantitative analyses, the precipitate is instead washed with dilute solutions of ammonium salts that act as flocculating agents.

PREPARATION BY CONDENSATION METHODS

Sulfur is insoluble in water but somewhat soluble in alcohol. When an alcoholic solution of sulfur is mixed with water, a bluish white colloidal dispersion results. In the absence of added stabilizing agents, the particles tend to agglomerate and precipitate upon standing. This technique of first dissolving a material in a water-miscible solvent such as alcohol or acetone and then producing a hydrosol by precipitation with water is applicable to many organic compounds. It has been used to prepare hydrosols of stearic acid, natural resins like mastic, and the so-called pseudolatexes. Another less common physical condensation method is to introduce a current of sulfur vapor into water, which produces colloidal particles. Alternatively, the very fine powder produced by condensing sulfur vapor onto cold solid surfaces (sublimed sulfur or flowers of sulfur) can be dispersed in water by the addition of a suitable surfactant to produce a hydrosol.

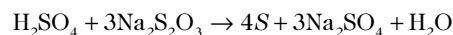
Organic compounds that are weak bases, such as the alkaloids, are usually much more soluble at lower pH values, where they are ionized, than at higher pH values, where they exist as the free base. Therefore, increasing the pH of their aqueous solutions above their pK_a may cause precipitation of the free base. Conversely, organic compounds that are weak acids, such as the barbiturates, are usually much more soluble at higher pH values, where they are ionized, than at lower pH values, where they exist as the free acid form. Therefore, lowering the pH of their aqueous solutions well below their pK_a usually causes precipitation of the free acid. Depending upon the supersaturation (defined below) of the nonionized bases or acids and the presence of stabilizing agents, the resultant dispersions may be within the colloidal range.

Chemical condensation methods include the reaction between hydrogen sulfide and sulfur dioxide (e.g., by bubbling H_2S into an aqueous SO_2 solution):



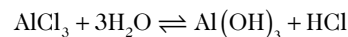
The same reaction occurs when aqueous solutions containing sodium sulfide and sulfite are acidified with an excess of sulfuric or hydrochloric acid. Another reaction is the decomposition of

sodium thiosulfate by sulfuric acid, using either very dilute or very concentrated solutions to obtain colloiddally dispersed sulfur:



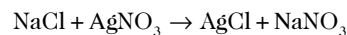
Both reactions also produce pentathionic acid ($\text{H}_2\text{S}_5\text{O}_6$) as a by-product. The preferential adsorption of the pentathionate anion onto the surface of the sulfur particles confers a negative electric charge to the particles, thereby stabilizing the sol.^{2,5} When powdered sulfur is boiled with a slurry of lime, it dissolves with the formation of calcium pentasulfide and thiosulfate. Subsequent acidification produces the colloidal "milk of sulfur," which upon washing and drying yields Precipitated Sulfur USP.

Sols of ferric, aluminum, chromic, stannic, and titanium hydroxides or hydrous oxides are produced by the hydrolysis of the corresponding chlorides or nitrates:



Hydrolysis is promoted by boiling the solution and/or adding a base to neutralize the formed acid.

Double decompositions that produce insoluble salts can also lead to colloidal dispersions. An example is silver chloride:



In addition, the reduction of gold, silver, copper, mercury, platinum, rhodium, and palladium salts with formaldehyde, hydrazine, hydroxylamine, hydroquinone, or stannous chloride forms hydrosols of the metals, which are strongly colored (e.g., red or blue).²

Kinetics of Particle Formation

When the solubility of a compound in water is exceeded, its solution becomes supersaturated and the compound may precipitate or crystallize. The rate of precipitation, the resulting particle size (whether colloidal or coarse), and the particle size distribution (which can be narrow for mono- or homodispersed particles or broad for poly- or heterodispersed particles) depends upon two successive and largely independent processes. These are *nucleation* and *crystallization* (i.e., growth of nuclei). When a solution of a salt or sucrose is supercooled or when a chemical reaction produces a salt in a concentration exceeding its solubility product, the separation of excess solid from the supersaturated solution is far from instantaneous. Clusters of ions or molecules called nuclei must exceed a critical size before they become stable and capable of growing into colloidal-size crystals. These embryonic particles have much more surface for a given weight of material than larger and more stable crystals. Therefore, they have a higher surface free energy and greater solubility.

The occurrence of *nucleation* depends upon the *relative supersaturation*. If C is the actual concentration of the solute before crystallization and C_s is its solubility limit, then $C - C_s$ is the supersaturation and $(C - C_s)/C_s$ is the relative supersaturation. Russian chemist Peter von Weymarn recognized early in the twentieth century that the rate or velocity of nucleation (number of nuclei formed per liter per second) is proportional to the relative supersaturation. Nucleation seldom occurs at relative supersaturations below 3. However, this statement refers to homogeneous nucleation, where the nuclei have the same chemical composition as the crystallizing phase. If the solution contains solid impurities, such as dust particles in suspension, these may act as nuclei or centers of crystallization (heterogeneous nucleation).

Once nuclei have formed, *crystallization* begins. Nuclei grow by the aggregation of ions or molecules from solution. Crystallization continues until the supersaturation is relieved (i.e., until $C = C_s$) and may result in the formation of either colloidal or coarse particles. The rate of crystallization or growth of nuclei is proportional to the supersaturation:

$$\frac{dm}{dt} = \frac{A_{sp}D}{\delta}(C - C_s) \quad (9)$$

This equation is similar to the Noyes-Whitney equation that governs particle dissolution, except that $C < C_s$ for the latter process, making dm/dt negative. In both equations, m is the mass of material crystallizing out in time t , D is the diffusion coefficient of the solute molecules or ions, δ is the length of the diffusion path or the thickness of the liquid layer adhering to the growing particles, and A_{sp} is their specific surface area. The presence of dissolved impurities may affect the rate of crystallization and even change the crystal habit, provided that these impurities are surface-active and become adsorbed onto the nuclei or growing crystals. For instance, 0.005 percent polysorbate 80 or octoxynol 9 significantly retards the growth of methylprednisolone crystals in aqueous media.

Von Weimarn found that the particle size of the crystals depends strongly upon the concentration of the precipitating substance. At very low concentrations and slight relative supersaturation, diffusion is quite slow because the concentration gradient that drives the process is very small. Sufficient nuclei will usually form to relieve the slight supersaturation. However, crystal growth is limited by the small amount of excess dissolved material available to each particle and, therefore, the particles cannot grow beyond colloidal dimensions. This condition is represented by points A, D, and G of the schematic plot of von Weimarn (Figure 20-8)¹⁷. At intermediate concentrations the extent of nucleation is somewhat greater and much more material is available for crystal growth. Therefore, coarse crystals form rather than colloidal particles. This condition is represented by points B, E, and H in Figure 20-8. At high concentrations, nuclei appear so quickly and in such large numbers that supersaturation is relieved before any appreciable diffusion can occur. The high viscosity of the medium also slows down the diffusion of excess dissolved ions or molecules, retarding crystal growth without substantially affecting the rate of nucleation. Therefore, a large number of very small particles results that, because of their proximity, tend to link and produce a translucent gel. This condition is represented by points C and F in Figure 20-8. Upon subsequent dilution with water, such gels usually yield colloidal dispersions. Thus, colloidal systems are usually produced at very low and high supersaturations.

Low solubility is a necessary condition for producing colloidal dispersions. If the solubility of the precipitate is increased, for instance by heating the dispersion, a new family of curves will result similar in shape to those shown in Figure 20-7 but displaced to the right (toward higher concentrations) and

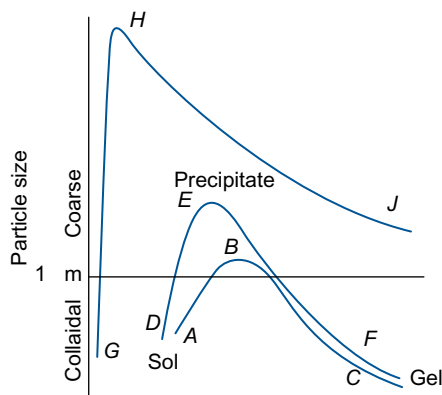


Figure 20-8. Effect of aging and concentration of the precipitating material upon particle size. Curves ABC, DEF, and GHI correspond to increasing aging. Both axes are on a logarithmic scale. (Data from von Weimarn PP. In: Alexander J, ed. *Colloid Chemistry*, vol I. New York: Chemical Catalog Co. (Reinhold), 1926.

upward (toward larger particle sizes).^{1,5,18,19} An additional phenomenon illustrated in Figure 20-7 is that aging increases particle size. Curves ABC, DEF, and GHI correspond to increasing time periods after mixing the reagents, namely, 10–30 minutes, several hours, and weeks to years, respectively. This gradual increase in the particle size of crystals in their mother liquor is a recrystallization process called *Ostwald ripening*. Very small particles have a higher solubility than large particles of the same substance due to their greater specific surface area and higher surface free energy. In a saturated solution containing precipitated particles with a wide range of particle sizes, the very smallest particles dissolve spontaneously and deposit onto the larger particles. The growth of the larger crystals at the expense of the very small ones occurs because this process lowers the free energy of the dispersion. Adding small amounts of surface-active compounds that adsorb at the particle surface slows down the process.

Increasing the solubility of the precipitate accelerates the spontaneous coarsening of colloidal dispersions upon aging. For instance, barium sulfate precipitated by mixing concentrated solutions of sodium sulfate and barium chloride is largely in the colloidal size range and passes through filter paper. The colloidal particles gradually grow in size by Ostwald ripening, forming large crystals that can be removed quantitatively by filtration. Heating the aqueous dispersion speeds recrystallization by increasing the solubility of barium sulfate in water.

Conversely, the addition of ethyl alcohol lowers the solubility of barium sulfate and slows Ostwald ripening, which allows the dispersion to remain in the colloidal state for years.

The relationship between particle size and solubility is given by the Ostwald-Freundlich or Kelvin equation, which, for non-ionic solutes, is^{2,5}:

$$\ln \frac{S}{S_\infty} = \frac{2\gamma M}{r d R T} \quad (10)$$

where S and S_∞ are the solubility of colloidal particles having a radius r and the solubility of large, flat particles ($r = \infty$), respectively. For electrolytes, the mean ionic activity is included. The solid-solvent interfacial free energy, γ , can only be determined indirectly (e.g., by means of this equation). The ratio of the molecular weight of the solute to its density (M/d) equals its molar volume. Assuming $M = 500$ g/mol, $d = 1.00$ g/mL, and $\gamma = 30$ erg/cm², and using the values of 8.314×10^7 erg/mol·K for the gas constant R and 298 K for the absolute temperature T , dispersed particles having radii of 1×10^{-6} cm (10 nm), 1×10^{-5} cm (0.1 micrometer), 1×10^{-4} cm (1 micrometer), and 1×10^{-3} cm (10 micrometers) correspond to S/S_∞ ratios of 3.36, 1.13, 1.012, and 1.0012, respectively. Therefore, while particles having sizes at the lower end of the colloidal range are appreciably more soluble than coarser particles of the same compound, the solubility of finely ground drug or excipient powders (particle radii typically in the range of 1 to 10 micrometers) is only increased by 1 percent or less.

Condensation methods generally produce polydisperse sols, because nucleation continues while established nuclei grow. However, monodispersed colloidal sols can be prepared by precipitation using a technique that involves the formation of all nuclei in a single, brief burst. A sufficiently brief period of homogeneous nucleation relieves the supersaturation to such an extent that no new nuclei can subsequently form. Therefore, the nuclei created during the initial burst grow uniformly as the remaining excess of precipitating material diffuses and deposits onto them. The supersaturation never again reaches sufficiently high values for forming new nuclei, because it is relieved by the continuous growth of the existing nuclei.^{2,5,18,19}

The controlled hydrolysis of salts of divalent and trivalent cations in aqueous solutions at elevated temperatures has been used to produce colloidal dispersions of metal (hydrous) oxides having uniform sizes in a variety of well-defined shapes (e.g., spheres or laths or cubes or disks). Complexation of the

cations, concentration, and temperature control the rate of hydrolysis, and therefore, the chemical composition, crystallinity, shape, and size of the dispersed phase.^{18,20}

STABILIZATION OF COLLOIDAL SYSTEMS

It should be reiterated that hydrosols of hydrophobic substances are intrinsically unstable. While mechanical disintegration may break up the dispersed phase into colloidal particles, flocculation or coagulation causes the dispersed particles to become progressively coarser and fewer, ultimately resulting in the complete separation of a macroscopic phase. The reduction in surface area and in surface free energy accompanying flocculation or coagulation is small, because irregular solid particles are rigid and only touch at a few points upon aggregation. However, these loose initial contacts may grow with time by sintering or recrystallization. Sintering is the “fusion” of primary particles into larger primary particles, which propagates from the initial small areas of contact. This recrystallization process is spontaneous, because it decreases the specific surface area of the dispersed solid and the surface free energy of the dispersion. Sintering is analogous to Ostwald ripening, the recrystallization process that transfers solids from colloidal to coarse particles. Low solubility and the presence of adsorbed surface-active substances retard both processes.

If aqueous dispersions of hydrophobic solids are to resist re-aggregation (i.e., flocculation and coagulation), they must be stabilized during or shortly after the dispersion process. Stabilizing factors include the presence of electrical charges at the particle surface and the presence of adsorbed macromolecules or nonionic surfactants. The presence of positive or negative charges may result from the dissociation of the solid's ionogenic groups or the adsorption of ions such as ionic surfactants. These stabilizing factors do not alter the intrinsic thermodynamic instability of lyophobic dispersions, and therefore, ΔG_s remains positive and phase separation or aggregation is still energetically favored over dispersal. However, they establish kinetic barriers that delay the aggregation processes almost indefinitely; the dispersed particles cannot come together close enough for van der Waals attractive forces to produce coagulation.²

Stabilization can be provided by adsorbed surfactants. In a flocculated dispersion, groups of several particles are agglomerated into flocs. Frequently, the particles of a floc are in physical contact. When a surfactant is added to a flocculated sol, the dissolved surfactant molecules adsorb onto the surface of the particles. Surfactant molecules also tend to pry apart the flocs by wedging themselves in between the particle contact points. This action frees additional surface area for surfactant adsorption. The breaking up of flocs or secondary particles is defined as deflocculation or peptization. Ophthalmic suspensions should be deflocculated, because the large particle size of flocs causes irritation to the eye. Parenteral suspensions should also be deflocculated to prevent the larger particles from clogging hypodermic syringes, causing tissue irritation, or blocking capillary blood vessels. However, deflocculated suspensions tend to cake (i.e., the sediment formed by gravitational settling is compact and may be hard to redisperse upon shaking). Caking in oral suspensions is prevented by controlled flocculation.

Surfactants tend to accumulate at interfaces because of their amphiphilic nature. This process is an *oriented physical adsorption*. Surfactant molecules arrange themselves at the interface between water and an organic solid or liquid of low polarity in such a way that the hydrocarbon chain is in contact with the surface of the solid particle or sticks inside the oil droplet while the polar head group is oriented toward the water phase. This orientation leaves the polar head group in contact with the water so that it can be hydrated, and removes the surfactant's hydrophobic hydrocarbon chain from the bulk of the water, where it is unwelcome because it interferes with the hydrogen bonding between water molecules. Figure 20-9 shows that at a low surfactant concentration and at low surface coverage, the hydrocarbon chains of the adsorbed surfactant molecules lie

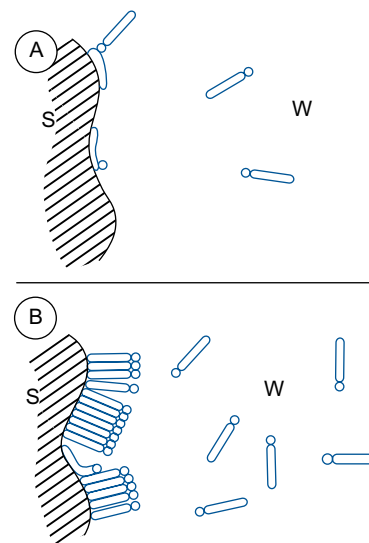


Figure 20-9. Schematic representation of the physical adsorption of surfactant molecules at a hydrophobic solid (S)–water (W) interface. Cylindrical portions and spheres represent the hydrocarbon chains and polar head groups of the surfactant molecules, respectively. (A) Low surfactant concentration, low surface coverage. (B) Near critical micelle concentration, surface coverage near saturation.

flat against the solid surface. At higher surfactant concentrations, the surfactant molecules are adsorbed in an upright position to allow for the greatest number of molecules per unit surface area. In this position, the terminal methyl groups of the hydrocarbon tails are in contact with the hydrophobic particle surface and the hydrocarbon tails are in lateral contact with each other. London dispersion forces promote the attraction between both types of adjoining groups.

The adsorption of ionic surfactants increases the charge density and the zeta potential of the dispersed particles. These two parameters are low for water-insoluble organic substances. The increase in electrostatic repulsion between nonpolar particles due to the adsorption of surface-active ions stabilizes the dispersion against coagulation. This “charge stabilization” is described by the DLVO theory.

Most water-soluble nonionic surfactants are polyoxyethylated. Each molecule consists of a hydrophobic hydrocarbon chain combined with a hydrophilic PEG chain – for example, $\text{CH}_3(\text{CH}_2)_{15}(\text{OCH}_2\text{CH}_2)_{10}\text{OH}$. Hydration of the ether groups and the terminal hydroxyl group renders the surfactant molecule water-soluble. It adsorbs at the interface between a hydrophobic solid and water, with the hydrocarbon moiety adhering to the solid surface and the PEG moiety protruding into the water, where it is hydrated. Therefore, the particle surface is surrounded by a thin layer of hydrated PEG chains. This hydrophilic shell forms a steric barrier that prevents close contact between particles and inhibits coagulation (“steric stabilization”). Nonionic surfactants also reduce the sensitivity of hydrophobic dispersions toward coagulation by salts (i.e., they increase the coagulation values).²¹

The adsorption of water-soluble polymers provides a second mechanism for the stabilization of hydrophobic dispersions. Water-soluble polymers that have some hydrophobic groups can be surface-active and adsorb at the interface between water and a hydrophobic organic solid, because their hydrophobic groups limit their water solubility and render them amphiphilic. Such polymers also tend to accumulate at the air-water interface and, therefore, lower the surface tension of the aqueous phase. Conversely, polyelectrolytes that have a high concentration of ionic groups are excessively water-soluble, which reduces or eliminates their surface activity and tendency to adsorb

at interfaces (e.g., sodium carboxymethylcellulose). Polyvinyl alcohol is another polymer that does not adsorb extensively at interfaces because of its high concentration of hydroxyl groups, which makes it very water-soluble. Polyvinyl alcohol is manufactured by the hydrolysis of polyvinyl acetate, which is water-insoluble. Hydrolysis of about 85 percent of the acetyl groups produces a copolymer that is both water-soluble and surface-active. Other surface-active polymers include methylcellulose, hydroxypropyl cellulose, high-molecular-weight polyethylene glycols or PEGs (polyethylene oxides), and proteins. The surface activity of proteins is due to the presence of hydrophobic groups at concentrations too low to cause insolubility in water. Proteins are denatured upon adsorption at air-water and solid-water interfaces.

As shown in Figure 20-10A, polymer molecules adsorb onto solid surfaces in the form of loops projecting into the aqueous phase rather than lying flat against the solid substrate. Only a small portion of an adsorbed polymer is in direct contact with the solid surface. However, because of its great chain length, there are enough of these contact points to anchor the adsorbed macromolecule firmly onto the surface. At sufficiently high concentrations, adsorbed polymers may form a layer surrounding the entire dispersed particles. This layer consists of the polymer chains as well as the water of hydration associated with them and any water mechanically trapped inside the chain loops. This sheath becomes an integral part of the particle surface and may prevent coagulation. The mechanisms by which adsorbed nonionic macromolecules prevent the coagulation of hydrophobic sols are the same ones operative in the stabilization of sols by nonionic surfactants. The hydrophilic PEG moieties of the adsorbed surfactant molecules that protrude into the aqueous phase resemble the chain ends of the adsorbed macromolecules rather than their looped segments.

The following protective mechanisms are operative:

1. The layer of adsorbed polymer and enmeshed water surrounding the particles forms a *mechanical or steric barrier* ("steric stabilization") that prevents the particles from approaching each other closely enough for the interparticle attraction of London dispersion forces to produce coagulation. These forces are only effective over interparticle distances smaller than twice the thickness of the adsorbed polymer layer. These layers are somewhat elastic; they may be dented by a collision between two particles but tend to return to their original shape.
2. When two particles approach so closely that their adsorbed polymer layers overlap, the chain loops of the

opposing layers compress and mix with or interpenetrate each other. The freedom of motion of the chain segments in the overlapped region becomes restricted, which produces a negative entropy change. Therefore, any reduction in interparticle distance, which is required for coagulation, results in a positive change in free energy. As a result, the reverse process of particle separation occurs spontaneously, because disentangling the two opposing adsorbed polymer layers is more energetically favorable. The particles are thus prevented from coagulating by *entropic repulsion* through the mechanism of *entropic stabilization* of the sol. This mechanism predominates when the concentration of polymer in the adsorbed layer is low.

3. As the adsorbed polymer layers on two approaching particles overlap, the polymer concentration in the overlap region causes a local increase in osmotic pressure, which is relieved by an influx of water. This influx of water to dilute the polymer loops pushes the two particles apart, preventing coagulation.
4. If the adsorbed polymer has some ionic groups, stabilization by electrostatic repulsion or charge stabilization, as previously described, is an additional factor that prevents a close interparticle approach and coagulation.
5. The adsorption of water-soluble polymers changes the nature of the surface of hydrophobic particles to hydrophilic, resulting in an increased resistance to coagulation by salts.²²

The water-soluble polymers whose adsorption stabilizes hydrophobic sols and protects them against coagulation are called *protective colloids*. *Gelatin* and *serum albumin* are the preferred protective colloids for stabilizing parenteral suspensions because of their biocompatibility. These two polymers, as well as casein (milk protein), dextrin (partially hydrolyzed starch), and vegetable gums like acacia and tragacanth, are metabolized in the human body. Cellulose derivatives and most synthetic protective colloids such as *povidone* are not biotransformed. Because of this and their large molecular size, these polymers are not absorbed but excreted intact when administered in an oral dosage form.

SENSITIZATION

Sensitization is the opposite of protective action (i.e., a decrease in the stability of the hydrophobic sols). At concentrations well below those at which it exerts a protective action, a protective colloid may flocculate a sol in the absence of added salts and/or lower the coagulation values of the sol. In the case of nonionic polymers and of polyelectrolytes having charges of the same sign as the sol particles, flocculation results from the bridging mechanism illustrated in Figure 20-10B. At very low polymer concentrations there are not nearly enough polymer molecules present to completely cover each sol particle. Because the particle surfaces are largely bare, a single macromolecule may be adsorbed onto two particles, thereby bridging the gap between them and pulling them close together. Flocs are formed when several particles become connected through polymer molecules that are adsorbed jointly onto two or possibly even three particles. Such flocculation usually occurs over a narrow range and at very low polymer concentrations. At higher concentrations bridging is unlikely to occur, because there is enough polymer to completely cover all of the particles and the adsorbed polymer stabilizes or peptizes the sol.^{5,22}

If the polymer contains ionic groups of charge opposite to that of the sol particles, a limited amount of polymer adsorption neutralizes the charge of the particles and reduces their zeta potential to nearly zero. This eliminates stabilization by electrostatic repulsion. In addition, steric stabilization is ineffective because of the low surface coverage of the adsorbed polymer. Therefore, the sol either coagulates by itself or may be coagulated with a very small amount of sodium chloride. At higher

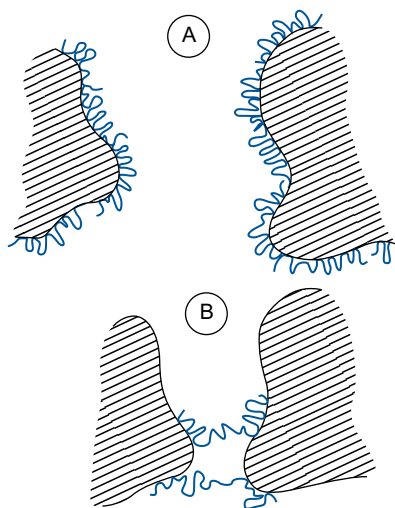


Figure 20-10. Protective action (A) and sensitization (B) of adsorbed polymer chains upon sols containing hydrophobic particles.

polymer concentrations, where adsorption is more extensive, the charge on the particles is converted to the sign of the polyelectrolyte, which reactivates charge stabilization and adds steric stabilization. As a result, the coagulation value of the sol increases well above the original value. For example, partly hydrolyzed polyacrylamide containing about 20 percent of ammonium acrylate repeat units is an anionic polyelectrolyte. Addition of this polyacrylamide to aluminum hydroxide sols at a polymer concentration of 1:1,000,000 and a pH of 6–7, where the sols are positively charged and the polyelectrolyte is fully ionized, results in flocculation. At a polymer concentration of 1:10,000 the sols become negatively charged because extensive polymer adsorption introduces an excess of -COO^- groups over the =Al^+ ions on the particle surface. This creation of negatively charged particles introduces electrostatic and steric stabilization, which makes the sols more stable against flocculation by salts than they were before the addition of the polyacrylamide.

Polymer *B* in Figure 20-11 illustrates this example. The curve in the lower plot indicates sensitization, with the coagulation value for sodium chloride lowered by as much as 60 percent. Zeta potential measurements can be used to distinguish between sensitization by bridging and sensitization by charge neutralization. The charge reversal caused by the adsorption of polymer *B* is illustrated in the upper plot and indicates that charge neutralization is the cause of sensitization. If polymer *B* had a zeta potential–polymer concentration plot similar to polymer *A*, sensitization would be ascribed to bridging. The nonionic polymer *A* in Figure 20-11 stabilizes the sol at all concentrations. Sensitization is not observed either by bridging or

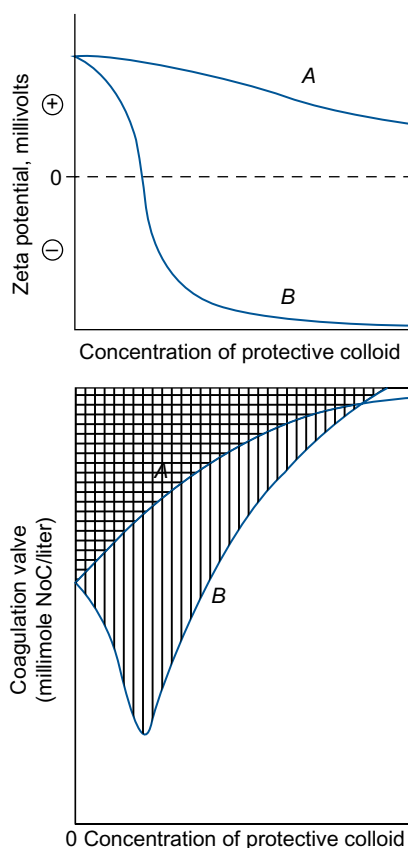


Figure 20-11. Protective action and sensitization. Polymer *A* exerts protective action at all concentrations, while polymer *B* sensitizes at low concentrations and stabilizes at high concentrations. Horizontal and vertical hatching indicates region of flocculation for a sol treated with various concentrations of polymers *A* and *B*, respectively. Clear region underneath indicates sol is deflocculated. (From Schott H, Martin AN. In Dittert LW, ed., *American Pharmacy*, 7th ed. Philadelphia: JB Lippincott, 1974.)

by charge neutralization. The reason that polymer *A* slightly lowers the positive zeta potential of the sol is that the increasing amounts of adsorbed polymer chains gradually shift the plane of shear outward and away from the positively charged surface. If polymer *A* were a cationic polyelectrolyte, the zeta potential–polymer concentration plot would gradually rise with an increase in polymer adsorption rather than drop.

Even water-soluble polymers that are too thoroughly hydrophilic to be adsorbed by hydrophobic sol particles can stabilize such sols. Their thickening action increases the viscosity of the sols. This slows brownian motion and sedimentation, giving the particles less opportunity to come in contact with each other and, therefore, decreasing flocculation.

ASSOCIATION COLLOIDS

Association colloids are formed by self-assembling enough small molecules to produce aggregates in the colloidal size range. This group of colloids includes surfactant micelles, microemulsions, and liposomes.

FORMATION OF SURFACTANT MICELLES

The dual or amphiphilic nature of surfactants or surface-active agents was discussed previously. Water attracts their polar head groups but repels their hydrocarbon tails. Consequently, surfactants tend to concentrate and adsorb at air-water, oil-water, and solid-water interfaces. The surface tension of aqueous surfactant solutions decreases with increasing surfactant concentration up to a point, beyond which it remains nearly constant (Figure 20-12). Curves *A* and *B* in Figure 20-12 illustrate the surface tension (against air) and the interfacial tension (against oil) of an aqueous surfactant solution as a function of surfactant concentration. Surface-active impurities may cause a minimum in the surface tension (shown as a dotted curve) rather than a mere leveling off. Abrupt changes occur not only to the surface and interfacial properties but also to the surfactant solution's bulk properties such as equivalent conductivity (curve *D*),

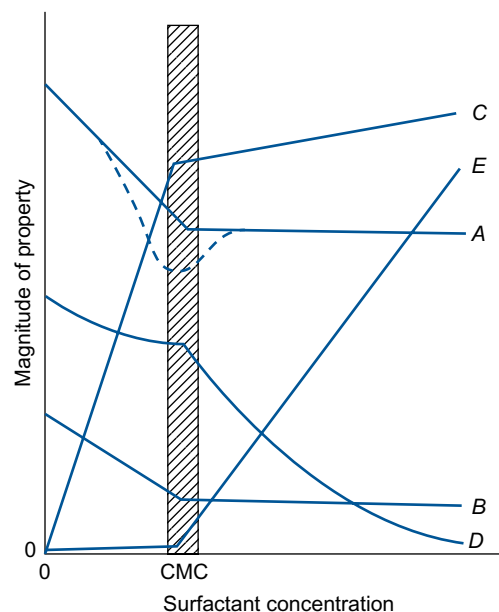


Figure 20-12. Effect of surfactant concentration and micelle formation on various properties of the aqueous solution of an ionic surfactant. Curve *A*, Surface tension; curve *B*, interfacial tension; curve *C*, osmotic pressure; curve *D*, equivalent conductivity; curve *E*, solubility of a compound with very low solubility in pure water. (From Schott H, Martin AN. In: Dittert LW, ed. *American Pharmacy*, 7th edn. Philadelphia: JB Lippincott, 1974.)

co-ion and counterion activities in the case of ionic surfactants, colligative properties like osmotic pressure (curve C), turbidity (but the increase is far too weak to be visible to the naked eye), refractive index, ultraviolet and nuclear magnetic resonance (NMR) spectra, partial molar volume, relative viscosity, and the diffusion coefficients and solubility of water-insoluble, oil-soluble compounds (curve E). All of these changes occur over a very narrow concentration range, which is shown as a crosshatched band and is referred to as the *critical micelle concentration* (CMC).

As the surfactant concentration in a liquid is increased, the amount of the surfactant adsorbed at the liquid-air and liquid-container interfaces increases and these interfaces become increasingly crowded. When the concentration is increased further, the surfactant molecules will continue to adsorb at these interfaces until tightly packed monolayers are formed and there is no longer any room for further surfactant adsorption. At this point, the surface and interfacial tensions reach their constant values,²³ and it would seem that the bulk solubility limit of the surfactant has been reached. However, if more surfactant is added to the solution, the excess surfactant molecules will begin to associate into small aggregates called *micelles* (at the CMC), while the concentration of nonassociated surfactant molecules remains nearly constant. Diluting the surfactant solution to below the CMC causes the micelles to disperse or break up into single or nonassociated surfactant molecules.

Micelles are not static aggregates; they dissociate, regroup, and reassociate rapidly. The half-life of ionic surfactant micelles in the absence of additives is a small fraction of a second. Furthermore, there is a dynamic equilibrium (i.e., an incessant exchange) between single surfactant molecules in solution, surfactant molecules adsorbed in monolayers at the interfaces, and surfactant molecules associated as micelles.

The shape of micelles in dilute aqueous surfactant solutions is approximately spherical (Figure 20-13A). The polar head groups of the surfactant molecules are arranged in an outer spherical shell, while their hydrocarbon chains are oriented toward the center where they form a spherical core. These hydrocarbon chains are randomly coiled and entangled. The micellar interior has a nonpolar, liquid-like character resembling a liquid normal paraffin such as dodecane. In nonionic surfactant micelles, the polyoxyethylene moieties are oriented outward and permeated by water, while the hydrocarbon moieties form an "oil droplet" core similar to ionic micelles (Figure 20-13B): this arrangement is energetically favorable. The hydrophilic head groups, located externally, are in contact with water and remain extensively hydrated. The hydrocarbon moieties are removed from the aqueous medium and partly shielded from contact with water by the polar head groups. Therefore, they no longer interfere with hydrogen bonding among the water molecules. This interference is the reason why surfactant molecules are pushed out of aqueous media toward interfaces. The hydrocarbon tails of the surfactant molecules, located in the micellar interior, attract one another by weak dispersion forces.^{21,24-26}

Representative CMC values and aggregation numbers (number of surfactant molecules per micelle) are listed in Table 20-4.^{26,27} Ionic surfactants have higher CMC values than nonionic surfactants, because electrostatic repulsion of the charged head groups makes micellization more difficult. The addition of simple salts reduces these repulsive forces and therefore lowers the CMC values of ionic surfactants. Within any homologous surfactant series, the CMC decreases regularly with increasing hydrocarbon chain length and therefore with increasing surface activity of the surfactant. As is seen in Table 20-4, each additional methylene group decreases the CMC by approximately one-half. This is a consequence of Traube's rule, which states that, 'in dilute aqueous solutions of surfactants belonging to an one homologous series, the molar concentrations required to produce equal lowering of the surface tension of water decreases threefold for each additional CH₂ group in the hydrocarbon chain of the solute'.

With nonionic surfactants, the polyoxyethylene content of the surfactant also influences their CMC (Table 20-4). As the percentage of polyoxyethylene increases, the CMC of the nonionic surfactant increases. Simply, this can be remembered by considering that the surfactant becomes more soluble with increasing polyoxyethylene content. Therefore there is less drive for the surfactants to form micelles and thus micelles form at a higher concentration (a higher CMC). Temperature can also influence the CMC of nonionic surfactants, with the CMC increasing as the temperature decreases. However the changes in CMC are quite small (compare Polyoxyl 12 dodecyl ether at 20°C vs 55°C; Table 20-4).

The micelles of the surfactants listed in Table 20-4 are either spherical or ellipsoidal. They are rather small, because their sizes were determined in relatively dilute solutions (containing only a few percent of surfactant) and mostly in pure water at room temperature. Their diameters are between 2 and 8 nm, which places them at the lower end of the colloidal size range. For this reason, surfactants are sometimes called *association colloids*. Adding salts increases the size of ionic micelles. Raising the temperature increases the size of nonionic micelles, especially if the temperatures are within 20°C of their respective cloud points. These factors reduce the water solubility of ionic and nonionic surfactants, thereby rendering them more surface-active.

As micelles become larger, they also become more asymmetric. Their shape changes from spherical or ellipsoidal to cylindrical and eventually to lamellar. In cylindrical micelles, the polar head groups form the periphery and the hydrocarbon tails fill the interior of the cylinders (Figure 20-13C). In lamellar micelles the surfactant molecules are arranged in parallel bimolecular sheets with a tail-to-tail orientation (i.e., the hydrocarbon tails form the inner layer). Water is stratified between the sheets, thereby hydrating the external polar head groups (Figure 20-13D). In both types of micelles the hydrocarbon tails are randomly coiled and in a liquid-like state.⁴ In concentrated aqueous solutions containing 20% or more of surfactant, cylindrical micelles often line up parallel to each other and arrange themselves in hexagonal arrays. Likewise, lamellar micelles are often packed parallel and equidistant from each other with the intervening water layers having a uniform thickness. These ordered solutions are liquid crystals or mesophases; they are birefringent and very viscous. Even though they are liquids, they have some of the properties of crystalline solids.^{2,21,25,26}

Oil-soluble surfactants (e.g., heavy-metal soaps, docusate sodium, and nonionic surfactants with hydrophilic-lipophilic balance values <7) form aggregates when dissolved in organic liquids having low polarity such as hydrocarbons and chlorinated hydrocarbons. These micelles are inverted or turned inside out; their hydrocarbon tails are oriented outward into the oil phase, while their polar head groups are in the center of the micelle, where water can be solubilized (Figure 20-13E). Because the bulky head groups are in the center, the aggregation numbers for *reverse micelles* are small, usually between 3 and 20.^{25,28}

MICROEMULSIONS

Microemulsions are liquid dispersions of water and oil that are made homogeneous, transparent, and stable by the addition of relatively large amounts of a surfactant and a co-surfactant. *Oil* is defined as any liquid having low polarity and low miscibility with water (e.g., toluene, cyclohexane, and mineral or vegetable oils). Microemulsions have intermediate properties between micelles containing solubilized oils and emulsions. Emulsions are lyophobic, unstable, and their preparation requires the input of considerable amounts of mechanical energy, which may be supplied by colloid mills, homogenizers, or ultrasonic generators. Conversely, microemulsions are on the borderline between lyophobic and lyophilic colloids. As microemulsions are thermodynamically stable, they generally form spontaneously when oil, water, surfactants, and co-surfactants are mixed together.

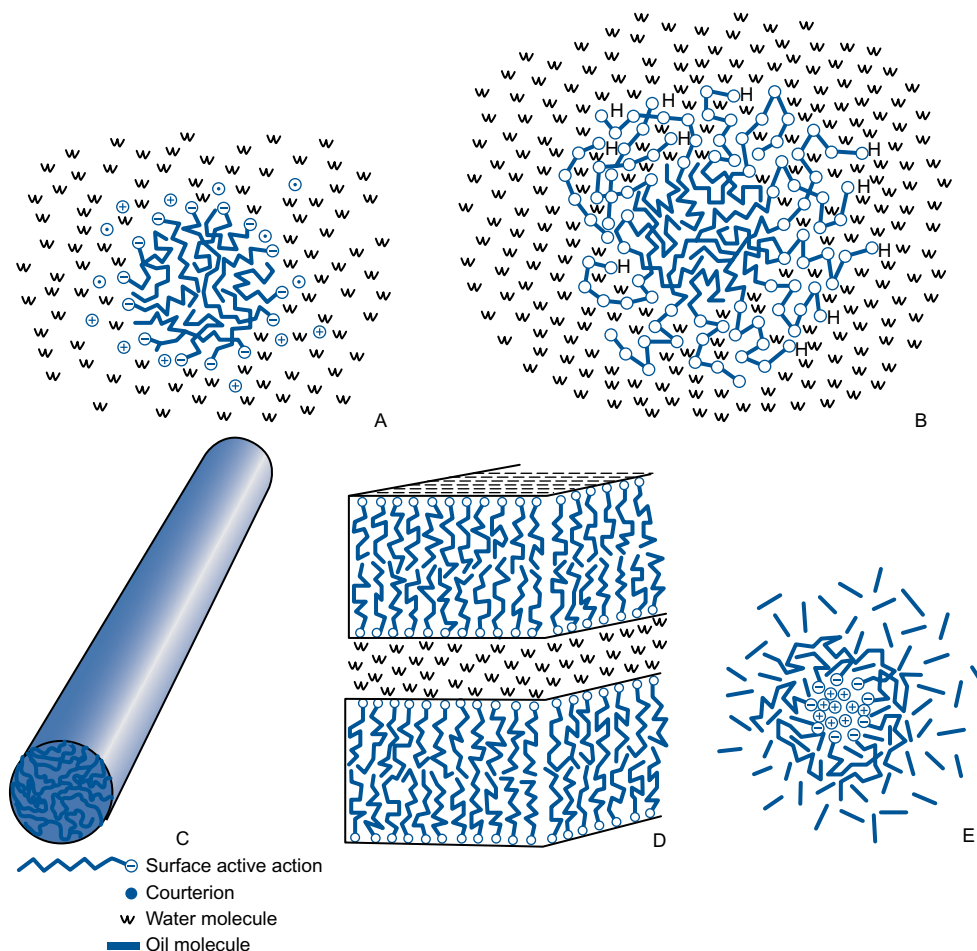


Figure 20-13. Different types of micelles. (A) Spherical micelle of an anionic surfactant; (B) spherical micelle of a nonionic surfactant; (C) cylindrical micelle of an ionic surfactant; (D) lamellar micelle of an ionic surfactant; (E) reverse micelle of an anionic surfactant in oil. (From Shinoda K, Nakagawa T, Tamamushi B-I, Isemura T. *Colloidal Surfactants*. New York: Academic Press, 1963.)

Both emulsions and microemulsions may contain high volume fractions of the internal phase. For instance, some O/W systems contain 75 percent (v/v) of oil dispersed in 25 percent water, although lower volume fractions of the internal phase are more common.

Microemulsion droplets have a mean diameter range of approximately 6 to 100 nm and a narrow droplet size distribution. Because the droplet diameters are less than one-fourth of the wavelength of light (420 nm for violet and 660 nm for red light), microemulsions scatter little light. Therefore, they are transparent or at least translucent. By contrast, emulsions have broad droplet size distributions and are generally opaque, because most of their droplets have diameters greater than the wavelength of light and most oils have higher refractive indices than water.

Emulsions contain three components, namely, oil, water, and surfactant; in comparison, microemulsions generally require a fourth component, a *co-surfactant*. Commonly used co-surfactants include linear alcohols of medium chain length that are sparingly miscible with water. The combination of surfactant and co-surfactant promotes the generation of extensive interfaces through the spontaneous dispersion of oil in water, or vice versa. The large interfacial area between the oil and water consists of a mixed interfacial film containing both surfactant and co-surfactant molecules. This film is called the *interphase*, because it is thicker than the typical surfactant monolayers formed at the oil-water interfaces in emulsions. The interfacial tension at the oil-water interface in microemulsions approaches zero, which also contributes to their spontaneous formation. According to

another viewpoint, microemulsions are regarded as micelles extensively swollen by large amounts of solubilized oil.

Micellar solutions, microemulsions, and emulsions can be of the O/W or W/O type. As mentioned previously, aqueous micellar solutions can solubilize oils in the hydrocarbon cores of the micelles. Conversely, oil-soluble surfactants like sorbitan monooleate and docusate sodium form reverse micelles in oils (Figure 20-13E) that are capable of solubilizing water in their polar centers. The solubilized oil in the former micelles and the solubilized water in the latter may in turn enhance the micellar solubilization of oil-soluble and water-soluble drugs, respectively.

Typical formulations for an O/W and a W/O microemulsion are shown in Table 20-5. The ratio of grams of surfactant to grams of solubilized or emulsified oil or water ranges from 2 to 20 for micellar solutions and 0.01 to 0.1 for emulsions. Microemulsions have intermediate values. For example, the ratios for the formulations in Table 20-5 are near unity. In industrial formulations, the ratios are closer to 0.1 to reduce costs. Microemulsions are used for a variety of applications including floor polishes, agricultural pesticides, tertiary petroleum recovery, and pharmaceutical delivery systems.

LIPOSOMES AND BILAYER CONSTRUCTS

Liposomes are synthetic spherical vesicles consisting of an aqueous core surrounded by one or more concentrically arranged bilayer membranes (Fig 20-14). These membranes are commonly composed of phospholipids, but can be prepared

Table 20-4. Critical Micelle Concentrations and Micellar Aggregation Numbers of Various Surfactants in Water at Room Temperature

Structure	Name	CMC (mmol/dm ³)	No. of Surfactant Molecules per Micelle
n-C ₁₁ H ₂₃ COOK	Potassium laurate	24	50
n-C ₈ H ₁₇ SO ₃ Na	Sodium octane sulfonate	150	28
n-C ₁₀ H ₂₁ SO ₃ Na	Sodium decane sulfonate	40	40
n-C ₁₂ H ₂₅ SO ₃ Na	Sodium dodecane sulfonate	9	54
n-C ₁₂ H ₂₅ OSO ₃ Na	Sodium dodecyl sulfate	8	62
n-C ₁₂ H ₂₅ OSO ₃ Na	Sodium dodecyl sulfate ^a	1	96
C ₂₀ H ₃₇ O ₄ SO ₃ Na	Docusate sodium	5	48
n-C ₁₀ H ₂₁ N(CH ₃) ₃ Br	Decyltrimethylammonium bromide	63	36
n-C ₁₂ H ₂₅ N(CH ₃) ₃ Br	Dodecyltrimethylammonium bromide	14	50
n-C ₁₄ H ₂₉ N(CH ₃) ₃ Br	Tetradecyltrimethylammonium bromide	3	75
n-C ₁₄ H ₂₉ N(CH ₃) ₃ Cl	Tetradecyltrimethylammonium chloride	3	64
n-C ₁₂ H ₂₅ NH ₃ Cl	Dodecylammonium chloride	13	55
C ₁₂ H ₂₅ (OCH ₂ CH ₂) ₈ OH	Polyoxyl 8 dodecyl ether	0.13	132
C ₁₂ H ₂₅ (OCH ₂ CH ₂) ₈ OH	Polyoxyl 8 dodecyl ether ^b	0.10	301
C ₁₂ H ₂₅ (OCH ₂ CH ₂) ₁₂ OH	Polyoxyl 12 dodecyl ether	0.14	78
C ₁₂ H ₂₅ (OCH ₂ CH ₂) ₁₂ OH	Polyoxyl 12 dodecyl ether ^b	0.091	116
C ₉ H ₁₉ C ₆ H ₄ (OCH ₂ CH ₂) ₁₀ OH	Nonoxynol 10	0.07	276
C ₉ H ₁₉ C ₆ H ₄ (OCH ₂ CH ₂) ₃₀ OH	Nonoxynol 30	0.24	44

CMC, critical micelle concentration.

^a Interpolated for physiological saline, 0.154 M NaCl.^b At 55°C instead of 20°C.**Table 20-5. Table .5. Microemulsion Formulations**

Compound	Function	Content in Microemulsions (%)	
		O/W	W/O
Sodium lauryl sulfate	Surfactant	13	10
1-Pentanol	Co-surfactant	8	25
Xylene	Oil	8	50
Water	Lipid	71	15

O/W, oil-in-water microemulsion; W/O, water-in-oil microemulsion.

from a range of lipids and/or surfactants. When lipid molecules are dispersed in an aqueous system, the resultant self-assembled structure is dependent (e.g. micelles versus liposomes) on the structure of the lipid, the lipid mixture, lipid concentration and temperature. In particular the molecular shape of the lipid has a strong driving force in determining the specific colloidal construct formed. The shape of the lipid molecule can be expressed as its critical packing parameter (which considers the ratio of the hydrophilic and hydrophobic volumes of the molecule). For example, single-chain lipids with large head-groups (e.g. sodium dodecyl sulphate) will form micelles, whereas double-chain lipids with large head-groups (e.g. phosphatidylcholine) will form bilayer vesicles.

Due to their bilayer structure (Fig 20-14) liposomes are able to accommodate water- and lipid-soluble molecules, with water

soluble drugs being incorporated into the aqueous regions of the liposomes, whilst lipophilic drugs can be solubilised within the bilayer of the liposomes, in a similar manner to drug solubilisation within micelles. In addition, molecules can be conjugated or electrostatically bound to the surface of the vesicles. By such means a large range of drugs, nucleic acid systems and protein-based therapies can be incorporated within the liposomes thereby offering protection to the drug, and improved delivery and targeting. Indeed, there is a number of clinically approved liposome products already commercially available (see Clinical applications of liposomes in the below Pharmaceutical Application section).

Whilst phosphatidylcholine is generally the most commonly used lipid, there are a range of lipids available with a choice of head-groups (including cationic, anionic and non-ionic surfactants) and also lipids with a variety of lipid acyl tails; the former can be used to manipulate the surface charge of the vesicles, and the latter influences the permeability of the liposomal bilayer. Both the lipid tail carbon chain length and its degree of saturation influence the phase transition temperature of the lipid. The phase transition temperature is defined as the temperature required to change the physical state of a lipid from an ordered gel phase, where the lipids hydrocarbon tails are fully extended and closely packed, to a disordered 'fluid' liquid crystalline state, where the lipid hydrocarbon tails are disordered. When choosing lipids, generally as the hydrocarbon tail length increases, the phase transition temperature increases (due to the increased van der Waals interactions between the tail region). Similarly, unsaturated cis bonds, prohibit effective van der Waals interactions and reduce the phase transition temperature

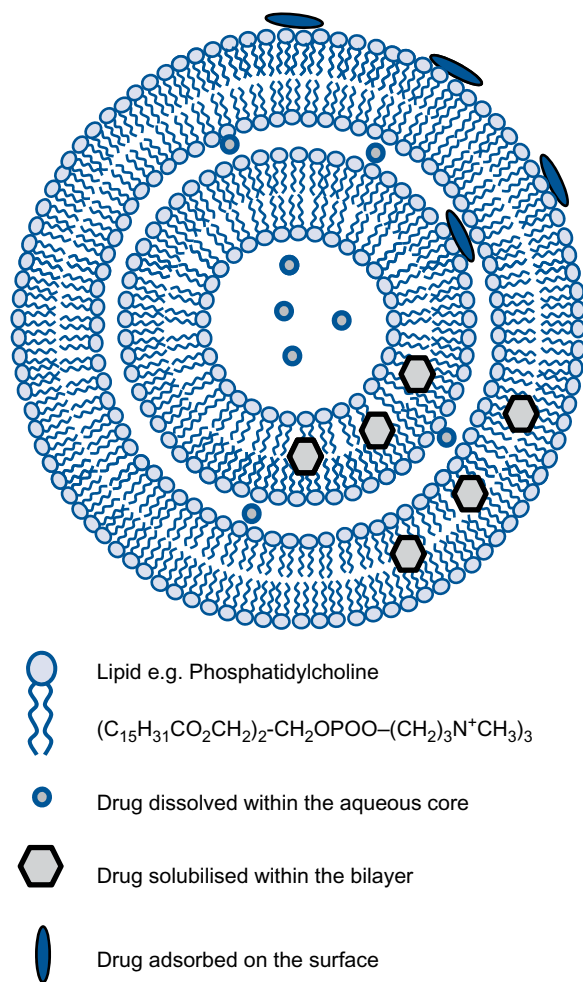


Figure 20-14. Schematic representation of a liposomes with drug entrapped in the aqueous phase, the bilayer or adsorbed to the surface.

of the lipid. These factors will impact on the bilayer permeability, and drug retention of liposomes, with liposomes in their fluid state having more permeable 'leaky' bilayers compared with high-transition temperature lipid bilayers. Therefore at biological temperature, liposomes formulated using a high transition temperature lipid, such as distearoyl phosphatidylcholine (with a transition temperature of 55°C and thus in the ordered state gel state at biological temperature) will show less drug leakage than liposomes prepared from dilauroyl phosphatidyl choline (which has a transition temperature of -1°C and in the disordered fluid state).

A second consideration in the choice of bilayer composition of liposomes is the inclusion of cholesterol. Whilst not able to form bilayers in its own right, due to its amphipathic nature, cholesterol can insert into bilayers and improve the packaging of the lipids within the bilayer, thereby reducing bilayer permeability and drug leakage.

Liposomes are generally classified by their size and number of bilayers: small unilamellar vesicles (SUV) have a single bilayer and are ~ 25 to 100 nm in size, large unilamellar vesicles (LUV) have a single bilayer but are > 100 nm in size and Large multilamellar vesicles (MLV) have several bilayers and are 100 nm to several micrometers in size. Liposomes can be produced by a range of methods, all of which basically involve the dispersion and hydration of the lipid in an aqueous solvent. The size of vesicles formed is generally dictated by the method used to produce the vesicles.

PHARMACEUTICAL APPLICATIONS OF COLLOIDAL SYSTEMS

Colloidal materials are used for a variety of pharmaceutical applications including therapeutic and diagnostic agents, drug delivery systems, and pharmaceutical excipients. With the recent advances in biotechnology and protein engineering, many new drug substances are colloids including recombinant human insulin, interferons, interleukins, and monoclonal antibodies. Drug substances may also be prepared as colloid-sized particles to improve bioavailability or therapeutic activity (e.g., colloidal sulfur).

RADIOACTIVE COLLOIDS

Colloidal dispersions containing radioactive isotopes are being used as diagnostic and therapeutic agents in nuclear medicine. *Colloid gold Au 198* is made by reducing a solution of gold (^{198}Au) chloride either by treatment with ascorbic acid or by heating with an alkaline glucose solution. Gelatin is added as a protective colloid. The particle size ranges from 5 to 50 nm with a mean of 30 nm, and the color of the sol is cherry-red in transmitted light. Violet or blue sols have excessively large particle sizes and should be discarded. Colloidal gold is used as a diagnostic and therapeutic aid. The half-life of ^{198}Au is 2.7 days. *Technetium 99m sulfur colloid* is prepared by reducing sodium pertechnetate $^{99\text{m}}\text{Tc}$ with sodium thiosulfate. The product, a mixture of technetium sulfide and sulfur in the colloidal particle size range, is stabilized with gelatin. It is primarily used in liver, spleen, and bone scanning, and has a half-life of 6.0 hours.

CROSSLINKED POLYMERS

When linear, water-soluble polymers are crosslinked, they swell in water but no longer dissolve. The crosslinks tie the macromolecular chains together by primary covalent bonds, transforming each particle into a single, giant molecule. The water-swollen grains of crosslinked polymers are permeable to low-molecular-weight solutes. Examples of crosslinked polyelectrolytes include the cation-exchange resin sodium polystyrene sulfonate copolymerized with divinylbenzene (used to reduce hyperkalemia by exchanging some of its Na^+ with K^+) and the anion-exchange resins cholestyramine and colestipol hydrochloride (which reduce hypercholesterolemia by binding bile salt anions). Polycarbophil, a lightly crosslinked polymer of acrylic acid, only ionizes and swells in the nearly neutral small intestine, where it absorbs water and reduces the fluidity of diarrheal stool.

COLLOIDAL DELIVERY SYSTEMS

Colloidal delivery systems include micelles, microemulsions, liposomes, parenteral emulsions, microspheres, nanoparticles, and drug-polymer conjugates.

Micelles and Solubilization

Micelles have been used to solubilize poorly water-soluble compounds. For example, *AquaMEPHYTON Injection* (Merck & Co.) contains phytonadione (vitamin K_1) dissolved in the core of micelles of a polyoxyethylated fatty acid derivative. *Fugisone®* (Squibb) is a mixed micellar formulation which is used to solubilise Amphotericin B, an anti-fungal agent used to treat invasive fungal infections such as systemic candidiasis and histoplasmosis. As illustrated in Figure 20-15, the interior of surfactant micelles formed in an aqueous media consists of hydrocarbon tails in a liquid-like, disordered state. Therefore, the micelles resemble miniscule pools of liquid hydrocarbon surrounded by shells of polar head groups. Compounds that are poorly soluble in water but soluble in hydrocarbon solvents can be dissolved inside these micelles, and thereby brought homogeneously into the overall aqueous medium.

Being oleophilic, the solubilized molecules are primarily located in the hydrocarbon core of the micelles (Figure 20-15A).

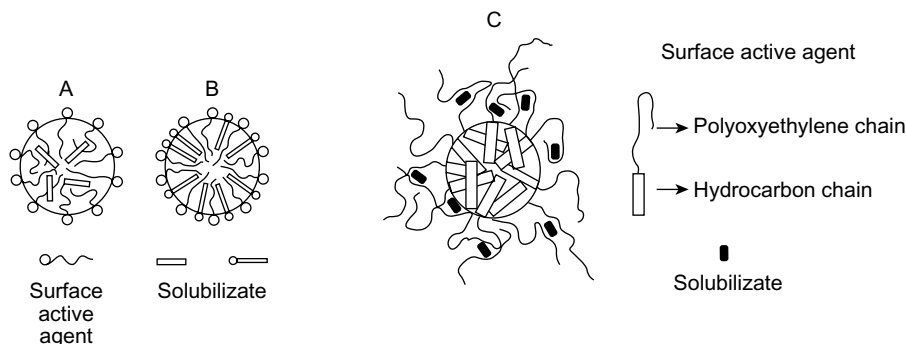


Figure 20-15. The locations of solubilizates in spherical micelles. (A) Ionic surfactant (solubilized molecule has no hydrophilic groups). (B) Ionic surfactant (solubilized molecule has a hydrophilic group). (C) Nonionic surfactant (polar solubilize). (From Shinoda K, Nakagawa T, Tamamushi B-I, Isemura T. *Colloidal Surfactants*. New York: Academic Press, 1963.)

However, many water-insoluble drugs also contain polar functional groups such as hydroxyl, carbonyl, ether, amino, amide, and cyano groups. Upon solubilization, these hydrophilic groups are located among the polar head groups of the surfactant in the periphery of the micelle so as to become hydrated (Figure 20-15B). For instance, when cholesterol or dodecanol is solubilized by sodium lauryl sulfate micelles, their hydroxyl groups penetrate between the sulfate ions and are even bound to them through hydrogen bonds, while their hydrocarbon portions are immersed among the dodecyl tails of the surfactant in the micelle core. Micelles of polyoxyethylated nonionic surfactants consist of an outer shell of hydrated PEG moieties and a core of hydrocarbon moieties. Compounds like phenol, cresol, benzoic acid, salicylic acid, and esters of *p*-hydrobenzoic and *p*-aminobenzoic acids have some solubility in water and oils but considerable solubility in liquids of intermediate polarity such as ethanol, propylene glycol, or aqueous solutions of PEGs. When solubilized by nonionic micelles, these compounds are located in the outer hydrated PEG shell as shown in Figure 20-15C. Since these compounds have hydroxyl or amino groups, they frequently form complexes with the ether oxygens of the surfactant through hydrogen bonding.^{21,24–26,28}

Micellar solubilization is generally nonspecific; any drug that is appreciably soluble in oils can be solubilized. Each compound has a solubilization limit, which depends upon temperature and the nature and concentration of the surfactant. There are two general categories of solubilizates. The first consists of comparatively large, asymmetric and rigid molecules such as steroids and dyes that form crystalline solids. Because of a dissimilarity in structure, these compounds do not blend in with the normal paraffin tails that make up the micellar cores but remain as distinct solute molecules. They are sparingly solubilized by micelles with only a few molecules per micelle at saturation (Table 20-6). The number of carbon atoms in the micellar hydrocarbon core required to solubilize one molecule of a steroid or dye at saturation is of the same order of magnitude as the number of carbon atoms in bulk liquid dodecane or hexadecane required to dissolve one molecule of steroid or dye at saturation.

Because solubilization depends on the presence of micelles, it does not take place below the CMC. Therefore, such solubilization can be used to determine the CMC, particularly when the solubilize is a dye or another compound easy to assay. Plotting the maximum amount of a water-insoluble dye solubilized by an aqueous surfactant, or the absorbance of its saturated solutions, versus the surfactant concentration produces a straight line that intersects the surfactant concentration axis at the CMC. Above the CMC, the amount of solubilized dye is directly proportional to the number of micelles, and therefore, proportional to the overall surfactant concentration. Below the CMC, no solubilization takes place. This is represented by curve *E* in Figure 20-12.

The second category of compounds that may be solubilized are often liquid at room temperature and consist of relatively small, symmetrical, and/or flexible molecules such as many constituents of essential oils. These molecules mix and freely blend in with the hydrocarbon portions of the surfactants in the core of the micelles and therefore become indistinguishable from them. Such compounds are extensively solubilized and in the process usually swell the micelles. They augment the volume of the hydrocarbon core and increase the number of surfactant molecules per micelle. Their solubilization frequently lowers the CMC.

Microemulsions

O/W microemulsions are also formulated as aqueous vehicles for oil-soluble drugs to be administered by the percutaneous, oral, or parenteral routes. Oil-soluble drugs are incorporated into O/W emulsions by dissolving them in the oil phase before emulsification. Similarly, oil-soluble drugs are incorporated into microemulsions by prior dissolution within the oil phase. The advantage of microemulsions as dosage forms as compared to conventional emulsions is their smaller droplet size, which increases drug release, and their superior physical stability. Neoral® (Novartis) is a self-microemulsifying formulation of the immunosuppressant ciclosporin. Ciclosporin has a very low water solubility, and was first formulated in an emulsion preconcentrate incorporated within a soft gelatin capsule, however this emulsion system showed large variation in bioavailability (between 10 and 60% in individual patients). Neoral® which is a microemulsion preconcentrate, leads to the formation of microemulsion droplets < 50 nm and offers less variability in bioavailability.

Liposomes

As mentioned liposomes can be used as vehicles to deliver a wide range of drugs including synthetic drugs, polypeptides, proteins (including enzymes and antibodies), and nucleic acid based therapies. Water-soluble drugs are incorporated into the aqueous phase of the liposomes and oil-soluble drugs can also be solubilized by the hydrocarbon chains of the lipid bilayers. There are a range of clinical products that use liposomes to enhance drug targeting so to reduce the side-effect profile of the drug and improve therapeutic outcomes. These include products used for the treatment of certain cancers (e.g. Daunoxome® (Gilead)), systemic fungal infections (e.g. AmBisome® (Gilead)), vaccines (e.g. Inflaxal V (Janssen-Cilag Ltd)) and macular degeneration (e.g. Visudyne® (Novartis)); the majority of the products are designed for intravenous injection.

After intravenous injection, liposomes (like many particulate delivery systems) interact with blood opsonins, which cause their removal from the blood circulation at rates that are dictated by their vesicle size, lipid composition and surface charge.

Table 20-6. Micellar Solubilization Capacities of Different Surfactants for Estrone

Surfactant	Concentration Range (M)	Temperature (°C)	Moles Surfactant per Moles Solubilized Estrone
Sodium laurate	0.025–0.23	40	91
Sodium oleate	0.002–0.35	40	53
Sodium lauryl sulfate	0.004–0.15	40	71
Sodium cholate	0.09–0.23	20	238
Sodium deoxycholate	0.007–0.36	20	476
Diamyl sodium sulfosuccinate	0.08–0.4	40	833
Dioctyl sodium sulfosuccinate	0.002–0.05	40	196
Tetradecyltrimethylammonium bromide	0.005–0.08	20	45
Hexadecylpyridinium chloride	0.001–0.1	20	32
Polysorbate 20	0.002–0.15	20	161
Polysorbate 60	0.0008–0.11	20	83

(Data from Shinoda K, ed. *Solvent Properties of Surfactant Solutions*. New York: Dekker, 1967.)

Due to this interaction, opsonin-coated liposomes accumulate (via opsonin recognition by appropriate cell receptors) in the mononuclear phagocytic system (MPS) and in particular the liver and spleen. This allows liposomes to be used for passive targeting of such sites. When administered via the intramuscular or subcutaneous route, liposomes end up in the lymphatic system including the local lymph nodes. This provides the opportunity for liposomes to be used for the delivery of vaccines.

This passive targeting is exploited in the liposome formulation Myocet® (Cephalon) - a liposomal formulation of doxorubicin. The liposomes within Myocet are ~180 nm in size and composed of egg phosphatidylcholine and cholesterol. Due to their larger size, the liposomes in Myocet® are rapidly taken by the MPS. This is thought to create a 'MPS depot' which produces a slow release of the drug into the blood circulation, mimicking a slow transfusion of doxorubicin.

To avoid uptake by the MPS, liposomes can be designed to avoid opsonin recognition by appropriate engineering of their surface. This allows liposomes to target sites other than MPS. To achieve this liposomes are coated with hydrophilic polymers, such as polysialic acids and polyethylene glycol (PEG). This hydrophilic polymer coating blocks opsonisation, increasing liposome circulation times, and promotes opportunities for the liposomes to interact with target cells other than those of the MPS. Through such means liposomes can accumulate in pathological sites with leaky vasculature, including tumour sites and sites of inflammation, through the 'Enhanced Permeability and Retention effect' which occurs at sites of inflammation and tumour sites. As these sites the integrity of the endothelial barrier is often disrupted by the presence of endothelial fenestrations (gaps) as large as 200–300 nm. The change in the tumour endothelial barrier is a result of angiogenesis occurring during tumour growth which results in defective hypervascularity (enhanced permeability) and a deficient lymphatic drainage system (enhanced retention). These changes in pathophysiology allow liposomes to escape from the circulation and become trapped in the tumour vasculature, hence passively targeting the PEGylated liposomes to tumour sites. Once trapped, the liposomes then slowly release the drug, resulting in a high local

concentration and a low systemic exposure. This increases its effectiveness and decreases its side effects. Examples of a products using this passive targeting to tumor sites is DaunoXome Liposomal Injection (Gilead), which contains daunorubicin citrate, and Caelyx®/DOXIL Liposome Injection (Janssen Biotech), which contains doxorubicin HCl.

Liposomes are also used to deliver the low solubility drug, Amphotericin B, in the product AmBisome® (Gilead). In this formulation the drug is solubilised in the liposomal bilayer of liposomes ~ 80 nm in diameter and composed of hydrogenated soy phosphatidylcholine, cholesterol, distearoylphosphatidylglycerol and α -tocopherol.

Liposomes can also be formulated to produce a sustained drug release, multi-vesicle liposomes can be used. Both DepoCyte® (Napp) and DepoDur® (Flynn Pharma Ltd) are large multi-vesicle vesicles which achieve sustained drug release through their slow clearance from the administration site and their slow breakdown. After a single epidural injection, DepoDur® can give relief from post-operative pain for up to 48 hours.

Liposomes dispersed in water are subject to degradation via hydrolysis of ester bonds and oxidation of unsaturated acyl chains, aggregation and fusion, as well as leakage of encapsulated drugs. Freeze-drying followed by rehydration and redispersion just before use is commonly used to extend their shelf-life.

Microspheres and Microcapsules

These are essentially spherical particles in the micrometer size range which can be manufactured to be solid or porous (microparticles or microspheres) or they can be hollow in nature (microcapsules). They can be formulated from polymers, lipids and proteins, with polymers such as poly(lactide-co-glycolide) (PLGA) and polylactic acid (PLA) being commonly used. Drugs can be incorporated within the matrix systems, and therefore drug release from the particles is dictated by the degradation rate of the matrix. There are a number of polymeric microparticles approved for clinical use. For example Lupron Depot® (Abbott) is a suspension of PLGA microspheres which are injected subcutaneously and act as a drug depot after administration, giving controlled-release release of leuprolide acetate for

up to 6 months. Decapeptyl® SR (Ipsen Ltd) is a sustained release PLGA microsphere formulation of triptorelin. Triptorelin acetate, a gonadorelin (LHRH) analogue and it is prescribed for the treatment of advanced prostate cancer and endometriosis. The microsphere formulation is injected intramuscularly and the bioavailability of triptorelin is reported as approximately 50%. Nutropin Depot for Injectable Suspension (Genentech), which is administered one to two times per month by subcutaneous injection and slowly releases somatotropin of recombinant DNA origin.

Nanoparticles

Nanoparticles is a term applied to many nano-sized structures and there is a strong overlap between nanotechnology and colloidal science, which is not surprising given that nanotechnology refers to pharmaceutical materials, structures and products that have one or more dimension between approximately 1 and 100 nm. However if we are to refine the description to solid nanoparticles, here we can consider solid constructs in the nanometer range. These can be prepared by two general methods either involving size reduction of particles (e.g. by milling) to within the nanoparticle range or molecular agglomeration (e.g. by precipitation methods) to form nanoparticles. Size reduction is used to prepare drug particles in the nanosize range where no carrier material is added, whilst the molecular agglomeration is more commonly used to prepare nanoparticle carriers in which the drug is loaded.

Considering nanoparticles as drug carriers, there are a range of excipients that can be used to form the carrier including polymers (e.g. poly(lactide-co-glycolide), polylactic acid and polysaccharides) and high melting point lipids. Polymeric nanoparticles can be designed to have PEGylated coatings, similar to liposomes, to facilitate drug targeting to tumour sites.

In addition to polymers and lipids, nanoparticles can also be prepared from proteins. The first commercial produced based on protein nanotechnology is Abraxane® (nabTM-paclitaxel; Abraxis). Abraxane® is currently approved for use in patients who don't respond to combination chemotherapy for metastatic disease or relapse within 6 months of adjuvant chemotherapy. Abraxane® consists of 130 nm particles of albumin-bound paclitaxel. By binding paclitaxel to albumin this overcomes the low solubility of paclitaxel and also promotes active targeting of the paclitaxel to tumour cells. As mentioned, drug targeting of nanoparticles to tumours results from the EPR effect. In addition, the albumin in the nanoparticles may use 'Albumin-activated gp60 pathway' to improve drug delivery: by albumin binding to gp60 albumin receptors found on the surface of vasculature endothelial cells the paclitaxel-albumin is then transported across blood vesicle walls. The binding of albumin to the gp60 receptors, activates the membrane protein Caveolin-1 and subsequently results in the internalisation of the cell membrane and the formation of transcytotic vesicles (known as Caveolae). These caveolae then transport their contents across the endothelial cell cytoplasm, and release their contents into the tumours interstitium. In the case of tumours this transport system is thought to be upregulated. Therefore Abraxane is able to exploit this this albumin-activated gp60 transport mechanism to target the tumour site. After entering the systemic circulation, the albumin-bound paclitaxel can bind to the gp60 albumin receptors and be carried across the endothelial cells via transcytosis. Once the albumin-bound paclitaxel reaches the interstitium, the albumin may bind to an extracellular matrix glycoprotein known as SPARC (secreted protein acid and rich in cysteine) which is also over-expressed in tumour cells. This can trigger the release of the paclitaxel from the albumin, allowing the free drug to diffuse to the nucleus of tumour cells initiating cell death.

Drug-Polymer Conjugates

Polymers have also been used to produce soluble drug-polymer conjugates, which are formed by chemical reactions to produce

covalent bonds between a drug and a polymeric molecule. These polymer conjugates include synthetic polymers as well as biological polymers such as globular and fibrous proteins, antibodies, and polysaccharides. For example, *Mylotarg for Injection* (gemtuzumab ozogamicin; Wyeth Pharmaceuticals) consists of a conjugate between an IgG monoclonal antibody and calicheamicin, a chemotherapeutic agent. These conjugates are administered by intravenous infusion, and the antibody targets the CD33 antigen found on the surface of leukemic blasts and immature normal cells of myelomonocytic lineage, but not on normal hematopoietic stem cells. The antibody-drug conjugate is believed to have a lysosomotropic mechanism of action; in other words, the antibody-antigen complex is internalized into the cell through endocytosis and entrapped within endosomal compartments. These endosomes merge with primary lysosomes to form secondary lysosomes, where the drug-antibody bonds are presumably broken by the acidic environment or cleaved by certain lysosomal enzymes. Once cleaved from the antibody, the drug is able to diffuse out of the lysosome and exert its pharmacological effect.

In-111 and Y-90 Zevalin (Spectrum Pharmaceuticals) consists of a conjugate between an ibritumomab antibody (a murine IgG) and either an indium-111 or an yttrium-90 radioisotope. These conjugates are administered by intravenous infusion and attach to the CD20 antigen found on the surface of normal and malignant B-lymphocytes. However, the antibody-antigen complex is not internalized. The attachment of the antibody induces cell apoptosis, and the close proximity of the radioisotope creates free radicals that damage nearby cells.

PEGIntron Powder for Injection (Schering) also consists of a conjugate. In this case, interferon alfa-2b is conjugated with a PEG derivative. The conjugates are injected subcutaneously. The PEG derivative prevents detection and therefore decreases the clearance of interferon alfa-2b from the body.

The actively targeted conjugates just described can be contrasted to passively targeted macromolecular conjugates for solid-tumor tissue. Passively targeted macromolecular conjugates have shown preferential accumulation in solid tumors because of the Enhanced Permeability and Retention (EPR) effect. The preferential accumulation reduces systemic toxicity by reducing damage to noncancerous organs. In addition, the EPR effect is more effective for macromolecules greater than 40 kDa but negligible for smaller molecules that are cleared more rapidly from the tumor interstitium.²⁹

In addition to the PEGylated proteins, OPAXIOTM (Cell Therapeutics, Inc.) is currently in clinical trials. In this polymer-drug conjugate, paclitaxel is conjugated to poly-L-glutamic acid (PGA). By conjugation of paclitaxel to the water soluble PGA, this improves the solubility of paclitaxel, and it can be infused into the body without the addition of solvents. This conjugate has a high drug content (~37% w/w) and the conjugate can passively target tumour sites via the EPR effect. The drug is then released intracellularly via degradation of PGA by lysosomal proteases and the ester linker is degraded by esterases or acid hydrolysis.

EXCIPIENTS

Most of the excipients, adjuvants, or nontherapeutic ingredients of dosage forms listed below are monographs in the NF or USP.

Colloids are also used as pharmaceutical excipients for a variety of purposes including thickening agents. Colloidal thickening agents or viscosity builders belong to four chemical categories.

Semisynthetic cellulose derivatives include methylcellulose, carboxymethylcellulose sodium, hydroxypropyl methylcellulose, and hydroxypropyl cellulose. *Natural polymers* include acacia, tragacanth, xanthan gum, sodium alginate, and carrageenan. *Synthetic polymers* include carbomer, a copolymer of acrylic acid; poloxamer, a block copolymer of ethylene oxide and propylene oxide; polyvinyl alcohol; and povidone (polyvinylpyrrolidone). *Particulate colloids* include bentonite,

colloidal silicon dioxide, and microcrystalline cellulose. These viscosity builders may be used to decrease the dissolution rate of controlled-release dosage forms, to decrease the sedimentation or creaming rates of dispersed systems, to improve the taste-masking abilities of liquid vehicles, and to provide consistency to ointments. Many of the water-soluble viscosity builders mentioned above are surface-active and are also used as emulsifying and suspending agents. Even particulate colloids are used to stabilize emulsions and suspensions.

Colloidal silicon dioxide is a white powder consisting of sub-microscopic spherical particles of fairly uniform size in the range of 5–50 nm or higher. It is used to thicken liquid dosage forms and in tablets. The surface of colloidal silicon dioxide particles contains siloxane (Si-O-Si) and silanol (Si-OH) groups. When colloidal silicon dioxide powder is dispersed in nonpolar liquids, the particles tend to adhere to one another through hydrogen bonds between these surface groups. The spherical particles of finer grade colloidal silicon dioxide are linked together into short chain-like aggregates as shown in Figure 20-4. This creates loose three-dimensional networks that increase the viscosity of the liquid vehicles even at levels as low as a few percent. The hydrogen-bonded structures are torn apart by stirring but rebuilt while at rest, conferring a thixotropic nature to the thickened liquids.

Aerosil 200 is the grade of colloidal silicon dioxide most widely used as a pharmaceutical adjuvant. Its primary spheres, which are extensively sintered together, have an average diameter of 12 nm. At levels of 8 to 10 percent, it thickens liquids of low polarity such as vegetable and mineral oils to the consistency of ointments, imparting considerable yield values to them. Hydrogen-bonding liquids such as alcohols and water solvate the silica spheres, thereby reducing the hydrogen bonding between particles. Therefore, the higher silica levels of 12 to 18 percent or more are required to gel these solvents.

The grades that consist of relatively large and unattached spherical particles, such as those in Figure 20-3, are less efficient thickening agents because they lack the high specific surface area and asymmetry of the finer grades. The consistency of ointments thickened with colloidal silicon dioxide is not appreciably reduced at higher temperatures. Incorporation of colloidal silicon dioxide into ointments and pastes, such as those of zinc oxide, also reduces the syneresis or *bleeding* of the liquid vehicles.

Colloidal silicon dioxide is also used in dry dosage forms. The spherical particles are nonporous and have a density of 2.13 g/cm³. However, the bulk density of their powder is a mere 0.05 g/cm³. Because the powder is extremely light, it is frequently used to increase the fluffiness or bulk volume of powder formulations. In addition, the high porosity of colloidal silica enables it to absorb a variety of liquids from fluid fragrances to viscous tars, transforming them into free-flowing powders that can be incorporated into tablets or capsules. The porosity in colloidal silicon dioxide is due entirely to the enormous void space between the particles, which themselves are solid. When these ultrafine particles are incorporated at levels as low as 0.1 to 0.5 percent into a powder consisting of coarse particles or granules, they coat the surface of the granules and act as tiny ball bearings and spacers. This improves the flowability of the powder and eliminates caking, which is important in tableting. In addition, colloidal silicon dioxide is used to improve tablet disintegration. It is also used as a glidant and as a moisture absorber.

Microcrystalline cellulose is manufactured by controlled hydrolysis of purified native cellulose, which dissolves the amorphous matrix but leaves the crystallites intact. The needle- or rod-shaped crystallites act as suspending agents in water, producing thixotropic structured vehicles. At concentrations of about 15 percent, the cellulose microcrystals gel water to an ointment consistency by swelling and producing a continuous network of rods that extends throughout the entire vehicle. Attraction between the elongated particles is presumably due to flocculation in the secondary minimum. Treatment of the

microcrystalline mass with sodium carboxymethylcellulose facilitates its disintegration into primary needle-shaped particles and enhances their thickening action.

Gelatinous precipitates of inorganic hydrophilic compounds such as *aluminum hydroxide gel*, *aluminum phosphate gel*, and *magnesium hydroxide* consist of coarse flocs produced by the agglomeration of colloidal particles formed in the initial stage of precipitation. They possess large internal surface areas, which is one of the reasons why the first two are used as substrates for adsorbed vaccines and toxoids. *Alumina and magnesia oral suspensions*, a mixture of gelatinous precipitated aluminum and magnesium hydroxides, as well as aluminum hydroxide gel, are used as antacids.

Gelation is used to manufacture the following suppository bases: glycerinated gelatin suppositories; glycerin suppositories (in which glycerin is solidified with sodium stearate that crystallizes out as a network of needles upon cooling the hot solution); and PEG suppositories (in which low-molecular-weight liquid PEGs, such as PEG 400, are stiffened by high-molecular-weight PEGs such as 3350 or 4000, which are waxy solids). A pharmaceutical application of gelation in a nonaqueous medium is the manufacture of *Plastibase* or *Jelene* (Squibb), which is prepared using 5 percent of a low-molecular-weight polyethylene and 95 percent of mineral oil. The polymer is soluble in mineral oil above 90°C, which is close to its melting point. When the solution is cooled below 90°C, the polymer precipitates and causes gelation. The mineral oil is immobilized in the network of entangled, adhering, insoluble polyethylene chains, which probably even associate into small crystalline regions. Unlike petrolatum, this gel can be heated to about 60°C without any substantial loss in consistency.

Crospovidone and croscarmellose sodium are crosslinked povidone and carboxymethylcellulose sodium, respectively. These crosslinked polymers swell rapidly and extensively in aqueous media and therefore are frequently used as tablet disintegrants. Starch performs the same function; its major constituent, amylopectin, is highly branched and insoluble in water but swells considerably. Because crosslinked hydrophilic polymers swell extensively without dissolution, they are also used as matrices for controlled-release dosage forms.

REFERENCES

1. Shaw DJ. *Introduction to Colloid and Surface Chemistry*, 4th edn. Oxford: Butterworth-Heinemann, 1992.
2. Hiemenz PC, Rajagopalan R. *Principles of Colloid and Surface Chemistry*, 3rd edn. New York: Dekker, 1997.
3. van Olphen H. *An Introduction to Clay Colloid Chemistry*, 2nd edn. New York: Wiley, 1977.
4. Schott H, Martin AN. In: Dittert LW, ed. *American Pharmacy*, 7th edn. Philadelphia: JB Lippincott, 1974.
5. Ross S, Morrison ID. *Colloidal Systems and Interfaces*. New York: Wiley, 1988.
6. Ross DA, et al. The determination of the mean and standard deviation of the size distribution of a colloidal suspension of submicron particles using the fiber optic Doppler anemometer, FODA. *J Colloid Interface Sci* 1978;64: 533–542.
7. Morawetz H. *Macromolecules in Solution*, 2nd edn. New York: Wiley-Interscience, 1975.
8. Veis A. *The Macromolecular Chemistry of Gelatin*. New York: Academic, 1964.
9. Ward AG, Courts A, eds. *The Science and Technology of Gelatin*. New York: Academic Press, 1977: Chapter 6.
10. Parks GA. The isoelectric point of solid oxides, solid hydroxides, and aqueous hydroxo complex systems. *Chem Rev* 1965; 65:177–198.
11. Schott H. Relationship between zero point of charge and solubility product for hydroxide of polyvalent cations. *J Pharm Sci* 1977; 66:1548–1550.
12. Adamson AW. *Physical Chemistry of Surfaces*, 5th edn. New York: Wiley, 1990.

13. Sonntag H, Strenge K. *Coagulation and Stability of Disperse Systems*. New York: Halstead, 1972.
14. Hunter RJ. *Zeta Potential in Colloid Science*. New York: Academic Press, 1981.
15. Davies JT, Rideal EK. *Interfacial Phenomena*, 2nd edn. New York: Academic Press, 1963.
16. *Primer on Capillary Electrophoresis*, vols I–VII. Fullerton CA: Beckman Instruments, 1993–1995.
17. von Weimarn PP. In: Alexander J, ed. *Colloid Chemistry*, vol I. New York: Chemical Catalog Co. (Reinhold), 1926.
18. Overbeek JTG. Monodisperse colloidal systems, fascinating and useful. *Adv Colloid Interface Sci* 1982; 15:251–277.
19. LaMer VK, Dinegar RH. Theory, production and mechanism of formation of monodispersed hydrosols *J Am Chem Soc* 1950; 72: 4847–4854.
20. Matijevic E. Monodispersed metal (hydrous) oxides—a fascinating field of colloid science. *Acc Chem Res* 1981; 14:22–29.
21. Schick MJ, ed. *Nonionic Surfactants—Physical Chemistry*. New York: Dekker, 1987.
22. Vincent B. The effect of adsorbed polymers on the stability of dispersions. *Adv Colloid Interface Sci* 1974; 4: 193–277.
23. Schott H. Saturation adsorption at interfaces of surfactant solutions. *J Pharm Sci* 1980; 69: 852–854.
24. Shinoda K *et al.*. *Colloidal Surfactants*. New York: Academic Press, 1963.
25. Attwood D, Florence AT. *Surfactant Systems*. London: Chapman & Hall, 1983.
26. Rosen MJ. *Surfactants and Interfacial Phenomena*, 2nd edn. New York: Wiley, 1989.
27. Mukerjee P, Mysels KJ. *Critical Micelle Concentrations of Aqueous Surfactant Systems*, NSRDS-NBS 36. Washington DC, National Bureau of Standards, 1971.
28. Shinoda K, ed. *Solvent Properties of Surfactant Solutions*. New York: Dekker, 1967.
29. Maeda H *et al.*. Tumor vascular permeability and the EPR effect in macromolecular therapeutics: a review. *J Control Rel* 2000; 65: 271–284.

Chapter 21

Coarse Dispersions

James Swarbrick, DSc, PhD; Joseph T. Rubino, PhD, RPh and Orapin P. Rubino, PhD

THE DISPERSION STEP 371

INTERFACIAL PROPERTIES 371

SUSPENSIONS 372

SETTLING AND ITS CONTROL 373

FORMULATION OF SUSPENSIONS 374

EMULSIONS 378

EMULSION TYPE AND MEANS OF DETECTION 379

FORMATION AND BREAKDOWN OF DISPERSED LIQUID DROPLETS 379

EMULSIFYING AGENT 380

PREPARATION OF EMULSIONS 385

STABILITY OF EMULSIONS 388

BIOAVAILABILITY FROM COARSE DISPERSIONS 389

This chapter includes the formation of suspensions and emulsions and the factors that influence their stability and performance as dosage forms. For the purpose of the present discussion, a dispersed system, or dispersion, will be regarded as a two-phase system in which one phase is distributed as particles or droplets in the second, or continuous, phase. In these systems, the dispersed phase frequently is referred to as the discontinuous or internal phase, and the continuous phase is called the external phase or dispersion medium. Discussion will be restricted to those solid-liquid and liquid-liquid dispersions that are of pharmaceutical significance, namely, suspensions and emulsions. However, more complicated phase systems (e.g., a combination of liquid and liquid crystalline phases) can exist in emulsions. This situation will be discussed in the section dealing with emulsions.

All dispersions may be classified into three groups based on the size of the dispersed particles. Chapter 20 deals with one such group—colloidal dispersions—in which the size of the dispersed particles is in the range of approximately 1 nm to 0.5 μm . Molecular dispersions, the second group in this classification, are discussed in Chapter 19. The third group, consisting of *coarse dispersions* in which the particle size exceeds 0.5 μm , is the subject of this chapter. Knowledge of coarse dispersions is essential for the preparation of both pharmaceutical suspensions (solid-liquid dispersions) and emulsions (liquid-liquid dispersions).

Table 21-1 contains examples of pharmaceutical suspensions and emulsions. Pharmaceutical coarse dispersions are administered via several routes of administration and include many different types of active ingredients.

THE DISPERSION STEP

The pharmaceutical formulator is concerned primarily with producing a smooth, uniform, easily flowing (pouring or spreading) suspension or emulsion in which dispersion of particles can be effected with minimum expenditure of energy.

In preparing suspensions, particle-particle attractive forces need to be overcome by the high shearing action of such devices as the colloid mill, or by use of surface-active agents. The latter greatly facilitate wetting of lyophobic powders and assist in the removal of surface air that shearing alone may not remove; thus, the clumping tendency of the particles is reduced. Moreover, lowering of the surface free energy by the adsorption of these agents directly reduces the thermodynamic driving force opposing dispersion of the particles.

In emulsification, shear rates are frequently necessary for dispersion of the internal phase into fine droplets. The shear forces are opposed by forces operating to resist distortion and subsequent breakup of the droplets. Again surface-active agents

help greatly by lowering interfacial tension, which is the primary reversible component resisting droplet distortion. Surface-active agents also may play an important role in determining whether an oil-in-water (O/W) or a water-in-oil (W/O) emulsion preferentially survives the shearing action.

Once the process of dispersion begins there develops simultaneously a tendency for the system to revert to an energetically more stable state, manifested by flocculation, coalescence, sedimentation, crystal growth, and caking phenomena. If these physical changes are not inhibited or controlled, successful dispersions will not be achieved or will be lost during shelf-life.

INTERFACIAL PROPERTIES

Because suspensions and emulsions are dispersions of one phase within another, the process of dispersion creates a tremendous increase in interfacial area between the dispersed particles or droplets and the dispersion medium. When considering the interfacial properties of dispersed particles, two factors must be taken into account, regardless of whether the dispersed phase is solid or liquid. The first relates to an increase in the free energy of the surface as the particle size is reduced and the specific surface increased. The second deals with the presence of an electrical charge on the surface of the dispersed particles.

SURFACE FREE ENERGY

When solid and liquid materials are reduced in size, they tend to agglomerate or stick together. This clumping, which can occur either in an air or liquid medium, is an attempt by the particles to reduce the excess free energy of the system. The increase in surface free energy is related to the increase in surface area produced when the mean particle size is reduced. It may be expressed as

$$\Delta F = \gamma \Delta A \quad (1)$$

where ΔF is the increase in surface free energy in ergs, ΔA is the increase in surface area in cm^2 , and γ is the interfacial tension in dyn/cm , between the dispersed particle or droplet and the dispersion medium. The smaller ΔF is, the more thermodynamically stable is the suspension of particles. A reduction in ΔF is effected often by the addition of a wetting agent (discussed in Chapter 19), which is adsorbed at the interface between the particle and the vehicle, thereby reducing the interfacial tension. This causes the particles to remain dispersed and settle relatively slowly. Unfortunately, in solid-liquid suspensions, the particles can form a hard cake at the bottom of the container when they eventually settle. Such a sediment, which can be extremely difficult to redisperse, can lead to dosing errors when the product is administered to the patient.

Table 21-1. Examples of Pharmaceutical/Cosmetic Suspension and Emulsion Products

Suspensions		Emulsions	
Name	Route of Administration	Name	Route of Administration
Milk of Magnesia® (Magnesium Hydroxide)	Oral	Mineral Oil	Oral
Nystatin Oral Suspension, USP	Oral	Restasis® (cyclosporine ophthalmic emulsion)	Eye
Cortisporin® Ophthalmic & Otic suspensions (neomycin/polymyxin, bacitracin, hydrocortisone)	Ear, eye	Intralipid®, Liposyn® (intravenous fat emulsion)	Intravenous
Ibuprofen oral suspension, USP	Oral	Lipomul®	Oral
Tobradex® (tobramycin, dexamethasone)	Eye	Simethicone emulsion, USP	Oral, -manufacturing aid
Tussionex® (chlorpheniramine/hydrocodone sustained release)	Oral	Propofol injectable emulsion, USP	Intravenous
Prednisolone acetate ophthalmic suspension, USP	Eye	LubriDerm® (emollient)	Topical
NPH Insulin	subcutaneous	Lidocaine/prilocaine cream (EMLA®)	topical
Sterile Procaine Penicillin G Injectable Suspension, USP	intramuscular	Diazemuls® (diazepam)	intravenous
Griseofulvin oral suspension, USP	oral		
Phenytoin oral suspension, USP	oral		
Sulfamethoxazole/trimethoprim oral suspension	oral		

SURFACE POTENTIAL

As discussed in Chapter 19, both attractive and repulsive forces exist between particles in a liquid medium. The balance between these opposing forces determines whether two particles approaching each other actually make contact or are repulsed at a certain distance of separation. Although much of the theoretical work on electrical surface potentials has been carried out on lyophobic colloids, the theories developed in this area have been applied to suspensions and emulsions.

SUSPENSIONS

A *pharmaceutical suspension* may be defined as a coarse dispersion containing finely divided insoluble material suspended in a liquid medium. Because some products occasionally are prepared in a dry form to be placed in suspension at the time of dispensing by the addition of an appropriate liquid vehicle, this definition is extended to include these products.

Suspension dosage forms are given by the oral route, injected intramuscularly or subcutaneously, instilled intranasally, inhaled into the lungs, applied to the skin as topical preparations, or used for ophthalmic or otic purposes in the eye or ear, respectively. They are an important class of dosage form which can offer several advantages. Suspensions offer an alternative oral dosage form for patients who cannot swallow a tablet or capsule such as pediatric and geriatric patients. Oral antibiotics, analgesic and antipyretic drugs are commonly administered as suspensions to these groups of patients. Suspensions are often used to deliver poorly water-soluble drugs which cannot be formulated as aqueous solutions. In early drug development studies of potentially new drug candidates, suspension dosage forms are frequently used to deliver drugs to animals, particularly when the drug is given orally. In addition, drugs that have an unpleasant taste may preferably be formulated as a suspension to reduce interaction of drug with taste receptors in the mouth. Because suspended drug must undergo a dissolution step prior to crossing biological membranes, suspensions offer a way to provide sustained release of drug by parenteral, topical, and oral routes of administration.

There are certain criteria that a well-formulated suspension should meet. The dispersed particles should be of such a size that they do not settle rapidly in the container. However, in the event that sedimentation does occur, the sediment must not form a hard cake. Rather, it should be capable of redispersion with a minimum of effort on the part of the patient. Finally, the product should be easy to pour, have a pleasant taste, and be resistant to microbial attack.

The three major concerns associated with suspensions are:

1. ensuring adequate dispersion of the particles in the vehicle.
2. minimizing settling of the dispersed particles.
3. preventing caking of these particles when a sediment forms.

Much of the following discussion will deal with the factors that influence these processes and the ways in which settling and caking can be minimized.

FLOCCULATION AND DEFLOCCULATION

Zeta potential, ζ , is a measurable indication of the potential existing at the surface of a particle. When ζ is relatively high (25 mV or more), the repulsive forces between two particles exceed the attractive London forces. Accordingly, the particles are dispersed and are said to be *deflocculated*. Even when brought close together by random motion or agitation, deflocculated particles resist collision due to their high surface potential.

The addition of a preferentially adsorbed ion whose charge is opposite in sign to that on the particle leads to a progressive lowering of ζ . At some concentration of the added ion, the electrical forces of repulsion are lowered sufficiently and the forces of attraction predominate. Under these conditions the particles may approach each other more closely and form loose aggregates, termed *flocs*. Such a system is said to be *flocculated*.

Some workers restrict the term "flocculation" to the aggregation brought about by chemical bridging; aggregation involving a reduction of repulsive potential at the double layer is referred to as *coagulation*. Other workers regard flocculation as aggregation in the secondary minimum of the potential energy curve of

two interacting particles and coagulation as aggregation in the primary minimum. In the present chapter the term *flocculation* is used for all aggregation processes, irrespective of mechanism.

The continued addition of the flocculating agent can reverse the above process, if the zeta potential increases sufficiently in the opposite direction. Thus, the adsorption of anions onto positively charged, deflocculated particles in suspension will lead to flocculation. The addition of more anions eventually can generate a net negative charge on the particles. When this has achieved the required magnitude, deflocculation may occur again. The only difference from the starting system is that the net charge on the particles in their deflocculated state is negative rather than positive. For further understanding of the processes involved with the electrical double layer, the reader is directed to the chapter on "Colloids." Some of the major differences between suspensions of flocculated and deflocculated particles are presented in Table 21-2.

FLOCCULATION KINETICS

The rate at which flocculation occurs is a consideration in the stability of suspended dispersions. Whether flocculation is judged to be rapid or slow depends on the presence of a repulsive barrier between adjacent particles. In the absence of such a barrier, and for a monodispersed system, rapid flocculation occurs at a rate given by the Smoluchowski equation.

$$\delta N / \delta t = -4\pi D R N^2 \quad (2)$$

where $\delta N / \delta t$ is the disappearance rate of particles/mL, R is the distance between the centers of the two particles in contact, N is the number of particles per mL, and D is the diffusion coefficient. Under these conditions the rate is proportional to the square of the particle concentration. The presence or absence of an energy barrier is influenced strongly by the type and concentration of any electrolyte present. When an energy barrier does exist between adjacent particles, the flocculation rate likely will be much smaller than predicted by Equation 2.

SETTLING AND ITS CONTROL

To control the settling of dispersed material in suspension, the pharmacist must be aware of those physical factors that will affect the rate of sedimentation of particles under ideal and nonideal conditions. Also important are the various coefficients used to express the amount of flocculation in the system and the effect flocculation will have on the structure and volume of the sediment.

SEDIMENTATION RATE

The rate at which particles in a suspension sediment is related to their size and density and the viscosity of the suspension

medium. Brownian movement may exert a significant effect, as will the absence or presence of flocculation in the system.

Stokes' Law

The velocity of sedimentation of a uniform collection of spherical particles is governed by *Stokes' law*, expressed as

$$v = \frac{2r^2(\rho_1 - \rho_2)g}{9\eta} \quad (3)$$

where v is the terminal velocity in cm/sec, r is the radius of the particles in cm, ρ_1 and ρ_2 are the densities (g/cm³) of the dispersed phase and the dispersion medium, respectively, g is the acceleration due to gravity (980.7 cm/sec²), and η is the Newtonian viscosity of the dispersion medium in poises (g/cm sec). Stokes' law holds only if the downward motion of the particles is not sufficiently rapid to cause turbulence. Micelles and small phospholipid vesicles do not settle unless they are subjected to centrifugation.

Although conditions in a pharmaceutical suspension are not in strict accord with those laid down for Stokes' law, Equation 3 provides those factors that can be expected to influence the rate of settling. Thus, sedimentation velocity will be reduced by decreasing the particle size, provided that the particles are kept in a deflocculated state. The rate of sedimentation will be an inverse function of the viscosity of the dispersion medium.

However, too high a viscosity is undesirable, especially if the suspending medium is Newtonian rather than shear-thinning (see Chapter 22), because it then becomes difficult to redisperse material that has settled. It also may be inconvenient to remove a viscous suspension from its container. When the size of particles undergoing sedimentation is reduced to approximately 2 μ m, random Brownian movement is observed and the rate of sedimentation departs markedly from the theoretical predictions of Stokes' law. The actual size at which Brownian movement becomes significant depends on the density of the particle as well as the viscosity of the dispersion medium.

Effect of Flocculation

In a deflocculated system containing a distribution of particle sizes, the larger particles naturally settle faster than the smaller particles. The very small particles remain suspended for a considerable length of time, with the result that no distinct boundary is formed between the supernatant and the sediment. Even when a sediment becomes discernible, the supernatant remains cloudy.

When the same system is flocculated (in a manner to be discussed later), two effects are immediately apparent. First, the flocs tend to fall together, so a distinct boundary between the sediment and the supernatant is readily observed; second, the supernatant is clear, showing that the very fine particles have been incorporated into the flocs. The initial rate of settling

Table 21-2. Relative Properties of Flocculated and Deflocculated Particles in Suspension

Deflocculated	Flocculated
1. Particles exist in suspension as separate entities.	1. Particles form loose aggregates.
2. Rate of sedimentation is slow, as each particle settles separately and particle size is minimal.	2. Rate of sedimentation is high, as particles settle as a floc, which is a collection of particles.
3. A sediment is formed slowly.	3. A sediment is formed rapidly.
4. The sediment eventually becomes very closely packed, due to weight of upper layers of sedimenting material. Repulsive forces between particles are overcome and a hard cake is formed that is difficult, if not impossible, to redisperse.	4. The sediment is packed loosely and possesses a scaffold-like structure. Particles do not bond tightly to each other and a hard, dense cake does not form. The sediment is easy to redisperse, so as to reform the original suspension.
5. The suspension has a pleasing appearance, as the suspended material remains suspended for a relatively long time. The supernate also remains cloudy, even when settling is apparent.	5. The suspension is somewhat unsightly, due to rapid sedimentation and the presence of an obvious, clear supernatant region. This can be minimized if the volume of sediment is made large. Ideally, volume of sediment should encompass the volume of the suspension.

in flocculated systems is determined by the size of the flocs and the porosity of the aggregated mass. Under these circumstances it is perhaps better to use the term *subsidence*, rather than sedimentation.

QUANTITATIVE EXPRESSIONS OF SEDIMENTATION AND FLOCCULATION

Frequently, the pharmacist needs to assess a formulation in terms of the amount of flocculation in the suspension and compare this with that found in other formulations. The two parameters commonly used for this purpose are outlined below.

Sedimentation Volume

The *sedimentation volume*, F , is the ratio of the equilibrium volume of the sediment, V_u , to the total volume of the suspension, V_0 . Thus,

$$F = V_u/V_0 \quad (4)$$

As the volume of suspension that appears occupied by the sediment increases, the value of F , which normally ranges from nearly 0 to 1, increases. In the system where $F = 0.75$, for example, 75% of the total volume in the container is apparently occupied by the loose, porous flocs forming the sediment. This is illustrated in Figure 21-1. When $F = 1$, no sediment is apparent even though the system is flocculated. This is the ideal suspension for, under these conditions, no sedimentation will occur. Caking also will be absent. Furthermore, the suspension is esthetically pleasing, there being no visible, clear supernatant.

Degree of Flocculation

A better parameter for comparing flocculated systems is the *degree of flocculation*, β , which relates the sedimentation volume of the flocculated suspension, F , to the sedimentation volume of the suspension when deflocculated, F_∞ . It is expressed as

$$\beta = F/F_\infty \quad (5)$$

The degree of flocculation is, therefore, an expression of the increased sediment volume resulting from flocculation. If, for example, β has a value of 5.0 (see Figure 21-1), this means that the volume of sediment in the flocculated system is five times that in the deflocculated state. If a second flocculated formulation results in a value for β of say 6.5, this latter suspension obviously is preferred, if the aim is to produce as flocculated a product as possible. As the degree of flocculation in the system decreases, β approaches unity, the theoretical minimum value.

FORMULATION OF SUSPENSIONS

The formulation of a suspension possessing optimal physical stability depends on whether the particles in suspension are to be flocculated or to remain deflocculated. One approach involves use of a structured vehicle to keep deflocculated particles in suspension; a second depends on controlled flocculation as a means of preventing cake formation. A third, a combination of

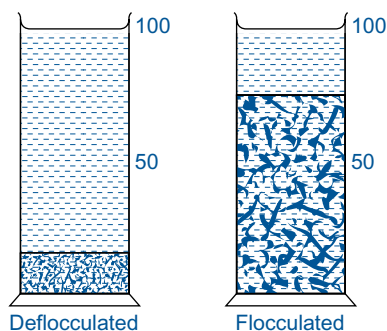


Figure 21-1. Sedimentation parameters of suspensions. Deflocculated suspension: $F_\infty = 0.15$. Flocculated suspension: $F = 0.75$; $\beta = 5.0$.

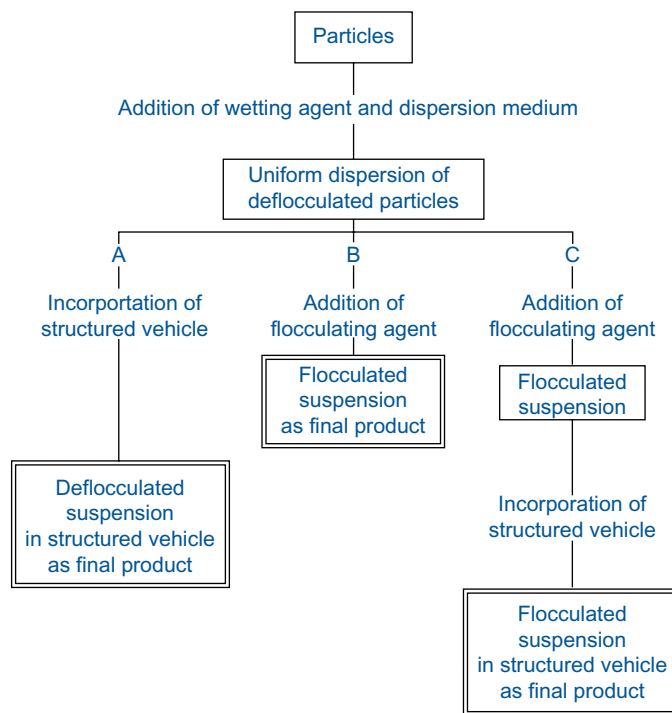


Figure 21-2. Alternative approaches to the formulation of suspensions.

the two previous methods, results in a product with optimum stability. The various schemes are illustrated in Figure 21-2.

Dispersion of Particles

The dispersion step has been discussed earlier in this chapter. Surface-active agents commonly are used as wetting agents; maximum efficiency is obtained when the HLB value lies within the range of 7 to 9. A concentrated solution of the wetting agent in the vehicle may be used to prepare a slurry of the powder; this is diluted with the required amount of vehicle. Alcohol and glycerin may be used sometimes in the initial stages to disperse the particles, thereby allowing the vehicle to penetrate the powder mass.

Only the minimum amount of wetting agent should be used, compatible with producing an adequate dispersion of the particles. Excessive amounts may lead to foaming or impart an undesirable taste or odor to the product. Invariably, as a result of wetting, the dispersed particles in the vehicle are deflocculated.

Structured Vehicles

Structured vehicles are generally aqueous solutions of polymeric materials, such as the hydrocolloids, that are usually negatively charged in aqueous solution. Typical examples are methylcellulose, carboxymethylcellulose, bentonite, and carbomer. The concentration employed will depend on the consistency desired for the suspension that, in turn, will relate to the size and density of the suspended particles. They function as viscosity-imparting suspending agents and, as such, reduce the rate of sedimentation of dispersed particles.

The rheological properties of suspending agents are considered elsewhere (Chapter 22). Ideally, these form pseudo-plastic or plastic systems that undergo shear-thinning. Some degree of thixotropy is also desirable. Non-Newtonian materials of this type are preferred over Newtonian systems because, if the particles eventually settle to the bottom of the container, their redispersion is facilitated by the vehicle thinning when shaken. When the shaking is discontinued, the vehicle regains its original consistency and the redispersed particles are held suspended. This process of redispersion, facilitated by a shear-thinning vehicle, presupposes that the deflocculated particles have not

yet formed a cake. If sedimentation and packing have proceeded to the point where considerable caking has occurred, redispersion is virtually impossible.

Controlled Flocculation

When using the controlled flocculation approach (see Figure 21-2B and C), the formulator takes the deflocculated, wetted dispersion of particles and attempts to bring about flocculation by the addition of a flocculating agent; most commonly, these are electrolytes, polymers, or surfactants. The aim is to *control* flocculation by adding that amount of flocculating agent that results in the maximum sedimentation volume.

Flocculation using Electrolytes

Electrolytes are probably the most widely used flocculating agents. They act by reducing the electrical forces of repulsion between particles, thereby allowing the particles to form the loose flocs so characteristic of a flocculated suspension. As the ability of particles to come together and form a floc depends on their surface charge, zeta potential measurements on the suspension, as an electrolyte is added, provide valuable information as to the extent of flocculation in the system.

This principle is illustrated by reference to the following example, taken from the work of Haines and Martin.² Particles of sulfamerazine in water bear a negative charge. The serial addition of a suitable electrolyte, such as aluminum chloride, causes a progressive reduction in the zeta potential of the particles. This is due to the preferential adsorption of the trivalent aluminum cation. Eventually, the zeta potential will reach zero and then become positive as the addition of AlCl_3 is continued.

If sedimentation studies are run simultaneously on suspensions containing the same range of AlCl_3 concentrations, a relationship is observed (Figure 21-3) between the sedimentation volume F , the presence or absence of caking, and the zeta potential of the particles. To obtain a flocculated, noncaking suspension with the maximum sedimentation volume, the zeta potential must be controlled so as to lie within a certain range (generally less than 25 mV). This is achieved by the judicious use of an electrolyte. A comparable situation is observed when a negative ion such as PO_4^{3-} is added to a suspension of positively charged particles such as bismuth subnitrate.

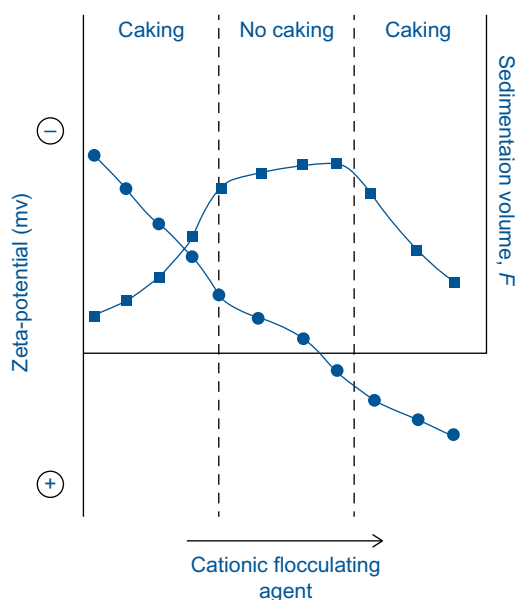


Figure 21-3. Typical relationship between caking, zeta potential, and sedimentation volume, as a positively charged flocculating agent is added to a suspension of negatively charged particles.: zeta potential.: sedimentation volume.

Work by Matthews and Rhodes³⁻⁵ involving both experimental and theoretical studies has confirmed the formulation principles proposed by Martin and Haines. The suspensions used by Matthews and Rhodes contained 2.5% w/v of griseofulvin as a fine powder together with the anionic surfactant sodium dioxyethylated dodecyl sulfate (10^{-3} molar) as a wetting agent. Increasing concentrations of aluminum chloride were added and the sedimentation height (equivalent to the sedimentation volume, see Chapter 20) and the zeta potential recorded. Flocculation occurred when a concentration of 10^{-3} molar aluminum chloride was reached. At this point the zeta potential had fallen from -46.4 to -17.0 mV. Further reduction of the zeta potential, to -4.5 mV by use of 10^{-2} molar aluminum chloride did not increase sedimentation height, in agreement with the principles shown in Figure 21-3.

Matthews and Rhodes then went on to show, by computer analysis, that the DLVO theory (see Chapter 20) predicted the results obtained—namely, that the griseofulvin suspensions under investigation would remain deflocculated when the concentration of aluminum chloride was 10^{-4} molar or less. Only at concentrations in the range of 10^{-3} to 10^{-2} molar aluminum chloride did the theoretical plots show deep primary minima, indicative of flocculation. These occurred at a distance of separation between particles of approximately $50 \text{ \AA}$, which led Matthews and Rhodes to conclude that coagulation had taken place in the primary minimum.

Schneider *et al.*⁶ have published details of a laboratory investigation (suitable for undergraduates) that combines calculations based on the DLVO theory carried out with an interactive computer program with actual sedimentation experiments performed on simple systems.

Flocculation by Polymers

In addition to electrolytes, *polymers* play an important role as flocculating agents in pharmaceutical suspensions. Many of the hydrocolloids that produce structured vehicles can also be used to flocculate drug particles. Flocculation by polymers is generally less well understood than flocculation by inorganic electrolytes and their successful use as flocculants may require trial and error in order to identify optimal selection of a polymer and an appropriate concentration. Nonetheless, the production of stable and elegant suspensions using polymers is common in pharmaceutical suspensions.

The effectiveness of a polymer as a stabilizing agent for suspensions primarily depends on the affinity of the polymer for the particle surface as well as the charge, size, and orientation of the polymer molecule in the continuous phase. Many pharmaceutically useful polymers contain polar functional groups that are separated by a hydrocarbon backbone. As a result of this structure, a polymer molecule may adsorb to particle surfaces while maintaining a degree of interaction with the solvent. As observed with inorganic ionic flocculating agents, polymers can produce both flocculated and deflocculated suspensions. An ionic polymer has the potential to affect the zeta potential of a particle in a manner similar to inorganic electrolytes. However, it is believed that polymers can also act as flocculating agents due to the bridging of the polymer between the surfaces of different particles.¹ The effect can be highly concentration dependent as illustrated in Figure 21-4.

At very low concentrations of polymer, bridging between particles occurs as a result of the simultaneous adsorption of a polymer molecule onto the surfaces of different particles. At somewhat higher concentrations of polymer, sufficient binding sites are still available on the particles, permitting additional interparticle attachments to form. It is these intermediate concentrations that result in optimum flocculation and sedimentation volume. At high concentrations of polymer, complete coverage of the particle surface with polymer occurs and insufficient binding sites remain on the particles to permit interparticle bridging. In this case, the degree of flocculation is low, but the close association of individual particles is inhibited by

Polymer concentration		
Low	Intermediate	High
Sedimentation Volume		
Similar to Neat Drug	High	Low

Figure 21-4. Flocculation by hydrophilic polymers. Optimal degree of flocculation and sedimentation volume occurs when a large number of interparticle bridges are formed. High concentrations of polymer result in a deflocculated suspension via steric repulsion.

a phenomenon known as steric stabilization. In general, *steric stabilization* refers to the ability of adsorbed polymers to prevent close approach and cohesion of dispersed particles due to the fact that the mixing of polymers adsorbed at the particle surfaces is energetically unfavorable. Suspensions formulated with relatively high concentrations of polymer would be deflocculated and therefore tend to have small sedimentation volumes. For further information on the practical considerations in the formulation of suspensions using polymers the reader is directed to an article by Scheer.⁷

The sedimentation volume achieved by addition of polymeric flocculating agents may or may not agree with DLVO theory. For example, Kellaway and Najib⁸ found that sulfadimidine suspensions stabilized with the anionic polymer sodium carboxymethylcellulose obeyed the expected relationship between electrophoretic mobility, a measurement that is proportional to zeta potential, and to sedimentation volume in agreement with Figure 21-3. However, stabilization of the suspension with the nonionic polymer polyvinylpyrrolidone (PVP) did not obey the expected relationship between electrophoretic mobility and sedimentation volume. At high concentrations of PVP, particles had a low zeta potential, but contrary to the predictions of Figure 21-3 a low sedimentation volume was observed. Although adsorption of the polymer reduced the charge at the plane of shear, flocculation did not occur. This is believed to be due to steric stabilization of the particles at high concentrations of PVP with few interparticle bonds resulting in a low degree of flocculation.

The conformation of the polymers in the continuous phase may also have an effect on the degree of flocculation. At concentrations where flocculation occurs, polymers that have a

linear conformation in the continuous phase will generally be more effective flocculants than polymers that are coiled in the continuous phase.

In some situations, a combination of polymeric and ionic flocculating agents have been used. In general, the sensitivity of the dispersed solid to flocculation by added electrolyte is enhanced by the presence of the polymer.

Table 21-3 contains a list of suspending agents that have been used in the formulation of pharmaceutical suspensions. Many of these can serve dual functions as flocculating/stabilizing and viscosity enhancing agents.

Flocculation using Detergents

Detergents can be used to produce flocculation in suspensions. Ionic detergents can produce flocculation by neutralization of the charges on the surface of the drug particle. Direct neutralization of particle surface charge occurs due to the adsorption of the detergent to the particle surface. As a result, repulsive forces between particles is reduced and close association of particles can occur. When used as flocculating agents, ionic detergents would need to have a charge that is opposite of the drug particle charge. The optimum concentration required to achieve flocculation is generally determined by experimentation.

Flocculation in Structured Vehicles

The ideal formulation for a suspension would seem to be when flocculated particles are supported in a structured vehicle. As shown in Figure 21-2 (under C), the process involves dispersion of the particles and their subsequent flocculation. Finally, a lyophilic polymer is added to form the structured vehicle. In developing the formulation, care must be taken to ensure the absence of any incompatibility between the flocculating agent and the polymer used for the structured vehicle. A limitation is that virtually all the structured vehicles in common use are hydrophilic colloids and carry a negative charge. This means that an incompatibility arises if the charge on the particles is originally negative. Flocculation in this instance requires the addition of a positively charged flocculating agent or ion; in the presence of such a material, the negatively charged suspending agent may coagulate and lose its suspendability. This situation does not arise with particles that bear a positive charge, as the negative flocculating agent that the formulator must employ is compatible with the similarly charged suspending agent.

A method that can be used to circumvent incompatibilities between an anionic suspending agent and a cationic flocculating agent is to reverse the charge on the particle through the use of a positively charged surface active material such as gelatin. Adsorption of gelatin to the surface of a negatively charged particle can reverse the particle charge when the continuous phase is adjusted to a relatively low pH. This may permit flocculation to be achieved with an anionic flocculating agent such as citrate ion or phosphate ion. Addition of these flocculating agents would be compatible with polymeric suspending agents that largely consist of molecules of anionic charge. Martin *et al.*⁹ have suggested that this effect can also be achieved using surface active amines, provided their toxicity does not prevent their use.

Particle Size and Distribution

Particle size is an important consideration for the physical stability of a suspension. As predicted by Stokes' law, particles of small diameter tend to settle more slowly compared to larger particles; however, small particles will have an increased tendency to cake upon settling if they are not flocculated. In addition, particle-particle interactions can also have a significant effect on suspension stability. For suspensions with a relatively high percentage of solids, interparticle interactions may produce more viscous or thixotropic dispersions. Smaller particles will have a high surface area/weight ratio that favors interactions between the particles and may produce desirable rheological characteristics.

Table 21-3. Suspending Agents Used in the Formulation of Pharmaceutical Suspensions

Type of Polymer	Examples	Structure	Commercial Names
<i>Cellulose derivatives</i>			
Anionic	Carboxymethylcellulose (CMC)	Cellulose ether	
	Microcrystalline cellulose blends	Crystalline cellulose + cellulose ether	Avicel
Nonionic	Methylcellulose (MC) Ethylcellulose (EC) Hydroxyethylcellulose (HEC) Hydroxypropylcellulose (HPC) Hydroxypropylmethylcellulose (HPMC)	Cellulose ether	MC - Methocel, Metocel, Tylopur, Culminol, Celocol, Walsroder EC - Ethocel HEC - Natrasol, Cellocize, Bermocol, Tylose, Blanose HPC-Klucel, Lacrisert HPMC - Methocel, Methlose, Pharmacoat, Culminol, Tylose, Celocol
<i>Natural polymers</i>			
Anionic	alginates, carageenan, xanthan gum, acacia, tragacanth	polysaccharide	
Nonionic	locust bean gum, guar gum	polysaccharide	
<i>Synthetic polymers</i>			
Anionic	Carbomers	Crosslinked polyacrylate	Carbopol
Nonionic	Polyvinyl pyrrolidone (PVP), polyvinyl alcohol (PVA) poloxamer		Plasdone, Povidone, Kollidon
Clays	Magnesium aluminum silicate (Veegum), bentonite, Hectorite	Hydrated aluminum silicate, Magnesium hectorite	

In addition to the effects on the physical properties of a suspension, particle size has important implications on the biopharmaceutical performance of the drug. Aqueous suspensions can effectively serve as a means to deliver poorly water-soluble drugs by the enteral, parenteral, and topical routes. For drugs whose solubility in water is low, the dissolution rate of the drug particles may be a primary factor that limits absorption of the drug. In these cases, the rate and extent of absorption of the drug may be enhanced through the use of small particles. Small particles dissolve faster than larger particles due to the increased surface area per unit weight of drug of the former. Lastly, the uniformity of dosing over the life of the product will be enhanced by ensuring that a relatively small particle size is achieved. This is especially true for suspensions whose individual doses are withdrawn from a larger container, such as suspensions for oral use. Additional information on the bioavailability of drug from suspensions is presented at the end of this chapter.

As most pharmaceutical powders are polydisperse rather than monodisperse, the distribution of particle sizes may also play an important role in the physical stability of a suspension. A relatively narrow distribution of particle sizes is desirable for good stability. A narrow particle size distribution provides a more uniform settling rate and allows for better predictability of suspension properties from batch to batch of finished suspension. In addition, the phenomenon of Ostwald ripening will be minimized when the distribution of particles is narrow. *Ostwald ripening* is the phenomenon in which larger particles grow in size due to the dissolution of smaller particles. This phenomenon could result in pharmaceutically unstable suspensions (caking) and alter the bioavailability of the product through an alteration in the dissolution rate. The use of an appropriate polymer with an affinity for the surface of the dispersed solid reduces or eliminates crystallization in suspensions that may occur due to Ostwald ripening or dissolution/crystallization phenomenon caused by temperature fluctuations. This effect

occurs at concentrations of polymer that provide complete surface coverage of the particles. Thus, a hydrophilic colloid, such as a cellulose derivative, with high affinity for the particle surface is often added initially to the suspension formulation to provide a protective action.

In flocculated suspensions, a narrow distribution of particles also tends to result in floccules with a more opened structure. If a flocculated suspension is prepared using a powder with a wide distribution of particles, the floccules would consist of links between larger particles with small particles filling the voids created by the interparticle links between larger particles. This would create a floccule that is more dense compared to the more open structure that would be expected from a floccule composed of particles of more uniform size. The more opened floc structure is desirable, as it may exhibit thixotropic properties in addition to a large sedimentation volume.

Nonaqueous Suspensions

Although most pharmaceutical suspensions have a primarily aqueous continuous phase, formulation of a drug in a nonaqueous continuous phase is occasionally required. Suspension of a water-soluble drug in a nonaqueous vehicle may provide a means to prepare a liquid formulation of a drug that has poor long-term stability in aqueous solution. Dispersions of drugs in oleaginous vehicles can also provide a sustained release form of drug as observed with certain depot injections and topical products.

Aerosols represent another important class of nonaqueous suspensions. The physical stability of suspended drugs in nonaqueous propellants for aerosol products can have a significant impact on the uniformity of dose and operation of the aerosol system. Caking of the suspended particles can cause clogging of the various mechanical components of the aerosol system.

According to Coulomb's law, the force between two charges is inversely proportional to the dielectric constant of the medium between the charges:

$$f_{\infty} \frac{q_1 q_2}{Dx^2} \quad (6)$$

where f is the force between the particles, q_1 and q_2 are the charges on the particles, D is the dielectric constant, and x is the distance between the charges.

In general, most nonaqueous pharmaceutical liquids have a dielectric constant that is lower than water. This would result in a greater attraction between ions or particles of opposite charge and greater repulsion between ions or particles of similar charge. The effect of a continuous phase of low dielectric constant can therefore affect a suspension formulation in different ways. The use of added electrolytes will be less useful due to their low degree of ionization and poor solubility in some nonaqueous media. In addition, the density of charges on the particle surfaces will be reduced, but repulsion between particles may be facilitated. The result is that controlled flocculation using electrolytes is difficult to achieve as with aqueous suspensions, and caking may occur upon settling. Thus, alternate means of producing pharmaceutically acceptable suspensions must be employed.

Nonionic surfactants of low HLB values can be used to improve the physical stability of the suspensions. Stearic and other aliphatic acids and stearate salts, particularly aluminum monostearate, have been used as suspending agents. These materials increase the viscosity of the oil and produce a structured medium that can hinder the settling of drug particles. Alternatively, thickening agents such as Avicel, colloidal silicon dioxide, and long-chain alcohols can be used to reduce the sedimentation rate in nonaqueous suspensions.

Few studies have been performed to predict formulation and physical stability of drugs in nonaqueous suspensions. Parsons *et al.*¹⁰ found that the suspension properties of a number of solids in a nonaqueous aerosol propellant depended on the surface properties of the solids. Solids that had relatively polar surfaces tended to aggregate to larger extents than solids with relatively nonpolar surfaces. The moisture content of the dispersed solid and continuous phase may also play an important role on the aggregation of the solid. Adsorbed moisture on the dispersed solid may help to create a liquid bridge between particles when dispersed in certain nonaqueous solvents. If carefully controlled, this could provide a means to obtain some degree of flocculation in certain nonaqueous vehicles. Examples are discussed by Hiestand.¹

Chemical Stability of Suspensions

Particles that are completely insoluble in a liquid vehicle are unlikely to undergo most chemical reactions leading to degradation. However, most drugs in suspension have a finite solubility, even though this may be of the order of fractions of a microgram per milliliter. As a result, the material in solution may be susceptible to degradation. However, Tingstad *et al.*¹¹ developed a simplified method for determining the stability of drugs in suspension. The approach is based on the assumptions that

1. degradation takes place only in the solution and is first order.
2. the effect of temperature on drug solubility and reaction rate conforms with classical theory.
3. dissolution is not rate-limiting on degradation.

Preparation of Suspensions

The small-scale preparation of suspensions may be undertaken readily by the practicing pharmacist with the minimum of equipment. The initial dispersion of the particles is best carried out by trituration in a mortar, the wetting agent being added in small increments to the powder. Once the particles have been wetted adequately, the slurry may be transferred to the final container. The next step depends on whether the deflocculated particles are to be suspended in a structured vehicle, flocculated, or flocculated and then suspended. Regardless of

Table 21-4. Commercially Available Vehicles used for the Extemporaneous Preparation of Pharmaceutical Suspensions

Commercial Name	pH	Supplier/Manufacturer
Ora-Plus	4.2	Paddock
Ora-Blend	4.2	Paddock
Suspendol-S	5.3–6.0	Paddock
Syrspend SF	4.2	Gallipot
Syrspend SF Alka*	Alkaline	Gallipot

* Supplied as dry powder to be reconstituted prior to use.

which of the alternative procedures outlined in Figure 21-2 is employed, the various manipulations can be carried out easily in the bottle, especially if an aqueous solution of the suspending agent has been prepared beforehand. Pre-formulated vehicles are commercially available to pharmacists to facilitate the extemporaneous preparation of suspensions. These products are essentially aqueous structured vehicles that contain cellulosic suspending agents, surfactants, buffering agents, and electrolytes. Some products may contain sweetening agents. Pre-formulated suspending vehicles are available at acidic or basic pH to match physical and chemical stability requirements of different drugs. Table 21-4 list some commercially available suspension vehicles that can be used for extemporaneous compounding.

It is most common in modern pharmacy practice to prepare suspensions using commercially available tablets or capsules as a source of the active ingredient. Examples of compounding using solid oral dosage forms as starting material can be found in the pharmaceutical literature¹²

For detailed discussion of the methods used in the large-scale production of suspensions, see the relevant section in Chapter 24.

EMULSIONS

An *emulsion* is a dispersed system containing at least two immiscible liquid phases. The majority of conventional emulsions in pharmaceutical use have dispersed particles ranging in diameter from 0.1 to 100 μm . As with suspensions, emulsions are thermodynamically unstable as a result of the excess free energy associated with the surface of the droplets. The dispersed droplets, therefore, strive to come together and reduce the surface area. In addition to this flocculation effect, also observed with suspensions, the dispersed particles can coalesce, or fuse, and this can result in the eventual destruction of the emulsion. To minimize this effect, a third component, the *emulsifying agent*, is added to the system to improve its stability. The choice of emulsifying agent is critical to the preparation of an emulsion possessing optimum stability. The efficiency of present-day emulsifiers permits the preparation of emulsions that are stable for many months and even years, even though they are thermodynamically unstable.

In recent years, it has been recognized that complex multiple-phase combinations can exist in emulsions. Thus, liquid crystalline phases and gel structures can form from the combination of the basic three-component mixture of water, oil, and surfactant (emulsifying agent). Often, these structures confer significant stability to the emulsion and therefore are to be desired. Such multiple-phase emulsions and their stability have been reviewed by Eccleston.¹³

Emulsions are widely used in pharmacy and medicine, and emulsified materials can possess advantages not observed in formulations in other dosage forms. For example, certain medicinal agents that have an objectionable taste have been made

more palatable for oral administration when formulated in an emulsion. The principles of emulsification have been applied extensively in the formulation of dermatological creams and lotions. Intravenous emulsions of contrast media have been developed to assist the physician in undertaking x-ray examinations of the body organs while exposing the patient to the minimum of radiation. Considerable attention has been directed towards the use of sterile, stable intravenous emulsions containing fat, carbohydrate, and vitamins all in one preparation. Such products are administered to patients unable to assimilate these vital materials by the normal oral route. Emulsions are also used to deliver nutrients via the enteral route in the form of nutritional supplements. More recently, emulsions have been used to deliver poorly water-soluble drugs, such as general anesthetics and anti-cancer compounds, via the intravenous route.

Emulsions offer potential in the design of systems capable of giving controlled rates of drug release and affording protection to drugs susceptible to oxidation or hydrolysis. There is still a need for well-characterized dermatologic products with reproducible properties, regardless of whether these products are antibacterial, sustained-release, protective, or emollient lotions, creams, or ointments. In addition, emulsions may provide a useful way to deliver poorly water-soluble drugs via enteral and parenteral routes. The principle of emulsification is also involved in an increasing number of aerosol products.

The pharmacist must be familiar with the types of emulsions and the properties and theories underlying their preparation and stability; such is the purpose of the remainder of this chapter. Microemulsions, which can be regarded as isotropic, swollen micellar systems are discussed in Chapter 24.

EMULSION TYPE AND MEANS OF DETECTION

A stable emulsion must contain at least three components: the *dispersed phase*, the *dispersion medium*, and the *emulsifying agent*. Invariably, one of the two immiscible liquids is aqueous, and the second is an oil. Whether the aqueous or the oil phase becomes the dispersed phase depends primarily on the emulsifying agent used and the relative amounts of the two liquid phases. Hence, an emulsion in which the oil is dispersed as droplets throughout the aqueous phase is termed an *oil-in-water (O/W) emulsion*. When water is the dispersed phase and an oil the dispersion medium, the emulsion is of the *water-in-oil (W/O) type*. Most pharmaceutical emulsions designed for oral administration are of the O/W type; emulsified lotions and creams are either O/W or W/O, depending on their use. Butter and salad creams are W/O emulsions.

So-called *multiple emulsions* have been developed with a view to delaying the release of an active ingredient. In these types of emulsions three phases are present: the emulsion has the form W/O/W or O/W/O. In these “emulsions within emulsions,” any drug present in the innermost phase must now cross two phase boundaries to reach the external, continuous phase.

It is important for pharmacists to know the type of emulsion they have prepared or are dealing with, because this can affect its properties and performance. Unfortunately, the several methods available can give incorrect results, so the type of emulsion determined by one method should always be confirmed by means of a second method.

DILUTION TEST

The dilution method depends on the fact that an O/W emulsion can be diluted with water and a W/O emulsion with oil. When oil is added to an O/W emulsion or water to a W/O emulsion, the additive is not incorporated into the emulsion and separation is apparent. The test is greatly improved if the addition of the water or oil is observed microscopically.

CONDUCTIVITY TEST

An emulsion in which the continuous phase is aqueous can be expected to possess a much higher conductivity than an

emulsion in which the continuous phase is an oil. Accordingly, it frequently happens that when a pair of electrodes, connected to a lamp and an electrical source, are dipped into an O/W emulsion, the lamp lights because of the passage of a current between the two electrodes. If the lamp does not light, it is assumed that the system is W/O.

DYE-SOLUBILITY TEST

The knowledge that a water-soluble dye will dissolve in the aqueous phase of an emulsion while an oil-soluble dye will be taken up by the oil phase provides a third means of determining emulsion type. Thus, if microscopic examination shows that a water-soluble dye has been taken up by the continuous phase, we are dealing with an O/W emulsion. If the dye has not stained the continuous phase, the test is repeated using a small amount of an oil-soluble dye. Coloring of the continuous phase confirms that the emulsion is of the W/O type.

FORMATION AND BREAKDOWN OF DISPERSED LIQUID DROPLETS

An emulsion exists as the result of two competing processes: the dispersion of one liquid throughout another as droplets, and the combination of these droplets to reform the initial bulk liquids. The first process increases the free energy of the system, while the second works to reduce the free energy. Accordingly, the second process is spontaneous and continues until breakdown is complete—that is, until the bulk phases are reformed.

It is of little use to form a well-dispersed emulsion if it quickly breaks down. Similarly, unless adequate attention is given to achieving an optimum dispersion during preparation, the stability of an emulsion system may be compromised from the start. Dispersion is brought about by well-designed and well-operated machinery, capable of producing droplets in a relatively short period of time. Such equipment is discussed in Chapter 24. The reversal back to the bulk phases is minimized by using those parameters that influence the stability of the emulsion once it is formed.

DISPERSION PROCESS TO FORM DROPLETS

Consider two immiscible liquid phases in a test tube. To disperse one liquid as droplets within the other, the interface between the two liquids must be disturbed and expanded to a sufficient degree so that “fingers” or threads of one liquid pass into the second liquid, and *vice versa*. These threads are unstable, and become varicose or beaded. The beads separate and become spherical, as illustrated in Figure 21-5. Depending on the agitation or the shear rate used, larger droplets are also deformed to give small threads, which in turn produce smaller drops.

The time of agitation is important. Thus, the mean size of droplets decreases rapidly in the first few seconds of agitation. The limiting size range is generally reached within 1 to 5 min, and results from the number of droplets coalescing being equivalent to the number of new droplets being formed. It is uneconomical to continue agitation any further.

The liquids may be agitated or sheared by several means. *Shaking* is employed commonly, especially when the components are of low viscosity. Intermittent shaking is frequently more efficient than continual shaking, possibly because the short time interval between shakes allows the thread that is forced across the interface time to break down into drops that are then isolated in the opposite phase. Continuous, rapid agitation tends to hinder this breakdown to form drops. A mortar and pestle is employed frequently in the extemporaneous preparation of emulsions. It is not a very efficient technique and is not used on a large scale. Improved dispersions are achieved by the use of high-speed mixers, blenders, colloid mills, or homogenizers. Ultrasonic techniques also have been employed and are described in Chapter 24.

The phenomenon of *spontaneous emulsification*, as the name implies, occurs without any external agitation. There

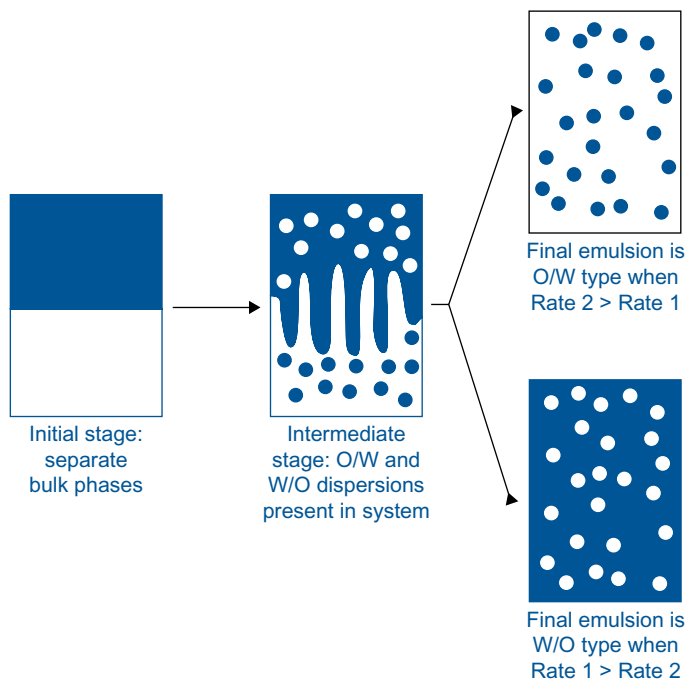


Figure 21-5. Effect of rate of coalescence on emulsion type. Rate 1: O/W coalescence rate. Rate 2: W/O coalescence rate.; oil.: water. For an explanation of Rates 1 and 2, refer to the discussion of Davies on page 329¹⁴.

is, however, an internal agitation arising from certain physico-chemical processes that affect the interface between the two bulk liquids. For a description of this process, see Davies and Rideal in the bibliography.

COALESCENCE OF DROPLETS

Coalescence is a process distinct from flocculation (aggregation), which commonly precedes it. Flocculation is the clumping together of particles, but coalescence is the fusing of the agglomerates into a larger drop, or drops. Coalescence usually is rapid when two immiscible liquids are shaken together, as there is no large energy barrier to prevent fusion of drops and reformation of the original bulk phases. When an emulsifying agent is added to the system, flocculation still may occur but coalescence is reduced to an extent depending on the efficacy of the emulsifying agent to form a stable, coherent interfacial film. It is therefore possible to prepare emulsions that are flocculated yet do not coalesce. In addition to the interfacial film around the droplets acting as a mechanical barrier, the drops also are prevented from coalescing by the presence of a thin layer of continuous phase between particles clumped together. Davies¹⁴ showed the importance of coalescence rates in determining emulsion type; this work is discussed in more detail on page 329.

EMULSIFYING AGENT

The process of coalescence can be reduced to insignificant levels by the addition of a third component—the emulsifying agent or *emulsifier*. The choice of emulsifying agent is frequently critical in developing a successful emulsion, and the pharmacist should be aware of

- the desirable properties of emulsifying agents.
- how different emulsifiers act to optimize emulsion stability.
- how the type and physical properties of the emulsion can be affected by the emulsifying agent.

DESIRABLE PROPERTIES

Some of the desirable properties of an emulsifying agent are that it should

1. be surface active and reduce surface tension to below 10 dynes/cm.
2. be adsorbed quickly around the dispersed drops as a condensed, nonadherent film that will prevent coalescence.
3. impart to the droplets an adequate electrical potential so that mutual repulsion occurs.
4. increase the viscosity of the emulsion.
5. be effective in a reasonably low concentration.

Not all emulsifying agents possess these properties to the same degree; in fact, not every good emulsifier necessarily possesses all these properties. Further, there is no one ideal emulsifying agent because the desirable properties of an emulsifier depend, in part, on the properties of the two immiscible phases in the particular system under consideration.

Interfacial Tension

Lowering of interfacial tension is one way in which the increased surface free energy associated with the formation of droplets, and hence surface area, in an emulsion can be reduced (Equation 1). Assuming the droplets to be spherical, it can be shown that

$$\Delta F = \frac{6\gamma V}{d} \quad (7)$$

where V is the volume of dispersed phase in milliliters, and d is the mean diameter of the particles. To disperse 100 mL of oil as 1- μm (10^{-4} -cm) droplets in water when $\gamma_{\text{O/W}} = 50$ dynes/cm, requires an energy input of

$$\begin{aligned} \Delta F &= \frac{6 \times 50 \times 100}{1 \times 10^{-4}} = 30 \times 10^7 \text{ ergs} \\ &= 30 \text{ joules or } 30/4.184 = 7.2 \text{ cal} \end{aligned}$$

In the above example the addition of an emulsifier that will reduce γ from 50 to 5 dynes/cm will reduce the surface free energy from 7.2 to around 0.7 cal. Likewise, if the interfacial tension is reduced to 0.5 dynes/cm (a common occurrence), the original surface free energy is reduced a hundredfold. Such a reduction can help to maintain the surface area generated during the dispersion process.

Film Formation

The major requirement of a potential emulsifying agent is that it readily form a *film* around each droplet of dispersed material. The main purpose of this film—which can be a monolayer, a multilayer, or a collection of small particles adsorbed at the interface—is to form a barrier that prevents the coalescence of droplets that come into contact with one another. For the film to be an efficient barrier, it should possess some degree of surface elasticity and should not thin out and rupture when sandwiched between two droplets. If broken, the film should have the capacity to reform rapidly.

Electrical Potential

The origin of an electrical potential at the surface of a droplet has been discussed earlier in the chapter. Insofar as emulsions are concerned, the presence of a well-developed charge on the droplet surface is significant in promoting stability by causing repulsion between approaching drops. This potential is likely to be greater when an ionized emulsifying agent is employed. Electrical potential has been shown to be a significant factor for maintaining the stability of intravenous fat emulsions that are stabilized with lecithin.

Concentration of Emulsifier

The main objective of an emulsifying agent is to form a condensed film around the droplets of the dispersed phase. An inadequate concentration will do little to prevent coalescence. Increasing the emulsifier concentration above an optimum level achieves little in terms of increased stability. In practice the aim is to use the minimum amount consistent with producing a satisfactory emulsion.

It frequently helps to have some idea of the amount of emulsifier required to form a condensed film, one molecule thick, around each droplet. Suppose we wish to emulsify 50 g of an oil, density = 1.0, in 50 g of water. The desired particle diameter is 1 μm . Thus,

$$\text{Particle diameter} = 1 \mu\text{m} = 1 \times 10^{-4} \text{ cm}$$

$$\text{Volume of particle} = (\pi d^3 / 6) = 0.524 \times 10^{-12} \text{ cm}^3$$

$$\text{Total number of particles in 50 g} = (50 / 0.524 \times 10^{-12}) = 95.5 \times 10^{12}$$

$$\text{Surface area of particle} = \pi d^2 = 3.142 \times 10^{-8} \text{ cm}^2$$

$$\text{Total surface area} = 3.142 \times 10^{-8} \times 95.5 \times 10^{12} = 300 \times 10^4 \text{ cm}^2$$

If the area each molecule occupies at the oil–water interface is 30 $\text{\AA}^2$ ($30 \times 10^{-16} \text{ cm}^2$), we require

$$\frac{300 \times 10^4}{30 \times 10^{-16}} = 1 \times 10^{21} \text{ molecules}$$

A typical emulsifying agent might have a molecular weight of 1000. Thus, the required weight is

$$\frac{1000}{6.023 \times 10^{23}} \times 10^{21} = 1.66 \text{ g}$$

To emulsify 10 g of oil would require 0.33 g of the emulsifying agent.

Although the approach is an oversimplification of the problem, it does at least allow the formulator to make a reasonable estimate of the required concentration of emulsifier.

Emulsion Rheology

The emulsifying agent and other components of an emulsion can affect the rheologic behavior of an emulsion in several ways, as summarized in Table 21-5.¹⁵ It should be borne in mind that the droplets of the internal phase are deformable under shear and that the adsorbed layer of emulsifier affects the interactions between adjacent droplets and also between a droplet and the continuous phase. The means by which the rheological behavior of emulsions can be controlled have been discussed by Rogers.¹⁶

MECHANISM OF ACTION

Emulsifying agents may be classified in accordance with the type of film they form at the interface between the two phases.

Monomolecular Films

Those surface-active agents that are capable of stabilizing an emulsion do so by forming a monolayer of adsorbed molecules or ions at the oil–water interface (Figure 21-6). In accordance with Gibbs' law (Chapter 19) the presence of an interfacial excess necessitates a reduction in interfacial tension. This results in a more stable emulsion because of a proportional reduction in the surface free energy. Of itself, this reduction is probably not the main factor promoting stability. More significant is the fact that the droplets are surrounded now by a coherent monolayer that prevents coalescence between approaching droplets. If the emulsifier forming the monolayer is ionized, the presence of strongly charged and mutually repelling droplets increases the stability of the system. With un-ionized, nonionic surface-active agents, the particles may still carry a charge; this arises from adsorption of a specific ion or ions from solution.

Table 21-5. Factors Influencing Emulsion Viscosity

1. Internal phase
 - a. Volume concentration (ϕ); hydrodynamic interaction between globules; flocculation, leading to formation of globule aggregates.
 - a. Viscosity (η_i); deformation of globules in shear.
 - b. Globule size, and size distribution, technique used to prepare emulsion; interfacial tension between the two liquid phases: globule behavior in shear; interaction with continuous phase; globule interaction.
2. Chemical constitution.
 - a. Continuous phase
 - b. Viscosity (η_o), and other rheological properties.
 - c. Chemical constitution, polarity, pH; potential energy of interaction between globules.
 - d. Electrolyte concentration if polar medium.
3. Emulsifying agent
 - a. Chemical constitution; potential energy of interaction between globules.
 - b. Concentration, and solubility in internal and continuous phases; emulsion type; emulsion inversion; solubilization of liquid phases in micelles.
 - c. Thickness of film adsorbed around globules, and its rheological properties, deformation of globules in shear; fluid circulation within globules.
 - d. Electroviscous effect.
4. Additional stabilizing agents
 - a. Pigments, hydrocolloids, hydrous oxides.
 - b. Effect on rheological properties of liquid phases, and interfacial boundary region.

(From Davies JT. In: *Proceedings of the International Congress on Surface Activity*, 2nd ed. London: Butterworth/Academic, 1957, p 426.)

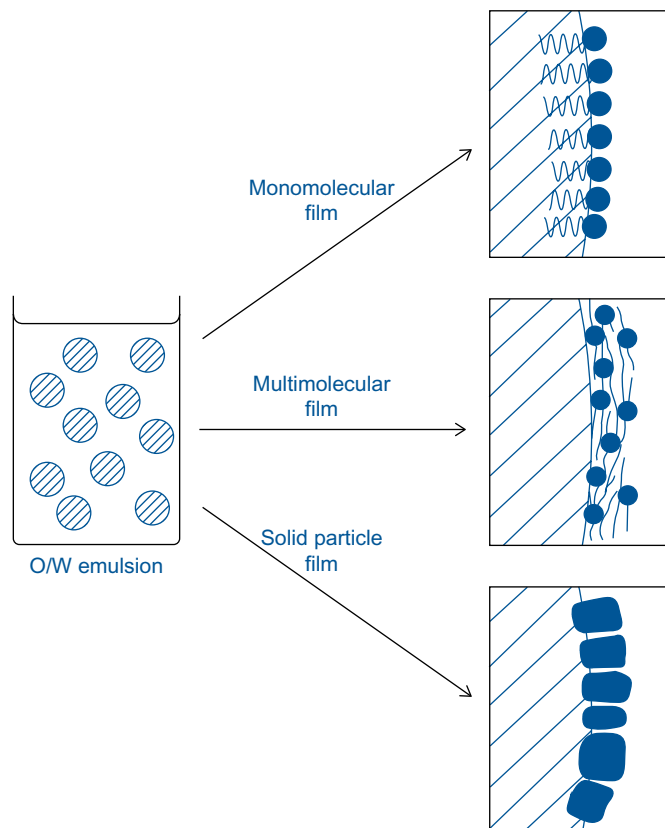


Figure 21-6. Types of films formed by emulsifying agents at the oil–water interface. Orientations are shown for O/W emulsions.: oil.: water.

Multimolecular Films

Hydrated lyophilic colloids form multimolecular films around droplets of dispersed oil (see Figure 21-6). The use of these agents has declined in recent years because of the large number of synthetic surface-active agents available that possess well-marked emulsifying properties. Although these hydrophilic colloids are adsorbed at an interface (and can be regarded therefore as surface active), they do not cause an appreciable lowering in surface tension. Rather, their efficiency depends on their ability to form strong coherent multimolecular films. These act as a coating around the droplets and render them highly resistant to coalescence, even in the absence of a well-developed surface potential. Furthermore, any hydrocolloid not adsorbed at the interface increases the viscosity of the continuous aqueous phase; this enhances emulsion stability.

Solid Particle Films

Small solid particles that are wetted to some degree by both aqueous and nonaqueous liquid phases act as emulsifying agents. If the particles are too hydrophilic, they remain in the aqueous phase; if too hydrophobic, they are dispersed completely in the oil phase. A second requirement is that the particles are small in relation to the droplets of the dispersed phase (see Figure 21-6).

CHEMICAL TYPES

Emulsifying agents also may be classified in terms of their chemical structure; there is some correlation between this classification and that based on the mechanism of action. For example, the majority

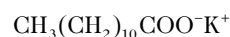
of emulsifiers forming monomolecular films are synthetic, organic materials. Most of the emulsifiers that form multimolecular films are obtained from natural sources and are organic. A third group is composed of solid particles, invariably inorganic, that form films composed of finely divided solid particles.

Accordingly, the classification, adopted divides emulsifying agents into *synthetic*, *natural*, and *finely dispersed solids* (Table 21-6). A fourth group, the *auxiliary materials* (Table 21-7) are weak emulsifiers. The list of agents is not meant to be exhaustive, but rather merely illustrates the various types available.

Synthetic Emulsifying Agents

Synthetic emulsifying agents, a group of surface-active agents that act as emulsifiers, may be subdivided into anionic, cationic, and nonionic, depending on the charge possessed by the surfactant.

Anionics—In the anionic subgroup, the surfactant ion bears a negative charge. The potassium, sodium, and ammonium salts of lauric and oleic acid are soluble in water and are good O/W emulsifying agents. They do, however, have a disagreeable taste and are irritating to the gastrointestinal (GI) tract; this limits them to emulsions prepared for external use. Potassium laurate, a typical example, has the structure



Solutions of *alkali soaps* have a high pH; they start to precipitate out of solution below pH 10 because the unionized fatty acid is now formed, and this has a low aqueous solubility.

Table 21-6. Classification of Emulsifying Agents

Type	Type of Film	Examples
Synthetic (surface-active agents)	Monomolecular	<i>Anionic</i>
		Soaps Potassium laurate Triethanolamine stearate Sulfates Sodium lauryl sulfate Alkyl polyoxyethylene sulfates Sulfonates Diocetyl sodium sulfosuccinate
		<i>Cationic</i> Quaternary ammonium compounds Cetyltrimethylammonium bromide Lauryldimethylbenzylammonium chloride
		<i>Nonionic</i>
		Polyoxyethylene fatty alcohol ethers Sorbitan fatty acid esters Polyoxyethylene sorbitan fatty acid esters Polyoxyethylene polyoxypropylene block copolymers (poloxamers) Lanolin alcohols and ethoxylated lanolin alcohols
Natural	Multimolecular	<i>Hydrophilic colloids</i> Acacia Gelatin
	Monomolecular	Lecithin Cholesterol
Finely divided solids	Solid particle	<i>Colloidal clays</i> Bentonite Veegum <i>Metallic hydroxides</i> Magnesium hydroxide

Table 21-7. Auxiliary Emulsifying Agents

Product	Source and Composition	Principal Use
Cetyl alcohol	Chiefly $C_{16}H_{33}OH$	Lipophilic thickening agent and stabilizer for O/W lotions and ointments
Glyceryl monostearate	$C_{17}H_{35}COOCH_2CHOHCH_2OH$	Lipophilic thickening agent and stabilizer for O/W lotions and ointments
Methylcellulose	Series of methyl ethers of cellulose	Hydrophilic thickening agent and stabilizer for O/W emulsions; weak O/W emulsifier
Sodium carboxymethylcellulose	Sodium salt of the carboxymethyl esters of cellulose	Hydrophilic thickening agent and stabilizer for O/W emulsions
Stearic acid	A mixture of solid acids from fats, chiefly stearic and palmitic	Lipophilic thickening agent and stabilizer for O/W lotions and ointments. Forms a true emulsifier when reacted with an alkali

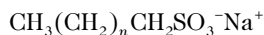
Further, the free fatty acid is ineffective as an emulsifier, so emulsions formed from alkali soaps are not stable at pH values less than about 10.

The calcium, magnesium, and aluminum salts of fatty acids, often termed the *metallic soaps*, are water insoluble and result in W/O emulsions.

Another class of soaps are salts formed from a fatty acid and an organic amine such as triethanolamine. These O/W emulsifiers also are limited to external preparations, but their alkalinity is considerably less than that of the alkali soaps and they are active as emulsifiers down to around pH 8. These agents are less irritating than the alkali soaps.

Sulfated alcohols are neutralized sulfuric acid esters of such fatty alcohols as lauryl and cetyl alcohol. These compounds are an important group of pharmaceutical surfactants. They are used chiefly as wetting agents, although they do have some value as emulsifiers, particularly when used in conjunction with an auxiliary agent.

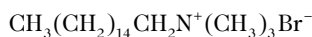
Sulfonates are a class of compounds in which the sulfur atom is connected directly to the carbon atom, giving the general formula



A frequently used compound is sodium lauryl sulfate. Sulfonates have a higher tolerance to calcium ions and do not hydrolyze as readily as the sulfates. A widely used surfactant of this type is dioctyl sodium sulfosuccinate.

Cationics—The surface activity in the cationic group resides in the positively charged cation. These compounds have marked bactericidal properties. This makes them desirable in emulsified anti-infective products such as skin lotions and creams. The pH of an emulsion prepared with a cationic emulsifier lies in the pH 4 to 6 ranges. Because this includes the normal pH of the skin, cationic emulsifiers are advantageous in this regard also.

Cationic agents are weak emulsifiers and generally are formulated with a stabilizing or auxiliary emulsifying agent such as cetostearyl alcohol. The only group of cationic agents used extensively as emulsifying agents are the quaternary ammonium compounds. An example is cetyltrimethylammonium bromide.

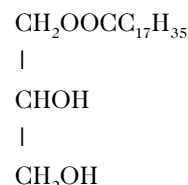


Cationic emulsifiers should not be used in the same formulation with anionic emulsifiers because they will interact. The incompatibility may not be immediately apparent as a precipitate, but virtually all of the desired antibacterial activity will generally have been lost.

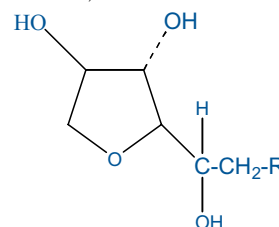
Nonionics—Nonionics, undissociated surfactants, find widespread use as emulsifying agents when they possess the proper balance of hydrophilic and lipophilic groups within the molecule. Their popularity is based on the fact that, unlike the anionic and cationic types, nonionic emulsifiers are

not susceptible to pH changes and the presence of electrolytes. The number of nonionic agents available is legion; the most frequently used are the glyceryl esters, polyoxyethylene glycol esters and ethers, and the sorbitan fatty acid esters and their polyoxyethylene derivatives. More recently, the polyoxyethylene/polyoxypropylene block copolymers have become popular surfactants and emulsifying agents.

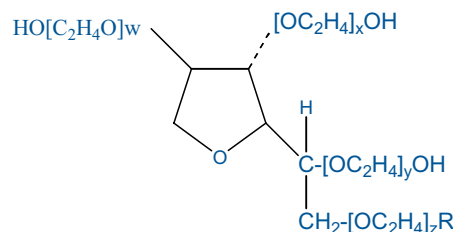
A glyceryl ester, such as glyceryl monostearate, is too lipophilic to serve as a good emulsifier; it is used widely as an auxiliary agent (see Table 21-7) and has the structure



Sorbitan fatty acid esters, such as sorbitan monopalmitate



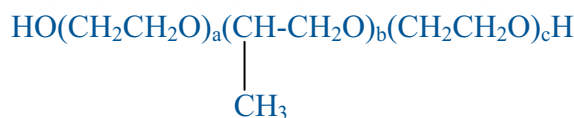
are nonionic oil-soluble emulsifiers that promote W/O emulsions. The polyoxyethylene sorbitan fatty acid esters, such as polyoxyethylene sorbitan monopalmitate



[R is $(C_{15}H_{31})COO$; Sum or w, x, y, and z is 20] are hydrophilic water-soluble derivatives that favor O/W emulsions.

Polyoxyethylene glycol esters, such as the monostearate, $C_{17}H_{35}COO(CH_2OCH_2)_nH$, also are used widely.

Polyoxyethylene/polyoxypropylene block copolymers



also known as *poloxamers* consist of combined chains of oxyethylene with oxypropylene where the oxyethylene portion imparts hydrophilicity and the oxypropylene portion imparts lipophilicity. The molecules are synthesized as long segments of the hydrophilic portions combined with long segments of the hydrophobic portions, with each portion referred to as a *block*. This organization produces hydrophilic and hydrophobic domains that impart the surface active character to these agents. Poloxamers have been used in the formulation of intravenous emulsions and can impart structure to vehicles and interfacial films that can protect the dispersed phase against coalescence. The polymeric nature of these surfactants protects emulsions against coalescence via steric stabilization at the droplet interface.

In general, for nonionic emulsifiers, emulsion stability is best when blends of emulsifiers are used. Thus, an O/W emulsifier customarily will be used in an emulsion with a W/O emulsifier. When blended properly, the nonionics produce fine-textured stable emulsions. For example, an emulsifier blend of polysorbates (O/W emulsifier) with sorbitan esters (W/O emulsifier) are frequently used as opposed to the use of either class of emulsifier alone. The rationale for the use of mixed emulsifiers is further discussed below in the discussion of HLB System and Mixed emulsifiers.

Natural Emulsifying Agents

Of the numerous emulsifying agents derived from natural (i.e., plant and animal) sources, consideration will be given only to acacia, gelatin, lecithin, and cholesterol. Many other natural materials are only sufficiently active to function as auxiliary emulsifying agents or stabilizers.

Acacia is a carbohydrate gum that is soluble in water and forms O/W emulsions. Emulsions prepared with acacia are stable over a wide pH range. Because it is a carbohydrate it is necessary to preserve acacia emulsions against microbial attack by the use of a suitable preservative.

Gelatin, a protein, has been used for many years as an emulsifying agent. Gelatin can have two isoelectric points, depending on the method of preparation. So-called Type A gelatin, derived from an acid-treated precursor, has an isoelectric point of between pH 7 and 9. Type B gelatin, obtained from an alkali-treated precursor, has an isoelectric point of approximately pH 5. Type A gelatin acts best as an emulsifier around pH 3, where it is positively charged; on the other hand, Type B gelatin is best used around pH 8, where it is negatively charged. The question as to whether the gelatin is positively or negatively charged is fundamental to the stability of the emulsion when other charged emulsifying agents are present. To avoid an incompatibility, all emulsifying agents should carry the same sign. Thus, if gums (such as tragacanth, acacia, or agar) that are negatively charged are to be used with gelatin, then Type B material should be used at an alkaline pH. Under these conditions the gelatin is similarly negatively charged.

Lecithin is an emulsifier obtained from both plant (e.g., soybean) and animal (e.g., egg yolk) sources and is composed of various phosphatides. The primary component of most lecithins is phosphatidylcholine and the term "lecithin" is often used to describe purified samples of phosphatidylcholine. Frequently, lecithins that are used as emulsifiers also contain mixtures of phosphatides, including phosphatidylserine, phosphatidylinositol, phosphatidylethanolamine, and phosphatidic acid in addition to phosphatidylcholine. Although phosphatidylcholine is a zwitterionic compound, the presence of other phosphatides such as phosphatidylinositol and phosphatidic acid, as well as small quantities of lysophosphatides, result in an emulsifier that imparts a net negative charge to dispersed particles.

Lecithin can be an excellent emulsifier for naturally occurring oils such as soy, corn, or safflower. Highly stable O/W emulsions can be formed with these oils. Purified lecithins from soy or egg yolk are the principal emulsifiers for intravenous fat emulsions. Lecithin provides stable emulsions with droplet sizes of

less than 1 μm in diameter. It is critical that a small, uniform particle size be maintained in these emulsions to eliminate the risks of fat embolism after intravenous injection. Commercially available intravenous fat emulsions generally resist coalescence for several years and are able to withstand high temperature sterilization by autoclaving. The excellent physical stability observed with these emulsions is believed to be the result of the large negative zeta potential that results from the small quantity of charged lipids present in lecithin as well as the ability of the lecithin to form mesophases resembling liposomes. During manufacture of the emulsions, homogenization produces small droplets that are surrounded by concentric layers of phospholipids. The latter may form a protective layer that prevents coalescence of the droplets. As an emulsifier, lecithin produces the best results at a pH of around 8.

As with any natural product, the content of lecithins will vary from source to source and their emulsifying properties and toxicity may also vary. For highly critical applications, such as intravenous emulsions, the source and composition of the lecithin must be carefully controlled and monitored.

Cholesterol is a major constituent of wool alcohols, obtained by the saponification and fractionation of wool fat. It is cholesterol that gives wool fat its capacity to absorb water and form a W/O emulsion.

Finely Dispersed Solids

Finely dispersed solids are emulsifiers that form particulate films around the dispersed droplets, producing emulsions that are coarse-grained but have considerable physical stability. It appears possible that any solid can act as an emulsifying agent of this type, provided it is reduced to a sufficiently fine powder. In practice, the group of compounds used most frequently are the colloidal clays.

Bentonite is a white to gray, odorless and tasteless powder that swells in the presence of water to form a translucent suspension with a pH of about 9. Depending on the sequence of mixing it is possible to prepare both O/W and W/O emulsions. When an O/W emulsion is desired, the bentonite is first dispersed in water and allowed to hydrate so as to form a magma. The oil phase is then added gradually with constant titration. Because the aqueous phase is always in excess, the O/W emulsion type is favored. To prepare a W/O emulsion, the bentonite is first dispersed in oil; the water is then added gradually.

Although *Veegum* is used as a solid particle emulsifying agent, it is employed most extensively as a stabilizer in cosmetic lotions and creams. Concentrations of less than 1% Veegum will stabilize an emulsion containing anionic or nonionic emulsifying agents.

Auxiliary Emulsifying Agents

Auxiliary emulsifying agents include those compounds that are normally incapable themselves of forming stable emulsions. Their main value lies in their ability to function as thickening agents and thereby help stabilize the emulsion. Agents in common use are listed in Table 21-7. Auxiliary emulsifying agents that are amphiphilic in nature are, in some cases, capable of forming gel or liquid crystalline phases with the primary emulsifying agent when combined with water and oil. This type of behavior may help to stabilize emulsions due to an increased viscosity, as observed in topical creams. Alternatively, gel or liquid crystalline phases may prevent coalescence by reducing van der Waals forces between particles or by providing a physical barrier between approaching particles of the internal phase. This latter effect is thought to be an important function in phospholipid-stabilized emulsions that must maintain a low viscosity to permit administration via the intravenous route. Additional information is provided by Eccleston.¹²

EMULSIFYING AGENTS AND EMULSION TYPE

For a molecule, ion, colloid, or particle to be active as an emulsifying agent, it must have some affinity for the interface

between the dispersed phase and the dispersion medium. With the monolayer and multilayer films, the emulsifier is in solution, and therefore it must be soluble to some extent in one or both of the phases. At the same time it must not be overly soluble in either phase; otherwise, it will remain in the bulk of that phase and not be adsorbed at the interface. This balanced affinity for the two phases also must be evident with finely divided solid particles used as emulsifying agents. If their affinity, as evidenced by the degree to which they are wetted, is either predominantly hydrophilic or hydrophobic, they will not function as effective wetting agents.

The great majority of the work on the relation between emulsifier and emulsion type has been concerned with surface-active agents that form interfacial monolayers. Thus, the present discussion will concentrate on this class of agents.

Hydrophile-Lipophile Balance

As the emulsifier becomes more hydrophilic, its solubility in water increases and the formation of an O/W emulsion is favored. Conversely, W/O emulsions are favored with the more lipophilic emulsifiers. This led to the concept that the type of emulsion is related to the balance between hydrophilic and lipophilic solution tendencies of the surface-active emulsifying agent.

Griffin¹⁷ developed a scale based on the balance between these two opposing tendencies. This so-called *HLB scale* is a numerical scale, extending from 1 to approximately 50. The more hydrophilic surfactants have high HLB numbers (in excess of 10), whereas surfactants with HLB numbers from 1 to 10 are considered to be lipophilic. Surfactants with a proper balance in their hydrophilic and lipophilic affinities are effective emulsifying agents because they concentrate at the oil-water interface. The relationship between HLB values and the application of the surface-active agent is shown in Table 21-8. Some commonly used emulsifiers and their HLB numbers are listed in Table 21-8. The utility of the HLB system in rationalizing the choice of emulsifying agents when formulating an emulsion will be discussed in a later section.

Rate of Coalescence and Emulsion Type

Davies¹⁴ indicated that the type of emulsion produced in systems prepared by shaking is controlled by the relative coalescence rates of oil droplets dispersed in the oil. Thus, when a mixture of oil and water is shaken together with an emulsifying agent, a multiple dispersion is produced initially that contains oil dispersed in water and water dispersed in oil (see Figure 21-5). The type of the final emulsion that results depends on whether the water or the oil droplets coalesce more rapidly. If the O/W coalescence rate (Rate 1) is much greater than W/O coalescence rate (Rate 2), a W/O emulsion is formed because the dispersed water droplets are more stable than the dispersed oil droplets. Conversely, if Rate 2 is significantly faster than Rate 1, the final emulsion is an O/W dispersion because the oil droplets are more stable.

According to Davies,¹⁴ the rate at which oil globules coalesce when dispersed in water is given by the expression

$$\text{Rate 1} = C_1 e^{-W_1/RT} \quad (8)$$

The term C_1 is a collision factor that is directly proportional to the phase volume of the oil relative to the water, and is an inverse function of the viscosity of the continuous phase (water). W_1 defines an energy barrier made up of several contributing factors that must be overcome before coalescence can take place. First, it depends on the electrical potential of the dispersed oil droplets, as this affects repulsion. Second, with an O/W emulsion, the hydrated layer surrounding the polar portion of emulsifying agent must be broken down before coalescence can occur. This hydrated layer is probably around 1 nm thick with a consistency of butter. Finally, the total energy barrier depends on the fraction of the interface covered by the emulsifying agent.

Equation 9 describes the rate of coalescence of water globules dispersed in oil:

$$\text{Rate 2} = C_2 e^{-W_2/RT} \quad (9)$$

Here, the collision factor C_2 is a function of the water-oil phase volume ratio divided by the viscosity of the oil phase. The energy barrier W_2 is, as before, related to the fraction of the interface covered by the surface-active agent. Another contributing factor is the number of $-\text{CH}_2-$ groups in the emulsifying agent; the longer the alkyl chain of the emulsifier, the greater the gap that has to be bridged if one water droplet is to combine with a second drop.

Davies¹⁴ showed that the HLB concept is related to the distribution characteristics of the emulsifying agent between the two immiscible phases. An emulsifier with an HLB of less than 7 will be preferentially soluble in the oil phase and will favor formation of a W/O emulsion. Surfactants with an HLB value in excess of 7 will be distributed in favor of the aqueous phase and will promote O/W emulsions.

PREPARATION OF EMULSIONS

Several factors must be taken into account in the successful preparation and formulation of emulsified products. Usually, the type of emulsion (i.e., O/W or W/O) is specified; if not, it probably will be implied from the anticipated use of the product. The formulator's attention is focused primarily on the selection of the emulsifying agent, or agents, necessary to achieve a satisfactory product. No incompatibilities should occur between the various emulsifiers and the several components commonly present in pharmaceutical emulsions. Finally, the product should be prepared in such a way as not to prejudice the formulation.

SELECTION OF EMULSIFYING AGENTS

The selection of the emulsifying agent or agents is of prime importance in the successful formulation of an emulsion. The pharmacist must ensure that, in addition to its emulsifying properties, the material chosen is nontoxic and that the taste, odor, and chemical stability are compatible with the product. Thus, an emulsifying agent that is entirely suitable for inclusion in a skin cream may be unacceptable in the formulation of an oral preparation due to its potential toxicity. This consideration is most important when formulating intravenous emulsions.

The HLB System

With the increasing number of available emulsifiers, particularly the nonionics, the selection of emulsifiers for a product was essentially a trial-and-error procedure. Fortunately, the work of Griffin^{17,18} provided a logical means of selecting emulsifying agents. Griffin's method, based on the balance between the hydrophilic and lipophilic portions of the emulsifying agent, is now widely used and has come to be known as the HLB system. It is used most in the rational selection of combinations of non-ionic emulsifiers, and we shall limit our discussion accordingly.

As shown in Table 21-8, if an O/W emulsion is required, the formulator should use emulsifiers with an HLB in the range of 8 to 18. Emulsifiers with HLB values in the range of 4 to 6 are given consideration when a W/O emulsion is desired. Some typical examples are given in Table 21-9.

Table 21-8. Relationship between HLB Range and Surfactant Application

HLB Range	Use
0-3	Antifoaming agents
4-6	W/O emulsifying agents
7-9	Wetting agents
8-18	O/W emulsifying agents
13-15	Detergents
10-18	Solubilizing agents

Table 21-9. Approximate HLB Values for a Number of Emulsifying Agents

Generic or Chemical Name	HLB	Water Dispersibility
Sorbitan trioleate	1.8	No dispersion
Sucrose distearate	3.0	
Propylene glycol monostearate	3.4	
Glycerol monostearate (non-self-emulsifying)	3.8	Poor dispersion
Propylene glycol monolaurate	4.5	
Sorbitan monostearate	4.7	
Glycerol monostearate (self-emulsifying)	5.5	
Sorbitan monolaurate	8.6	Milky dispersion
Polyoxyethylene-4-lauryl ether	9.5	
Polyethylene glycol 400 monostearate	11.6	Translucent to clear
Polyoxyethylene-4-sorbitan monolaurate	13.3	Clear solution
Sucrose stearate	14.5	
Polyoxyethylene-20-sorbitan monopalmitate	15.6	
Polyoxyethylene-40-stearate	16.9	
Sodium oleate	18.0	
Sodium lauryl sulfate	40.0	

Table 21-10. Required HLB Values for Some Common Emulsion Ingredients

Substance W/OO/W		
Acid, stearic	—	17.0
Alcohol, cetyl	—	13.0
Lanolin, anhydrous	8	15.0
Oil, cottonseed	—	7.5
Mineral oil, light	4	10–12.0
Mineral oil, heavy	4	10.5
Wax, beeswax	5	10–16.0
Microcrystalline	—	9.5
Paraffin	—	9.0

Another factor is the presence or absence of any polarity in the material being emulsified, because this will affect the polarity required in the emulsifier. Again, as a result of extensive experimentation, Griffin evolved a series of “required HLB” values—that is, the HLB value required by a particular material if it is to be emulsified effectively. Some values for oils and related materials are contained in Table 21-10. Naturally, the required HLB value differs depending on whether the final emulsion is O/W or W/O.

Fundamental to the utility of the HLB concept is the fact that the HLB values are algebraically additive. Thus, by using a low HLB surfactant with one having a high HLB it is possible to prepare blends having HLB values intermediate between those of the two individual emulsifiers. The following formula serves as an example.

O/W EMULSION

Liquid petrolatum (Required HLB 10.5)	50 g
Emulsifying agents	5 g
Sorbitan monooleate	(HLB 4.3)
Polyoxyethylene 20 sorbitan monooleate	(HLB 15.0)
Water, qs.	100 g

Table 21-11. Nonionic Blends Having HLB Values of 10.5

Surfactant Blend	HLB	Required Amounts (%) to Give HLB = 10.5
Sorbitan tristearate	2.1	34.4
Polyoxyethylene 20 sorbitan monostearate	14.9	65.6
Sorbitan monopalmitate	6.7	57.3
Polyoxyethylene 20 sorbitan monopalmitate	15.6	42.7
Sorbitan sesquioleate	3.7	48.5
Polyoxyethylene lauryl ether	16.9	51.5

By simple algebra it can be shown that a mixture containing 42% of sorbitan monooleate blended with 58% of polyoxyethylene 20 sorbitan monooleate will result in a mixed emulsifying agent having the required HLB of 10.5. Because the formula calls for 5 g, the required weights are 2.1 and 2.9 g, respectively. The oil-soluble sorbitan monooleate is dissolved in the oil and heated to 75°; the water-soluble polyoxyethylene 20 sorbitan monooleate is added to the aqueous phase that is heated to 70°. At this point the oil phase is mixed with the aqueous phase and the whole is stirred continuously until cool.

The formulator is not restricted to these two agents to produce a blend with an HLB of 10.5. Table 21-11 shows the various proportions required, using other pairs of emulsifying agents, to form a blend of HLB 10.5. When carrying out preliminary investigations with a particular material to be emulsified, it is advisable to try several pairs of emulsifying agents. Based on an evaluation of the emulsions produced, it becomes possible to choose the best combination.

Occasionally, the required HLB of the oil may not be known, in which case it becomes necessary to determine this parameter. Various blends are prepared to give a wide range of HLB mixtures and emulsions are prepared in a standardized manner. The HLB of the blend used to emulsify the best product, selected on the basis of physical stability, is taken to be the required HLB of the oil. The experiment should be repeated using another combination of emulsifiers to confirm the value of the required HLB of the oil to within, say, ± 1 HLB unit.

There are methods for finding the HLB value of a new surface-active agent. Griffin¹⁷ developed simple equations that can be used to obtain an estimate with certain compounds. It has been shown that the ability of a compound to spread at a surface is related to its HLB. In another approach a linear relation between HLB and the logarithm of the dielectric constant for a number of nonionic surfactants has been observed.

An interesting approach, developed by Davies,¹⁴ is related to his studies on the relative rates of coalescence of O/W and W/O emulsions (page 329). According to Davies, hydrophilic groups on the surfactant molecule make a positive contribution to the HLB number, whereas lipophilic groups exert a negative effect. Davies calculated these contributions and termed them HLB Group Numbers (Table 21-12). Provided the molecular structure of the surfactant is known, one simply adds the various group numbers in accordance with the following formula:

$$\text{HLB} = \sum (\text{hydrophilic group numbers}) - \sum (\text{lipophilic group numbers}) + 7$$

Poor agreement is found between the HLB values calculated by the use of group numbers and the HLB values obtained using the simple equations developed by Griffin. However, the student should realize that the absolute HLB values per

Table 21-12. HLB Group Numbers

	Group Number
Hydrophilic groups	
—SO ₄ ^{&-} Na ^{&+}	38.7
—COO ^{&-} K ^{&+}	21.1
—COO ^{&-} Na ^{&+}	19.1
N (tertiary amine)	9.4
Ester (sorbitan ring)	6.8
Ester (free)	2.4
—COOH	2.1
Hydroxyl (free)	1.9
—O—	1.3
Hydroxyl (sorbitan ring)	0.5
Lipophilic groups	
—CH—	
—CH ₂ —	
CH ₃ —	-0.475
—CH—	
Derived groups	
—(CH ₂ —CH ₂ —O)—	+0.33
—(CH ₂ —CH ₂ —CH ₂ —O)—	-0.15

(From Wedderburn DL. In: *Advances in Pharmaceutical Sciences*, vol 1. London: Academic Press, 1964, p 195.)

se are of limited significance. The utility of the HLB approach (using values calculated by either Griffin's or Davies' equations) is to:

1. provide the formulator with an idea of the relative balance of hydrophilicity and lipophilicity in a particular surfactant.
2. relate that surfactant's emulsifying and solubilizing properties to other surfactants. The formulator still needs to confirm experimentally that a particular formulation will produce a stable emulsion.

Later, Davies and Rideal¹⁹ attempted to relate HLB to the $C_{\text{water}}/C_{\text{oil}}$ partition coefficient and found good agreement for a series of sorbitan surfactants. Schott showed, however, that the method does not apply to polyoxyethylated octylphenol surfactants. Schott concluded that "so far, the search for a universal correlation between HLB and another property of the surfactant that could be determined more readily than HLB has not been successful."²⁰

The HLB system gives no information as to the *amount* of emulsifier required. Having once determined the correct blend, the formulator must prepare another series of emulsions, all at the same HLB, but containing increasing concentrations of the emulsifier blend. Usually, the minimum concentration giving the desired degree of physical stability is chosen.

When varying the amounts of emulsifier in an emulsion it is useful to consider the use of a phase diagram to select the proper ratio of oil/water/surfactant. The use of the phase diagram to aid in the formulation of emulsions has been discussed by Swarbrick.²¹ This approach can provide a systematic way to optimize an emulsion formulation and help to identify the existence of liquid crystalline phases that, when present in an emulsion formulation, can enhance the stability. Because liquid crystals exhibit birefringence, observation of prototype emulsions under polarized light microscopy can be a useful tool to identify combinations of water-oil and emulsifier that produce liquid crystals. It should be noted that liquid crystals are often formed when relatively high concentrations (e.g., 20% or more) of surfactant are used in a formulation. The toxicity of the

emulsifier for the intended use (e.g., topical, oral, or parenteral) must be considered in addition to the physical characteristics.

Mixed Emulsifying Agents

Emulsifying agents are frequently used in combination because a better emulsion usually is obtained. This enhancement may be due to several reasons, one or more of which may be operative in any one system. Thus, the use of a blend or mixture of emulsifiers may

1. produce the required hydrophile-lipophile balance in the emulsifier.
2. enhance the stability and cohesiveness of the interfacial film.
3. affect the consistency and feel of the product.

The first point has been considered in detail in the previous discussion of the HLB system.

With regard to the second point, Schulman and Cockbain in 1940 showed that combinations of certain amphiphiles formed stable films at the air-water interface. It was postulated that the complex formed by these two materials (one, oil-soluble; the other, water-soluble) at the air-water interface was also present at the O/W interface. This interfacial complex was held to be responsible for the improved stability. For example, sodium cetyl sulfate, a moderately good O/W emulsifier, and elaidyl alcohol or cholesterol, both stabilizers for W/O emulsions, show evidence of an interaction at the air-water interface. Furthermore, an O/W emulsion prepared with sodium cetyl sulfate and elaidyl alcohol is much more stable than an emulsion prepared with sodium cetyl sulfate alone.

Elaidyl alcohol is the *trans* isomer. When oleyl alcohol, the *cis* isomer, is used with sodium cetyl sulfate, there is no evidence of complex formation at the air-water interface. Significantly, this combination does not produce a stable O/W emulsion either. Such a finding strongly suggests that a high degree of molecular alignment is necessary at the O/W interface to form a stable emulsion. This high degree of molecular alignment may be a prerequisite event for the formation of lamellar liquid crystalline or gel phases. As illustrated in Figure 21-7, the combination of certain long chain acids and alcohols with water can result in the formation of micelles and liquid crystals. It has also been observed that when liquid crystals or gels form in an emulsion, increased stability is generally observed. As discussed

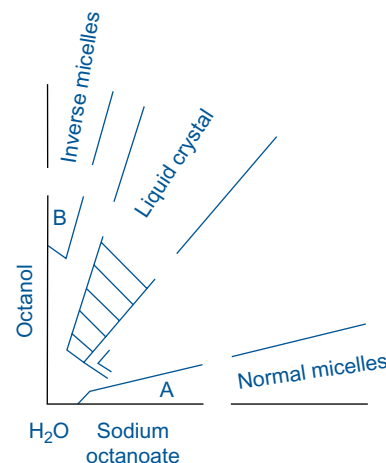


Figure 21-7. Phase diagram illustrating the formation of micellar and liquid crystalline phases in mixtures of a long-chain alcohol, long-chain acid, and water. Compositions that form the lamellar liquid crystalline phase can provide enhanced emulsion stability. (From Friberg S, Larson K. In: Brown GH, ed. *Advances in Liquid Crystals*, vol 2. New York: Academic Press, 1976, p 173.)

previously, gel or liquid crystalline phases can have an important effect in inhibiting coalescence in emulsions.

When using combinations of emulsifiers, care must be taken to ensure their compatibility, as charged emulsifying agents of opposite sign are likely to interact and coagulate when mixed.

Steric Stabilization

Many useful nonionic surfactants consist of hydrophobic portions composed of fatty acids or other lipophilic organic compounds and hydrophilic portions composed of polyoxyethylene chains. When used to prepare O/W emulsions, the oxyethylene chains protrude into the aqueous side of the O/W interface while the hydrophobic portion of the emulsifier will be primarily located in the oil side. As in the case of suspensions, approaching oil droplets will be influenced by van der Waals attractive forces as well as repulsive forces. For an emulsion that is stabilized by a non-ionic surfactant, the repulsive forces consist of electrostatic and non-electrostatic forces. The electrostatic repulsive forces are similar to those discussed for suspensions and depend largely upon the zeta potential of the oil droplets.

Non-electrostatic forces may also arise from a phenomenon that is frequently described as *steric stabilization*. This effect has been explained as follows. First, as emulsion droplets approach, the adsorbed layers of surfactant on each droplet begin to mix. The hydrophilic oxyethylene chains behave as soluble polymers; as their concentration increases in the region of interfacial mixing, segments of the polymers from separate droplets compete for water molecules. This results in restricted movement of the polymer chains or a loss of entropy. Likewise, a positive heat of solution (enthalpy) may result from the mixing of the polymers in the interfaces. The loss of entropy and/or increase in enthalpy results in an increase in the free energy of mixing, meaning that spontaneous mixing in the interfacial region is not favorable. The particles will tend to separate in order to reverse the temporary increase in the free energy of mixing.

An additional effect that causes repulsion of the droplets may be a result of the increased osmotic pressure that results in the area of contact between the two emulsion droplets. The concentration of oxyethylene groups in the region of overlap between the two droplets increases, necessitating an influx of water into the region. This increase in osmotic pressure has the effect of forcing the droplets apart. Thus, in addition to their favorable effect of reducing interfacial tension, nonionic surfactants that possess long, hydrophilic chains provide additional emulsion stabilization via the energetically unfavorable result of mixing of polymer chains at the droplet-droplet interface.

METHOD OF PREPARATION

Different methods are employed, depending on the type of emulsifying agent used and the scale of manufacture. Traditionally, the mortar and pestle was used for the small scale preparation of emulsions stabilized by the presence of such agents as acacia and tragacanth. However, the use of these agents has declined drastically in recent years; as a result, the use of the mortar and pestle has declined as well. (Refer to the 18th edition of this text, page 306, for details of the mortar and pestle method.)

An increasing number of emulsions are being formulated with synthetic emulsifying agents, especially of the nonionic type. The components in such a formulation are separated into those that are oil-soluble and those that are water-soluble. These are dissolved in their respective solvents by heating to about 70° to 75°. When solution is complete, the two phases are mixed and the product is stirred until cool. This method, which requires nothing more than two beakers, a thermometer, and a source of heat, is necessarily used in the preparation of emulsions containing waxes and other high-melting-point materials that must be melted before they can be dispersed in the emulsion. The relatively simple methodology involved in the use of synthetic surfactant-type emulsifiers is one factor that has led to their

widespread use in emulsion preparation. This, in turn, has led to a decline in the use of the natural emulsifying agents.

With hand homogenizers, an initial rough emulsion is formed by trituration in a mortar or shaking in a bottle. The rough emulsion then is passed several times through the homogenizer. A reduction in particle size is achieved as the material is forced through a narrow aperture under pressure. A satisfactory product invariably results from the use of a hand homogenizer and overcomes any deficiencies in technique. Should the homogenizer fail to produce an adequate product, the formulation, rather than the technique, should be suspected.

For a discussion of the techniques and equipment used in the large-scale manufacture of emulsions, see Chapter 24.

STABILITY OF EMULSIONS

Several criteria must be met in a well-formulated emulsion. Probably the most important and most readily apparent requirement is that the emulsion possess adequate physical stability; without this, any emulsion soon will revert back to two separate bulk phases. In addition, if the emulsified product is to have some antimicrobial activity (e.g., a medicated lotion), care must be taken to ensure that the formulation possesses the required degree of activity. Frequently, a compound exhibits a lower antimicrobial activity in an emulsion than, say, in a solution. Generally, this is because of partitioning effects between the oil and water phases, which cause a lowering of the *effective* concentration of the active agent. Partitioning has also to be taken into account when considering preservatives to prevent microbiological spoilage of emulsions. Finally, the chemical stability of the various components of the emulsion should receive some attention, as such materials may be more prone to degradation in the emulsified state than when they exist as a bulk phase.

In the present discussion, detailed consideration will be limited to the question of physical stability. Reviews of this topic have been published by Garrett²² and Kitchener and Mussellwhite.²³ For information on the effect that emulsification can have on the biologic activity and chemical stability of materials in emulsions, see Wedderburn,²⁴ Burt²⁵ and Swarbrick.²¹

The theories of emulsion stability have been discussed by Eccleston²⁶ in an attempt to understand the situation in both a simple O/W emulsion and complex commercial systems. A recent review by the same author¹³ has discussed the stability of multiple phase emulsions and the role of bilayer gels and liquid crystalline phases on the physical stability of these systems.

The three major phenomena associated with physical stability are

1. the upward or downward movement of dispersed droplets relative to the continuous phase, termed *creaming* or *sedimentation*, respectively.
2. the *aggregation* and possible *coalescence* of the dispersed droplets to reform the separate, bulk phases.
3. *inversion*, in which an O/W emulsion inverts to become a W/O emulsion and *vice versa*.

CREAMING AND SEDIMENTATION

Creaming is the upward movement of dispersed droplets relative to the continuous phase; *sedimentation*, the reverse process, is the downward movement of particles. In any emulsion one process or the other takes place, depending on the densities of the disperse and continuous phases. This is undesirable in a pharmaceutical product where homogeneity is essential for the administration of the correct and uniform dose. Furthermore, creaming, or sedimentation, brings the particles closer together and may facilitate the more serious problem of coalescence.

The rate at which a spherical droplet or particle sediments in a liquid is governed by Stokes' law (Equation 3). Other equations have been developed for bulk systems, but Stokes' equation is still useful because it points out the factors that influence

the rate of sedimentation or creaming. These are the diameter of the suspended droplets, the viscosity of the suspending medium, and the difference in densities between the dispersed phase and the dispersion medium.

Usually, only the use of the first two factors is feasible in affecting creaming or sedimentation. Reduction of particle size contributes greatly toward overcoming or minimizing creaming, because the rate of movement is a square-root function of the particle diameter. There are, however, technical difficulties in reducing the diameter of droplets to below about 0.1 μm . The most frequently used approach is to raise the viscosity of the continuous phase, although this can be done only to the extent that the emulsion still can be removed readily from its container and spread or administered conveniently.

AGGREGATION AND COALESCENCE

Even though creaming and sedimentation are undesirable, they do not necessarily result in the breakdown of the emulsion, as the dispersed droplets retain their individuality. Furthermore, the droplets can be redispersed with mild agitation. More serious to the stability of an emulsion are the processes of aggregation and coalescence. In *aggregation* (flocculation) the dispersed droplets come together but do not fuse. *Coalescence*, the complete fusion of droplets, leads to a decrease in the number of droplets and the ultimate separation of the two immiscible phases. Aggregation precedes coalescence in emulsions; however, coalescence does not necessarily follow from aggregation. Aggregation is, to some extent, reversible. Although it is not as serious as coalescence, it will accelerate creaming or sedimentation, because the aggregate behaves as a single drop.

Aggregation is related to the electrical potential on the droplets, but coalescence depends on the structural properties of the interfacial film. As discussed previously, it has been recognized that combinations of emulsifiers produce more stable emulsions than a single emulsifier alone. One reason for this synergy, as suggested by Shulman and Cockbain, is that appropriate combinations of surfactants form densely packed complex films at the oil–water interface. Additional beneficial effects of mixed emulsifier films could result from an increase in viscosity of the interfacial emulsifier film. A viscous interfacial film could enhance emulsion stability because thinning of the film at the points of droplet to droplet contact would be inhibited. An additional explanation for the beneficial effect of mixed-film emulsifiers suggests that appropriate mixtures of surfactants provide a more elastic interfacial film. A more elastic interfacial film would resist rupture upon collision of emulsion droplets.

It has also been observed that when emulsifiers are combined in certain concentrations and proportions, liquid crystalline phases can be formed. The preparation of emulsions with surfactants that form liquid crystalline states can have greater stability against coalescence compared to emulsions that are formulated in the absence of liquid crystalline states. Friberg and Larson²⁷ have explained the enhanced stability of emulsions due to liquid crystals in terms of a reduced van der Waals attraction between emulsion droplets. Such an effect depends upon the formation of layers or lamellae around the emulsion droplets. Each layer of liquid crystal contributes to a further reduction in the van der Waals attractive force.

An additional effect of liquid crystals may be related to the high viscosity that often is observed upon their formation. Liquid crystals possess a viscosity that is on the order of 100-fold greater than most oil–water interfaces. The high viscosity may result in reduced rates of coalescence. A key factor that may be important for the stabilizing effect of liquid crystals is the location of the liquid crystalline phase in relation to the dispersed droplets. To effectively inhibit coalescence, the liquid crystals should concentrate at the interface between the droplet and the continuous phase. This may not occur with all oil–water–surfactant combinations.

Particle-size analysis can reveal the tendency of an emulsion to aggregate and coalesce long before any visible signs of

instability are apparent. The methods available have been reviewed by Groves and Freshwater.²⁸

INVERSION

An emulsion is said to *invert* when it changes from an O/W to a W/O emulsion, or *vice versa*. Inversion sometimes can be brought about by the addition of an electrolyte or by changing the phase-volume ratio. For example, an O/W emulsion having sodium stearate as the emulsifier can be inverted by the addition of calcium chloride, because the calcium stearate formed is a lipophilic emulsifier and favors the formation of a W/O product.

Inversion often can be seen when an emulsion, prepared by heating and mixing the two phases, is being cooled. This takes place presumably because of the temperature-dependent changes in the solubilities of the emulsifying agents. The phase inversion temperature (PIT) of nonionic surfactants has been shown by Shinoda and Kunieda²⁹ to be influenced by the HLB number of the surfactant—the higher the PIT value, the greater the resistance to inversion.

Apart from work on PIT values, little quantitative work has been carried out on the process of inversion; nevertheless, it would appear that the effect can be minimized by using the proper emulsifying agent in an adequate concentration. Wherever possible, the volume of the dispersed phase should not exceed 50% of the total volume of the emulsion.

BIOAVAILABILITY FROM COARSE DISPERSIONS

All dosage forms must be capable of releasing the drug in a known and consistent manner following administration to the patient. Both the rate and extent of release are important. Ideally, the extent of release should approach 100%, while the rate of release should reflect the desired properties of the dosage form. For example, with products designed to have a rapid onset of activity, the release of drug should be immediate. With a long-acting product, the release should take place over several hours or days, depending on the type of product used. The rate and extent of drug release should be reproducible from batch to batch of the product, and should not change during shelf-life.

The principles on which biopharmaceutics is based are dealt with in some detail in Chapters 38 to 39. Although most published work in this area has been concerned with the bioavailability of solid dosage forms administered by the oral route, the rate and extent of release from both suspensions and emulsions are also important and so must be considered in some detail.

BIOAVAILABILITY FROM SUSPENSIONS

Suspensions of a drug may be expected to demonstrate improved bioavailability compared to the same drug formulated as a tablet or capsule. This is because the suspension already contains discrete drug particles, whereas tablet dosage forms must invariably undergo disintegration in order to maximize the necessary dissolution process. Frequently, antacid suspensions are perceived as being more rapid in action and therefore more effective than an equivalent dose in the form of tablets. Bates *et al.*³⁰ observed that a suspension of salicylamide was more rapidly bioavailable, at least during the first hour following administration, than two different tablet forms of the drug; this study was also able to demonstrate a correlation between the initial *in vitro* dissolution rates for the several dosage forms studied and the initial rates of *in vivo* absorption. A similar argument can be developed for hard gelatin capsules, where the shell must rupture or dissolve before drug particles are released and can begin the dissolution process. Such was observed by Antal *et al.*³¹ in a study of the bioavailability of several doxycycline products, including a suspension and hard gelatin capsules. Sansom *et al.*³² found that mean plasma phenytoin levels were higher after the administration of a suspension than when an equivalent dose was given as either tablets or capsules. It

was suggested that this might have been due to the suspension having a smaller particle size.

In common with other products in which the drug is present in the form of solid particles, the rate of dissolution, and thus potentially the bioavailability of the drug in a suspension, can be affected by such factors as particle size and shape, surface characteristics, and polymorphism. Strum *et al.*³³ conducted a comparative bioavailability study involving two commercial brands of sulfamethiazole suspension (Product A and Product B). Following administration of the products to 12 normal individuals and blood samples taken at predetermined times over a period of 10 h, the Strum study found no statistically significant difference in the extent of drug absorption from the two suspensions. The absorption rate, however, differed, and from *in vitro* studies it was concluded that product A dissolved faster than Product B, and that the former contained more particles of smaller size than the latter, differences that may be responsible for the more rapid dissolution of particles in Product A. Product A also provided higher serum levels during *in vivo* tests 0.5 h after administration. The results showed that the rate of absorption of sulfamethiazole from a suspension depended on the rate of dissolution of the suspended particles, which in turn was related to particle size. Previous studies^{34,35} had shown the need to determine the dissolution rate of suspensions to gain information as to the bioavailability of drugs from this type of dosage form.

The viscosity of the vehicle used to suspend the particles has been found to have an effect on the rate of absorption of nitrofurantoin but not the total bioavailability. Thus Soci and Parrott³⁶ were able to maintain a clinically acceptable urinary nitrofurantoin concentration for an additional 2 h by increasing the viscosity of the vehicle.

BIOAVAILABILITY FROM EMULSIONS

There are indications that improved bioavailability may result when a poorly absorbed drug is formulated as an orally administered emulsion. However, little research appears to have been done to directly compare emulsions and other dosage forms such as suspensions, tablets, and capsules; thus, it is not possible to draw unequivocal conclusions as to advantages of emulsions. If a drug with low aqueous solubility can be formulated so as to be in solution in the oil phase of an emulsion, its bioavailability may be enhanced. It must be recognized, however, that the drug in such a system has several barriers to pass before it arrives at the mucosal surface of the GI tract.

For example, with an O/W emulsion, the drug must diffuse through the oil globule and then pass across the oil-water interface. This may be a difficult process, depending on the characteristics of the interfacial film formed by the emulsifying agent. In spite of this potential drawback, Wagner *et al.*³⁷ found that indoxole, a nonsteroidal anti-inflammatory agent, was significantly more bioavailable in an O/W emulsion than in either a suspension or a hard gelatin capsule. Bates and Sequeira³⁸ found significant increases in maximum plasma levels and total bioavailability of micronized griseofulvin when formulated in a corn O/W emulsion. In this case, however, the enhanced effect was not due to emulsification of the drug in the oil phase *per se*, but more probably because of the linoleic and oleic acids present having a specific effect on GI motility.

QUALITY CONTROL TESTING FOR SUSPENSIONS AND EMULSIONS

Most dosage forms containing active pharmaceutical ingredients require chemical and physical tests to ensure potency, purity and stability of the active pharmaceutical ingredient. Suspension and emulsion dosage forms typically require various physical tests to ensure consistent particle size and distribution as well as in-use tests to ensure that patients will be able to use the drug product effectively over the shelf-life of the product. Suspensions and fluid emulsions are generally tested for average particle size as well as some measure of particle size distribution. The latter could be a statistical parameter, such as standard deviation, but may also

be a limit on the upper or lower particle size, such as "99% of particles must be less than X micrometers." Rheological properties such as viscosity are also useful because they indicate ease of use, such as spreadability for a topical emulsion or suspension. For a parenteral suspension or emulsion, ease of injectability may also be a valuable test because the product may need to consistently be injected through a 21–27 Ga needle. Changes in viscosity, particle size or floccule size may alter the ability to effectively use the product. Other physical tests may include sedimentation rate/volume, zeta potential, redispersibility, pH, color, and possibly microscopic appearance. Microbiological testing may also be conducted, not only for sterile suspensions and emulsions, but whenever the product is capable of supporting microbial growth. When an antimicrobial preservative is required for the product, preservative effectiveness testing should be conducted and the product must be tested to be sure adequate concentrations of the preservative are present throughout the product shelf-life. It is also important to demonstrate dose uniformity of the suspension or emulsion by testing the drug content of aliquots of product that simulate dose volumes used by the patient. For suspension dosage forms, a dissolution test will most likely need to be performed because dissolution rate is an important step controlling rate and extent of absorption for poorly water soluble drugs.

As with other pharmaceutical dosage forms, chemical, physical and microbiological tests are performed at initial release of the product and during product stability studies. Key physical tests, such as particle size and viscosity, can also be conducted during manufacture of the product to ensure that the manufacturing process is producing product with acceptable physical characteristics.

REFERENCES

1. Hiestand EN. Theory of coarse suspension formulation. *J Pharm Sci* 1964; 53: 1–18.
2. Haines BA, Martin A. Interfacial properties of powdered material; caking in liquid dispersions. *J Pharm Sci* 1961; 50: 228–232.
3. Matthews BA, Rhodes CT. Coagulation and flocculation in suspensions of griseofulvin and polystyrene latex. *J Pharm Pharmacol* 1968; 20(Suppl.): 204S.
4. Matthews BA, Rhodes CT. Some studies of flocculation phenomena in pharmaceutical suspensions. *J Pharm Sci* 1968; 57(4): 569–573.
5. Matthews BA, Rhodes CT. Use of the Derjaguin, Landau, Verwey, and Overbeek theory to interpret pharmaceutical suspension stability. *J Pharm Sci* 1970; 59(4): 521–5.
6. Schneider W *et al.* Pharmaceutical suspensions and the DLVO them. *Am J Pharm Ed* 1978; 42: 280–289.
7. Scheer AJ. *Drug Cosmet Ind* 1981 (Apr): 40.
8. Kellaway IW, Najib NM. Hydrophilic polymers as stabilisers and flocculants of sulphadimidine suspensions. *Int J Pharm* 1981; 9: 59.
9. Martin AN *et al.* *Physical Pharmacy*, 3rd edn. Philadelphia: Lea & Febiger, 1983: 551.
10. Parsons GE *et al.* The use of surface energy and polarity determinations to predict physical stability of non-polar, non-aqueous suspensions. *Int J Pharm* 1992; 83: 163–170.
11. Tingstad J *et al.* Simplified method for determining chemical stability of drug substances in pharmaceutical suspensions. *J Pharm Sci* 1973; 62(8): 1361–1363.
12. Allen LV, Allen's *Compounded Formulations: The complete U.S. Pharmacist Collection*. Washington D.C.: American Pharmaceutical Association, 2003.
13. Eccleston GM. In: Swarbrick J, Boylan JC eds. *Encyclopedia of Pharmaceutical Technology*, vol 5. New York: Dekker, 1992: 137–188.
14. Davies JT. A quantitative kinetic theory of emulsion type I. Physical chemistry of the emulsifying agent. In: Gas/Liquid and Liquid/Liquid Interfaces. *Proceedings of the 2nd International Congress on Surface Activity*. London: Butterworth/Academic, 1957: 426–438.

15. Sherman P. Rheology of emulsions. In: Sherman P, ed. *Emulsion Science*. New York: Academic Press, 1968: Chap 4.
16. Rogers JA. *Cosmet Toiletries* 1978; 93 (7): 29.
17. Griffin WC. Classification of surface-active agents by HLB. *J Soc Cosmet Chem* 1949; 1: 311–326.
18. Griffin WC. Calculation of HLB values of non-ionic surfactants. *J Soc Cosmet Chem* 1954; 5: 249–256.
19. Davies JT, Rideal EK. *Interfacial Phenomena*. New York: Academic Press, 1961: Chapter 8.
20. Schott J. Hydrophilic-lipophilic balance and distribution coefficients of non-ionic surfactants. *J Pharm Sci* 1971; 60: 648–649.
21. Swarbrick J. Phase equilibrium diagrams: an approach to the formulation of solubilized and emulsified systems. *J Soc Cosmet Chem* 1968; 19: 187–209.
22. Garrett ER. Stability of oil-in-water emulsions. *J Pharm Sci* 1965; 54: 1557–1570.
23. Kitchener JA, Mussellwhite PR. The theory of stability of emulsions. In: Sherman P, ed. *Emulsion Science*. New York: Academic Press, 1968, Chapter 2.
24. Wedderburn DL. Preservation of emulsions against microbial attack. In: Bean HA et al. eds. *Advances in Pharmaceutical Sciences*, vol 1. London: Academic Press, 1964: 195–268.
25. Burt BW. An Approach to Emulsion Formulation. *J Soc Cosmet Chem* 1965; 16: 465–477.
26. Eccleston GM. *Cosmet Toiletries* 1986; 101 (11): 73.
27. Friberg S, Larson K. In: Brown GH, ed. *Advances in Liquid Crystals*, vol 2. New York: Academic Press, 1976: 173.
28. Groves MJ, Freshwater DC. Particle-size analysis of emulsion systems. *J Pharm Sci* 1968; 57(8): 1273–1291.
29. Shinoda K, Kunieda H. In: Becher P ed. *Encyclopedia of Emulsion Technology*. New York: Dekker, 1983: Chapter 5.
30. Bates TR et al. Correlation between the rate of dissolution and absorption of salicylamide from tablet and suspension dosage forms. *J Pharm Sci* 1969; 58: 1468–1470.
31. Antal EJ et al. Bioequivalency of doxycycline products. *J Pharm Sci* 1975; 64: 2015–2018.
32. Sansom LN et al. Plasma phenytoin levels produced by various phenytoin preparations. *Med J Aust* 1975; 2: 593–595.
33. Meyer MC et al. Bioavailability of sulfadiazine solutions, suspensions, and tablets in humans. *J Pharm Sci* 1978; 67(12): 1659–1661.
34. Bates TR et al. Inconsistencies in rationale underlying official USP dissolution rate specifications for nitrofurantoin. *J Pharm Sci* 1973; 62: 2057–2058.
35. Howard SA et al. Dissolution profiles for multisized prednisolone acetate suspensions. *J Pharm Sci* 1977; 66: 557–559.
36. Soci MM, Parrott EL. Influence of viscosity on absorption from nitrofurantoin suspensions. *J Pharm Sci* 1980; 69: 403–406.
37. Wagner JG et al. The effect of the dosage form on serum levels of indoxole. *Clin Pharmacol Ther* 1966; 7: 610–619.
38. Bates TR, Sequeira JA. Bioavailability of micronized griseofulvin from corn oil-in-water emulsion, aqueous suspension, and commercial tablet dosage forms in humans. *J Pharm Sci* 1975; 64: 793–797.

BIBLIOGRAPHY

- Adamson AW. *Physical Chemistry of Surfaces*, 4th edn. New York: Wiley-Interscience, 1980.
- Attwood D, Florence AT. In: *Surfactant Systems; Their Chemistry, Pharmacy and Biology*. London: Chapman & Hall, 1983: 469.
- Becher P. *Emulsions: Theory and Practice*, 2nd edn. New York: Reinhold, 1965.
- Becher P. *Encyclopedia of Emulsion Technology*, vols 1–3. New York: Dekker, 1983–1988.
- Davies JT, Rideal EK. *Interfacial Phenomena*. New York: Academic Press, 1963.
- Eccleston GM. In: *Encyclopedia of Pharmaceutical Technology*, vol 5. New York: Dekker, 1992:137.
- Hiemenz PC. *Principles of Colloidal and Surface Chemistry*, 2nd edn. New York: Dekker, 1986.
- Matijevic E, ed. *Surface and Colloid Science*, vols 1–4. New York: Wiley, 1971.
- Osipow LI. *Surface Chemistry*. New York: Reinhold, 1962.
- Parfitt G. *Dispersion of Powders in Liquids*. New York: Applied Science, 1973.
- Sherman P. *Emulsion Science*. New York: Academic Press, 1964.
- Sherman P. *Rheology of Emulsions*. New York: Macmillan, 1963.
- Vold RD, Vold MJ. *Colloid and Interface Chemistry*. Reading MA: Addison-Wesley, 1983.

Chapter 22

Rheology

Lawrence H. Block, PhD

INTRODUCTION 393

FUNDAMENTALS 393

ELASTIC SOLIDS 395

NEWTONIAN FLUIDS 395

NON-NEWTONIAN FLUIDS 397

OTHER EMPIRICAL EQUATIONS AND MODELS 402

COMPUTATIONAL RHEOLOGY 404

RHEOLOGICAL MEASUREMENTS 405

GENERAL VISCOMETER AND RHEOMETER TYPES 405

CAPILLARY AND SLIT VISCOMETERS AND RHEOMETERS 406

ROTATIONAL VISCOMETERS AND RHEOMETERS 407

FALLING BALL AND ROLLING BALL VISCOMETERS AND RHEOMETERS 408

COMPARISON BETWEEN INSTRUMENTS 409

NON-NEWTONIAN CORRECTIONS 409

INTRODUCTION

Rheology is the branch of physics that deals with deformation, including flow, of matter. Although this definition was proposed in 1929, the recognition of rheological phenomena dates back to antiquity.¹ The earliest application of rheology (ca. 1600 BCE) is associated with the Egyptian Amenemhet, who made a 7° correction to the drainage angle of a water clock in order to account for the temperature-dependent variation in water flow during the course of a day. Archimedes's claim (ca. 250 BCE)—“Give me but one firm spot on which to stand, and I will move the earth.”—was based on the application of solid mechanics, the oldest branch of the physical sciences.¹

Reiner² describes a simple mechanical experiment in which he lets three different materials—a pencil, a ball of plasticine, and a known mass of water—fall from some height onto the surface of a table. Newton's second law tells us that $F = ma$, where F is the force acting upon each of these materials of mass m , and a is the acceleration of the center of mass of each material. Since F is proportional to m , a is the same for each of these materials. Consequently, these three bodies fall toward the table in exactly the same manner. Their material differences do not become apparent until they reach the tabletop. At that point, the pencil rebounds somewhat, the plasticine stays put, and the water spreads over the tabletop and, on reaching the edge, flows off. These very different outcomes—which mechanics is unable to explain—are the focus of rheology.

The ubiquity of rheological phenomena in pharmacy is evident in the levigation or mixing of ointments on slabs, the use of a mortar and pestle to prepare suspensions and emulsions, the flow of emulsions through colloid mills and pumps, the use of roller mills for compacting powders or processing ointments, and the mechanical properties of glass or plastic containers and of rubber or polymeric closures. Squeezing ointments, creams, or toothpaste from a collapsible tube, spreading lotion on the skin, or spraying liquids from atomizers or aerosol cans all involve rheological phenomena. The fluidity of solutions to be injected by syringe or infused intravenously, the flexibility of tubing used in catheters, and the strength of sutures and ligatures are important rheological properties. Drug release from dosage forms and delivery systems is often controlled or modulated by the rheological properties of the formulation matrix. Diffusion—a process occurring at the molecular level—is governed, in part, by the rheological behavior of the molecule's environment. Rheological principles govern the circulation of blood and lymph through capillaries and large vessels, the flow of mucus, the transit of the luminal contents through the gastrointestinal

tract, the bending of bones, the stretching of cartilage, and the contraction of muscles.

The centrality of rheological behavior to many unit operations in pharmaceutical manufacturing, to drug product functionality and stability, and to patient or consumer use of dosage forms and delivery systems necessitates a thorough understanding of rheology and the measurement of rheological behavior. The fundamentals of rheology are presented in the following section in the sequence that underscores their temporal recognition and application in pharmacy rather than their historic development in physics.

FUNDAMENTALS

The jargon of rheology can be problematic for the uninitiated. For example, as Scott Blair³ notes, stress and strain, in everyday English, have virtually the same meaning. Rheologists, however, use the word *stress* to refer to a system of forces, whether applied in a compressive, extensional, or shear mode, and use *strain* to refer to a change in size or shape.

Rheological principles stem from two fundamental laws derived in the late seventeenth century: Robert Hooke's law of elasticity (ca. 1676) and Isaac Newton's law of flow (1687). The corresponding equations, which embody these laws, characterize Hookean and Newtonian materials, respectively. When a force is applied to a body, the two rheological extremes of behavior are the pure elastic deformation of a Hookean solid and the pure, viscous flow of a Newtonian liquid. Pure (ideal) elasticity means that the body returns to its original form once the stress is removed, while pure (ideal) viscosity means that the liquid flows even under the smallest stress and does not return to its original shape or form once the stress is removed.⁴ The resistance to deformation, or flow, is described by the modulus of elasticity or Young's modulus, E , for an elastic body undergoing extension, and by η , the coefficient of viscosity for a liquid.

Elastic deformation of solids is described by Hooke's law:

$$dl = \frac{\sigma}{E} \quad (1)$$

where dl is the elastic deformation or extension in length l caused by the application of stress σ . This is illustrated in Figure 22-1.

Viscous deformation, i.e., viscous flow, occurs in accordance with Newton's law,

$$\sigma = \eta \dot{\gamma} \quad (2)$$

wherein the applied stress σ results in flow with a velocity gradient $\dot{\gamma}$, or rate of shear. The proportionality constant η is

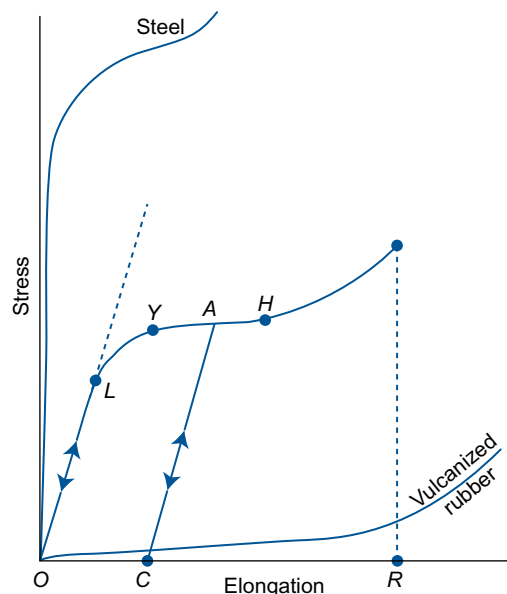


Figure 22-1. Elastic deformation in accordance with Hooke's law.

termed viscosity, while its reciprocal is called fluidity. Viscosity has also been described as the internal friction in the fluid as it corresponds to the resistance of the fluid to the relative motion of adjacent layers of liquid. This is illustrated in Figure 22-2.

Imagine a liquid contained between two very large, parallel plates as being divided into a stack of very thin, parallel layers, much like a deck of cards, as shown in Figure 22-2. Shear is applied to the liquid by pulling or pushing the top plate with a constant force F per unit area A , i.e., F/A , or σ , while holding the bottom plate stationary. The top liquid layer, in contact with the moving plate, adheres to it and moves with the same velocity as the plate. The second layer, adjacent to the top one, is dragged along by friction, but its velocity is reduced somewhat by the resistance of the layers beneath it. Each layer is pulled forward by the layer moving above it but is held back by the layer underneath it, over which it moves and which it drags along. The farther the liquid layers are from the moving plate, the smaller their velocities. The bottom layer adheres to the stationary plate and has zero velocity. Thus, the velocity of the liquid layers increases in the direction x perpendicular to the direction of flow y . The shear strain or deformation in shear, γ , is the displacement y divided by the height, x , of the sheared or deformed portion of the liquid, as shown in Figure 22-2. It equals the tangent of the displacement angle θ in radians that, at low values, is approximately equal to 0, i.e.,

$$\tan \theta = \gamma = \frac{y}{x} \approx 0$$

In due time, all layers except the bottom one undergo infinite deformation. What distinguishes one liquid from another is the rate at which the deformation increases with time. This is called the rate of (deformation in) shear, $\dot{\gamma}$ or $d\gamma/dt$, the derivative of γ with respect to time, t . An equivalent definition for $\dot{\gamma}$ is the velocity gradient, i.e., the rate at which the velocity, v , changes with the distance, x , perpendicular to the direction of flow:

$$\dot{\gamma} = \frac{d\gamma}{dt} = \frac{dv}{dx} \quad (3)$$

The rate of shear or velocity gradient, $\dot{\gamma}$, indicates how fast the liquid flows when a shear stress is applied to it. Its unit according to both definitions is s^{-1} , since γ is dimensionless, velocity is expressed in m/sec, and x in m.

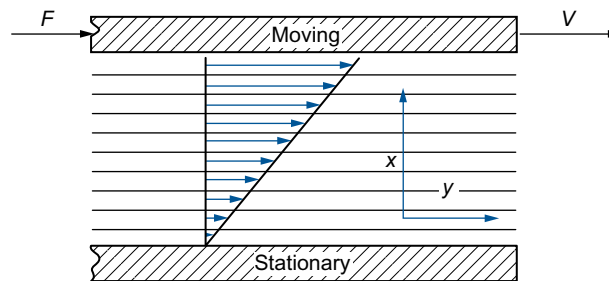


Figure 22-2. Laminar flow of a liquid contained between two parallel plates.

It should be noted that the symbols often used in the past in the pharmaceutical literature for the rate of shear, D , and for the shearing stress, τ , have been replaced by $\dot{\gamma}$ and σ , respectively, in accordance with more widely accepted nomenclature recommendations.^{5,6}

Characteristic shear rates for pharmacy-related activities are listed in Table 22-1. Even for a given process, the shear rate can vary within wide limits, depending upon the scale of the process and the processing rate. Thus, when a lotion is rubbed into the skin, if the hand (moving surface) slides across the skin (stationary surface) with a velocity $v = 45 \text{ cm/s}$ and if the thickness of the lotion film is $x = 0.05 \text{ cm}$, then, according to Equation 3, the rate of shear is $\dot{\gamma} = (45 \text{ cm/s})/(0.05 \text{ cm}) = 900 \text{ s}^{-1}$. For a given force and a constant viscosity, the rate of shear is uniform throughout the layer of lotion.

The flow of liquids by parallel layers, moving past each other and dragging adjacent layers along (as in Figure 22-2), is called *laminar* or *streamline* flow. At higher velocities and/or if the plates have rough surfaces, eddies or swirls develop, whereby mass transfer occurs from one layer or lamina to another. Theoretically, this complex phenomenon—referred to as turbulent flow—may be described by a set of partial differential equations, known as the Navier-Stokes equations, which govern fluids in motion. However, explicit solutions of these nonlinear equations, originally derived in the 1840s on the basis of laws of conservation of mass, momentum, and energy, remain elusive.

From a historic rheological viewpoint, deformation of matter was first described in ideal terms. Precise differentiations were made among perfect, rigid *Euclidean* bodies (solids), ideal

Table 22-1. Approximate Shear Rates for Pharmaceutical Operations

Operation	Rate of Shear, s^{-1}
Particle sedimentation in a suspension	10^{-3}
Film drainage	10^{-1} – 10^1
Pouring liquid from a bottle	50
Extrusion	1–100
Pumping fluids	1–3,000
Mixing or stirring	10–1,000
Spreading lotion or cream on skin	400–1,000
Levigating ointment on a slab with spatula	400–1,000
Injection via syringe	4,000
Dispensing nasal spray from plastic squeeze bottle	20,000
Processing in colloid mill	10^5 – 10^6
Film-coating	10^4 – 10^6
Spray drying	10^5 – 10^6

Hookean elastic solids, *Pascalian*, or inviscid fluids, and *Newtonian* liquids. For ideal Euclidean solids, only mass (or density) is relevant; rigid bodies do not undergo deformation under stress, irrespective of the amount of force applied. When stress is applied to an ideal Hookean elastic solid, the deformation induced is fully recovered when the stress is removed. Inviscid fluids (e.g., gases at low pressure) exhibit no resistance to flow when stressed, whereas Newtonian liquids undergo flow at a rate that is proportional to the stress applied.

Unfortunately, most solids and fluids encountered in pharmacy do not exhibit ideal behavior consistent with the classical models that evolved with Hooke, Pascal, or Newton. By the nineteenth century, evidence for more complex, nonideal rheological behavior began to accumulate, and the clear-cut dividing line between *Hookean elastic solids* and Newtonian or viscous liquids became increasingly blurred. Some systems that behave as elastic solids when subjected to small stresses or to moderate stresses of short duration will undergo permanent deformation, resembling very viscous liquids if the stresses are larger and/or applied for longer periods of time. For many materials, the temporal dependence of their rheological properties necessitates careful consideration of their handling prior to and during the process of rheological evaluation. Nonetheless, an understanding of ideal rheological behavior is necessary before deviations from ideality can be considered.

ELASTIC SOLIDS

In the stretching or extension of an elastic solid, the deformation is said to be in tension. The deformation or strain of the stretched body, or its elongation, is the difference between its length while under tension, l_s , and its original length, l , which is equal to the length after the stress is released, expressed as a fraction of the original length, namely, $(l_s - l)/l$. Other modes of deformation are by bending or flexure, torsion, compression, and shear.

For an ideal elastic solid, Hooke's law (Equation 1) states that the stress is directly proportional to the strain. This relationship is obeyed by real solids at moderate stresses and strains sustained for short periods of time. The *modulus of elasticity* or *Young's modulus*, E , is a measure of the stiffness, hardness, or resistance to elongation. There is also a modulus of shear or rigidity and a compression or bulk modulus. *Tensile compliance* is the reciprocal of Young's modulus, or the ratio of strain to stress.

In the CGS system, the units of stress are dynes/cm² or, since force = mass $\times$ acceleration, (g-cm/sec²)/cm² = g/(cm sec²). To convert dynes/cm² to the SI unit, Newton/m² or Pascal, divide by 10. Since strain is dimensionless, Young's modulus has the same dimensions as stress. Modulus values for a range of solids of pharmaceutical or biomedical interest are listed in Table 22-2.

Figure 22-1. shows representative stress-strain curves in tension, also called load-elongation curves. The cross-sectional area A of the solid becomes smaller as it is stretched. Therefore, to calculate the actual or true tensile stresses, the forces are divided by A_s , the cross-sectional area at each appropriate elongation. Stress-strain curves often are plotted with the strain or extension, the dependent variable, on the abscissa, while consistency or flow curves (see below) usually are plotted with stress, the independent variable, on the abscissa. The practice followed here is to plot stress on the ordinate for both stress-strain and consistency curves in order to make modulus and viscosity, respectively, the slopes of these curves.

The characteristic portions in the representative stress-strain curve *OLYAHB* in Figure 22-1 are as follows. Hooke's law of proportionality between stress and strain is obeyed throughout the linear portion *OL*. The elastic modulus of the solid is the slope of *OL* or the tangent of the angle *LOC*. The material behaves elastically up to the yield point Y , where the stress is called yield stress. When stresses below the yield stress are applied to

Table 22-2. Modulus of Elasticity^a of Representative Solids of Pharmaceutical or Biomedical Interest

Material	Young's Modulus, Pa
Steel	2.2×10^{11}
Glass	6×10^{10}
Potassium chloride	2.3×10^{10}
Silk, viscose rayon	1.5×10^{10}
Microcrystalline cellulose	1.3×10^{10}
Polyethylene, low density	2.4×10^8
Rubber, vulcanized	2×10^6
Tooth enamel	4.7×10^{10}
Bone	2.2×10^{10}
Tendon	1.3×10^9
Muscle	6×10^5
Soft tissue	7.5×10^3
Gelatin gels – 10% solids	2.4×10^4
Gelatin gels – 20% solids	1.0×10^5
Gelatin gels – 30% solids	1.5×10^5

the sample and then released, it stretches and contracts along the same curve *OLY*.

Beyond Y , the material behaves as a plastic rather than as an elastic solid. Along the (nearly) horizontal portion *YAH*, the material is ductile; it flows or creeps under practically constant stress like a viscous liquid. If the stress is released at A , the sample retracts along *AC*. The nonrecoverable deformation *OC* is called *permanent set*. Many materials undergoing such "cold flow" are strengthened by some change in structure, causing an upturn *HB* in the stress-strain curve. This is called *work (or strain) hardening*. It may result from the elimination of flaws, from a reduction in crystal size (as in the case of metals), or from reversible crystallization on stretching (as in the case of homopolymer elastomers).

At B , the sample ruptures; R is the elongation at the break or the ultimate elongation, and the stress corresponding to B is the *ultimate strength* or *tensile strength*. These values, as well as the load-elongation curve beyond Y , depend on the rate at which the sample is stretched.

The area *OLYAHBRCO* under the stress-strain curve is the energy or work required to break or rupture the material. It measures the material's toughness or brittleness. Glass is hard because of its high elastic modulus. Owing to the absence of a yield point and to a very low elongation to break, it is brittle, as opposed to steel, which undergoes work hardening, has a high elongation to break, and is tough. Plastics are medium-hard or soft. Those that exhibit comparatively high elongations at break, like polyethylene but unlike polystyrene, are tough. Vulcanized rubbers are tough even though they are soft (low elastic modulus) because their elongation to break is very high, namely, 600 to 800 percent.

NEWTONIAN FLUIDS

The viscosity of simple liquids (i.e., pure liquids consisting of small molecules and solutions where solute and solvent are small molecules) depends only on composition, temperature, and pressure. It increases moderately with increasing pressure and markedly with decreasing temperature. For solutions of solid solutes, the viscosity usually increases with concentration. Simple liquids follow Newton's law (Equation 2) of direct proportionality between shear stress and rate of shear, so that their viscosity is independent of the shear stress or the rate of shear. Their flow behavior is thus referred to as Newtonian. Representative Newtonian viscosities are listed in Table 22-3. Fluid flow through cylindrical pipes or capillaries is laminar,

Table 22-3. Newtonian Viscosities and "Activation Energies", E_a , for Viscous Flow

Material	Temperature, °C.	Viscosity, mPa•s	E_a for Viscous Flow, Kcal/mole
Water	20	1.0	4.2
	50	0.55	3.4
	99	0.28	2.8
Ethanol, absolute	20	1.2	3.3
	50	0.7	3.3
	20	2.9	6.8
40% w/w	50	1.13	5.3
	20	0.24	1.65
Ethyl ether	20		
Glycerin			
anhydrous	20	1,500	12.5
95% w/w	20	545	10.6

i.e., Newtonian, at low velocities, for small tube radii or for liquids of high viscosity. The liquid layers are very thin cylinders concentric with the duct.² During flow they telescope past one another, as shown in Figure 22-3A. The arrows in Figure 22-3B represent the velocity v of the individual cylindrical layers of radius r ; v is maximal in the center of the tube and decreases in the radial direction, i.e., in the direction r (previously x) perpendicular to the direction of flow y . The velocity is zero in the outermost liquid layer adjacent to and adhering to the wall, whose radius is equal to the inside radius of the tube R . In the center of the tube, where v is maximum, the velocity gradient $dv/dr = \gamma$ is zero. This is shown in Figure 22-3C, where the arrows represent γ and the velocity gradient is maximum at the wall.

If V is the volume of liquid flowing through a cylindrical tube of radius R in time t , the volumetric flow rate is V/t , and the shear rate at the wall is

$$\dot{\gamma}_{\text{wall}} = \frac{4}{\pi R^3} (V/t) \quad (4)$$

The shear stress is zero in the center of the tube and maximum at the wall:

$$\sigma_{\text{wall}} = \frac{R\Delta P}{2l} \quad (5)$$

The liquid flows through the tube due to pressure, either caused by its own weight (hydrostatic) or produced by a pump. This pressure exceeds the innate viscous friction of the liquid and is converted into heat. The pressure drop, ΔP , along a length l of the tube is the difference between the pressure at the beginning and at the end of the tube.

As viscosity is shear stress divided by rate of shear, and as both vary in the x -direction perpendicular to the direction of flow, both must be evaluated at the same location. Using the values at the wall of a cylindrical tube, dividing Equation 5 by Equation 4, and rearranging gives

$$\left(\frac{V}{t}\right) = \frac{R^4 \Delta P}{8l\eta} \quad (6)$$

This is Poiseuille's law, found experimentally by this French physician while studying the flow of liquids through capillary tubes representative of blood vessels. (The poise is also named in his honor.)

In the human body, the pumping action of the heart supplies the driving pressure for the flow of blood, which is the difference between the arterial and venous pressures. Digitalis

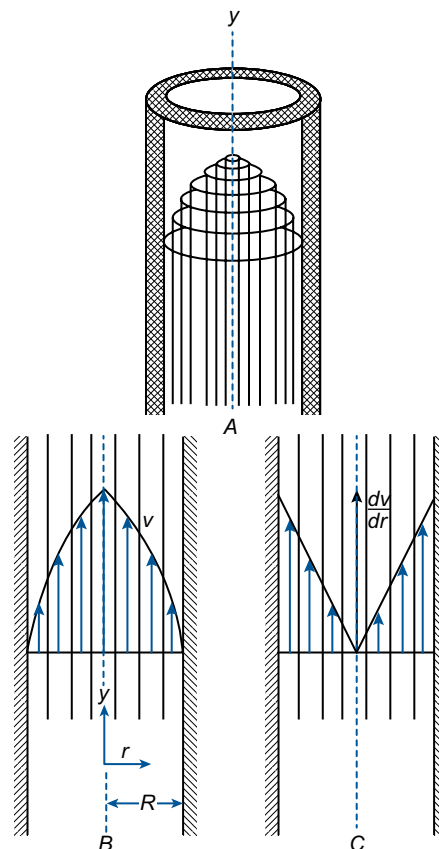


Figure 22-3. Laminar flow of a liquid through a cylindrical duct. A, Three-dimensional view of telescoping layers; B, cross-section showing radial distribution of velocity; C, cross-section showing radial distribution of velocity gradient.

glycosides increase the force of contraction of the heart muscle and make the heart a more efficient pump. This increases ΔP and, hence, the rate of flow of blood V/t . Vasodilator drugs like nitroglycerin or hydralazine hydrochloride increase the radius of blood vessels by relaxing the vascular smooth muscles. Since the flow rate varies with the fourth power of the radius of the blood vessel, a mere 5 percent increase in radius causes a 22 percent increase in the flow rate at constant blood pressure because $(1.05)^4 = 1.22$.

Plots of shear stress (on the y axis) as a function of the rate of shear (on the x axis) are referred to as flow curves or *rheograms*. The *rheograms* of typical Newtonian liquids, like those of Figure 22-4, are straight lines going through the origin. Viscosity is the slope of such a line or the tangent of the angle it makes with the horizontal axis. Of the two liquids shown in Figure 22-4, A has a higher viscosity than B because $\alpha > \beta$, so that $\eta_A (= \tan \alpha) > \eta_B (= \tan \beta)$; $\eta_A = \sigma_2/\gamma_2 = \sigma_1/\gamma_1$ and $\eta_B = \sigma_1/\gamma_3 = \sigma_3/\gamma_2$. A given shear stress, σ_1 , produces a greater rate of shear, γ_3 , in the more fluid Liquid B than γ_1 in the more viscous Liquid A. Alternatively, to produce a given rate of shear, γ_2 , in the two liquids requires a higher shear stress, σ_2 , for the more viscous Liquid A than σ_3 for the more fluid Liquid B.

In the CGS system, viscosity is defined as the tangential force per unit area, in dynes/cm², required to maintain a difference in velocity of 1 cm/s between two parallel layers of liquid 1 cm apart. Its unit is, therefore, dynes/cm²•sec⁻¹ or g/cm•s, which is called a *poise*. Because many common liquids, including water, have viscosities on the order of 1/100 of a poise, their viscosity is often expressed in *centipoise*. In the SI system, the unit of viscosity is Newton/m²•s⁻¹ or Pascal•s, which equals 10 poise. Typical Newtonian viscosities are listed in Table 22-3.

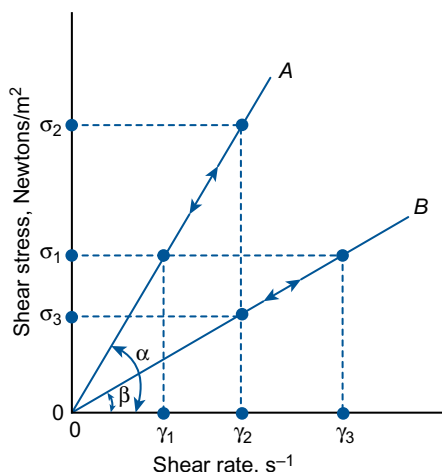


Figure 22-4. Rheograms or flow curves of two Newtonian liquids.

The variation of viscosity with temperature often is described by an Arrhenius equation:

$$\begin{aligned}\eta &= Ae^{E_a/RT} \\ \ln \eta &= \ln A + \left(\frac{E_a}{RT} \right)\end{aligned}\quad (7)$$

where A and E_a are constants, T is the absolute temperature, and R is the molar gas constant. Values of E_a , the activation energy for viscous flow, are listed in Table 22-3. Large values of E_a indicate that the viscosity decreases substantially with rising temperature. According to Equation 7, plots of $\ln \eta$ as a function of the reciprocal of the absolute temperature should be straight lines with slopes of E_a/R . For associated, e.g., hydrogen-bonded, liquids, such plots are often somewhat curved. According to Eyring's "hole theory," liquids contain vacancies or holes that are essential to flow; the activation energy is used largely to form these holes.^{7a} E_a is about 1/3 to 1/4 of the latent heat of vaporization for nonassociated liquids. Even though the viscosity of liquids often shows an inverse dependence on temperature, the concept of an activation energy and barrier associated with viscous flow (and diffusion), at least in simple liquids, has been disputed.^{7b}

NON-NEWTONIAN FLUIDS

Fluids that do not obey Newton's law (Equation 2) are described as non-Newtonian fluids. The rheological behavior of non-Newtonian fluids may be characterized either as time-independent or time-dependent non-Newtonian fluids.

TIME-INDEPENDENT NON-NEWTONIAN FLUIDS

Shear-thinning Fluids. Many colloidal systems, especially polymer solutions and flocculated solid/liquid dispersions, become more fluid the faster they are stirred. This shear-thinning behavior is often referred to as *pseudoplasticity*, but the latter term is outdated and potentially misleading. Shear-thinning behavior is an example of non-Newtonian flow because the viscosity, at constant temperature and composition, is not constant, as required by Newton's law of viscous flow (Equation 2), but decreases with increasing shear. As the increase in shear rate is greater than the increase in the corresponding shear stress, the flow curve of Figure 22-5 is concave toward the shear-rate axis.

There is an apparent viscosity for each value of shear rate or shear stress, which can be expressed in two different ways. At point P in Figure 22-5, the apparent viscosity can be taken as the slope of the secant to the flow curve at P , or $\tan \theta$, which is the viscosity of a Newtonian liquid whose flow curve passes

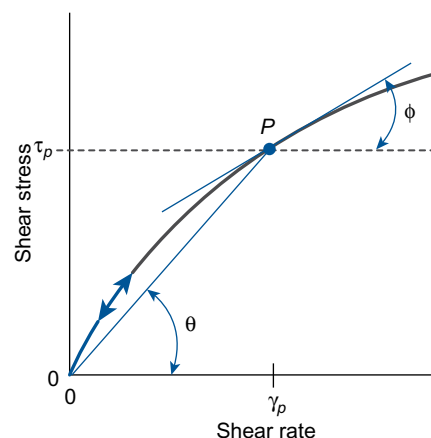


Figure 22-5. Flow curve of a shear-thinning liquid.

through P . This is equal to the ratio σ_P/γ_P . The second method defines the apparent viscosity as the slope of the tangent to the flow curve at P , i.e., $d\sigma_P/d\gamma_P = \tan \phi$. Since both θ and ϕ decrease with increasing shear stress or shear rate, so does the viscosity.

The shear-thinning behavior of polymer or macromolecule solutions arises from the alignment of neighboring macromolecules and the degree of their entanglement and concomitant immobilization of solvent. In aqueous solution, for example, the flexible, thread-like macromolecules are buffeted constantly by the surrounding water molecules in thermal agitation. This causes continuous random motion of chain segments by translation and by rotation around bonds between the atoms that make up the macromolecular backbone. These thermal fluctuations result in the formation of loose, roughly spherical coils that are permeated by water. The macromolecule chains are encased in sheaths of water of hydration. Additional water is mechanically trapped inside the open coils. The coiled macromolecules, in constant segmental motion, become entangled (Figure 22-6A). Upon the application of shear, a unidirectional laminar motion is superimposed on the random thermal motion of the water molecules and chain segments. The randomly coiled, entangled macromolecules tend to disentangle themselves and to align themselves in the direction of flow, as shown in Figure 22-6B. The viscosity of the solution—its resistance to flow—depends on the size and shape of the flow units. The imposition of increasing shear in these systems enables the macromolecule "chains" to uncoil progressively and become streamlined or elongated, thereby offering less resistance to flow than the original, approximately spherical, shapes. At the same time, the amount of water trapped inside the coils and dragged along decreases. Furthermore, the chains become gradually more disentangled. Reduced entrapment of water and decreased entanglement of the macromolecules reduce the size of the flow unit, thereby reducing the viscosity. A further reduction in viscosity results from shear-induced uncoiling of the macromolecules. Thus, the apparent viscosity at a given rate of shear reflects the degree of randomization, coiling, entanglement, and alignment of the macromolecules, and the extent to which solvent molecules are associated with the macromolecules.

Dispersions of flocculated solid particles exhibit shear-thinning if the particle-particle bonds are too weak to withstand the applied shear stresses. Examples of weak interparticle bonds include weakly flocculated particles, in a secondary minimum, or electrostatically attracted lamellar clay platelets with positively charged edges and negatively charged faces, which produce a "house-of-cards" structure in an aqueous suspension.

Shear progressively breaks up these aggregates at a rate that increases with increasing shear stress, releasing increasing amounts of trapped water. Brownian motion tends to rebuild the aggregates at a rate that is independent of shear. There is an

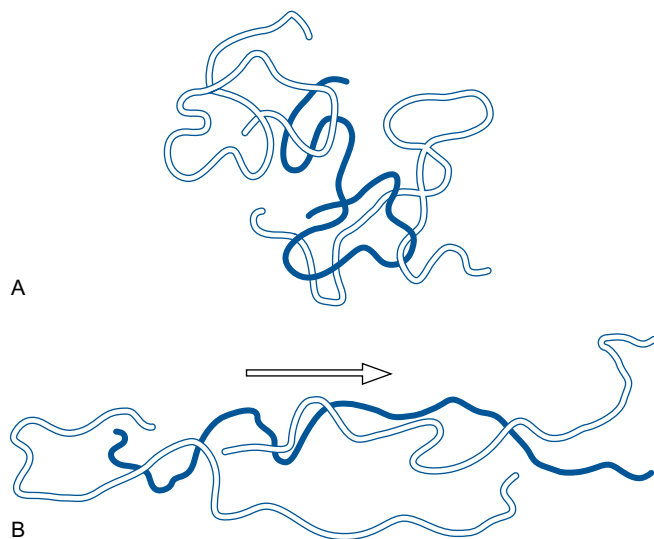


Figure 22-6. Three randomly coiled polymer chains in solution. (A) At rest; (B) In shear field.

average equilibrium size for the aggregates at each rate of shear that decreases with increasing shear, resulting in a decrease in the resistance to flow, or viscosity, as the shear increases.

At extremely low shear rates, well below 1 s^{-1} , the rate of disentanglement and alignment of polymer chains and the rate of breaking up of aggregates of particles under the influence of shear are negligible, compared to the rate of entanglement and randomization of polymer chains and to the rate of aggregation of particles produced by Brownian motion, respectively. Hence, the flow units are neither noticeably deformed nor reduced in size by shear, and the systems exhibit Newtonian flow, with a constant and high viscosity designated as the lower Newtonian or zero-shear viscosity, η_0 .

At very high shear rates, the dissolved polymer chains are wholly disentangled and well aligned in the direction of flow, and the aggregates of particles are broken up as far as possible. There is no residual structure left that can be broken up by further increments in shear rate: The viscosity levels off at a constant value called the upper Newtonian viscosity, η_∞ . Turbulent flow and shear-induced rupture of polymer chains may set in before the upper Newtonian regime is reached. As can be seen in Figure 22-7, η_∞ is considerably lower than η_0 . The value of the non-Newtonian viscosity observed at intermediate shear rates, including those encountered in most practical situations, depends on the amount of residual structure. It is, therefore, called *structural viscosity*.

Dilatancy. In contrast to shear-thinning, *shear-thickening* or *dilatancy*, i.e., an increase in viscosity with increasing shear, is rare. It is shown by concentrated dispersions of particles that do not tend to aggregate or stick together, provided the amount of liquid present is not much larger than that needed to fill the voids between the particles. Sediments of suspensions from which the supernatant liquid has been decanted are sometimes dilatant. When such a concentrated suspension is poured or stirred slowly, there is just enough liquid to lubricate the slipping of one particle past another, and the viscosity is low. When stirred fast, the particles get into each other's way, block each other, and bunch up rather than slipping past each other. Large voids form between the unevenly clustered particles, and as the liquid seeps into these, the suspension appears dry—as if the suspended solids had expanded or become dilated. This phenomenon, which results in progressive viscosity increases, becomes more severe with increasing shear. When high shear is followed by low shear or rest, the particles that had been crowded together separate again, the interparticle void volume decreases, and the viscosity drops as the suspension appears

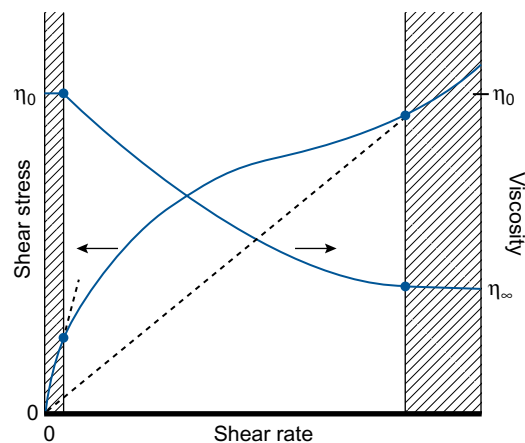


Figure 22-7. The three flow regions of a shear-thinning liquid. Shaded areas refer to lower (*left*) and upper (*right*) Newtonian regions; center area represents shear-thinning behavior.

wet again. Wet sand offers small resistance to slow flow or penetration but stiffens and appears dry when deformed fast.

Among the few systems reported⁸ to exhibit dilatant flow are suspensions of starch in water, aqueous glycerin, or ethylene glycol containing about 40 to 50 percent v/v starch, and concentrated suspensions of inorganic pigments in water and in nonpolar liquids with enough surfactant added to deflocculate the disperse phase completely, e.g., red iron oxide (12 percent v/v in water), zinc oxide (30 percent v/v in water), barium sulfate (39 percent v/v in water), and titanium dioxide (30–50 percent v/v in water).

Plasticity. Semisolids that do not flow at low shear stresses (exhibiting reversible deformation like elastic solids) but flow like liquids above their yield value (i.e., yield stress) are termed *plastics* or *Bingham bodies*. This type of rheological behavior is called *plasticity*. Plasticity is often exhibited by semisolids characterized as structured media, i.e., semisolids that have a cross-linked three-dimensional network of polymers, macromolecules, or particulates extending throughout the system.

Figure 22-8 shows the flow curves for two plastic systems. System *B* has a lower yield value than System *A* and Newtonian behavior at stresses above the yield value; $BC\sigma_{\text{yield}}$ is a straight line of inclination θ , so that the *plastic viscosity* of *B*, i.e., its viscosity above the yield value, is the slope of this line or $\tan \theta$:

$$\eta_{\text{plastic}} = \frac{\sigma - \sigma_{\text{yield}}}{\dot{\gamma}} \quad (8)$$

This is equivalent to moving the origin of the flow curve from zero stress to the yield stress and treating System *B* as a Newtonian liquid at stresses beyond. Semisolids with high yield values are described as “hard.” When their plastic viscosity is high, they are described as “stiff.”

Some Bingham bodies have flow curves that deviate from straight lines at stresses close to the yield stress, such as the portion *CD* in the flow curve of System *B*, where flow occurs even below the yield stress. This phenomenon is called *plug flow* because the material moves in chunks or as a plug rather than by laminar motion, often through slippage at the wall of the duct. In such cases, the yield value usually is obtained by extrapolating the linear portion *BC* to the stress axis. System *A* is shear-thinning above its yield stress. This type of flow behavior is observed frequently with suspensions thickened with dissolved polymers, where the vehicle itself is shear-thinning.

TIME-DEPENDENT NON-NEWTONIAN FLUIDS

In the previous discussion, shear-thinning and plastic behavior were seen to arise from competition between the detachment

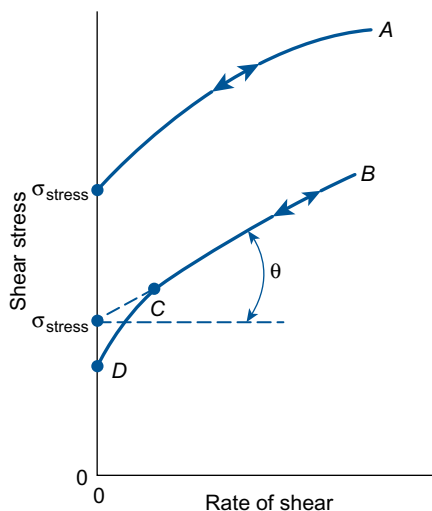


Figure 22-8. Flow curves of two plastic systems.

of entangled dissolved macromolecules, or the rupturing of van der Waals links among dispersed particles by shear, and the re-establishment of such links by Brownian motion. The balance between breakdown and restoration of links shifts more and more toward breakdown as the shear increases. Reduction in interchain or interparticle links results in smaller flow units and lower apparent viscosity. It was assumed tacitly that the system adapts itself to changing shear “instantaneously,” i.e., so fast that by the time the instrumental conditions had been changed to higher or lower shear and readings then taken, the equilibrium between breakdown and restoration of links at the new shear already had been reached, producing flow units of the new average equilibrium size and the corresponding new apparent viscosity. Points representing pairs of $\dot{\gamma}$, σ values, determined at increasing and at decreasing shear rates or shear stresses in Figures 22-4, 22-5, and 22-8, fall on the same single curves. It is immaterial whether a given shear rate was reached by increasing or decreasing the speed of the viscometer. This is the meaning of the double arrows on these curves.

Thixotropy. If the suspension is viscous and/or the particles are large and heavy, their Brownian motion is too slow to restore the broken interparticle links “instantaneously.” Likewise, the entanglements of polymer chains are slow to be reestablished by Brownian motion if their solution is viscous. If the rate of link restoration by Brownian motion is lower than the rate of link breakdown by shear, the apparent viscosity decreases even while the system is under constant shear, as the size of the particle aggregates or the extent of macromolecular entanglement is progressively reduced. Furthermore, the apparent viscosity at a given shear rate is lower if the system was stirred recently at high speeds than if that shear rate was approached from low speeds or from rest.

Materials whose consistency depends on the duration of shear as well as on the rate of shear are said to be *thixotropic* or to exhibit *thixotropy*. Their apparent viscosity depends not only on temperature, composition, and rate of shear or shear stress, but on the previous shear history and time under shear.

The extreme behavior is an isothermal, reversible sol-gel transformation produced by rest and by shear, respectively. For example, an aqueous dispersion of 8% w/w sodium bentonite sets to a gel within an hour or two after preparation when undisturbed, but flows and can be poured within many minutes after it had been stirred above the yield value. After prolonged rest it reverts to a gel as Brownian motion rebuilds the “house-of-cards” structure throughout the material.

Thixotropy in a shear-thinning liquid is shown in Figure 22-9. Starting with the system at rest (at the origin O) and gradually increasing the speed of the viscometer produces the “up”

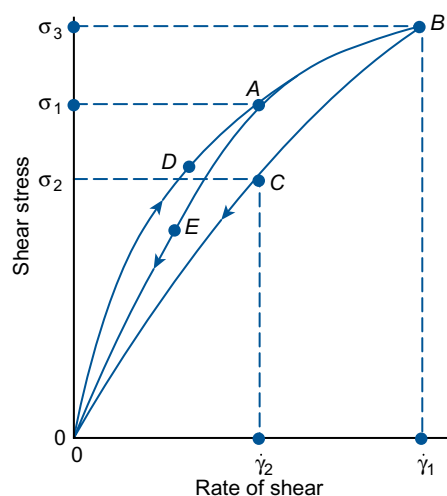


Figure 22-9. Flow curves of a shear-thinning liquid exhibiting thixotropy.

branch $ODAB$ of the flow curve. After the maximum shear rate $\dot{\gamma}_1$ and shear stress σ_3 corresponding to point B have been reached, the speed of the instrument is reduced. If there is not enough time for Brownian motion to regenerate completely the structure torn down at high speed, the liquid will be less viscous, and the “down” branch of the flow curve, BCO , will be lower than the “up” branch. Thus, the shear stress required to maintain the rate of shear $\dot{\gamma}_2$ has been reduced from σ_1 to σ_2 , and the apparent viscosity has dropped from $\sigma_1/\dot{\gamma}_2$ to $\sigma_2/\dot{\gamma}_2$. This contrasts with the flow curve of Figure 22-6, where the “up” and “down” branches coincide.

When starting from rest, if the speed is not increased all the way up to $\dot{\gamma}_1$ but only to $\dot{\gamma}_2$, corresponding to point A in Figure 22-9, and then decreased, the “down” branch is AEO . Since the maximum speed is lower than previously, less structure is broken down, and the apparent viscosity is not reduced by as much.

If the liquid in the instrument is kept at rest for a sufficient time period after it was subjected to the shear cycle $OD-ABCO$, Brownian motion rebuilds its structure, restoring its original high consistency. Starting from rest, the flow curve is again $ODABCO$. If no rest period is allowed and the shear cycle is repeated as soon as the “down” branch is completed, the next “up” branch is below $ODAB$, say, OEB in Figure 22-10. A third shear cycle following immediately after the second may give the “up” branch OGB . The “down” branch BCO may be curved, as in Figure 22-9, or straight, as in Figure 22-10. If the buildup of structure is very slow, there may be no structure left after the third shear cycle. In that case, the “up” branch coincides with the straight “down” branch BCO , and the liquid has become Newtonian. This is only temporary because the flow curve reverts to $OABCO$ of Figure 22-10 after a prolonged rest period.

Thixotropy frequently is superimposed on plastic flow behavior. The yield value may disappear after one or more shear cycles, as in curve C of Figure 22-11; it may be reduced, as in curve B (sometimes called false body behavior), or it may remain unaltered, as in curve A .

The difference between the up and down branches of a flow curve illustrates a common phenomenon called *hysteresis*. The area enclosed by the two branches (e.g., areas $ODAEO$ and $ODABCO$ in Figure 22-9) or by the two branches and the stress axis (as in Figure 22-11B and C) is called the hysteresis loop. Its size is a measure of the extent of thixotropic breakdown in the structure of the system. In Figure 22-10, the areas enclosed by the two branches of the flow curves representing successive shear cycles become progressively smaller: $OABCO > OFBCO > OGBCO$. This parallels a decrease in the amount of structural

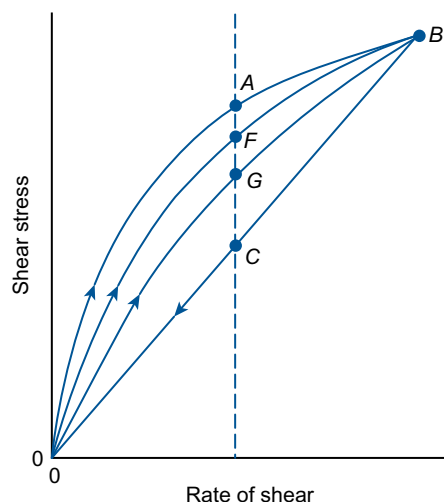


Figure 22-10. Flow curves representing successive shear cycles for a thixotropic, shear-thinning liquid.

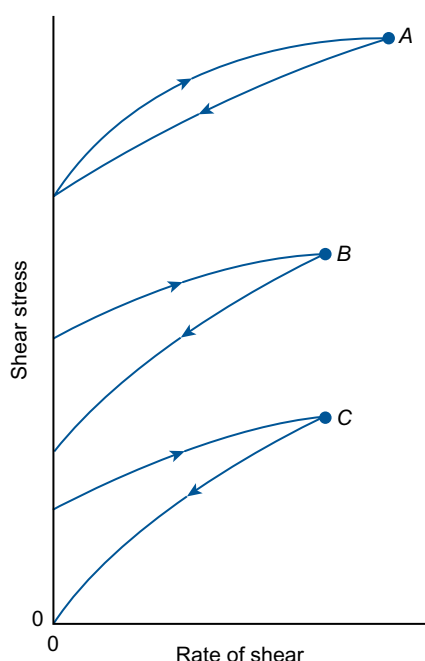


Figure 22-11. Flow curves of plastic systems exhibiting thixotropy (see text).

breakdown of the system as each cycle leaves intact less residual structure to be broken down in the next cycle. When no structure remains, the Newtonian flow curve *OCBCO* of Figure 22-10 results. The absence of hysteresis in the flow curves of Figures 22-5 and 22-8 is due to another cause: The rebuilding of structure by Brownian motion is as fast as, or faster than, the shear-induced structural breakdown or the response time of the viscometer.

Thixotropy may be represented quantitatively by the area of the hysteresis loop, by a thixotropic coefficient, or by index (*T.I.*) at a specific $\dot{\gamma}$ for a set period of time, e.g.,

$$T.I. = \frac{\sigma|_{\text{before shearing}}}{\sigma|_{\text{after shearing at } 0.01\text{s}^{-1}}} \quad (9)$$

or by shearing samples at various $\dot{\gamma}$ s as a function of time and fitting all of the data to an equation of the form

$$\eta = A\dot{\gamma}^B t^C \quad (10)$$

where *A* is a proportionality constant, and the exponents *B* and *C* are indices of shear-thinning and thixotropy, respectively. As *B* increases, the fluid becomes more shear-sensitive; as *C* increases, the fluid becomes more time-dependent.⁹ A simple alternative method involves the measurement of the decay of shear stress or apparent viscosity as a function of time at a constant rate of shear. This method is illustrated in Figure 22-12. When a system is stirred at a constant shear rate, it eventually reaches constant or equilibrium values for shear stress and apparent viscosity. This is shown by the leveling off of the curve. Equilibration times at a given shear rate may range from minutes to hours, or longer, depending upon the nature of the structured system.

Thixotropy is particularly useful in the formulation of pharmaceutical suspensions and emulsions. These must be poured easily from containers, indicating low viscosity. Low viscosity, however, causes rapid settling of solid particles in suspensions and rapid creaming of emulsions. Solid particles that have settled out frequently stick together, producing a sediment difficult to redispense ("caking" or "claying"). Creaming in emulsions is a first step toward coalescence. Thixotropy can be used to resolve this dilemma. A thixotropic agent, such as sodium bentonite magma, other colloidal clays (magnesium bentonite, attapulgite), colloidal silicon dioxide, or microcrystalline cellulose, is incorporated into the suspensions or emulsions to confer a high apparent viscosity or even a yield value. High viscosities retard sedimentation and creaming since, according to Stokes' law, the rate of sedimentation or creaming is inversely proportional to the viscosity of the medium. If the system possesses a yield value, sedimentation or creaming is prevented altogether since there is no flow below the yield stress, i.e., the apparent viscosity at low shear becomes infinite. When it is desired to pour some of the suspension or emulsion from its container, it is shaken well, at shear stresses considerably above the yield value. The agitation temporarily breaks down the thixotropic structure, such as the house-of-cards scaffold of bentonite, reducing the yield value to zero and lowering the apparent viscosity. This makes for easy pouring. Back on the shelf, the viscosity slowly increases again, and the yield value is restored as Brownian motion rebuilds the house-of-cards structure of bentonite. This prevents sedimentation and caking of the suspended particles and creaming of the emulsion droplets; the dispersed particles again become trapped in the plastic matrix. The optimum flow curve for such formulations is that of Figure 22-11C.

Rheopexy and Negative Thixotropy. Once the links among suspended particles or the entanglements among dissolved polymer chains have been broken by shear, their restoration by Brownian motion is slow if the suspensions or solutions are viscous. In such cases slow flow, gentle agitation, or moderate

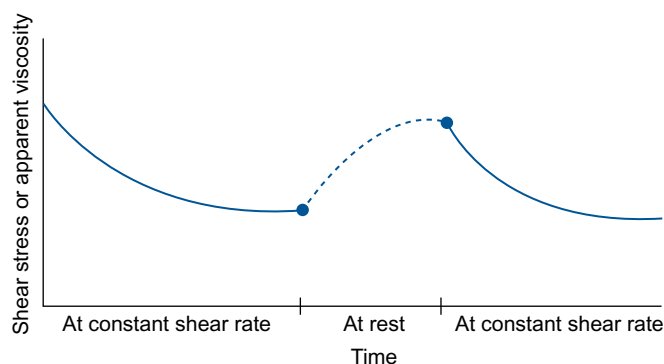


Figure 22-12. Time dependence of shear stress or of apparent viscosity of a thixotropic system.

and rhythmic vibration may accelerate the rebuilding of the structure, i.e., the restoration of the links between particles or macromolecules by Brownian motion. Low shear rates thus hasten the reappearance of high apparent viscosities or onset of gelation in thixotropic sols. In the case of sheared dispersions of bentonite, gentle vibration or rotation of the beaker speeds up the rebuilding of the house-of-cards structure. The material's recovery of some of its presheared viscosity at a faster rate when it is gently sheared, compared to when it is allowed to stand, is called *rheopexy*. Synovial fluid—responsible for lubricating and cushioning the joints—has been shown to exhibit considerable rheopectic behavior, apparently as a result of the interaction between hyaluronic acid and proteins (primarily albumin). Rheopexy is not to be confused with negative thixotropy (i.e., *antithixotropy*), which is defined as a reversible time-dependent increase in viscosity at a particular shear rate as a result of shear-induced buildup of structure over time.

VISCOELASTICITY

Normal condensed matter is either solid or liquid. The molecules of an ideal solid are fixed in place while those of an ideal liquid are mobile. Soft condensed matter occupies a middle ground between the solid and liquid states as it typically possesses a structure that is on a substantially larger scale than atomic or molecular dimensions, i.e., the mesoscopic scale. (Soft condensed matter, soft matter, and nanostructured systems are, in effect, colloids or colloidal dispersions, but the latter terms are being increasingly supplanted in the scientific literature by the former terms.) As a result, the macroscopic rheological behavior of soft condensed matter is determined by the structure and dynamics at the mesoscopic scale and is often described as viscoelastic in nature. When stressed, viscoelastic materials simultaneously exhibit some of the properties of elastic solids and some of the properties of viscous liquids: some deformation occurs instantaneously upon the application of stress and continues as long as the stress is applied. Upon removal of the stress, there is partial recovery of the original shape. In effect, viscoelastic systems are capable of storing part of the deformation energy elastically and reversibly. The relative proportions of elastic deformation and viscous flow are dependent upon the duration of time that stress is applied. In general, over short periods of time, a viscoelastic material will exhibit predominantly elastic behavior while, over longer periods of time, viscous flow will predominate. In effect, the rheological characterization of viscoelastic materials depends, in part, on the experimental methodology employed. One useful parameter is the dimensionless Deborah number, N_{De} , defined by Reiner¹⁰ as

$$N_{De} = \frac{\lambda}{t_p} \quad (11)$$

where λ is the stress relaxation time (defined as the time required for the stress to decay to $1/e \approx 36.8$ percent of its initial value) and t_p is the process or observation time. When material is instantaneously deformed, the structure is perturbed as viscous flow occurs, and its microstructural elements, whether molecules or particles, are in a higher energy state. It takes some time for these molecules or particles to diffuse to a location where their energy state is equivalent to the pre-stress level.¹¹ The relaxation time for viscoelastic materials is of the same order of magnitude as the observation time ($N_{De} \approx 1$), while the relaxation time for a Newtonian fluid would be 0 (i.e., relaxation time is instantaneous) and that for a Hookean solid would be ∞ (i.e., no relaxation occurs). For nonideal materials, then, $N_{De} \ll 1$ indicative of liquids and $N_{De} \gg 1$, of solids.

Oscillatory test methods have been particularly useful for characterizing viscoelastic materials and separating the elastic from the viscous components of their rheological behavior. In typical small amplitude oscillatory testing, stress is measured as a cone or plate, in contact with the sample, and is made to oscillate with low amplitude at a sinusoidal angular velocity. For elastic solids,

shear stress will be in phase with shear strain; for Newtonian fluids, shear stress will be 90° out of phase with shear strain. Viscoelastic materials will exhibit a shear stress that lags behind the shear strain by an angle—the phase angle or phase shift, Δ —that ranges between 0° and 90° . The tangent of Δ is a measure of the energy lost per oscillatory cycle to the energy stored per cycle, i.e., $\tan \Delta = G''/G'$, the ratio of the loss (viscous) modulus to the storage (elastic) modulus. At the “crossover” point, when these moduli are equal, i.e., when $G'' = G'$ or $\tan \Delta = 1$, the material is equally viscous and elastic. $\tan \Delta$ values < 1 are associated with gels or semisolids; $\tan \Delta > 1$ are associated with fluids.

Large amplitude oscillatory testing methods—wherein oscillation with increasing peak strain is employed—are gaining some traction. These approaches allow the determination of the yield strain associated with structured systems.

Silicone putty (*Silly Putty*) is an example of a viscoelastic material. It has a comparatively short mean relaxation time (~ 1 s) at room temperature. It bounces, behaving like an elastic solid when the time of observation or of application of stress is short, but flows and shows little elasticity when slowly stretched. However, viscoelasticity is widespread even among liquids and plastic materials that seem to lack elasticity or stringiness to the touch, especially if they are tested at small deformations. Higher deformations or rates of shear approaching use conditions frequently rupture the elastic network in these materials, causing the loss of the elastic components of their rheological properties. For instance, fluid emulsions are often slightly viscoelastic at very low shear due to flocculation of the disperse droplets and interlinking of the flocs; they flow readily and lose all recovery properties under slightly higher shear.¹² Davis¹³ determined the viscoelastic properties of oleaginous, emulsion, and absorption-type ointment bases by creep measurements. Radebaugh and Simonelli¹⁴ evaluated the viscoelastic properties of powder-filled semisolids, using starch dispersions in lanolin as their model. As gels, pastes, and polymer or macromolecule solutions often exhibit substantial viscoelasticity, this aspect of rheological behavior should not be ignored.

Most biological fluids (e.g., blood, mucus, saliva, cervical fluid) exhibit substantial viscoelastic behavior, due to their multicomponent nature. Early investigators conceptualized blood as a viscous fluid ($\eta \approx 4 \text{ mPa} \cdot \text{s}$), assuming that the viscosity controls its flow properties.¹⁵ However, blood—as a fluidized suspension of elastic cells—is a non-Newtonian fluid with viscoelastic properties. At low shear rates, blood viscosity is higher because of the tendency of erythrocytes to aggregate. At high shear rates, which are typical for the arterial side and capillaries, blood viscosity is lower and constant because of erythrocyte deformation.^{16–18} Comprehensive studies show that the shear dependence of the viscosity of blood may be attributed exclusively to the reversible formation or disintegration of erythrocyte aggregates, which increases at decreasing shear rates.^{19–21} In the circulation, the rheological behavior of blood is determined by the interactions of the erythrocyte cells with the vessel walls.^{16,22} Below a critical vessel caliber of 1 mm, blood viscosity becomes dependent upon vessel radius. Viscosity drastically decreases when the vessel caliber is approximately 12 to 15 μm , the result, in part, of blood's lower hematocrit in capillaries and the biconcavity of red blood cells, the average diameter (d) of which is 7.5 μm . Their size is by no means negligible, compared to the radius (R) of the capillaries. This leads to a reduction in the apparent viscosity by a factor of $(1 + d/R^2)$.²³ In addition, blood viscosity is affected by the tubular pinch effect, which consists of an accumulation of red cells in an annular region of cylindrical capillaries, located at a distance of about 60 percent of the capillary radius from the capillary axis during laminar flow of blood. Almost colorless plasma flows in the vicinity of the capillary wall. Blood flowing in the center of the capillary is also deficient in red cells. This phenomenon commonly is observed when suspensions of spherical or asymmetric particles flow through ducts whose diameter is only a low multiple of the particle size.²³

The flow properties of blood are determined by the hematocrit (*Hct*) value, plasma viscosity, red cell aggregation, and deformability.^{16,22} The studies of blood viscosity factors present an important mechanism to better understand the pathways of cardiovascular disorders. It also allows utilization of blood viscosity tests in diagnostic, prognostic, and preventive medicine.¹⁹ Elevated *Hct* levels have been associated with adverse cardiovascular outcomes, including arteriosclerosis, coronary heart disease, angina pectoris, myocardial infarction, and CHD incidence.^{24–26} Individuals who exercise regularly have been found to have reduced blood and plasma viscosity compared with nonexercisers.²⁵ On the other hand, after heavy exercise or during severe asthma attacks, serum mean lactate levels increase. In high concentrations, lactic acidosis produces erythrocyte swelling, increasing the *Hct* level and increasing whole-blood viscosity at both high and low shear rates.¹⁷

Unusual rheological phenomena associated with viscoelastic flow include the *Weissenberg effect*, in which fluid climbs up a shaft or impeller rotating in the fluid, and the *die swell effect*, wherein fluid exiting from a tube or capillary expands to two or more times the diameter of the tube. The Weissenberg effect is occasionally encountered in mixing operations, while the die swell effect may be experienced during extrusion processes. Both phenomena result from the so-called normal stress effect in viscoelastic flow, i.e., the tendency of some viscoelastic fluids to flow in a direction normal to the direction of shear. Experimentally observable stresses normal to the direction of shear arise from fluid motion and the isotropic hydrostatic pressure in the system. In Newtonian fluids, the stresses generated by the flow act parallel to the direction of shear; Figure 22-13 presents a schematic view in *x*, *y*, and *z* coordinates of the stresses operating within a system subjected to shear.

The stresses in the system are denoted by σ_{xx} , σ_{yy} , σ_{zz} , σ_{yx} , σ_{yz} , σ_{zy} , σ_{xy} , σ_{xz} , and σ_{zx} , where the first subscript letter indicates the direction of the plane in which the shear stress lies and the second gives the direction in which it acts. The stresses σ_{xx} , σ_{yy} , and σ_{zz} are the normal stresses in the system. In a Newtonian system in simple shear in the *x*-direction, there are only six independent stresses in the system.²⁷ It is customary to eliminate the isotropic pressure in the system by taking the differences between normal stresses²⁸:

$$\begin{aligned}\sigma_{xx} - \sigma_{yy} &= N_1 \\ \text{and} \\ \sigma_{yy} - \sigma_{zz} &= N_2\end{aligned}\quad (12)$$

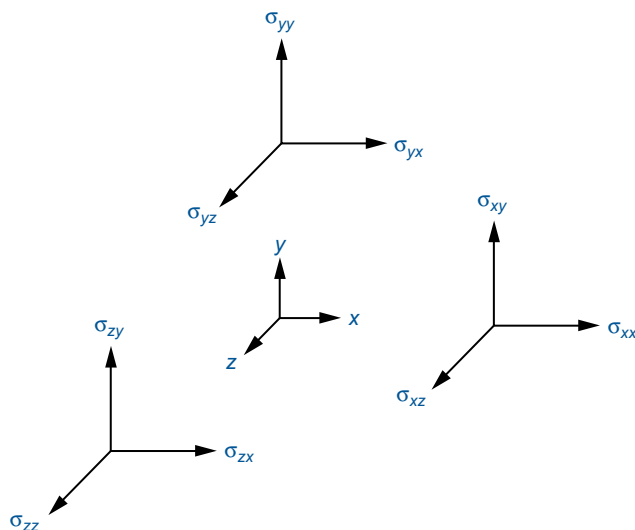


Figure 22-13. Schematic representation in *x*, *y*, and *z* coordinates of stresses operating within a system subjected to shear.

where N_1 and N_2 are the first and second normal stress differences, respectively.

In effect, stresses normal to the direction of shear are different from those in the parallel direction. Fluids that exhibit no normal stress effects but only time-independent non-Newtonian viscous flow are sometimes characterized as generalized Newtonian fluids. However, when normal stress effects are evident, the fluids are viscoelastic in nature. Another indicator of viscoelasticity related to the Deborah number, N_{De} , is the dimensionless Weissenberg number, N_{Wi} , that characterizes the importance of elasticity in the flow by expressing the ratio of the first elastic normal stress difference to the shear stress:

$$N_{wi} = \frac{N_1}{\sigma} \quad (13)$$

For Newtonian fluids $N_{Wi} \equiv 0$; for viscoelastic fluids $N_{Wi} > 0$. In the course of addressing scale-up issues in biotechnology, Zlokarnik²⁹ has estimated N_{Wi} for various aqueous hydrocolloid solutions (carboxymethylcellulose sodium, 1–2% w/v; xanthan gum, 0.05–0.2% w/v) used as model fluids in cell culture studies; typical N_{Wi} values range from 1 to 10. In effect, the substantial viscoelastic nature of these systems must not be ignored if scale-up is to be successful.

In the realm of pharmaceutical solids, plasticity and viscoelasticity are observed during the course of tableting. This is not unexpected, given the conditions of extreme stress used in the compaction of compressed tablets. The viscoelastic parameters of a number of drugs and excipients have been measured under various conditions during the stress-unloading phase of the tablet compaction cycle in a rotary tablet press.^{30,31}

RHEOLOGICAL MODELS

Liquids

Many liquids of pharmaceutical interest follow the empirical power law, or Ostwald-de Waele equation, over a wide range of shear rates, where

$$\begin{aligned}\dot{\gamma} &= K\sigma^n \\ \text{or} \\ \ln \dot{\gamma} &= \ln K + n \cdot \ln \sigma\end{aligned}\quad (14)$$

In many references in the literature, the power law is given as $\sigma = k\dot{\gamma}^n$, so that values of $n > 1$ correspond to dilatant or shear-thickening behavior and values of $n < 1$ to shear-thinning behavior. Of course, for Newtonian liquids $n = 1$.

For so-called *power-law liquids*, a log-log plot of $\dot{\gamma}$ as a function of σ yields a straight line of slope n . The power law equation has the advantage of representing flow behavior in terms of only two constants, K and n . On the other hand, it has the disadvantage of all power laws, namely, the dimensions of the intercept K depend on the value of n , the specific shear rate at which K is evaluated, and the nature of the rheometer.²

For $n = 1$, $K = 1/\eta$, and Newton's law (Equation 2) results. Thus, the exponent, n , is an index of the deviation from Newtonian flow behavior. The more n differs from unity, the more non-Newtonian is the flow behavior, i.e., the more substantial the viscosity decrease or increase with increasing shear. Among pharmaceutical liquids, the most commonly encountered deviants from time-independent Newtonian behavior are those described as shear-thinning fluids for which the power law exponent $n > 1$; less commonly encountered are dilatant fluids for which $n < 1$. Shear-thinning and dilatant liquids frequently follow this empirical power law (or Equation 8) over a wide range of shear rates.

OTHER EMPIRICAL EQUATIONS AND MODELS

Many empirical equations and models have been developed over the years in an effort to describe the flow behavior of non-

Newtonian systems. One of the more successful relationships is the Herschel-Bulkley model:

$$\sigma = k\dot{\gamma}^n + \sigma_0 \quad (15)$$

in which σ is the yield stress, and k is a consistency coefficient. For dilatant or shear-thickening systems, $k > 0$, $1 < n < \infty$, and $\sigma_0 = 0$; for shear-thinning systems, $k > 0$, $0 < n < 1$, and $\sigma_0 = 0$; and, for Bingham plastics, $k > 0$, $n = 1$, and $\sigma_0 > 0$.

Viscoelastic Materials. Viscoelastic behavior is often represented in terms of a mechanical model. Two of the basic elements used in such a model are a helical spring (which obeys Hooke's law and is characterized by a modulus E) and a dashpot (i.e., a cylindrical container with a loosely fitting piston filled with a Newtonian liquid, characterized by its viscosity, η). When the deformation is in shear rather than in tension, Young's modulus E is replaced with the rigidity or shear modulus G . When a spring and a dashpot are connected in series, they form a Maxwell element (Figure 22-14A); when they are connected in parallel, they form a Voigt-Kelvin element (Figure 22-14B). Several Maxwell and/or Voigt-Kelvin elements can be combined in parallel and/or in series to represent the complex viscoelastic behavior of solutions and semisolids. A simple combination is Burgers' model, which consists of a Maxwell and a Voigt-Kelvin element in series (Figure 22-14C) and is characterized by two elastic moduli and two viscosities.

When a constant load or stress, σ_0 , is applied to a Maxwell element, the elastic spring extends immediately to the recoverable strain or elongation, $\gamma_{el} = \sigma_0/E$ (Figure 22-15A). The piston in the dashpot pulls upward gradually; this permanent deformation, γ_{vis} , is directly proportional to time, t . The two deformations are additive: $\gamma = \gamma_{el} + \gamma_{vis}$. At time D , $\gamma = BD = BC + CD = OA + CD$. When the stress is removed at time D (Point B), the spring retracts immediately and fully, and the specimen contracts from B to C by a length, $\gamma_{el} = BC = OA$. The permanent or nonrecoverable deformation, or creep, is

$$\gamma_{vis} = CD = \frac{\sigma_0 t}{\eta}$$

In plots like those of Figure 22-15, compliance (i.e., strain per unit stress) often is used instead of strain. Compliance (e.g., shear or tensile) is the reciprocal of modulus.

If the Maxwell element is stretched to a given γ_0 , the stress required to maintain this deformation constant decreases gradually. As the piston of the dashpot is pulled gradually upward and the

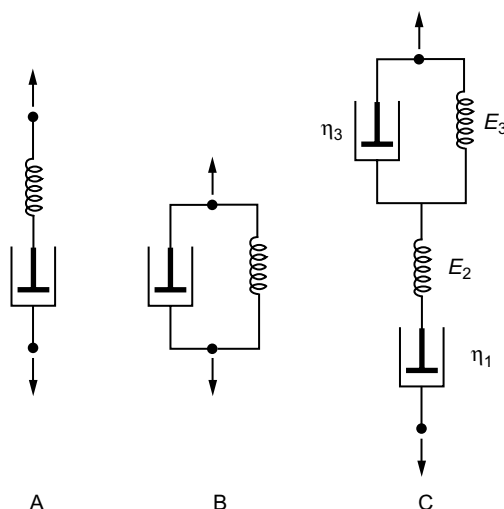


Figure 22-14. Elements of mechanical models for viscoelastic behavior. A, Maxwell element; B, Voigt-Kelvin element; C, Burgers' model. Arrows show applied force or load.

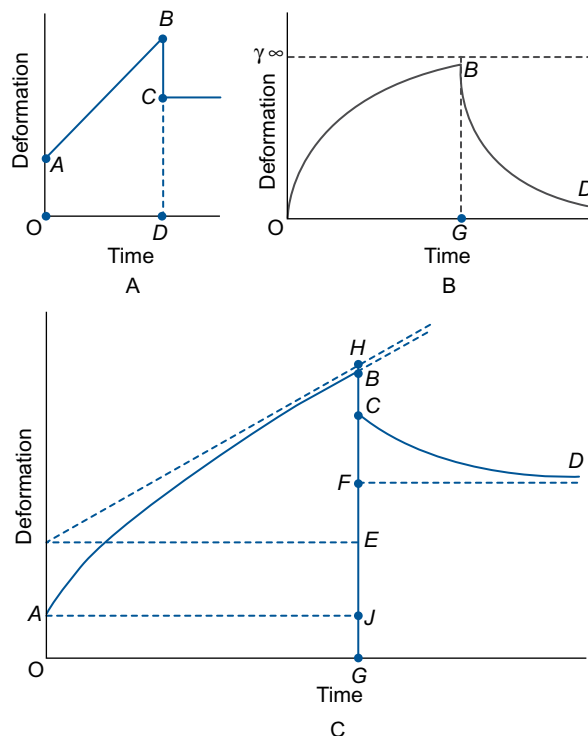


Figure 35-15. Deformation of three rheological models at constant applied stress. A, Maxwell element; B, Voigt-Kelvin element; C, Burgers' body.

dashpot extended, it increasingly relieves the stress on the spring, which gradually contracts. After a long time, as $\gamma \rightarrow \gamma_{vis}$, $\sigma \rightarrow 0$.

If the initial stress is σ_0 and the stress at time, t , is σ , the stress relaxation is

$$\sigma = \sigma_0 e^{-(Et/\eta)} = \sigma_0 e^{-(t/\theta)} \quad (16)$$

The exponent Et/η is dimensionless, and the ratio $\theta = \eta/E$ is the relaxation time.

When a constant stress σ_0 is applied to a Voigt-Kelvin element (Figure 22-15B), the spring can stretch only as fast as the slow extension of the viscous dashpot permits. The greater the viscosity of the liquid in the dashpot, the greater is this retardation. The stress is shared by spring and dashpot, i.e., $\sigma_0 = E\gamma + \eta\dot{\gamma}$. As σ_0 stretches the spring-dashpot assembly, the retarded elastic deformation of the specimen increases with time until, at $t = \infty$, the spring reaches the full extension corresponding to the applied stress: $\gamma_0 = \sigma_0/E$. No additional deformation then takes place. When the stress is removed at time G , the specimen retracts fully to its original shape where $\gamma = 0$, because of the elasticity of the spring, but the motion is damped along the exponential curve BD , which is the mirror image of OB , because the plunger is pulled back only slowly to its original position through the viscous liquid in the dashpot. A retardation time, θ_d , analogous to the relaxation time, is the time required for strain to relax to $1/e$ of its initial value when stress is removed and is defined as $\theta_d = E/\eta$. Along the retarded elastic deformation branch OB

$$\gamma = \left(\frac{\sigma_0}{E} \right) (1 - e^{-t/\theta_d}) = \gamma_\infty (1 - e^{-t/\theta_d}) \quad (17)$$

When the stress is removed, the exponential curve CD is described by

$$\gamma = \left(\frac{\sigma_0}{E} \right) e^{-t/\theta_d} = \gamma_\infty e^{-t/\theta_d} \quad (18)$$

Under constant load, Voigt-Kelvin elements reach a constant deformation; Maxwell elements continue to deform in creep as long as the load is applied. Upon removal of the load, Maxwell elements recover instantly, but not completely, while Voigt-Kelvin elements recover gradually, but completely. Most materials require more than one Maxwell or Voigt-Kelvin element to characterize their rheological behavior and to describe their load-deformation curves. The most suitable models frequently have a range or spectrum of relaxation or retardation times.

The creep deformation-time curve of Burgers' body (see Figures 22-14C and 22-15C) approximates the behavior of many viscoelastic materials, such as disperse and polymeric systems. In Figure 22-15C, the *OAB* portion of the curve, corresponding to the period when the model is under a constant stress, σ_0 , consists of two segments. When the load is applied, *spring 2* stretches instantly and the specimen is elongated from *O* to *A*. On a molecular level, this corresponds to the elastic stretching of bonds between primary structural units, such as primary particles aggregated into flocs or crystallites in a semicrystalline polymer above its glass transition temperature. If the stress is removed at *A*, the specimen would recover its original structure completely.

The second segment, *AB*, results from the combination of the recoverable deformation of *spring 3*, retarded by *dashpot 3*, which is connected in parallel, and the non-recoverable creep of *dashpot 1*. The recoverable deformation predominates in the initial, strongly curved region of *AB*. In this region, interparticle bonds break and reform. The remainder of *AB*, which approaches a straight line, represents mainly the creep of *dashpot 1*. Here, some of the bonds that break are too slow to reform within the test period. The rupture of such interparticle bonds releases some structural units, which flow past one another to produce the permanent deformation.

At time *G*, the overall deformation is the sum of the instantaneous deformation of *spring 2* (*BC* or *AO* or *JG*), of *spring 3* damped by *dashpot 3* (*CF*), and of *dashpot 1* (*HE*). The first two deformations are completely recoverable; the third is not recovered at all:

$$\gamma = BG = JG + EJ + BE = \frac{\sigma_0}{E_2} + \left(\frac{\sigma_0}{E_2} (1 - e^{-t/\theta_d}) \right) + \frac{\sigma_0}{\eta_1} t \quad (19)$$

where the retardation time $\theta_d = \eta_3/E_3$.

The recovery, *BCD*, follows a pattern similar to the deformation. When the stress is removed at time *G*, *spring 2* retracts instantly, and the specimen contracts along *BC* = *OA*. The retraction of *spring 3* is retarded by *dashpot 3* along *CD*. The non-recoverable part of the deformation, due to *dashpot 1*, is represented by *FG* = *HE*.

Disperse Systems. Many pharmaceutical preparations are dispersions of solids or liquids in liquid or semisolid vehicles, and their usefulness often depends on their flow properties. Few disperse systems are Newtonian. Most exhibit non-Newtonian flow behavior, some of it time-dependent, often in conjunction with elastic deformation.

$$\text{Einstein's law of viscosity } \eta_{sp} = \left(\frac{\eta_{12}}{\eta_1} \right) - 1 = \frac{\eta_{12} - \eta_1}{\eta_1} = 2.5\phi \quad (20)$$

is the simplest equation derived to describe the flow behavior of dispersions. Unfortunately, it applies only to Newtonian and idealized systems. The Newtonian viscosities η_{12} and η_1 are those of the dispersion and of the liquid vehicle or solvent, respectively; η_{sp} represents the specific viscosity of the dispersion, i.e., the increase in viscosity of the dispersion over that of the solvent, expressed as a multiple of the viscosity of the solvent; ϕ is the volume fraction of the disperse phase. The viscosity of a dispersion obeying Einstein's law depends only on the viscosity of the solvent and on the volume of solvent replaced by the disperse phase, not on the size of its particles.

Assumptions operative in Einstein's law of viscosity include negligible gravitational and inertial effects and the absence of turbulence. Particles of the dispersion are large compared to the solvent molecules (i.e., the discontinuities between the solvent molecules are negligible, compared to the size of the dispersed particles) but small compared to the dimensions of the viscometer (e.g., the diameter of the capillary or the gap between the coaxial cylinders). The particles of the dispersion neither attract nor repel one another. (In reality, as most dispersions consist of particles of like charge, the viscosity of such dispersions increases due to interparticle electrostatic repulsion [the *electroviscous effect*]. In aqueous dispersions, viscosity increases can be minimized by adding electrolytes.) In addition, Einstein's law assumes that the solvent is continuous and that the particles are unsolvated, smooth, and rigid spheres (e.g., glass beads, polymer latex particles, and many spores and fungi). However, emulsion droplets are deformable and the liquid inside them can circulate. This decreases the distortion of the flow pattern around the droplets and reduces the numerical constant in Equation 20 below 2.5. Rigid anisometric particles offer increased resistance to flow, raising the constant above 2.5. If the solvation layer of solvated spherical particles is included in ϕ , their dispersions may obey Equation 20. Examples of the latter are solutions of globular proteins at their isoelectric point, where their net electrical charge is zero.

Furthermore, the "ideal" dispersions addressed by Einstein's law are considered to be so dilute that the distortion of the laminar streamlines of the solvent at the surface of one particle does not overlap and reinforce the distortions around its neighbors. However, at higher disperse phase concentrations, the perturbation of laminar flow produced by one particle reaches into the fields of other particles. This produces additional resistance to flow and increases η_{sp} and η_{12} above the values given by Equation 20.

Deviations from these conditions result in higher dispersion viscosities than those calculated by Einstein's law except that, when the disperse phase is fluid, the calculated viscosity is too high. An example of an extreme positive deviation is found in aqueous sodium bentonite dispersions. Their specific viscosity is about 70 times greater than that calculated from Equation 20. The particles are thin plates, deviating considerably from spherical shape. They are hydrated, and their negatively charged faces attract the positively charged edges but repel the negatively charged faces of other particles. Polymer solutions with their thread-like, highly solvated, entangled macromolecules also deviate considerably from Einstein's law. Several variations on Einstein's law express the specific viscosity as a polynomial in ϕ , thereby broadening its applicability, for example, to more concentrated dispersions.

One of the more successful relationships applied to dispersions with a high solids content is the Casson model³²:

$$\sigma^{1/2} = k_0 + k_1 \dot{\gamma}^{1/2} \quad (21)$$

where k_0 and k_1 are constants that depend on the properties of the dispersion medium and the disperse phase. Although the Casson equation has been used empirically in modeling the rheological properties of a wide range of concentrated dispersions, it was originally derived from basic principles with the assumption that the disperse phase behaved as rigid rods.

COMPUTATIONAL RHEOLOGY

Empirical relationships aside, numerical methods for the characterization of non-Newtonian flow were developed in the 1960s, but computational rheology has emerged to address previously intractable problems, such as three-dimensional transient flows of polymeric liquids, non-isothermal non-Newtonian flows, or turbulent flow of generalized Newtonian and viscoelastic materials.³³ Computer software, in consort with modern rheometer design, has also facilitated the development

of more complex models of deformation and flow under flow regimes, ranging from the laminar to the turbulent, even encompassing the transitional flow regime, in which flow is neither completely laminar nor completely turbulent. In all likelihood, the net effect of these advances is to demystify rheological principles and allow the *a priori* estimation of the mechanical properties of the living and nonliving systems with which we contend.

RHEOLOGICAL MEASUREMENTS

Many of the instruments designed to measure steady shear flow and the corresponding viscosity do not enable the precise delineation of sample dimensions, the forces exerted, or the resultant deformation or flow. These devices are referred to in this text as viscometers. Those instruments that enable a thorough, geometrically accurate analysis of deformation and flow to be made are referred to as rheometers. The term *viscometer* should be reserved for a subset of rheometers that measure only Newtonian viscosity. When flow or deformation is not well defined or is indeterminate, the resulting measurements are likely to correspond, at best, to approximations of the true rheological behavior of the sample.

A wide array of viscometers and rheometers is available commercially; they vary in regard to mode of generation of simple shear flow, measurement geometries, range of shear rate and shear stress, sample size, ease of operation, reproducibility, and cost. It is necessary to select an instrument that provides the information desired for a specific application. Information on selected suppliers along with websites is listed in Table 22-4. This section outlines the basic aspects of some of the types of viscometers and rheometers available.

All of the equations routinely used with viscometers yield viscosities, shear rates, and shear stresses based on the assumption that the fluid is Newtonian. The viscosity, in particular, is often an *apparent* viscosity because the corresponding shear stress and shear rate may not be precisely defined. For non-Newtonian fluids—and in order to compare data from one

viscometer to another—the pseudo-Newtonian data must be corrected for non-Newtonian behavior.

One very important potential problem is measuring the apparent viscosity of a material at a single rate of shear instead of covering a wide range of shear rates. A Newtonian, a shear-thinning, a plastic, and a dilatant fluid may all have the same apparent viscosity if their flow curves have a common intersection point since apparent viscosity is defined as shear stress/shear rate, whereas their flow curves may demonstrate very dissimilar behavior.

Measuring the apparent viscosity over a range of shear rates but maintaining the material for only short times at each shear rate also can give misleading results by missing thixotropic or time-dependent effects. The latter usually are detected by measuring the flow curve first at increasing shear rates and, after reaching the desired maximum value, at decreasing shear rates. An alternate technique is to keep the material at a constant shear rate for a given period of time and to observe the decay, if any, of the shear stress with time.

GENERAL VISCOMETER AND RHEOMETER TYPES

At the outset it should be noted that simple shear flow can be generated by (a) *pressure flow*, in which a sample is forced through a channel, pipe, or slit; or (b) *drag flow*, in which motion is caused by moving one of the walls in contact with the sample to be evaluated. A third methodology, based on the velocity of an object rolling or falling through the liquid or rising through the liquid around the object, involves substantially more complex flow patterns. Viscometers or rheometers embodying these methods are exemplified by capillary or slit devices, rotational instruments, and falling or rolling ball units, respectively.

The apparent viscosity can be determined following calibration of the instrument with a standard Newtonian fluid:

$$\eta = K_{\eta}(\dot{\gamma}, \dot{\sigma}) \quad (22)$$

Table 22-4. Selected Viscometer and Rheometer Companies with Web Addresses

Company	Instrumentation	Web Address
Anton Paar USA, Inc.	Rotational rheometers; rolling ball viscometers	http://www.anton-paar.com/
Brookfield Engineering Laboratories, Inc.	Brookfield viscometers/rheometers	http://www.brookfieldengineering.com
Cannon Instrument Co.	Capillary viscometers	http://www.cannoninstrument.com
Grace Instrument Co.	Viscometers/rheometers, especially for high pressures & temperatures	http://www.graceinstrument.com
Malvern Instruments	Rosand capillary rheometers; Kinexus rheometers; Bohlin rheometers	http://www.malvern.com
proRheo GmbH	Rheomat viscometers/rheometers	http://www.prorho.de
RheoSense, Inc.	VROC viscometers/rheometers	http://www.rheosense.com
TA Instruments	ARES-G2 rheometers; Discovery Hybrid rheometers	http://www.tainstruments.com
Texture Technologies	TA.HD Plus; TA.XT Plus, TA.XT Express texture analyzers	http://www.texturetechnologies.com
Thermo Fisher Scientific, Inc.	Haake RheoStress, RotoVisco rotational rheometers; Höppler rolling ball viscometer	http://www.thermoscientific.com
Vilastic Scientific, Inc.	V-E System; Bioprofiler; Vilastic-3 viscoelasticity analyzers	http://www.vilastic.com

Table 22-5. Viscometer/Rheometer Parameters

Instrument Type	$\chi\gamma_{\dot{\alpha}}$	$\chi\sigma$
Pressure flow	Time of fluid flow	Fluid density
Drag flow	Rotational speed	Torque
Falling sphere or object	Velocity of falling sphere or object	Sphere or object diameter; density difference between sphere and fluid
Rolling sphere	Velocity of rolling sphere; angle of inclination	Sphere or object diameter; density difference between sphere and fluid

where K_{η} is the viscometer or rheometer constant for determining viscosity, and $x_{\dot{\gamma}}$ and x_{σ} are experimental parameters, related to shear rate and shear stress, respectively. Alternately, shear rate and shear stress can be determined separately as follows:

$$\dot{\gamma} = K_{\dot{\gamma}} x_{\dot{\gamma}} \quad (23)$$

and

$$\sigma = K_{\sigma} x_{\sigma} \quad (24)$$

where $K_{\dot{\gamma}}$ and K_{σ} are the viscometer constants for determining shear rate and shear stress, respectively. $K_{\dot{\gamma}}$ and K_{σ} are related to K_{η} as follows:

$$K_{\eta} = \frac{K_{\dot{\gamma}}}{K_{\sigma}} \quad (25)$$

The experimental parameters $x_{\dot{\gamma}}$ and x_{σ} for three basic types of viscometers and rheometers are listed in Table 22-5.

In shear flow, the temperature of the sample rises progressively as the energy used to overcome its viscous resistance is transformed into heat. The higher the viscosity, the greater the heat buildup. Since the viscosity of liquids depends strongly on temperature, accurate temperature control is essential. Despite the minimal heat buildup in capillary viscometers, since only a small portion of the test liquid is sheared at any given moment and the measurements are intermittent, capillary viscometer use typically employs provisions for temperature control. Accurate temperature control can be accomplished through the use of a constant-temperature water or oil bath for immersion of a capillary instrument or the sample holder (e.g., cup) of a rotational device, or the incorporation of a Peltier plate or its concentric cylinder equivalent.

CAPILLARY AND SLIT VISCOMETERS AND RHEOMETERS

CAPILLARY INSTRUMENTS

The glass Ostwald, Cannon-Fenske, and Ubbelöhde capillary viscometers are among the most popular capillary instruments in use. The duct is a cylindrical capillary of length L , and the driving force causing the liquid to flow through it is its weight. Thus, ΔP in Poiseuille's law (Equation 6) is replaced by the hydrostatic pressure $h\rho g$ of a liquid column of height h and density ρ ; g is the acceleration of gravity:

$$\eta = \left(\frac{\pi R^4 h g}{8LV} \right) \rho t = K_{\eta} \rho t \quad (26)$$

A standard volume of the liquid is transferred into the viscometer. Liquid is then drawn into the upper reservoir bulb of the instrument by suction (Figure 22-16). The efflux time, t ,

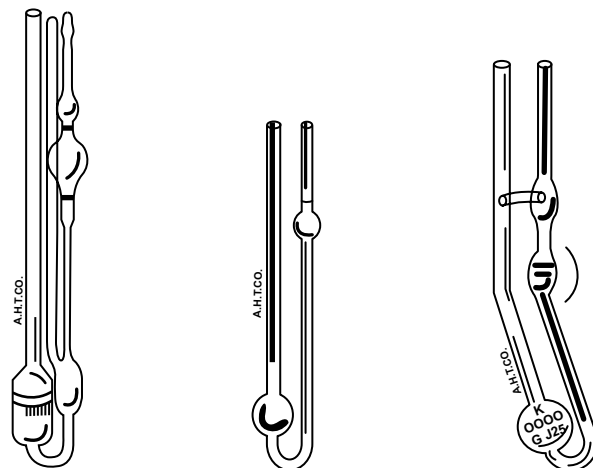


Figure 22-16. Capillary viscometers: Ubbelöhde, Ostwald, and Cannon-Fenske.

required for the liquid level (meniscus) to fall from the upper to the lower mark (the height, h), emptying the upper reservoir, is determined. The height decreases as liquid flows through the capillary, but its time-averaged value is constant for a given viscometer containing a constant volume of liquid.

Calibration of the viscometer consists of determining the constant K_{η} experimentally with a liquid of known viscosity and density by measuring the efflux time t according to Equation 26. The viscosity for an unknown liquid is then determined from Equation 27:

$$\eta = K_{\eta} \rho t \quad (27)$$

The liquid used to calibrate the viscometer should have approximately the same flow time t as the unknown, in order to minimize the divergence from Equation 27. The major portion of the potential energy represented by the hydrostatic pressure head is dissipated (as heat) in overcoming the viscous resistance to flow in the capillary tube, i.e., the friction of layer slipping past concentric layer. A small portion of the potential energy is required to accelerate the liquid as it enters the capillary from the reservoir (*kinetic energy correction*). Another small amount is used up in the streamlines converging from the broad reservoir into the narrow capillary and in spreading the streamlines upon issuing from the capillary (the so-called *entrance* or *end effects*). These two corrections are included experimentally in the constant K_{η} .

It is not necessary to evaluate K_{η} if the viscosity of one liquid relative to a reference liquid is determined. It suffices to measure the flow time t_1 for the reference liquid of known viscosity η_1 and density ρ_1 and to compare it with the flow time t_2 for the liquid of density ρ_2 whose viscosity η_2 is to be determined. Thus, the viscosity of the unknown is

$$\eta_2 = \left(\frac{t_2 \rho_2}{t_1 \rho_1} \right) \eta_1 \quad (28)$$

If the viscosity of the reference liquid is not known, a *relative viscosity* η_r can be determined, as defined in the following equation:

$$\eta_r = \frac{\eta_2}{\eta_1} = \frac{t_2 \rho_2}{t_1 \rho_1} \quad (29)$$

If the density of the liquid is not known, a *kinematic viscosity* ν , defined as the absolute viscosity divided by the fluid density, can be determined:

$$\nu = \frac{\eta}{\rho} \quad (30)$$

Kinematic viscosity has the units of stoke(s) or centistoke(s) and is not numerically or dimensionally equivalent to the viscosity in mPa·s.

Shear rate and shear stress can be determined explicitly from Equations 23 and 24 as they apply to capillary viscometers. It can be shown that Equation 26 can be separated into shear stress and shear rate as follows:

$$\dot{\gamma} = \frac{4VL}{\pi R^3 t} = K_{\dot{\gamma}} \left(\frac{L}{t} \right) \quad (31)$$

and

$$\sigma = \left(\frac{h g R}{2L} \right) \rho = K_{\sigma} \rho \quad (32)$$

Thus, $K_{\dot{\gamma}}$ and K_{σ} in Equations 31 and 32 could be calculated directly since all of the parameters in their definition appear to be measurable. In reality, the length of the capillary, L , cannot be measured precisely because of the corrections discussed previously. Consequently, calibration of the capillary viscometer with a Newtonian standard is required.

A range of shear rates and shear stresses can be obtained for a given liquid by using a series of glass capillary viscometers of different capillary diameters since the usual glass capillary viscometer affords viscosity measurements at only one time-averaged value of shear rate. The efflux times should exceed 200 s to minimize the kinetic energy correction and possible errors in determining the efflux time. Given the small ΔP attained with glass capillary viscometers, their use in measurements of high-viscosity fluids is impractical. Alternatively, then, various *capillary extrusion* instruments operating under pressure, using an external pressurized gas reservoir or other source of constant pressure, are commercially available. In addition, capillary extrusion viscometers may allow the operator to vary the shear rate or shear stress, thereby facilitating non-Newtonian fluid characterization.

SLIT VISCOMETERS AND RHEOMETERS

Slit viscometers and rheometers, like capillary extrusion devices, involve external pressure-driven flow. However, flow typically occurs through a thin rectangular channel or slit. In one microfluidic embodiment of this design (employing a software-driven precision temperature-controlled microsyringe), developed by RheoSense, Inc., an array of pressure sensors embedded in the floor of a microchannel enable the direct determination of the pressure gradient dp/dx , which is proportional to the apparent viscosity. For Newtonian liquids, the following relationships can be employed:

$$\dot{\gamma}_{app} = \frac{6Q}{\varpi h^2} \quad (33)$$

$$\sigma = - \left(\frac{dp}{dx} \right) \left(\frac{\varpi h}{2\varpi + 2h} \right) \quad (34)$$

and

$$\eta = \frac{\sigma}{\dot{\gamma}_{app}} \quad (35)$$

where Q is the flow rate, ϖ is the width of the channel, and h is the depth of the channel. For non-Newtonian liquids, the true shear rate in rectangular slit flow is calculated by applying the Weissenberg-Rabinowitsch correction to the data:

$$\dot{\gamma} = \left(\frac{\dot{\gamma}_{app}}{3} \right) \left(2 + \frac{d \ln \dot{\gamma}_{app}}{d \ln \sigma} \right) \quad (36)$$

The development of this *microfluidic* approach to rheometry confers advantages such as precise control of flow rates, use of low sample volumes (e.g., <50 μ L) and flow rates, and the attainment of a wide range of shear rates (e.g., 10^{-1} to 10^6 s $^{-1}$).

ROTATIONAL VISCOMETERS AND RHEOMETERS

These instruments depend on the fact that the rotation of a solid body in contact with a liquid is impeded by a retarding force as a result of the viscous drag, which is proportional to the viscosity of the liquid. The advantages of rotational viscometers and rheometers are that the shear rate or shear stress can be varied over a wide range of values, and that continuous measurements at a given shear rate or shear stress can be made for extended periods of time, affording measurements of the time dependency as well as of the shear dependency of the viscosity.

Classical rotational viscometers of the *MacMichael* type—with an outer sample-containing cup rotated at a constant, though adjustable speed—and the *Stormer* type—with a stationary cup and a bob or rotor driven by weights linked to the shaft of the bob—are passé, given the advances in instrument engineering and integrated software. In most modern rotational viscometers or rheometers, the cup, as with the Stormer viscometer, is fixed. The bob is rotated at a constant speed that can be varied over a wide range of rpm or shear rates. The torque on the rotating bob required to maintain a constant speed of rotation against the viscous drag of the liquid is detected and instantaneously converted to appropriate rheological parameters.

INFINITE GAP VISCOMETERS

This is a type of concentric-cylinder viscometer in which the viscous traction is measured on a spindle or bob rotating in the liquid, which is contained in a beaker or similar container. The size of the container is such that $R_c^2 \gg R_b^2$. The viscometer spindle can be inserted not only into beakers in the laboratory but also into kettles, reactors, and mixing tanks in the plant. Thus, the viscometer can be adapted for continuous in-line viscosity measurements. A guard can be mounted around the spindle to prevent it from being deflected laterally, thereby causing misalignment of the shaft. The guard also ensures that the condition, $R_c^2 \gg R_b^2$ is maintained. The apparent viscosity is given by

$$\eta_{app} = \frac{3M}{32R^3\Omega} \quad (37)$$

where M is the torque, R is the radius of the spindle or bob, and Ω is the angular velocity of the spindle or bob in radians s $^{-1}$.

COAXIAL-CYLINDER RHEOMETERS

The geometry of a coaxial-cylinder viscometer is shown in Figure 22-17. The viscosity is calculated by means of the *Margules* equation:

$$\eta = \left(\frac{R_c^2 - R_b^2}{4\pi h R_b^2 R_c^2} \right) \left(\frac{M}{\Omega} \right) = K'_{\eta} \left(\frac{M}{\Omega} \right) \quad (38)$$

where R_b and R_c are the radii of the bob and cup, respectively, h is the height of the bob immersed in the liquid, M is the torque, and Ω is the angular velocity of the bob in radians s $^{-1}$. The angular velocity of the bob is usually expressed in terms of rpm ($\Omega = 2\pi/60$ rpm). The calibration factor K_{η} can be determined experimentally for each combination of cup and bob by means of appropriate viscosity standards.

Equation 38 was derived for two coaxial cylinders of infinite length. The *end effect* is the traction on both end surfaces of the bob if it is completely immersed in the liquid, or on its bottom surface if it is only partly immersed. Thus, h in Equation 38 is the “effective” height of the bob. One way to correct for the end effect is by adding an increment Δh to the height h of the bob

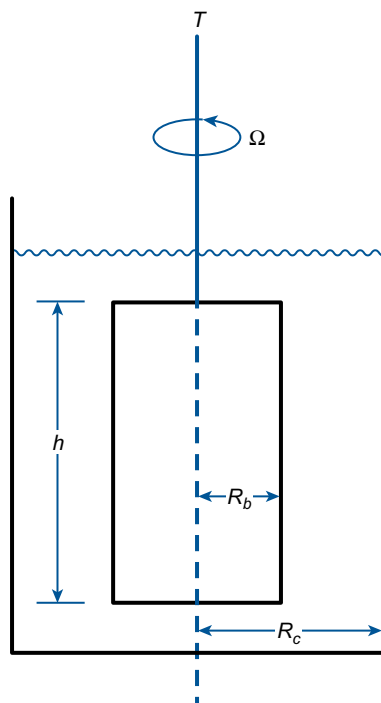


Figure 35-17. Geometry of a coaxial-cylinder (cup and bob) viscometer.

to arrive at an “effective height.” For a partly immersed bob with a flat bottom, Δh is frequently on the order of $0.1 h$. The added height can be determined experimentally for each material by filling the annular gap to different depths of immersion of the bob. The ratio M/Ω is plotted against the height or depth of immersion h . The negative intercept of this usually straight line with the h axis represents Δh . The end effect is more conveniently accounted for by calibrating the viscometer with a viscosity standard as for capillary viscometers.

Shear rate and shear stress can be determined explicitly from Equations 23 and 24 as they apply to a coaxial-cylinder viscometer. It can be shown that Equation 38 can be separated into shear stress and shear rate as follows:

$$\sigma = \frac{M}{2\pi R_b^2 h} \quad (39)$$

$$\dot{\gamma} = \left(\frac{2R_c^2 \Omega}{R_c^2 - R_b^2} \right) \quad (40)$$

CONE-AND-PLATE VISCOMETERS AND RHEOMETERS

These instruments consist of a rotating cone with a very obtuse angle and a stationary lower flat plate. The plate is raised until the apex of the cone just touches its surface. The liquid fills the narrow triangular gap between cone and plate (Figure 22-18). Its surface tension prevents it from spreading on the plate. The cone is driven at controlled speeds that can be varied continuously. The viscous drag on the rotating cone exerts a torque that is proportional to the shear stress. The angle θ formed by cone and plate is usually less than 3 degrees, the average gap width is less than 2 mm, and sample volumes smaller than 0.5 mL can be used.

For small values of θ in radians, the viscosity is determined as

$$\eta = \left(\frac{3\theta}{2\pi R_b^2} \right) \left(\frac{M}{\Omega} \right) \quad (41)$$

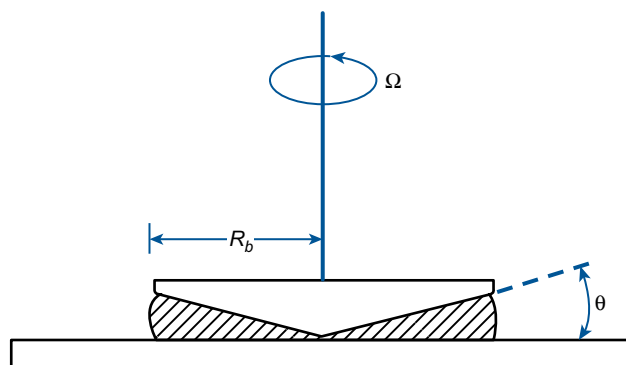


Figure 35-18. Geometry of a cone-and-plate viscometer.

where R_b is the cone radius. Equation 41 can be expressed in terms of shear stress and shear rate as follows:

$$\sigma = \frac{3M}{2\pi R_b^2} \quad (42)$$

$$\dot{\gamma} = \frac{\Omega}{\theta} \quad (43)$$

Equation 43 is applicable when $\theta < \sim 3^\circ$, i.e., the rate of shear across the conical gap is relatively constant.

PARALLEL PLATE VISCOMETERS AND RHEOMETERS

A parallel plate or parallel disk instrument is similar to a cone-and-plate instrument except that the upper plate or disk of radius R is rotated at a constant angular velocity parallel to a fixed plate. Sample introduction, especially for very viscous or soft semisolid materials, is far less problematic for a parallel plate instrument than for concentric cylinder or cone-and-plate devices. An additional advantage is that variations in the shear rate can be achieved not only by altering the rate of rotation but also by varying the size of the distance or gap between the plates. The equations, with h equal to the distance between the plates, are as follows:

$$\eta = \left(\frac{h}{\pi R^4} \right) \left(\frac{M}{\Omega} \right) \quad (44)$$

$$\sigma = \left(\frac{3M}{2\pi R^3} \right) \left(3 + \frac{d \ln M}{d \ln \gamma_R} \right) \quad (45)$$

$$\dot{\gamma} = \frac{R\Omega}{h} \quad (46)$$

One disadvantage is that the flow between parallel plates is not homogeneous so that strain varies from 0 at the center to a maximum at the edge of the rotating plate. Consequently, torque should be evaluated at various rotational speeds. Hence, the correction in Equation 45 is shown for the change in torque relative to change in the rate of shear at the edge of the plate.

FALLING BALL AND ROLLING BALL VISCOMETERS AND RHEOMETERS

Falling Ball Viscometers. With these instruments, viscosities are determined by measuring the velocity of a ball falling in the liquid being studied. This method is best suited for Newtonian liquids because it measures viscosities at a single shear rate. When a sphere of radius R and density ρ_2 descends vertically through a liquid of density ρ_1 , the driving force is the effective weight of the sphere, i.e., the weight of the sphere minus the

weight of the liquid it displaces. It equals the volume of the sphere multiplied by the density difference ($\rho_2 - \rho_1$) and by the acceleration of gravity g , i.e., $(4\pi R^3/3)(\rho_2 - \rho_1)g$. The viscous resistance of the liquid is given by Stokes's law, namely, $6\pi\eta Rv$. When the sphere attains the terminal or constant velocity v (which occurs soon after it is dropped into the liquid column), the two opposing forces are equal, so that the viscosity is, in general, governed by Stokes's law:

$$\eta = \frac{2R^2(\rho_2 - \rho_1)g}{9v} \quad (47)$$

Stokes assumed that the velocity of sedimentation was very low and that the liquid medium extended an infinite distance from the sphere. Among the factors requiring correction, therefore, is the proximity of the wall. The viscosity can then be determined from the following:

$$\eta = \left(\frac{2R^2g}{9} \right) \left(\frac{\rho_2 - \rho_1}{v} \right) = K_\eta \left(\frac{\rho_2 - \rho_1}{v} \right) \quad (48)$$

where v is the velocity in cm/sec of the falling, rolling, or rising object. Most viscometers of this type have a known distance marked on the instrument so that one merely measures the time for the object to move between the marks. Equation 48 then becomes:

$$\eta = \left(\frac{2R^2g}{L9} \right) \left(\frac{\rho_2 - \rho_1}{v} \right) t = K_\eta (\rho_2 - \rho_1) t \quad (49)$$

where L is the distance between the two marks on the viscometer, and t is the time for the object to travel between the marks. Calibration of the viscometer involves determining K_η using a viscosity standard for each ball or object.

For very viscous liquids, values of the Newtonian or the apparent viscosity at a single shear rate can be measured with a metal rod plunger immersed concentrically in a vertical cylindrical glass tube filled with the liquid. The tube is closed at the bottom and thermostatted. The diameter of the metal plunger is ~68 percent of the inside diameter of the glass tube. The weight of the plunger forces the liquid upward through the narrow, annular space between plunger and tube. The terminal or steady-state velocity of descent of the plunger is proportional to the viscosity of the liquid.

All of these instruments have guides to ensure that the probes descend along the vertical axis of the cylindrical containers.

Rolling Ball Viscometers and Rheometers. In rolling ball instruments, the ball rolls down the side of a tilted tube so that it follows the same path each time a measurement is made. In contrast to falling ball devices, the path is not always precisely the same. An additional advantage of rolling ball instruments is that the operator can adjust the angle at which the tube is tilted, thereby changing the effective shear stress and the time for the ball to pass between two marks. Furthermore, the use of balls of different composition, with different densities and/or different diameters (relative to the diameter of the tube), adds to the versatility and range of these instruments.

The apparent viscosity is given by Equation 50:

$$\eta = \frac{k(\rho_2 - \rho_1)}{v} \quad (50)$$

where k is an instrument constant, ρ_1 and ρ_2 are the densities of the ball and the fluid, respectively, and v is the velocity of the rolling sphere. The instrument constant k can be determined by calibration with known viscosity standards at a specific temperature, angle of inclination, ball and tube dimensions, and ball composition.

TENSILE AND TORSION TESTERS; PENETROMETERS

In the case of semisolids or very viscous liquids, a cone or needle attached to a holding rod is released and plunges vertically

into the sample under the influence of its own or added weight. The depth of penetration within a given time interval, e.g., 10 s, is used to rate the consistency of the material. The results cannot be translated into absolute viscosity or yield values. A modern variation of this principle is illustrated by the "Texture Analyzer" (www.texturetechnologies.com), which can provide a wide range of static and dynamic measurements of penetration and stress.

COMPARISON BETWEEN INSTRUMENTS

When a material is to be studied over a wide range of shear rates, more than one viscometer or rheometer may need to be used because each individual instrument may have too limited a range. It is advisable to use instruments in the range of overlapping shear rates to ensure that the corresponding flow curves do indeed coincide. When the flow curves are plotted as shear stress versus rate of shear, instruments of different dimensions, and even based on different principles, produce a single curve for a given material at a given temperature, *if the material is Newtonian*. Flow curves plotted in units other than shear stress and shear rate, such as torque units versus rpm, depend on the geometry of the viscometer and are *not* directly comparable.

In addition, the preceding discussion provides Newtonian parameters, even if the material being evaluated is non-Newtonian. In this case non-Newtonian corrections must be applied in order for flow curves from different viscometers to be comparable.

NON-NEWTONIAN CORRECTIONS³⁴⁻³⁶

All of the preceding equations have been derived based on Newtonian behavior, which means that the shear rate is constant everywhere in the viscometer, e.g., in a cup-and-bob viscometer, at the bob surface and at the cup surface. This is not true for non-Newtonian fluids. Comparison of non-Newtonian fluids requires that viscosity data be corrected to a common reference point. The result is a correction to the shear rate term in reference to a fixed point in the viscometer, i.e., the bob surface in a cup-and-bob viscometer or the wall in a capillary viscometer. It depends on both the viscometer or rheometer and the fluid being tested. In general, the correction takes the form

$$\dot{\gamma}_{\text{corrected}} = \dot{\gamma} \cdot f(n) \quad (51)$$

where $f(n)$ is the correction factor and n is the slope of a log-log plot shear stress as a function of shear rate. For many non-Newtonian fluids of pharmaceutical interest, n is a constant. If the log-log plot is not linear, n must be determined numerically at each data point. Flow curves plotted as shear stress as a function of corrected shear rates are then comparable between viscometers.

REFERENCES

1. Doraiswamy D. The origins of rheology: a short historical excursion. *Rheol Bull* 2002; 71(1): 1-9.
2. Reiner M. *Deformation, Strain, and Flow*, 2nd edn. New York: Interscience, 1960.
3. Scott Blair GW. *Elementary Rheology*. London: Academic Press, 1969: 2.
4. Houwink R, de Decker HK, eds. *Elasticity, Plasticity and Structure of Matter*, 3rd edn. London: Cambridge University Press, 1971.
5. Dealy JM. Official nomenclature for material functions describing the response of a viscoelastic fluid to various shearing and extensional deformations. *J Rheol* 1995; 39(1): 253-265.
6. Hackley VA, Ferraris CF. *Guide to Rheological Nomenclature: Measurements in Ceramic Particulate Systems*. NIST Special Publication 946. Washington DC: National Institute of Standards and Technology, 2001.

7. (a) Ree T, Eysring H, in Eirich FR, eds. *Rheology*, vol. 2. New York: Academic Press, 1958; (b) Alder BJ, Hildebrand JH. *Ind Eng Chem Fundam* 1973; 12: 388.
8. Fisher EK. *Colloidal Dispersions*. New York: Wiley, 1950.
9. Braun DB, Rosen MR. *The Rheology Modifiers Handbook: Practical Use and Application*. New York: William Andrew, 1999.
10. Reiner M. The Deborah number. *Physics Today* 1964; 17(1): 62.
11. Goodwin JW, Hughes RW. *Rheology for Chemists: An Introduction*. Cambridge: Royal Society for Chemistry, 2000.
12. Sherman P. *Emulsion Science*. London: Academic, 1968.
13. Davis SS. Viscoelastic properties of pharmaceutical semisolids. I. Ointment bases. *J Pharm Sci* 1969; 58: 412–418.
14. Radebaugh GW, Simonelli AP. Application of dynamic mechanical testing to characterize the viscoelastic properties of powder-filled semisolids. *J Pharm Sci* 1984; 73: 590–594.
15. Lowe GDO. Nature and clinical importance of blood rheology. In: Lowe GDO, edn. *Clinical Blood Rheology*. Boca Raton, FL: CRC Press, 1988: 1–10.
16. Gross DR, Hwang NHC. *The Rheology of Blood, Blood Vessels and Associated Tissues*. NATO Advanced Study Institute on Biorheology, 1980.
17. Reinhart WH *et al.* Acidosis induced by lactate, pyruvate, or HCl increases blood viscosity. *J Crit Care* 2002; 17(1): 68–73.
18. Patton HD *et al.* *Textbook of Physiology*, vol. 2, 21st edn. Philadelphia: WB Saunders, 1989.
19. Dintenfass L. *Rheology of Blood in Diagnostic and Preventive Medicine: An Introduction to Clinical Haemorheology*. London: Butterworths, 1976.
20. Gomes N *et al.* Shear stress modulates tumour cell adhesion to the endothelium. *Biorheology* 2003; 40: 41–45.
21. Chmiel H Walitza E. *On the Rheology of Blood and Synovial Fluids* Chichester: Wiley, 1980.
22. Chien S *et al.* *Clinical Hemorheology*. Dordrecht: Martinus Nijhoff, 1987.
23. Ruch TC, Patton HD. *Physiology and Biophysics*. Philadelphia: WB Saunders, 1965.
24. Brown DW *et al.* Hematocrit and the risk of coronary heart disease mortality. *Am Heart J* 2001; 142(4): 657–663.
25. Church TS *et al.* Improvements in blood rheology after cardiac rehabilitation and exercise training in patients with coronary heart disease. *Am Heart J* 2002; 14(2): 349–355.
26. Basaria S *et al.* Effects of methyl testosterone administration on plasma viscosity in postmenopausal women. *Clin Endocrinol* 2002; 57(2): 209–214.
27. Rielly CD. In: Fryer PJ *et al.* eds. *Chemical Engineering for the Food Industry*. London: Chapman & Hall, 1997.
28. Prud'homme RK. In: Schulz DN, Glass JE, eds. *Polymers as Rheology Modifiers*. Washington, DC: American Chemical Society, 1991.
29. Zlokarnik M. Scale-up of processes using material systems with variable physical properties. *Chem Biochem Eng Q* 2001; 2 (15): 43–47.
30. Rippie EGDanielson DW. Viscoelastic stress/strain behavior of pharmaceutical tablets: analysis during unloading and postcompression periods. *J Pharm Sci* 1981; 70(5): 47–82.
31. Danielson DWMorehead WTRippie EG. Unloading and postcompression viscoelastic stress versus straining behavior of pharmaceutical solids. *J Pharm Sci* 1983; 72(4): 342–345.
32. Casson N. In: Mill CC, edn. *Rheology of Disperse Systemology*. New York: Pergamon, 1959.
33. Owens RG, Phillips TN. *Computational Rheology*. Singapore: World Scientific, 2002.
34. Steffe JF. *Rheological Methods in Food Process Engineering*, 2nd edn. East Lansing, MI: Freeman Press, 1992. Available at [steffe/freebook](http://www.egr.msu.edu/~steffe/freebook)>www.egr.msu.edu/~steffe/freebook.
35. Van Wazer JR *et al.* *Viscosity and Flow Measurement*. New York: Interscience Publishers, 1962.
36. Macosko CW. *Rheology: Principles, Measurements, and Applications*. New York: Wiley-VCH, 1998.

INTRODUCTION 411

PROPERTIES OF POWDERS 411

BULK PROPERTIES OF POWDERS 419

PRODUCTION METHODS 425

PARTICLE SIZE REDUCTION 426

POWDER HANDLING 428

POWDERS AS A DOSAGE FORM 432

INTRODUCTION

Powder applications are expanding in a variety of highly developed fields, such as foods, cosmetics, chemicals, and many other fundamental fields, particularly in pharmaceuticals. The majority of active pharmaceutical ingredients (APIs) are administered as solid dosage forms, prepared by processing and formation of powders. A powder is defined as a dry, solid substance, consisting of a large number of finely divided particles (varying 10nm-1000µm), and typically obtained by crushing, grinding, or comminuting.¹ Powders have a large specific surface area and surface free energy, therefore, exhibiting some physical and chemical properties. Some of the properties related to pharmaceuticals, such as particle size, shape, surface area, density, porosity, flowability, will influence the forming, packing, and processing of a variety of dosage forms, e.g. granules, tablets, capsules, suspensions, etc. Additionally, the solubility and bioavailability of a drug formulation can be affected by some basic characteristics of powders as well. Besides acting as pharmaceutical excipients, such as diluents, disintegrating agents, and lubricants, powders can also present as a pharmaceutical dosage form. In summary, powders are fundamental to pharmaceuticals, which have a significant impact on the development, manufacture, quality control, and packing of a variety of dosage forms, including tablets, capsules, granules, suspensions, microcapsules, microspheres, and dry powder inhalers (DPIs). This chapter is designed to provide a comprehensive understanding of powders, including fundamental properties of powders, primary methods of powder production, handling, and powders as a dosage form.

PROPERTIES OF POWDERS

PARTICLE CHARACTERIZATION

Particle Size

Particle size is a basic property of a powder, which is essential to powder technology and dosage form design. Particles with different size and size distributions exhibit different behaviors. For instance, smaller particles have greater specific surface area and their bulk density, porosity, flowability, and solubility significantly differ from that of larger particles.

DEFINITION AND EXPRESSION OF PARTICLE SIZE

Particle size, typically expressed as particle diameter, is essential to characterizing powders. A powder is a plurality of solid particles; powders prepared using different methods, or by the same method, vary greatly in size and shape. The size of a single particle, having an ordinary spherical or cube shape, can be expressed by the diameter of a sphere or the length of a cube side. However, the majority of particles in a powder are non-spherical and irregular. As particle shape becomes more irregular, expressing the size of an individual particle is even more difficult. To fulfill the needs of production and research, scientists use a number of different expressions for particle size, based on size measurement.

The physical meaning of different particle size expressions is listed in Table 23-1.^{2,3} Herein Feret's and Martin's diameter are intended to express the geometric size of powder particles. Feret's

diameter refers to a diameter measured between imaginary parallel lines tangential to a randomly oriented particle and perpendicular to the scale of measurement; Martin's diameter refers to a diameter measured from the length of the particle at the point that divides a particle into two equal projected areas.⁴ These are statistical diameters that depend on both the orientation and measuring direction of particles. For irregularly shaped particles, a sphere with the same projected surface area or volume could represent as the equivalent sphere of the irregular particles, and the diameter of the irregular particles can be expressed by the diameter of the equivalent sphere, known as the surface area equivalent diameter, or volume equivalent diameter. When a particle is circumscribed by a circle, the diameter of the particle projection is called circumscribed equivalent diameter. There are other types of equivalent diameter, depending on the method used to measure the particle size, which is discussed in this chapter in detail.

MEAN DIAMETER AND PARTICLE SIZE DISTRIBUTION

Mean diameter represents the average size of a collection of particles; therefore, it is more practical and representative than a single particle diameter. Mean diameter can be expressed in different ways, for instance, the mean of a particle population, including arithmetic mean, geometric mean, weight mean, and volume mean, the median, and the mode (Fig. 23-1). Mean refers to the average of all the particles sizes; Median is the diameter for which 50% of the total are smaller and 50% are larger, the diameter corresponds to a cumulative fraction of 50%; and Mode represents the most frequent particle size. Different representations of mean diameter may have different numerical values, which must be specified correctly.

Arithmetic means are obtained by dividing the summarized parameters of all the particles in a powder by the total number of particles. These parameters include diameter, surface area, mass, and volume. In the following equations, D_d , D_s , D_m , and D_v refer to mean diameter based on diameter, surface area, mass, and volume, respectively. $\sum dd$ is the sum of the diameter of all the particles, likewise for $\sum ds$, $\sum dm$, and $\sum dv$, and $\sum dn$ is the total particle number in the powder.

$$D_d = \frac{\sum dd}{\sum dn} \quad (1)$$

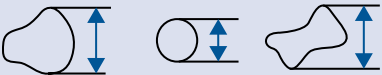
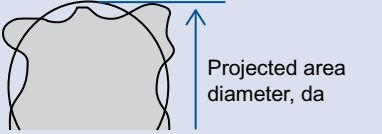
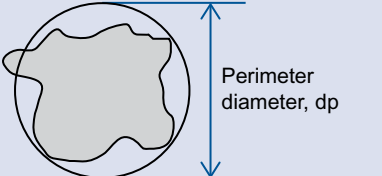
$$D_s = \frac{\sqrt{\sum ds}}{\sqrt{\sum dn}} \quad (2)$$

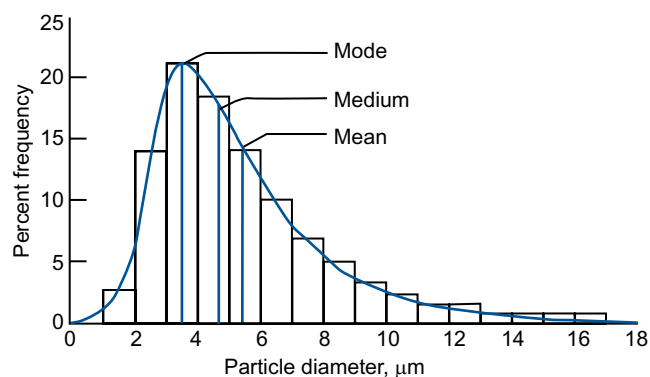
$$D_m = \frac{\sum dm}{\sum dn} \quad (3)$$

$$D_v = \sqrt[3]{\frac{\sum dv}{\sum dn}} \quad (4)$$

The majority of powders consist of particles with a range of different diameters. Particle size distribution refers to the percentage of the particles within a certain size range. When referring to the properties of powders, particle size is not the only important parameter to be aware of; size distribution must also be considered as an essential factor. Particle size distribution is important for understanding powder uniformity, and it has a

Table 23-1. Representation of Particle Size^{2,3}

Expression of Diameter	Definition	Equation or Graphical Representation	Corresponding Determination Method
Feret's diameter, d_F	Distance between pairs of parallel tangents to the projected outline of the particle in some fixed direction		Microscopic method
Martin's diameter, d_M	Length of a chord equally dividing the projected outline of the particle in some fixed direction		Microscopic method
Projected area diameter, d_a	Diameter of a circle having the same area (A) as the projected area of the particle resting in a stable position		Light extinction method
Perimeter diameter, d_p	Diameter of a circle having the same perimeter as the projected outline of the particle		Light extinction method
Equivalent volume diameter, d_v	Diameter of a sphere having the same volume (V) as the particle	$V = \frac{\pi}{6} d_v^3$	Coulter counter method
Equivalent surface diameter, d_s	Diameter of a sphere having the same surface (S) area as the particle	$S = \pi d_s^2$	Surface method (Absorption, permeability)
Sieve diameter, d_A	Width of the minimum square aperture through which the particle can pass	$\frac{1}{2}(a_1 + a_2)$ or $\sqrt{a_1 a_2}$ a_1, a_2 : Openings of sieves	Sieving method
Stokes diameter, d_{st}	Diameter of a sphere having the same gravitational settling velocity as the particle	$d_{st}^2 = \frac{d_v^3}{d_d}$	Sedimentation method
Equivalent light-scattering diameter, d_{ls}	Diameter of a sphere giving the same intensity if the light scattering as the particle	None	Light scattering method
Aerodynamic diameter, d_{Aero}	Diameter of a sphere having the same terminal settling velocity in the gas as the particle	None	Impactor method (Anderson impactor, NGI)

**Figure 23-1.** Mean, medium, and mode of a particle diameter. (From Silverman L, Billings CE, First MW. *Particle Size Analysis in Industrial Hygiene*. New York: Academic Press, 1971.)

great impact on the other properties of powders, including the drug dissolution profile. The cumulative percent represents the integral of the distribution function up to a certain value, or, in other words, the cumulative percent is the percentage of particles in the distribution that are less than that certain value.⁵

Particle size distribution can be expressed as the frequency distribution and the cumulative frequency distribution. To express size distribution in a more intuitive way, a histogram is often built based on frequency of particles within different size ranges. For frequency distribution, a histogram is constructed with the size, number, or weight percentage (i.e., frequency) of particles plotted on the y-axis and diameter range on the x-axis (Fig. 23-2). Such histograms represent the particle size distribution, and the percentage of particles having a given diameter can be determined from the histogram. For instance, the symmetrical distribution presented in Fig. 23-2a is a normally distributed histogram, where the particle size is distributed uniformly around a mean value. However, most powder particles will not exhibit a normal distribution. Their curves are usually skewed to one side of the histogram, depending on which size range occupies a larger portion of the powder solids, the smaller particles or the larger particles. The frequency distribution, shown in Fig. 23-2b, is referred to as positively skewed with a prolonged tail to the larger size particle range; the converse of this situation is referred to as being negatively skewed (Fig. 23-2c). Sometimes, a positively or a negatively skewed distribution can be transformed into a log normal distribution by replotting the x-axis with the logarithmic value of the diameter.

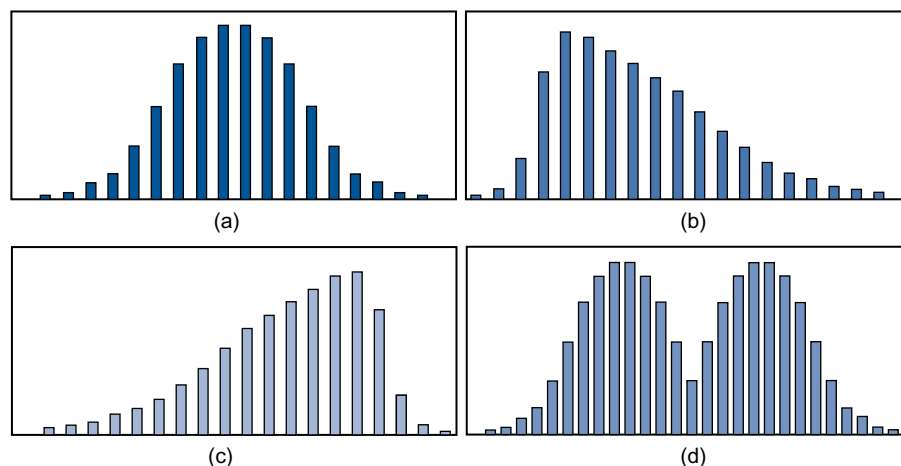


Figure 23-2. Histograms of particle size distribution, (a) a normal size distribution; (b) a positively skewed size distribution; (c) a negatively skewed size distribution; (d) a bimodal size distribution. (From Williams AC *et al* Research in primary dental care. *Br Dental J* 2004;197: 67-73).

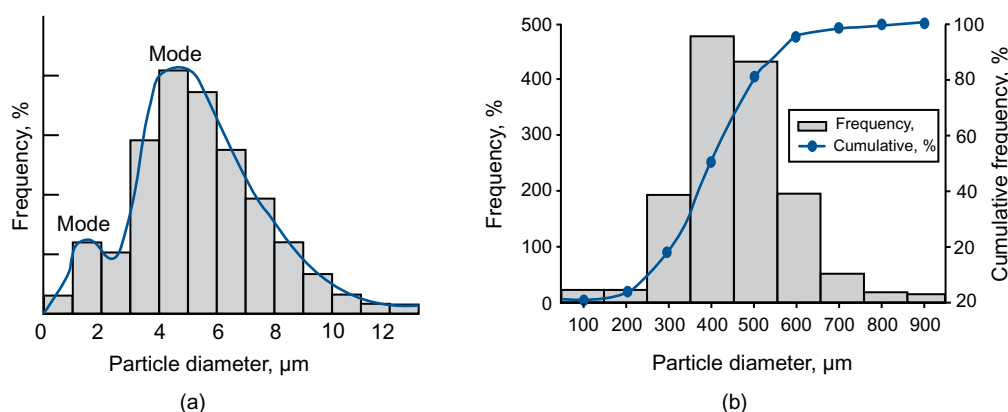


Figure 23-3. Histograms of (a) a bimodal size distribution and (b) cumulative size distribution. (From Silverman L, Billings CE, First MW. *Particle Size Analysis in Industrial Hygiene*. New York: Academic Press, 1971.)

In some cases, more than one mode within a size distribution curve exists (Fig. 23-3a), for example, consisting of some unmilled coarser particles and some fractured fine particles. Herein, an alternative to the frequency distribution, the cumulative frequency distribution, is used to make comparisons between powders. For the cumulative frequency distribution, the histogram is constructed by plotting the percentage of particles that are smaller or larger than a specified particle size limit on the y-axis and the diameter of the particles on the x-axis (Fig. 23-3b). By using the cumulative frequency distribution, the geometric mean can be acquired from the plot.

Powders with the same mean particle size may vary greatly in size distribution. Generally, a narrow size distribution yields more uniform particles.

Determination of Particle Size

As discussed, it is difficult to measure the size of irregular particles accurately. To solve the problem, a factor (F) is introduced to adjust for the differences between the equivalent diameter (D) and the actual size (S), as shown by the following equation:

$$D = FS \quad (5)$$

F is a shape factor; when visual methods are used to determine the particle size, F can be a combination of both shape and extinction coefficient. However, in most cases, F is unknown, so different particle size analysis methods must be conducted to demonstrate the particle size of a powder.⁶

Commonly used methods for determination of particle size include microscopy, separation, sedimentation, electrical

stream sensing zone (Coulter counter), laser light scattering, and specific surface area (adsorption, permeability).

Microscopic Method

Microscopy is one of the most precise particle size determination methods, since individual particles are measured by direct viewing of the particles, instead of statistically averaging the diameter of a group of particles. Microscopy is often used to verify another particle sizing method, during development and validation of a particle sizing method. Microscopy measures the projection of the particles, rather than the particles themselves. For spherical particles, the particle size descriptor is the diameter; for irregularly shaped particles, the diameter of a sphere, which has the same projected area, is used to describe the size of the particles. Other parameters describing irregular particles, such as the perimeter diameter, Feret's diameter and Martin's diameter, can be used to describe the particle size of irregularly shaped particles. A sample size of at least 300–600 particles should be chosen, and, if the size difference between particle-to-particle is comparatively large, the sample size should be increased to get a more significant determination.

The fundamental principle of microscopy is based on the analysis of two-dimensional images of projected particles on a projection screen. The projected particles are compared with circles that are generated by the graticule of the microscope.² The diameter of the circle, which corresponds most similarly to the projected particle, is used to describe the size of the particle. The projection screen is divided into small sections to achieve simultaneous analysis of multiple particles. Light

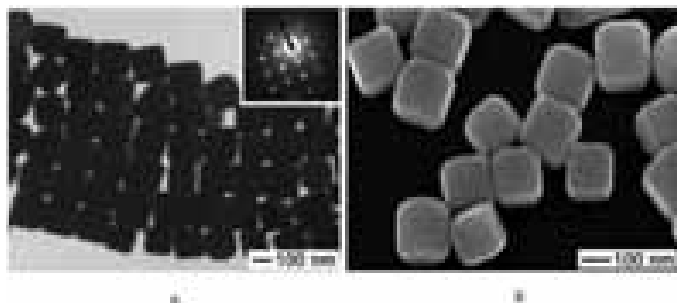


Figure 23-4. TEM (A) and SEM (B) of silver nanotubes. (From Sun Y, Xia Y. Shape-controlled synthesis of gold and silver nanoparticles. *Science* 2002; 298: 2176–2179.)

microscopy can be used to characterize particle sizes between about 1 to 150 microns in diameter.

In addition, scanning electron microscopy (SEM) and transmission electron microscopy (TEM) are used for particle size measurement as well. Both SEM and TEM can analyze particles having a much smaller size range than light microscopy, typically less than 1 micron in diameter (Fig. 23-4).⁷ Additionally, SEM can acquire a three-dimensional image of particles, providing information about particle shape and thickness. When preparing powder samples for microscopic analysis, light microscopy requires a monodispersed layer on the slide to avoid agglomeration; for SEM, samples are placed on aluminum stubs and coated with a thin film of gold (i.e., nanometer thickness); and for TEM, samples are often embedded in resin, which is then cut by a microtome, set on a grid and coated with metal.³

Since hundreds of particles must be viewed in the microscopic method, it is very time-consuming. To overcome the difficulties brought about by manual analysis of particles, a variety of semiautomatic technologies have been incorporated into the microscopic methods to reduce the workload. One of the most popular techniques, the particle comparator, was introduced. Here, the instrument is designed to accept a photographic image of the particle projection, produced either by a light or an electron microscope. The photograph is projected onto a translucent screen, and the particles are compared with a superimposed spot of light projected from the other side. A computer, connected with the variable aperture iris, automatically ensures that the particle is counted and placed into the appropriate particle size category.⁸ These semiautomatic technologies have many advantages, including easier and faster analysis, time savings, and no need for professional operators.

Separation Methods

Separation methods rely on measuring the percentage by number or mass of the extracted size group corresponding to the feed concentration from the powder sample. Sieving (also referred to as screening) is one of the most conventional and simplest separation methods, having the advantage of low-cost and ease of operation. A predetermined weight of dry powder is passed through wire mesh screens, having openings of different sizes, and the weight of powder that is retained on each sieve is measured. The nest of sieves is arranged from coarser mesh opening to smaller mesh opening, top to bottom. As a result of a standardized period of agitation (i.e., by vibration, rotation/tapping, or ultrasound), particles pass through the nest of screens based on two-dimensional size in which the smallest lateral dimension of each particle determines its ability to pass through a given wire mesh opening of the sieve (Fig. 23-5). Vibration provides both horizontal motion (i.e., to loosen the packing of the powder) and vertical motion (i.e., to mix the particles). Vibration time is considered complete, when the weight contained on any of the sieves does not change by more than 5%.



Figure 23-5. Sieve shaker. (Courtesy of Fritsch Sieve Shaker Analysette 3 Spartan.)

US standard sieves are designed by a sieve mesh size, which corresponds to the pore size of the sieves in inches for larger sieve openings and by the number of pores per linear inch for sieves with small openings. Table 23-2 lists the particle size conversion for US standard sieves.

If the powders are suspended in a liquid suspension or the particles of the powder tend to agglomerate, dry sieving is not the best choice. In this case, a process called wet sieving can be used. Water or an organic solvent (i.e., alcohol or acetone) is chosen, depending on the wettability of the materials. The addition of water or organic solvent is designated to reduce surface tension, remove contamination, and dissolve unwanted binders, which can oppose the flow of the particles through the sieves. In the wet sieving process, a spray of the liquid is applied to both the sieving surface and the powder substance or the suspension of powder particles is poured directly onto the sieve. This wet sieving method is particularly designed for micron particles, usually 6–150 μm .²

As shown in Table 23-1, the sieve diameter is defined as the width of the minimum square aperture through which particles can pass. Sources of error in the sieving method include the shape of the particles, weight of sample loaded onto the sieve, vibration intensity, and the phenomenon of static electricity caused by friction.

Impaction is another separation method commonly used to measure the aerodynamic diameter of aerosols. Aerosols are characterized by aerodynamic diameter, rather than geometric diameter. As shown in Table 23-1, aerodynamic diameter is an expression of a particle's aerodynamic behavior and diameter, as if it was a regularly shaped sphere with uniform density, which has the same terminal velocity. A cascade impactor is a typical type of impactor, based on the particle's inertial phenomenon for aerodynamic size distribution measurement of aerosols. Figure 23-6 shows the fundamental principle of the

Table 23-2. Particle Size Conversion of Stand Sieves and Mesh Sizes

Sieve Designation		Nominal Sieve Opening		
Standard	Mesh	Inches	mm	Microns
25.4mm	1 in.	1	25.4	25400
22.6mm	7/8 in.	0.875	22.6	22600
19.0mm	3/4 in.	0.75	19	19000
16.0mm	5/8 in.	0.625	16	16000
13.5mm	0.530 in.	0.53	13.5	13500
12.7mm	1/2 in.	0.5	12.7	12700
11.2mm	7/16 in.	0.438	11.2	11200
9.51mm	3/8 in.	0.375	9.51	9510
8.00mm	5/16 in.	0.312	8	8000
6.73mm	0.265 in.	0.265	6.73	6730
6.35mm	1/4 in.	0.25	6.35	6350
5.66mm	No. 3 1/2	0.223	5.66	5660
4.76mm	No. 4	0.187	4.76	4760
4.00mm	No. 5	0.157	4	4000
3.36mm	No. 6	0.132	3.36	3360
2.83mm	No. 7	0.111	2.83	2830
2.38mm	No. 8	0.0937	2.38	2880
2.00mm	No. 10	0.0787	2	2000
1.68mm	No. 12	0.0661	1.68	1680
1.41mm	No. 14	0.0555	1.41	1410
1.19mm	No. 16	0.0496	1.19	1190
1.00mm	No. 18	0.0394	1	1000
841µm	No. 20	0.0331	0.841	841
707µm	No. 25	0.0278	0.707	707
595µm	No. 30	0.0234	0.595	595
500µm	No. 35	0.0197	0.5	500
420µm	No. 40	0.0165	0.42	420
345µm	No. 45	0.0139	0.354	354
297µm	No. 50	0.0117	0.297	297
250µm	No. 60	0.0098	0.25	250
210µm	No. 70	0.0083	0.21	210
177µm	No. 80	0.007	0.177	177
149µm	No. 100	0.0059	0.149	149
125µm	No. 120	0.0049	0.125	125
105µm	No. 140	0.0041	0.105	105
88µm	No. 170	0.0035	0.088	88
74µm	No. 200	0.0029	0.074	74
63µm	No. 230	0.0025	0.063	63
53µm	No. 270	0.0021	0.053	53
44µm	No. 325	0.0017	0.044	44
37µm	No. 400	0.0015	0.037	37

cascade impactor. Briefly, for the cascade impactor, particles flowing in the air will pass through a series of successively smaller orifices consisting of rectangle slits or round holes and a collection plane surface close to the exit of the orifices, which is perpendicular to the direction of the flow. At each stage, the particles are supposed to make a right angle change of direction to follow the air stream. Larger particles, which are unable to make the right angle change in direction will impact on the collection surface; smaller particles will keep flowing in the air stream, until they interact with the correct surface. Usually, a filter is designated after the final stage to capture the smaller particles that successfully pass through the impactor. A suitable mathematical function is fitted with the amount of particles deposited on each collection surface and the cutoff size of

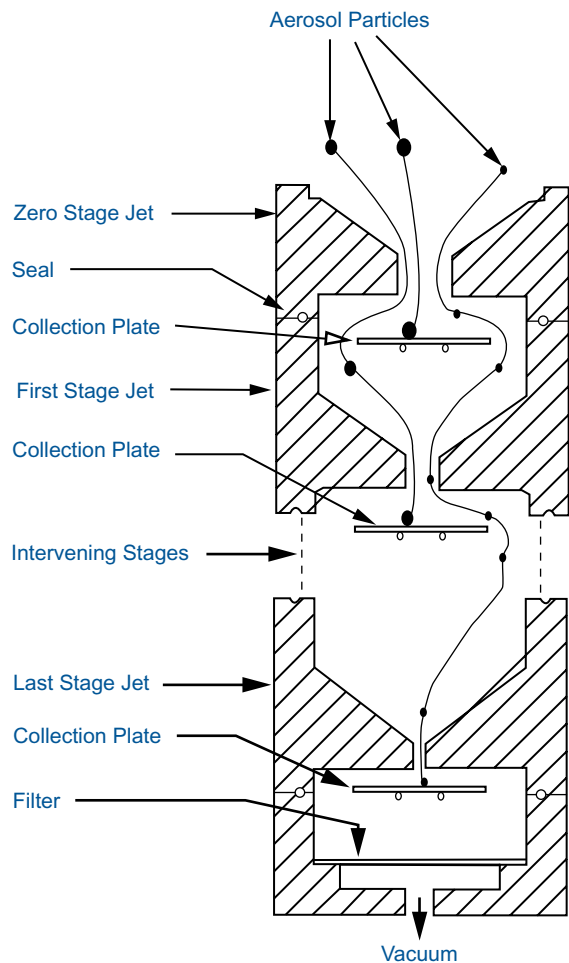


Figure 23-6. Schematic representation of the principle of cascade impactor. (From United States Pharmacopeial Convention. United States Pharmacopeia, 34th ed., and National Formulary, 29th ed. Baltimore, MD: United Book Press, 2001: p 319.)

each stage to describe the aerodynamic size distribution. A new cascade impactor, called the Next Generation Impactor (NGI), has been designed specifically for measuring the aerodynamic diameter of aerosols and is shown in Figure 23-7. The NGI has seven stages and a micro-orifice collector (MOC), which can capture the extremely small particles regularly collected on the filter of other impactors. The amount of particles deposited on the MOC can be analyzed in the same manner as other stages. The inlet flow rate of the NGI can be varied from 30 L/min up to 100 L/min. The orifice collection surfaces that are held in a tray can be readily removed from the NGI, so multiple trays are usually employed to facilitate fast sample turnaround times. Additionally, a maximum of approximately 40 mL of solvent can be added to the collection pans to accomplish appropriate drug recovery. All of the friendly features of the NGI provide a satisfactory impactor to the inhalation testing community.^{9,10}

Sedimentation Method

Sedimentation refers to the settling of a single particle in a liquid medium under the influence of gravitational and centrifugal forces. Typically, the particle size distribution is determined by either placing a thin layer of powder on the surface of the liquid or the powder if homogeneously dispersed in the liquid. If the particles are spherical, the particle concentration is dilute, and their sedimentation is unaffected by interaction between particles, then particle settling in a liquid medium can be described by Stokes' law, when given the settling velocity of particles, the particle size can be calculated according to Stokes' equation as following:

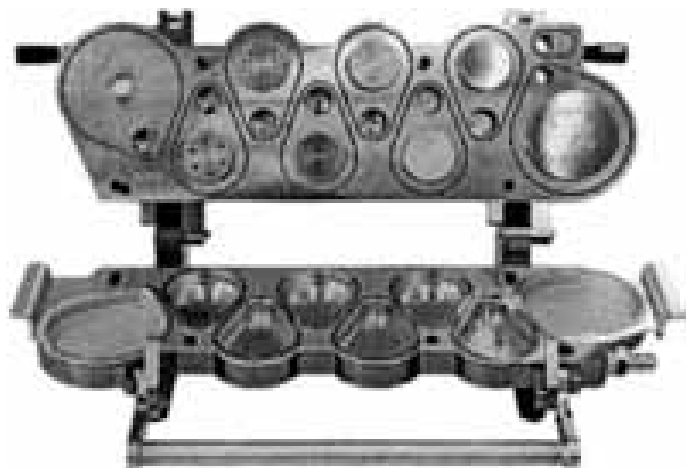


Figure 23-7. Next Generation Impactor (NGI). (From Marple VA *et al* Next generation pharmaceutical impactor (a new impactor for pharmaceutical inhaler testing). Part I: Design. *J Aerosol Med* 2003; 16: 283–299.)

$$V = \frac{d^2 (\rho_1 - \rho_2) g}{18\eta} \quad (6)$$

$$V = \frac{h}{t} \quad (7)$$

$$d = \sqrt{\frac{18\eta h}{(\rho_1 - \rho_2) g t}} \quad (8)$$

where V is the particle settling velocity (cm/s); r is particle radius (cm); ρ_1 and ρ_2 are the densities of the particles and dispersion medium, respectively; η is the viscosity of the dispersion medium (P: g/cm·s, 1 P=0.1 Pa·s); g is the gravity acceleration constant (cm / s²); t is the sedimentation time (s); and h is the sedimentation height (cm).

One of the most commonly used instruments for the sedimentation method is Andreasen pipette (Fig. 23-8). The apparatus is composed of a 200 mm graduated cylinder with a pipette located in the center. First, samples are mixed with the liquid to form a suspension. Next, the suspension is placed into the Andreasen pipette, sedimentation medium is added, and the mixture is stirred by shaking the pipette. After that, the pipette is left undisturbed for a certain period of time. Then, 10 mL of the suspension is collected from a position of 200 mm below the surface of the suspension, and the weight of the solid obtained, after drying the sample, is measured to calculate the diameter according to the Stokes' equation.¹¹ One can also calculate a cumulative distribution from samples removed at appropriate time intervals.

Centrifugal sedimentation is used to measure the particle size between about 5–10 μm in diameter. Upon centrifugation of a liquid medium containing the powder particles, the fraction of the total suspended powder that reaches the bottom of the centrifugation tube at different times is measured. The times are converted to their corresponding diameters, using equations based on the Stokes' equation, describing spherical particles moving through viscous media.

Electrical Sensing Zone Method (Coulter counter)

Figure 23-8 shows the fundamental principle of Coulter counter method. The testing tube is surrounded by a highly conducting electrolyte solution (i.e., about 0.1 M ionic strength), with particles dispersed in it. There is a small aperture on that tube, with two electrodes set at each side. Solution flow is initiated by the height difference between the liquid surfaces, which is caused by applying vacuum to the tube. When the particles pass

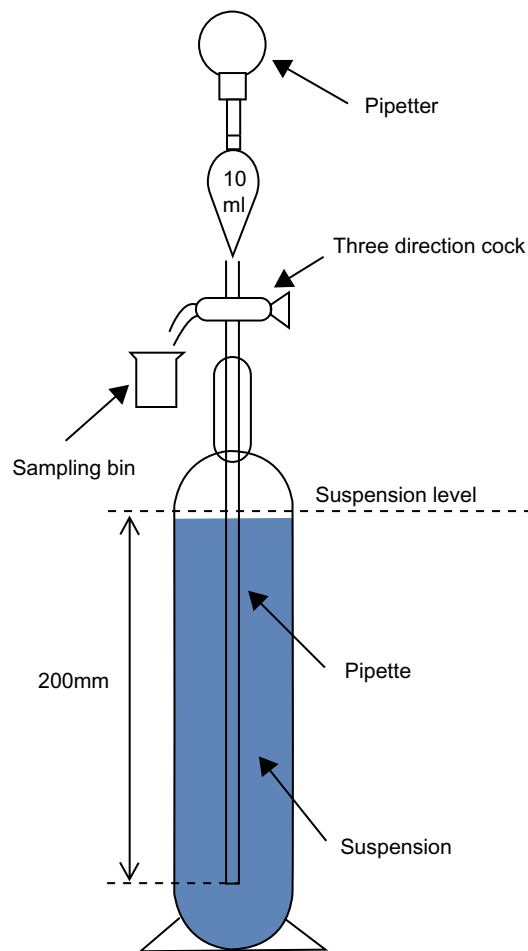


Figure 23-8. Schematic of Andreasen pipette. (From Kawachi S, Kasama S. Quantitative effect of micro-particles in iron ore on the optimum granulation moisture. *ISIJ International* 2009; 49: 637–644.)

through the aperture, the instant change in resistance leads to voltage pulses, the extent of which is proportional to particle volume. These pulses are amplified and counted electronically, the number of the particles and the particle size distribution is displayed on the computer.¹² To obtain the size distribution over a large diameter range, making a change to the aperture size may be necessary.

A typical particle size range from about 0.6 to 120 μm can be measured by the Coulter counter method, depending on how many apertures and tubes are used in the apparatus. There are some problems associated with this method. When multiple apertures and tubes are used, large particles may block the pathway or more than one particle may pass through the aperture at one time; particles smaller than the detection limit passing through the aperture may not be detected; and non-uniformity of the electrical field may lead to the abnormal voltage pulses.

Laser Light-Scattering Methods (Laser diffraction)

Laser light scattering methods are based on the fact that particles scatter light. The particle size of both large and small particles can be measured by this method.

Figure 23-10 shows low-angle laser light-scattering, also referred to as Fraunhofer diffraction. Light scattered by larger particles, those with a much larger size than the wavelength of light, has a forward direction and is slightly changing in angle. When the light emitted by a laser hits the samples, diffraction occurs and light-intensity is collected by a detector, after focusing of the scattered light by the lens. The light-intensity is

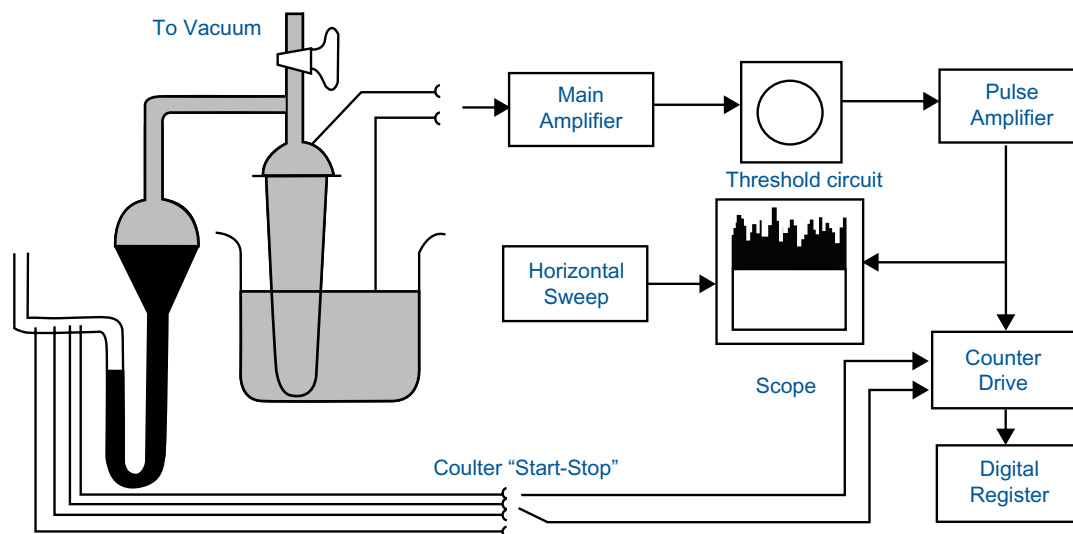


Figure 23-9. Diagram of Coulter counter apparatus. (From Masuda H. *Powder Technology Handbook*, 3rd ed. Boca Raton, FL: Taylor & Francis, 2006.)

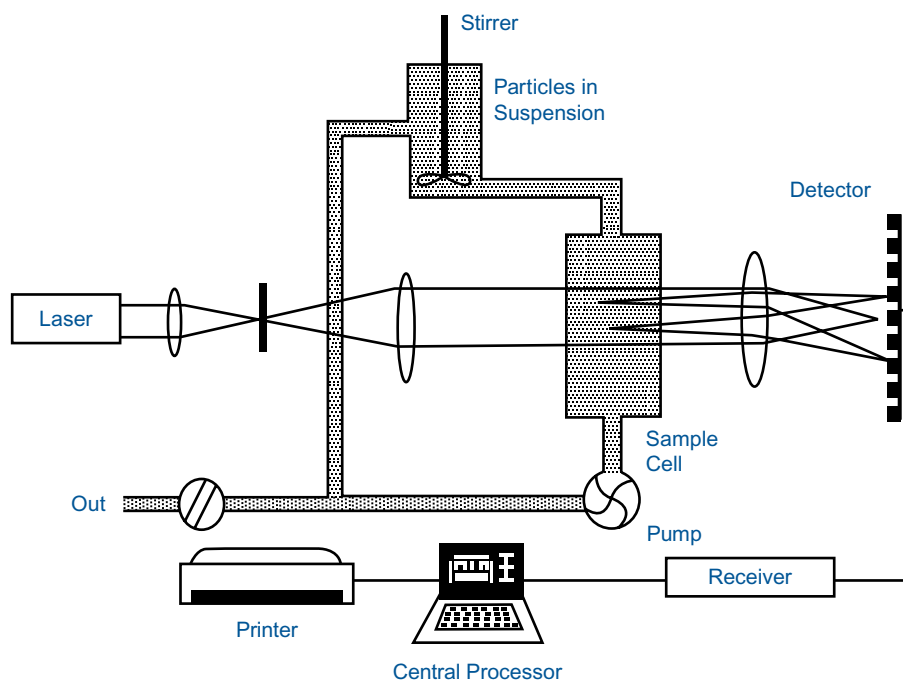


Figure 23-10. Schematic of a low-angle laser light-scattering instrument. (From Masuda H. *Powder Technology Handbook*, 3rd ed. Boca Raton, FL: Taylor & Francis, 2006.)

proportional to the particle diameter. In some cases, a micro-processor is incorporated with the photo detector to analyze the particle size distribution data directly.¹³

Photo correlation spectroscopy (PCS) is used to measure small particles in liquid. Brownian motion is the fundamental theory of PCS. The collisions between small or large particles with relatively smaller molecules of the suspension medium (liquid or gas) lead to randomized movement of the particles. This phenomenon is known as Brownian motion and is independent of particle shape or the liquid medium. The following equation describes the calculation of the particle diameter using PCS:

$$D = 1.38 \times 10^{-12} T 3 \pi d \eta s - 1, \quad (9)$$

where T is the absolute temperature (Kelvin), η is the viscosity of the liquid, and d is the hydrodynamic diameter. Figure 23-11

shows a PCS instrument: a laser beam passes through the sample suspension cell, and the light scattered perpendicular to the beam is collected and auto-correlated by a computer. This generates an average particle size with a polydispersity factor, which refers to degree of variability around the average size value.¹⁴

Surface Area Methods

Specific surface area refers to the surface area of particles expressed per unit weight or volume. Surface area increases as the particle size decreases. Therefore, the particle size can be calculated from the specific surface area of a powder. Determination of the surface area is by gas adsorption (0.03–1 μm) and permeability (1–100 μm).

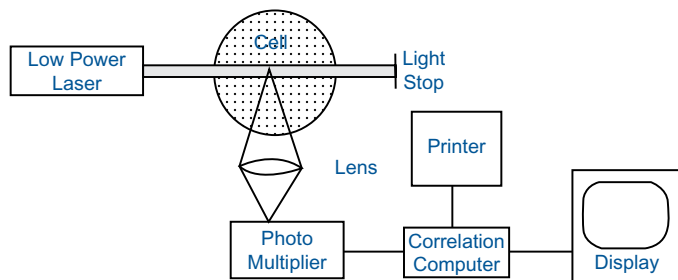


Figure 23-11. Schematic of a Photo correlation spectroscopy (PSC) apparatus. (From Masuda H. *Powder Technology Handbook*, 3rd ed. Boca Raton, FL: Taylor & Francis, 2006.)

When a solid is exposed to a gas, the gas molecules in contact with the surface may not be elastically bounced off, but, instead, may remain adsorbed onto the surface of the particle for a certain time at low temperature. This is called gas adsorption. When the powder sample is returned to room temperature, desorption of the gas molecules occurs. The sorption isotherms obtained in this process are demonstrated by the equation developed by Brunauer, Emmett, and Teller, typically referred as the BET method. Mostly nitrogen and krypton are the preferable gases used in BET. The gas to be adsorbed (adsorbate) is mixed with another inert, non-condensable, carrier gas, usually helium. A range of 5–30% adsorbate in carrier gas is often used. The surface area is calculated by the following equation:

$$\frac{P}{V(P_0 - P)} = \frac{1}{V_m C} + \frac{(C - 1)P}{V_m C P_0} \quad (10)$$

where V is volume of gas adsorbed at pressure P ; P is partial pressure of adsorbate; V_m is volume of gas absorbed in monolayer; P_0 is saturation pressure of adsorbate at experimental temperature; and C is a constant relating to the heats and condensation of the adsorbate.¹⁵

Permeability occurs when a gas or a liquid is passed through a powder bed. There are many factors that may affect the resistance of flow, including the surface area of the powder, the surface area of the bed, the pore size, and the viscosity of the fluid. The specific surface area can be calculated by the following equation:

$$\frac{\Delta p}{L} = \frac{150 V_0 \mu (1 - \varepsilon)^2}{\phi_s^2 D_p^2 \varepsilon^3} \quad (11)$$

where Δp is the pressure drop, L is the total height of the bed, V_0 is the superficial or “empty-tower” velocity, μ is the viscosity of the fluid, ε is the porosity of the bed, ϕ_s is the sphericity of the particles in the packed bed, and D_p is the diameter of the related spherical particle.¹⁶ Although this method does not provide size distribution data, it is very beneficial for rapidly estimating the mean particle size in some industrial applications.

Particle Shape

Particle shape can be described as follows: in General Chapter <776> of the *US Pharmacopoeia* (31st Edition), the following particle shape descriptors are provided, including: a) Acicular, which means slender, needle-like particle of similar width and thickness; b) Columnar, which means long, thin particle with a width and thickness that are greater than those of the acicular particle; c) Flake, which means thin, flat particle of similar length and width; d) Plate, which means flat particles of similar length and width but with greater thickness than flakes; e) Lath, which means long, thin, and blade-like particles; and f) Equant,

which means particles of similar length, width and thickness, and include cubical and spherical particles.⁴

The shape of powder particles in pharmaceutical applications is very complex, due to the method of particle formation and the milling method used to reduce the size of the particles. Therefore, the particle shape is often irregular; the surface is very rough, making it difficult to use appropriate descriptor terms to describe the shape of these irregular powder particles. Also, particles in a sample may have aggregated or agglomerated into more complex structures, which must be accounted for.

The most commonly used method to characterize particle shape is the microscopic method, which has been described in detail in this chapter. Along with rapidly developing computer technologies, imaging technology has been well established over the past few decades. It is not difficult to obtain image information and geometric structure of powder particles. The significant challenge for scientists is how to transfer the image information into a more quantitative description of particle shape.

Particle Density

Definition Density, another physical property of powders, is the ratio of mass to volume. To obtain the density of particles or powders, it is necessary to acquire the volume first. For non-porous solids, the estimation of the exact volume is straight forward. However, for powders with rough surfaces, cracks, and pores, the determination of volume is extremely complicated. According to different powder volume representations, the density for particles and powder can be classified as follows:

1. True density refers to the ratio of the mass of the particle to its actual volume, excluding pore volume and the volume of the gap between particles. A commonly used method for the determination of true density is helium pycnometry. Helium can readily penetrate into small pores; therefore, it is considered that the density measured by helium pycnometry approximates the true density. Conversely, liquids, such as water and alcohol, cannot penetrate small pores, so the density measured by these fluids is often slightly smaller than the true density. If a powder is compacted into tablets using high pressure, then the density of the tablet calculated from the weight and volume ratio (also referred as high pressure density), and the resultant density, is very close to the true density.
2. Particle density is defined as the ratio of particle mass to the particle volume, including the pores within the particle, but still excluding the gaps between particles. A commonly used method for determination of particle density is the mercury intrusion method. Since mercury has a high surface tension, it cannot pass through the small cracks or pores of the powder particles (<20 μm) under constant pressure, but can penetrate the gap between powder particles, so the volume measured by this method is sum of the volume of powder particles themselves and the volume of internal pores. Other than mercury, benzene, water, and carbon tetrachloride have been used to determine the particle density.
3. Bulk density is defined as the ratio of powder bed mass to volume of that powder bed, including both the pores and gap volume. Determination of powder bulk density is straight forward. The powder is filled into a cylinder, and, after vibration for a given time period and intensity, the volume of powder bed is measured directly from the cylinder. The bulk density is then calculated from the ratio of powder weight to volume.

A pycnometer is a typical apparatus for the liquid replacement method (Fig. 23-12). In this method, the particle density (ρ_p) is calculated according to the following equation:

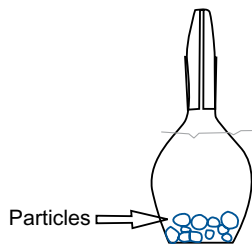


Figure 23-12. Pycnometer.

$$\rho_p = \frac{\rho_l (m_s - m_0)}{(m_1 - m_0) - (m_{s1} - m_s)} \quad (12)$$

where m_0 is empty mass of the pycnometer; m_1 is the pycnometer mass containing liquid; m_s is the pycnometer mass including sample particles; m_{s1} is the pycnometer mass, including both sample particles and liquid; and ρ_l is the density of liquid. A non-volatile liquid is necessary for this method, in which the particles are wettable and insoluble. Additionally, air bubbles must be removed by heating the liquid under reduced pressure, before the start of the measurement process.¹⁷

The constant gas volume method is a conventional gas replacement method. Figure 23-13 shows the fundamental mechanism of this method. When there are no particles in the cylinder, the piston is moving from A to B, causing the pressure to change from P_a to $P_a + P_1$; by addition of particles to the cylinder, the procedure is repeated, and the pressure change is acquired from P_a to $P_a + P_2$, from which the particle volume is calculated as follows:

$$V_s = v_0 \left(\frac{P_a}{\Delta P_1} - \frac{P_a}{\Delta P_2} \right) \quad (13)$$

where v_0 is the volume between position A and B, and V is the volume of the cylinder when piston is at position A.²

There are many other methods available for measurement of particle density, such as laser interferometer,¹⁸ diffraction method,¹⁹ differential mobility analyzer,²⁰ and sedimentation.²¹

Particle Agglomeration and Aggregation

The terms “aggregation” and “agglomeration” are frequently used in powder science to describe assemblies of powder particles in the dry state or in liquid suspension. However, the confusion brought about by interchanging the use of the two terms requires that the terms be defined.

Particles form assemblages as a result of fusion, sintering, interparticle bonding, and growth (caking). They may combine during production (milling, crystallization), drying, or storage. The assemblages are often polycrystalline, with particles joined together in some random manner. Particles subjected to a milling process have very high surface energy, so small amounts of water may be adsorbed. The adsorbed water can induce partial

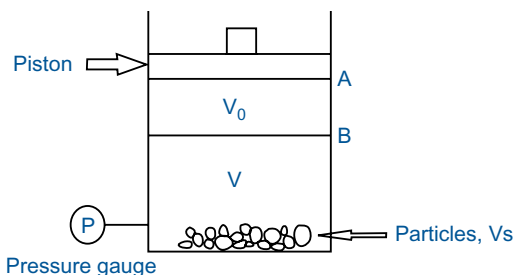


Figure 23-13. Constant gas volume method.

dissolution, leading to recrystallization at the surface, since it tends to lower its surface energy. Conversely, the glass transition temperature of amorphous solids can be reduced, due to the plasticizing effect of water, resulting in recrystallization.

According to the USP, an agglomerate is defined as fused or cemented particles, whereas an aggregate is defined as mass of adhered particles.⁴ In a paper published by the Particle Characterization subgroup of the Pharmaceutical Analytical Sciences Group (PASG) operating in the United Kingdom (UK), agglomeration and aggregation are discussed and differentiated. The paper proposes to discontinue the use of “aggregate,” when that term refers to particle assemblages. The term “agglomerate” should be used exclusively to refer to particle assemblages. Before crystallization, the molecules of the particles may combine with each other, leading to a supramolecular state in which the term “aggregate” can be used to describe the process, including when the molecules combine into prenucleation structures, which have the potential to grow into crystals. The term “aggregate” should be restricted to describe this process; in addition, aggregates may turn into agglomerates at some later point.²²

BULK PROPERTIES OF POWDERS

SPECIFIC SURFACE AREA

The specific surface area is an essential property for a powder, and it represents the void space on the surface of individual particles or agglomerates of particles. Moreover, the dissolution rate of a powder is partially impacted by its specific surface area. The most commonly used methods for determination of surface area for a powder are gas adsorption and permeability. There are a variety of other methods used for surface area determination. For instance, adsorption from solution, flow microcalorimeters, estimation of surface area from size distribution data, turbidity methods, and chemisorption have all been reported.

POROSITY

As mentioned, pores of powder particles include the area of the pore interior and the gap between particles; hence, the porosity of powder particles refers to the ratio of the volume of the pore interior and the space between particles to the total volume of the powder, including the pore interior, the gaps between particles, and the inherent volume of particles. Powder porosity is affected by many factors, such as particle shape, size, and surface properties. When granules and tablets are made from fine powders, they tend to be porous, and the porosity of tablets has a great impact on the disintegration time. For tablets with higher porosity, water can penetrate the tablets, more readily resulting in a faster disintegration time. Another parameter to express the porosity is the pore size distribution. Many techniques have been developed to characterize the pore size distribution of a solid powder; the two most useful are gas adsorption and mercury porosimetry.²³

As discussed above, gas adsorption is widely used for specific surface area determination. It is achieved by measuring the volume of gas adsorbed or desorbed at certain temperatures. Adsorption and desorption isotherms are generated by changing the pressure significantly in the range. BET surface area, along with pore size distribution information can be obtained with this method. However, pore size distributions measured by gas adsorption are limited by a narrower pore diameter range, as compared to mercury porosimetry. Therefore, mercury porosimetry is a preferred method for determination of pore size distribution of a porous solid powder.

Figure 23-14 shows a schematic of mercury porosimetry. The fundamental principle of this method is based on the capillary rise phenomenon, which means to push a non-wetting liquid (i.e., mercury) and cause it to enter into an extremely small volume (i.e., pores of powder particles), which requires extremely high pressure. First, the sample is weighed into the cell,

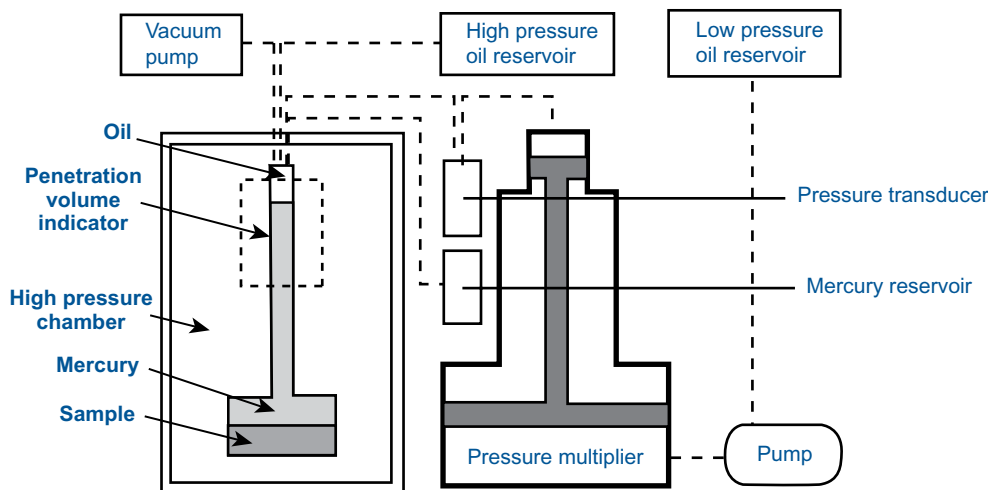


Figure 23-14. Diagram of a mercury porosimetry. (From Allen T. *Particle Size Measurement*, 5th ed. London: Chapman & Hall, 1997.)

degassed, and surrounded by mercury, and then the pressure is gradually increased. The volume of mercury forced into the sample at each pressure is measured. For a circularly shaped pore, the following equation demonstrates the relationship between pressure and pore size:

$$\Delta P = - \left(\frac{2\gamma}{r} \right) \cos \theta \quad (14)$$

where P is the pressure difference across the interface; γ is surface tension of the liquid; θ is the contact angle of the liquid (about 140° for mercury); and r is the pore radius. γ and θ are constant, so the equation can be simplified as follows: if P and r are expressed in psi and μm , respectively:

$$r = \frac{89.5}{P} \quad (15)$$

A curve is plotted of the pore radius versus pressure, according to this equation, and the pore size distribution can be obtained.¹⁶

CRYSTALLIZATION AND AMORPHOUS STATES

Definition and Characteristic

A powder is either in a crystalline or an amorphous state. As shown in Fig. 23-15a, crystalline means the molecules in the solid are arranged in a fixed order. Crystalline solids exhibit a melting point. The melting point is the temperature of a substance at which it changes state from solid to liquid. During the melting process, the molecules in the crystalline material obtain enough energy to break down the attractive forces that keep the molecules in the crystal structure.²⁴ The formation of crystals contributes greatly to powder production technology, which is discussed in detail in this chapter. The internal packing of molecules in a substance is important to its properties. Properties of the powder will change based on the internal packing of molecules, as well as the external shape of a crystal, which refers as the crystal habit. The crystal habit is defined as the outer appearance of a crystal. The terms used to describe the crystal habit include sphere, cube, and needle. Different crystal habits may have different surface properties, which leads to differences in powder properties, such as dissolution rate and powder flowability.^{25,26}

The amorphous state of a material is defined as when the molecules in the substance are packed in a non-fixed, random order (Fig. 23-15b). The properties of an amorphous material differ greatly from a crystalline material. An amorphous solid does not have a melting point, as the crystalline form does.

Characteristic of an amorphous material is the glass transition temperature (T_g), which refers to the temperature at which the state of the solid transforms between rubbery and glassy. When the temperature of the material is greater than the T_g , the substance exists in the rubbery state in which the mobility of molecules are great enough to lead to a rapid crystallization; when the temperature is lower than the T_g , the substance exists in the glassy state in which the mobility of molecules is low and the amorphous form can exist for a longer time, compared with the rubbery state. It is possible to lower the T_g of an amorphous solid by addition of a plasticizer. Water is an excellent plasticizer for most of the materials and will reduce the T_g . The water molecules can penetrate into the amorphous form of a material, due to the large void space left in the structure.²⁷ Hence, in most cases, amorphous solids tend to adsorb large amounts of water, thus, causing a stability challenge for pharmaceutical scientists. The amorphous state of a substance can be produced, when the solidification process is faster than the crystallization process, such as spray drying and ultra-rapid freezing, or by breaking the crystal form, such as by milling.²⁸ The state of a material, whether amorphous, crystalline, or both, has a great impact on the solubility or dissolution rate of a solid powder, which is discussed in the dissolution rate section.^{29,30}

Polymorphism refers to crystals with different packing orders (Fig. 23-15c). Polymorphic forms are obtained by changing the crystallization conditions, for example, by changing to a different solvent or changing the stirring method. There are numerous examples of organic materials, including drugs and excipients that exhibit polymorphism.³¹⁻³³ They may exhibit many different polymorphic forms; however, only one form is most stable, and any other polymorphic forms will, eventually, convert to the more stable form. The most stable polymorphic form has the highest melting point and lowest dissolution rate. All of the other polymorphic forms are referred to as metastable, meaning they may exist for a certain period time, varying from seconds to months, based on the storage conditions. Since metastable polymorphic forms have a relatively high dissolution rate, they

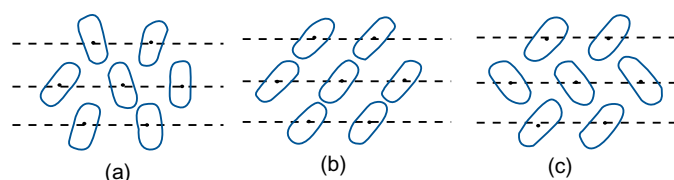


Figure 23-15. Representation of (a) crystallization (b) polymorphism and (c) amorphous.

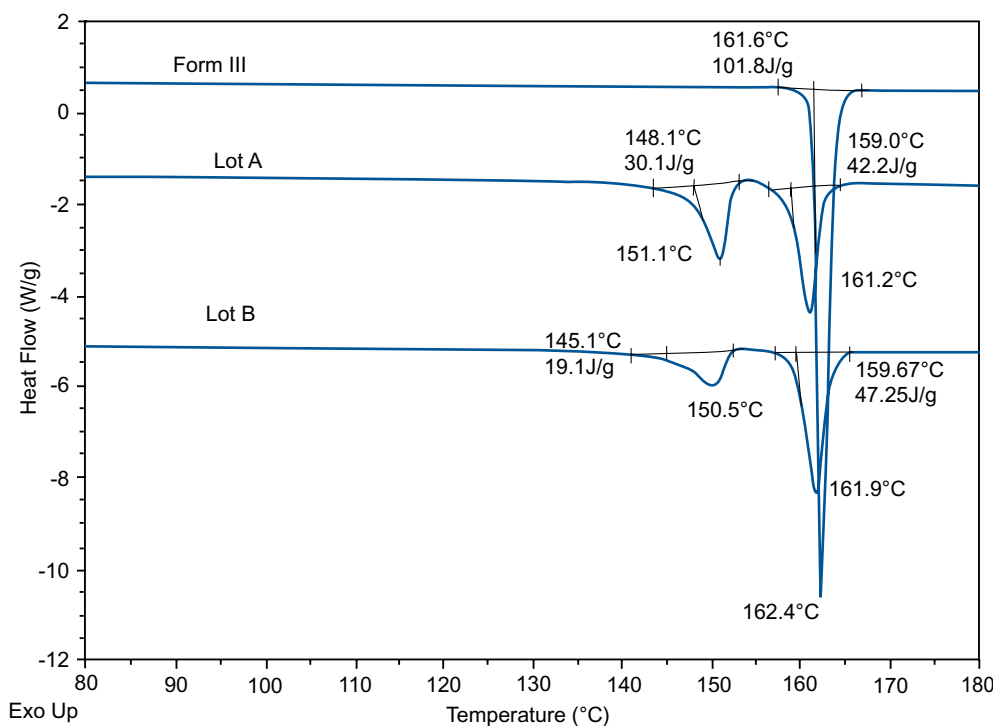


Figure 23-16. DSC pattern for celecoxib bulk powder (Form III) and polymorphic forms of celecoxib (Lot A and Lot B). (From Lu GW, *et al.* Characterization of a novel polymorphic form of celecoxib. *J Pharm Sci* 2006; 95: 305–317.)

might be a candidate for enhancing the bioavailability of poorly water soluble drugs. However, stabilization of the metastable form must be addressed to prevent it from converting to the stable polymorphic form. For instance, Lu *et al.*³⁴ prepared a polymorphic form of celecoxib in the presence of Polysorbate 80 and HPMC and characterized it with *in vitro* and *in vivo* studies. The DSC pattern, shown in Figure 23-16, indicates that the new polymorphic form has a melting onset at 145–148°C, which was approximately 12–15°C less than the melting point of bulk powder of celecoxib. Also, there was a faster dissolution rate (>2 times) of the polymorphic form observed, compared to the more stable form of celecoxib (Fig. 23-17). Correspondingly, the celecoxib suspension containing the polymorphic form had a significantly higher bioavailability (>4 times) than the marketed capsules and the suspension containing bulk powder of celecoxib (Fig. 23-18).

MEASUREMENT

X-ray diffraction (XRD) has been widely used to determine the crystallinity of a solid powder. The basic principle of XRD is the measurement of identical intensity patterns, composed of peaks at different scattering angles, after diffraction of different powder samples. The XRD method is based on Bragg's law:

$$n\lambda = 2d\sin\theta \quad (16)$$

where n is the order of the diffraction pattern; λ is the wavelength of the incident beam; and d is the distance between the planes in the crystal and θ is the angle of beam diffraction. As an example of the XRD pattern (Fig. 23-19), bulk itraconazole (ITZ) exhibits intense crystalline peaks at a variety of angles; conversely, the XRD patterns for wet-milled ITZ and URF-ITZ showed a reduced crystallinity.³⁰

Thermal analysis methods are also widely used to evaluate the crystallinity of a solid powder. Differential scanning calorimetry (DSC) is frequently used. Sample and reference are maintained at the same temperature, and the heat flow required to maintain the sample and reference at the same temperature is measured. A DSC diffractogram is plotted by differentiating the rate of heating versus temperature. For instance, the crystal-

linity of bulk ITZ, wet-milled ITZ, and URF-ITZ were evaluated by DSC, as shown in Figure 23-20. For both wet-milled ITZ and bulk ITZ, there is one endothermic melting peak exhibited at about 163.4°C, indicating its crystallinity. The URF-ITZ powder exhibited a glass transition at approximately 50°C, one exothermic recrystallization peak, and a single endothermic melting peak at 165.7°C, confirmed its amorphous nature. These results agreed with the XRD results discussed previously.³⁰

Powder Flowability

Definition and Significance

Flowability refers to the ease with which a powder will flow under a specified set of conditions. Flowability is a complex

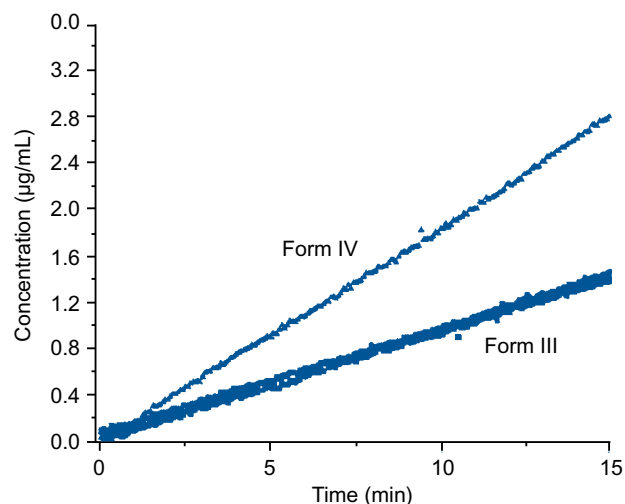


Figure 23-17. Rotating disk dissolution for Form IV (polymorphic form) and Form III (bulk powders) of celecoxib. (From Lu GW, *et al.* Characterization of a novel polymorphic form of celecoxib. *J Pharm Sci* 2006; 95: 305–317.)

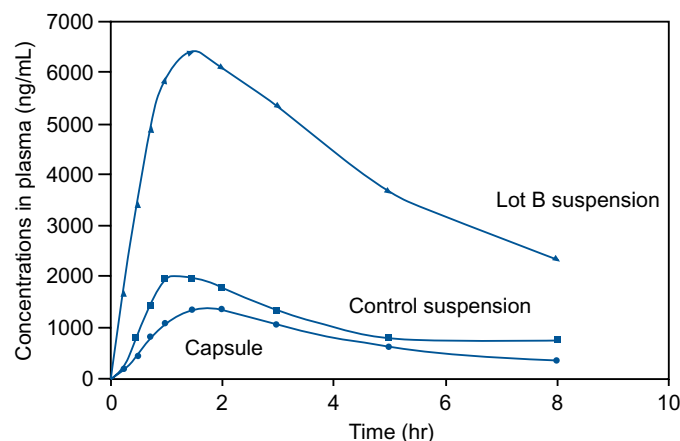


Figure 23-18. Pharmacokinetics of celecoxib capsules, control suspension and polymorphic form suspension in dogs ($n = 6$). (From Lu GW, *et al.* Characterization of a novel polymorphic form of celecoxib. *J Pharm Sci* 2006; 95: 305–317.)

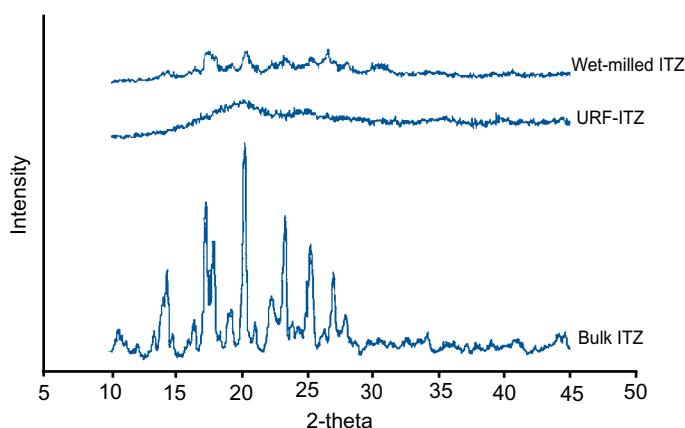


Figure 23-19. X-ray powder diffraction patterns (from the top to bottom): Wet-milled ITZ, URF-ITZ, and bulk ITZ. (From Yang W, *et al.* Comparison of bioavailability of amorphous versus crystalline itraconazole nanoparticles via pulmonary administration in rats. *Eur J Pharm Biopharm* 2010; 75: 33–41.)

phenomenon, referring to both physical properties affecting powder flow and the equipment required for handling, storage, and processing. Powder flowability is an essential property, which has a great impact on many pharmaceutical processes, such as blending, transfer, storage, compression, and handling, as well as the quality of pharmaceutical dosage forms. For example, when filling capsules with powders, powder flowability influences the accuracy of the amount filled into each capsule; when granules are compressed into tablets, the powder flowability and stratification can cause weight differences between tablets within the same batch; high speed tableting machines and direct compression methods require better powder flowability; reproducible sampling of powders and granulations is also associated with flow properties.³⁵

In the case of pharmaceutical dry powder inhalers (DPIs), a size range of 1–5 μm must be obtained for the APIs to reach the deep lung and compose an extremely cohesive powder. The adjustment and control of powder flowability is of key importance during formulation and process development of DPIs. For instance, one of the challenges is the mixing of such cohesive powders; also, the dry powder must be filled into the inhaler with high speed and precise control. In addition, the powder flowability may have a great impact on the perfor-

mance of DPIs. Emitting of the formulation from the inhaler and subsequent aerosolization of the powder into respirable particles upon inhalation is influenced by powder flowability. There are two ways to improve the flowability of DPIs: first is to form spherical API aggregates by adding micronized excipients; another more frequently used way is to mix the APIs with larger carrier particles, such as lactose, in which case the small APIs particles adhere to the surface of the carrier particles.

DETERMINATION METHODS

There are numerous techniques that have been developed to characterize powder flowability. Only the most frequently reported are subsequently discussed.³⁵

The angle of repose (α) has been widely used to describe powder flowability. For an accumulation of static powders, the angle between the free surface of the powder body and the horizontal plane is called the angle of repose. For instance, when adding additional powder to a powder mound, the angle between the side and the bottom surface reach a constant value, due to a balance between gravity and powder friction ($\alpha = \tan^{-1} \mu$). Figure 23-21 shows some methods. A smaller angle of repose indicates smaller frictional forces and greater flowability. For a free-flowing powder, the angle of repose is equal to or less than 30° ; powders with an angle of repose equal to or less than 40° have satisfactory flowability; and, when the angle of repose is greater than 40° , the powder does not flow freely. Different magnitudes for flowability data may be obtained, depending on which method is chosen. In addition, the reproducibility is typically poor for this method, so angle of repose is not always considered a perfect representation for powder flowability.³⁶

Flowability can be estimated from the bulk density of the powder as well. Figure 23-22 is a typical tapped density measuring device. After the powder sample is tapped by the rotating cam, the bulk density of the powder increases from $\rho_{\min}$ (poured bulk density) to $\rho_{\max}$ (tapped bulk density). The percentage compressibility (Carr's Index) can be calculated by the following equation:

$$\% \text{compressibility} = \frac{\text{Tapped density} - \text{Poured density}}{\text{Tapped density}} \times 100 \quad (17)$$

The relationship between powder flowability and compressibility is shown in Table 23-3.³⁷

Hopper flow rate is a direct method to determine powder flowability, by measuring the rate of powder discharging from the hopper (Fig. 23-23). At the center of the hopper bottom, there is a small hole that can be covered by a shutter. By releasing and opening the shutter, the powder is discharged from the hopper. The flow rate is calculated by dividing the mass of the powder by the total time taken to remove all the powder in the container. Higher flow rate is a representation of better flowability.

FACTORS AFFECTING POWDER FLOWABILITY

There are many factors that affect the flowability of powder solids, such as particle size, shape, porosity, density, and state of particles.³⁸⁻⁴⁰

Table 23-3. Relationship Between Powder Flowability and Compressibility³

Flowability	Extremely Poor	Very Poor	Poor	Fair	Good	Excellent
Compressibility (%)	>38	35–38	23–35	18–23	12–18	5–12

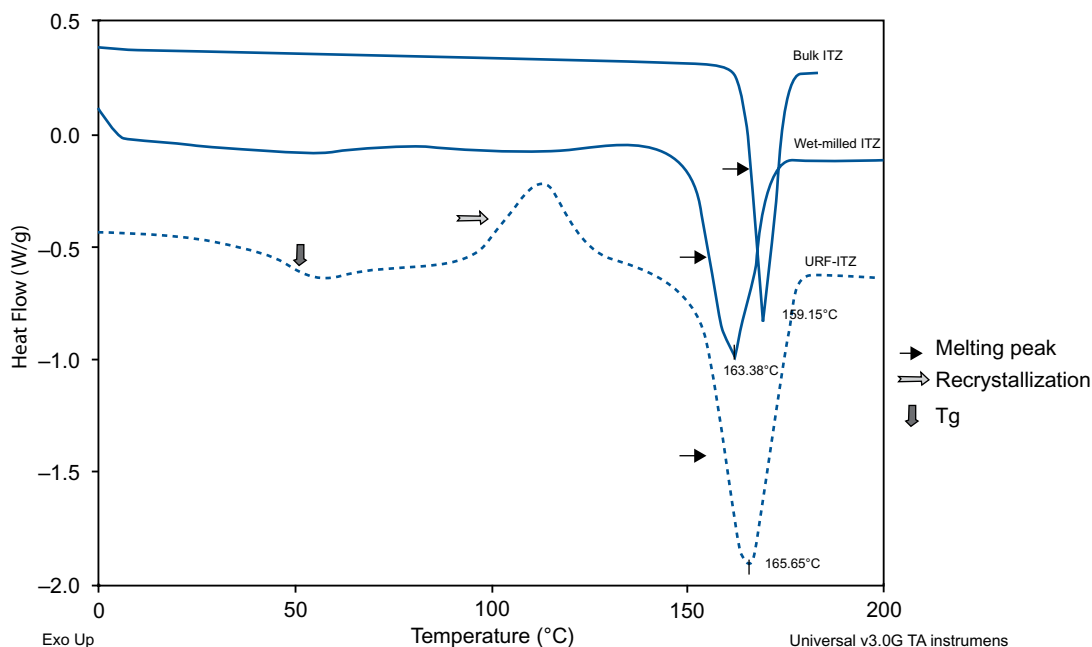


Figure 23-20. DSC profiles of bulk ITZ, wet-milled ITZ, and URF-ITZ. (From Yang W, *et al.* Comparison of bioavailability of amorphous versus crystalline itraconazole nanoparticles via pulmonary administration in rats. *Eur J Pharm Biopharm* 2010; 75: 33–41.)

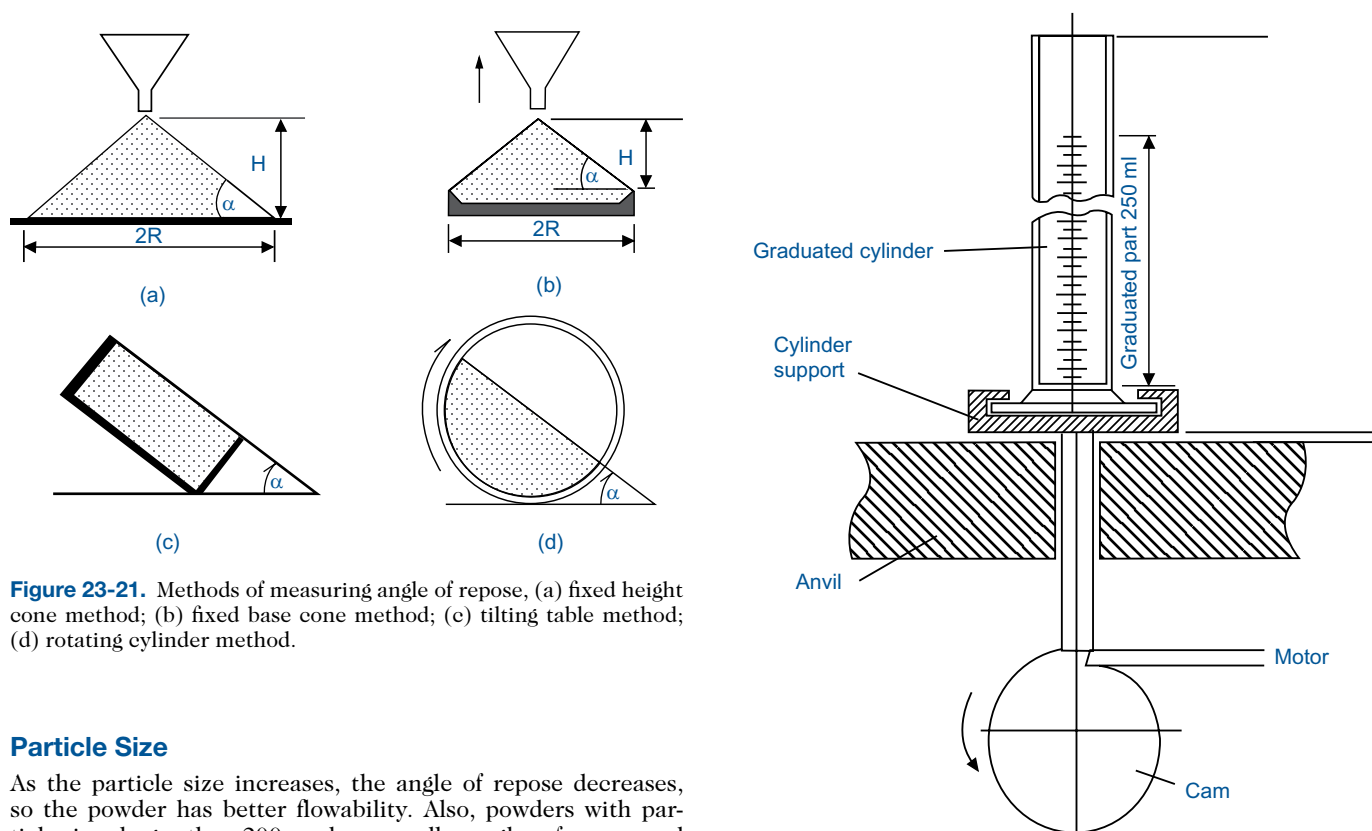


Figure 23-21. Methods of measuring angle of repose, (a) fixed height cone method; (b) fixed base cone method; (c) tilting table method; (d) rotating cylinder method.

Particle Size

As the particle size increases, the angle of repose decreases, so the powder has better flowability. Also, powders with particle sizes larger than 200 μm have smaller angles of repose and good flowability; for particles sized between 100 and 200 μm , cohesion and friction forces increase, and the angle of repose increases as well, leading to the reduction of flowability; for particles less than 200 μm , agglomeration occurs, and particles exhibit tackiness or stickiness.

Particle Shape and Surface Roughness

When the particle shape becomes more irregular, the surface is rougher, hence, more work must be done to overcome the frictional forces, and the powder has poor flowability.

Figure 23-22. Tap density tester.

Water Content

Cohesive forces are related to the amount of water contained in the powder. Within a certain range, the angle of repose increases with increasing moisture content, due to increasing cohesive forces, and flowability decreases. However, when the water content reaches a critical peak value, the angle of repose starts to

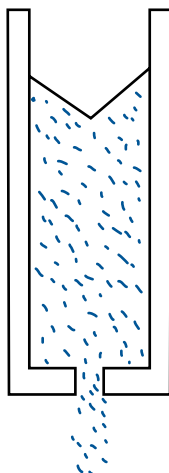


Figure 23-23. Hopper flow rate measuring device.

decrease, because a lubrication effect occurs when the gaps between particles are filled with water, and flowability increases.

Addition of Lubricants or Glidants

Lubricants and glidants are powders consisting of very fine particles. When adding lubricants and glidants to a powder with rough surfaces, the fine particles of lubricants and glidants can fill the hollow parts on the powder surfaces. The surface properties are improved, and particles are separated from each other, leading to a reduction of frictional forces between particles and, thus, enhancing powder flowability. If too much lubricants and glidants are added to the powder, there are too many fine particles in the powder, and the powder flowability decreases instead.

SOLUBILITY AND DISSOLUTION RATE

Definitions and Significances

For a solid powder, solubility and dissolution rate are extremely important properties, which have great influence on many pharmaceutical applications. But most essential, the bioavailability of solid dosage forms is influenced by solubility and dissolution. When molecules or ions are transferred from a solid into solution, this phenomenon is called dissolution. The extent of dissolved substance is known as the solubility of the solute in the solvent. The solubility of a substance is defined as the amount of substance in the solution when equilibrium is achieved between the dissolved and non-dissolved substance. When the concentration of solution is higher than the equilibrium concentration, a supersaturated concentration exists. All of the solubility and dissolution rates mentioned in this chapter are specified for solids in liquids.

Mechanism of Dissolution

The dissolution process consists of three steps. First, the molecules of solid are removed from the surface and replaced by liquid molecules. There is a phase transfer involved in this stage, the attraction forces between solid and liquid molecules must endure the cohesive forces between solid molecules; second, molecules of the solid enter into the pores (free spaces) between liquid molecules, after leaving the solid surface; and finally, diffusion occurs, which means the molecules of the solid pass through the boundary layers and into the bulk solution. According to this process, Fick's first law describes the dissolution rate of the solid powder:

$$\frac{dC}{dt} = k \frac{S}{h} (C_s - C) \quad (18)$$

where constant k is the diffusion coefficient; S is the surface area of the solid; h is the thickness of the diffusion layer; and

C_s and C are drug concentration at the solid surface (saturated) and in the bulk solution, respectively. Intrinsic dissolution rate (IDR) is the dissolution rate, which is independent of agitation speed, surface area, diffusion layer thickness, and the volume of solvent.³ Thus,

$$IDR = kC_s \quad (19)$$

Measurement

There are many reported methods that have been developed to measure the dissolution rate of solids in liquids. The most commonly used are based on the flask-stirrer method, rotating basket method, paddle method, and rotating and static disc method.⁴¹ These methods only differ in respect to the sample holder or agitation method. For all the methods, solute concentrations at different time points are measured and plotted as a function of time; the curve obtained is often referred to as the dissolution profile. The dissolution rate of different formulations can be compared from the curves visually, or quantitative parameters related to dissolution rate can be obtained by computer modeling.

Improvement of Solubility and Dissolution Rate of Powder Solids in Pharmaceutical Applications

A variety of methods have been employed to improve the dissolution rate of solid powder in pharmaceutical science, which are listed subsequently:

PARTICLE SIZE REDUCTION

According to the previous equation, the dissolution rate is proportional to the surface area of the solid, and the surface area of the particles is inversely proportional to particle size. Hence, by particle size reduction techniques (i.e., milling), the dissolution rate of the solid can be enhanced. For example, when the surface area of griseofulvin was increased from 0.3 to 2.4 m/g,² the relative absorption of griseofulvin at the same dose increased 1.5 times; for chloramphenicol powder with a particles size of 50 μm , the peak plasma concentration was reached at 1 hour, whereas, for 800 μm powder particles of chloramphenicol, the peak plasma concentration was reached at 3 hours.

INCREASING THE POROSITY OF POWDER

The dissolution medium may be able to penetrate the pores in the powder to yield an increase in the dissolution rate.

ADDITION OF WETTING AGENT

Powder particles may agglomerate in the dissolution medium. By adding wetting agents, the dispersibility of the solid is improved; agglomeration of particles can be prevented.

COSOLVENTS

In most cases, drug is more soluble in a mixture of solvents (i.e., water with water-miscible organic solvent) than in a single solvent (i.e., water).

SURFACTANTS

Surfactants can reduce the surface tension between solid particles and the air/liquid interface to improve the dissolution rate.

POLYMORPHISM AND AMORPHOUS

As previously discussed, polymorphism is a metastable form that can achieve supersaturated dissolution profiles. The supersaturated concentration can be formulated to last longer to improve the bioavailability of a poorly water-soluble drug. Similarly, for an amorphous form of the drug (see Fig. 23-24), improvement in the dissolution of two different morphologies (one is crystalline, the other is amorphous) at supersaturated conditions is shown. Clearly, the amorphous URF-ITZ displayed a much higher dissolution rate than the crystalline wet-milled ITZ.^{29,30}

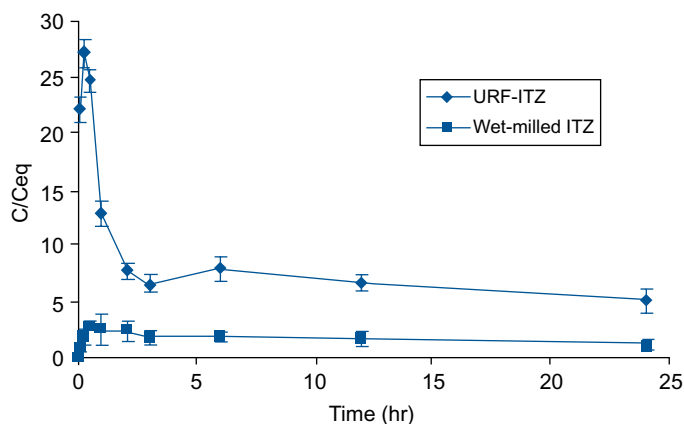


Figure 23-24. Dissolution profiles of crystalline wet-milled ITZ and amorphous URF-ITZ at supersaturation conditions. (From Yang W, *et al.* Comparison of bioavailability of amorphous versus crystalline itraconazole nanoparticles via pulmonary administration in rats. *Eur J Pharm Biopharm* 2010; 75: 33–41.)

PRODUCTION METHODS

MOLECULAR AGGREGATION

Precipitation and Crystallization

Pharmaceutical powders can be produced by precipitation and crystallization. They are relatively similar processes, by making a solid solute precipitate out of solution. To achieve this, first, a supersaturated solution of the solute is a necessary condition, which means the concentration of the solute in the solution is higher than the equilibrium solubility of the solute. Supersaturation does not yield a thermodynamically stable system; the exceeded solute tends to precipitate out. Second, after it reaches supersaturated concentration in the solution, the crystal or amorphous form starts to experience nucleation and growth. Nucleation is known as the process whereby solute molecules are gathered into a small mass on which the crystals or amorphous form can continue growing; growth means there are more solute molecules coming onto this small mass.

Many approaches have been developed to accomplish precipitation or crystallization:

- Raising the concentration of solute in the solvent to a higher level, by removing a certain amount of solvent, such as by solvent evaporation.
- In most cases, the solubility of solid powder decreases, when the temperature of the solution is lowered, so, by cooling the

solution, precipitation or crystallization can be achieved.

- Mixing the solution with another antisolvent in which the solid powder is insoluble or has very low solubility. For instance, for a poorly water-soluble drug, precipitation or crystallization can be obtained by adding water to the drug organic solution. In fact, there is a relatively new technology, called supercritical fluid method (SCF), based on this mechanism. Briefly speaking, a supercritical fluid is a substance maintained at a certain temperature and pressure above its critical point, which has both the properties of liquid and gas phases. It can penetrate like a gas and dissolve solids like a liquid. Supercritical fluid is a substitute for organic solvents to dissolve poorly water soluble drugs for powder generation, because the solubility of that powder in SCF can be easily altered by changing the temperature and pressure.⁴²

Spray Drying

Spray drying is a technique to generate powders by transforming the feed from a liquid state into a dry form, by spraying the feed into a hot drying medium (Fig. 23-25). The feed can be a solution, suspension, dispersion, or emulsion. The spray drying process mainly consists of five steps:⁴³

1. **Concentration**—Before introduced to the spray dryer, feedstock is usually concentrated.
2. **Atomization**—Favors evaporation to a dry powder, by having optimum properties.
3. **Droplet-Air Contact**—In the chamber, atomized liquid contacts with hot gas, leading to evaporation of a majority of the water or solvent contained in the droplets within a few seconds.
4. **Droplet Drying**—Water or solvent evaporation takes places.
5. **Separation**—Cyclones, bag filters, and electrostatic precipitators may be used for the final separation stage.

Spray drying technology has many applications in the pharmaceutical industry, especially for powder technology.⁴⁴ Spherical or non-spherical particles, hollow or solid particles, the size of the particles (frequently ranging 10–600 μm), and the uniformity of a powder can be altered by changing the parameters of the spray drier. By spray drying a poorly water-soluble drug, a coprecipitate with a polymer in a stable amorphous solid dispersion can be made. Therefore, the dissolution rate and bioavailability of many drugs can be greatly improved, such as tolbutamide, indomethacin, and ibuprofen.⁴⁵ In addition, particles produced by spray drying can be controlled to have very a good aerodynamic performance, making them suitable for inhalation.⁴²

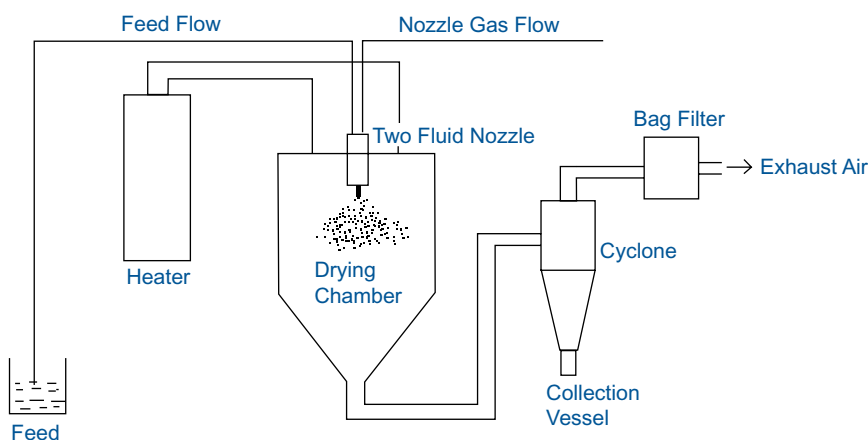


Figure 23-25. Schematic of spray drying. (From Patel RP, *et al.* *Indian J Sci Technol* 2009; 2: 44–47.)

Spray Freeze Drying (SFD)

Spray freeze drying (SFD) is a technique based on the principle that the rapid solidification process induced by freezing prevents the molecules from packing in a certain order (refers to crystallization). SFD combined with freeze-drying, consists of the following steps (Fig. 23-26 shows a typical SFD apparatus):

1. Atomization of feed using ultrasound, droplets are generated by one or two fluid nozzles or vibrating orifice (feed can be solution, suspension, or emulsion).
2. Freezing of the droplets in a cryogenic liquid or cryogenic vapor, usually liquid nitrogen.
3. Sublimation of solvent at low temperature and pressure by lyophilization or atmospheric freeze-drying using a cold desiccant gas stream.
4. SFD has similar pharmaceutical applications as spray drying⁴⁶.

PARTICLE SIZE REDUCTION

DEFINITION AND MECHANISM

Comminution is defined as a process that breaks larger particles of solid into smaller particles of appropriate size, by application of mechanical forces.⁴⁷ The purpose of comminution includes:

1. Reduction of drug particle size, increasing surface area, improve bioavailability.
2. Alteration of powder flow, improve mixing efficiency of different drugs and reduce powder irritation propensity.
3. Acceleration of dissolution rate of powder.
4. Further facilitate formulation of powder into various dosage forms, such as suspensions, tablets, and capsules.

In general, the fundamental mechanism of powder comminution is by employment of external mechanical forces to decrease the cohesion forces between the powder particles, which lead to a size reduction and surface area enhancement. Therefore, comminution is a process that transfers mechanical energy into surface energy. Comminution consists of a variety of operations, such as cutting, chopping, crushing, grinding, and milling, and is dependent on the type of equipment

selected. The choice of equipment is determined by the properties of the solid powder, such as the initial particle size and the desired size reached after size reduction. In most cases, solid powders have a relatively large particle size initially, resulting in a need for crushing or cutting operations required prior to grinding.

It has been proposed that less than 2% of the total mechanical energy is effective on the size reduction process. All of the remaining energy is lost in different ways, including elastic or plastic deformation of particles, deformation of cracks and metal equipment parts, interparticle friction, particle-equipment wall friction, heat, sound, and vibration. Three fundamental theories have advanced to describe the relationship between energy input and size reduction performance:

$$\text{Rittinger's theory: } E = k_R (S_n - S_i) \quad (20)$$

$$\text{Kick's theory: } E = \frac{K_K \log d_i}{d_n} \quad (21)$$

$$\text{Bond's theory: } E = 2K_B \left(\frac{1}{d_n} - \frac{1}{d_i} \right) \quad (22)$$

where K_R , K_K and K_B are Rittinger's constant of energy per unit area, Kick's constant of energy per unit mass, and Bond's work index, respectively; S_i is the initial surface area and S_n is the new surface area; and d_i is the initial particle diameter and d_n is the new particle diameter.^{36,48}

METHODS CLASSIFICATION

Mixing and Single Comminution

Drugs are generally processed by a single comminution, especially for drugs prone to oxidation and reduction reactions, which may cause an explosion. For expensive drugs, to reduce losses and ease of operation, single comminution should be used.

Dry and Wet Comminution

Drugs are usually processed by dry comminution. In the dry comminution process, the moisture content in the solid powder is lowered to a certain limit (<5%), prior to comminution. Conversely, wet comminution is when the comminuting process includes an amount of liquid (water or ethanol) incorporated into the dry powder to make the comminution process easy to operate. Wet comminution can avoid dust generation. Some poorly water soluble chemical agents, such as cinnabar, calamine, and talc, require smaller particle sizes and are often processed by wet comminution methods.

Comminution at Low Temperature

Some materials are difficult to comminute at room temperature, such as resins, gums, and dry extract. They can be comminuted at low temperatures, at which they have increased brittleness.

Rapid Expansion of Supercritical Solution (RESS)

RESS is a novel technology in which fine powder is produced by rapid expansion of supercritical solution. RESS is a combination of liquid-liquid extraction, based on material differences in solubility, and evaporation, based on material differences in volatility, and was developed from supercritical fluid extraction (SFE). Generally, a solution is formed when the solids are dissolved in a supercritical fluid, and then the solution is sprayed from a nozzle in an extremely short time (10^{-8} – 10^{-5} s), rapidly expanding to atmospheric pressure. Due to the decreased pressure, the solubility of solid in the supercritical fluid is reduced significantly. The concentration of solution reaches a supersaturated level and precipitation occurs. The advantage of RESS is the achievement of both narrow size distribution and finely divided particles (nanometer level). The most commonly used supercritical fluid is CO_2 , due to its non-toxic,

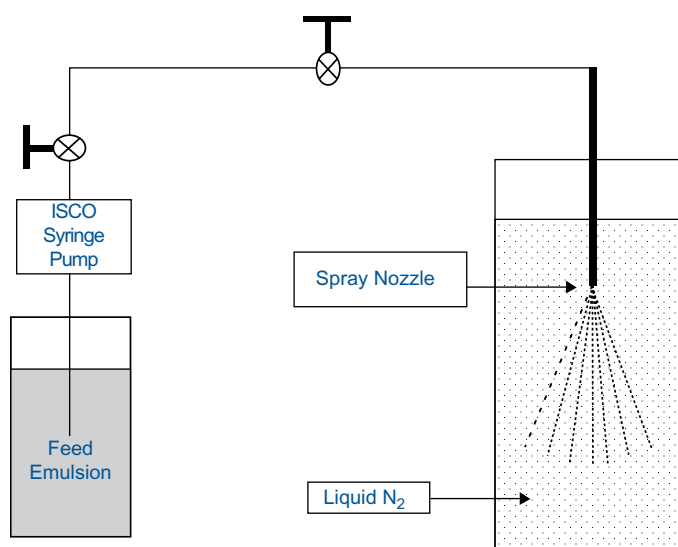


Figure 23-26. Diagram of spray-freezing into liquid. (From Rogers TL, et al. Micronized powders of a poorly water soluble drug produced by a spray-freezing into liquid-emulsion process. *Eur J Pharm Biopharm* 2003; 55: 161–172.)

non-corrosive, non-flammable, readily available, inexpensive, inert to most substances, and low critical parameter.⁴⁹

COMMINUTION EQUIPMENT

There are numerous types of large-scale comminution equipment available in pharmaceutical industry, which can be classified into three categories based on feed and size:^{2,50}

- Coarse crushers (e.g., jaw, gyratory, roll, and impact crushers), with feed materials ranging from 1.5 to 60 inches.
- Intermediate grinders (e.g., rotary cutters, disk, hammer, roller, and chaser mills), with produce provided between 20 and 200 mesh.
- Fine grinding mills (e.g., ball, rod, hammer, colloid, and fluid-energy mills; high-speed mechanical screen and centrifugal classifier), in which particles can pass through 200 mesh.

Only the most frequently used are subsequently discussed. For comminution equipment, three fundamental actions are involved in the comminution process:³

1. Attrition—breaking down of the material by a rubbing action of two surfaces;
2. Rolling—uses heavy rolling to crush and pulverize the material; and
3. Impact—the operation of hammers at high-speed impacts the lumps of material and breaks large particles to smaller ones, until the desired size is achieved.

Frequently, a combination of these three actions might occur, rather than only one action.

Crushers

JAW CRUSHER

A diagram is shown in Fig. 23-27a. It has two crushing surfaces; one is fixed, the second reciprocates by changing the pressure applied to it. Products need to be small enough to pass through the small hole at the bottom of the machine.

GYRATORY CRUSHER

Similar to the Jaw crusher, but having two moving surfaces, which approach each other and crush the powder (Fig. 23-27b).

CRUSHING ROLL

A typical crushing roll, shown in Fig. 23-27c, is composed of two rotating cylinders on horizontal shafts, driving in opposite directions. The distance between the cylinders can be adjusted to achieve the desired particle size.

Intermediate Grinding Mills

ROLLER MILLS

Fig. 23-27d shows a typical roller mill in which the feed is placed in the central rotating table, and material is ground on the concave table and by many rollers.

HAMMER MILLS

As shown in Fig. 23-27e, hammer mills consist of many hammers placed on the rotating shafts. When the feed enters the machine, the rapid moving hammers impact the solid material and break it down to relatively small particles.

CUTTER MILLS

In a cutter mill (Fig. 23-27f), many knives are attached to the rotating shafts, and, when the feed contacts with the knives, they cut the solid material into small fragments based on fracture of particles.

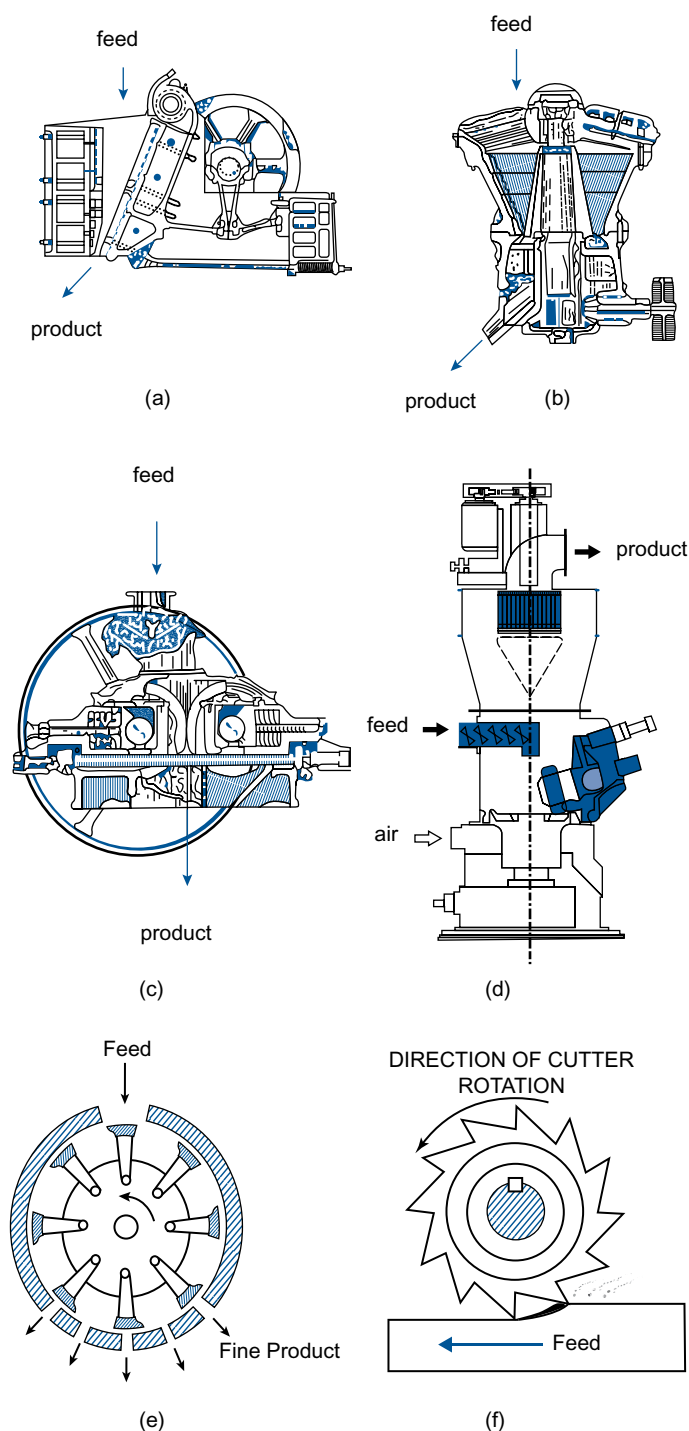


Figure 23-27. (a) Jaw crusher, (b) Gyratory crusher, (c) Crushing roll, (d) roller mill, (e) hammer mill and (f) cutter mill. (From Masuda H. *Powder Technology Handbook*, 3rd ed. Boca Raton, FL: Taylor & Francis, 2006.)

Fine Grinding Mills

BALL MILLS:

A horizontal rotating cylinder contains balls that occupy about 30-50% of the total volume of the cylinder. The balls have different sizes, so the void gaps between the balls are reduced. Large balls can break down the solid into pieces, and the small balls can break the solid into fine particles with the desired size. Both impaction and attrition are involved in this process (Fig. 23-28).

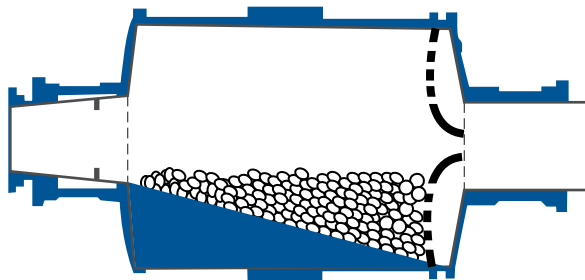


Figure 23-28. Ball mill.

VIBRATION MILLS

Vibration mills contain steel balls that occupy about 80% of its volume. When milling starts, the whole chamber of the vibration mill is vibrated, and the particle size of the solid is reduced by impaction.

FLUID ENERGY MILL

Also refer as jet mill (Fig. 23-29), the fluid energy mill contains a hollow housing having a diameter of approximately 20–200 cm. A fluid (usually air) is injected into the loop at very high pressure, which forms the turbulence zone. When the feed passes through the turbulence zone, impaction of particle to particle is achieved, due to the high energy of air. Size reduction occurs, based on impaction and attrition.

STIRRED MILL

Recently, to fulfill the demand for ultrafine particles (submicron), some grinding equipment that can produce particles less than 1 μm has been developed in the pharmaceutical industry, this includes the stirred mill. It has an agitating tank, and the balls and pebbles in the tank can facilitate agitation in the vertical direction.

To select the appropriate crushing and grinding equipment, many factors should be considered, including the starting particle size of the solid powder, the desired particle size for final product, and particle toughness and hardness. In some cases,

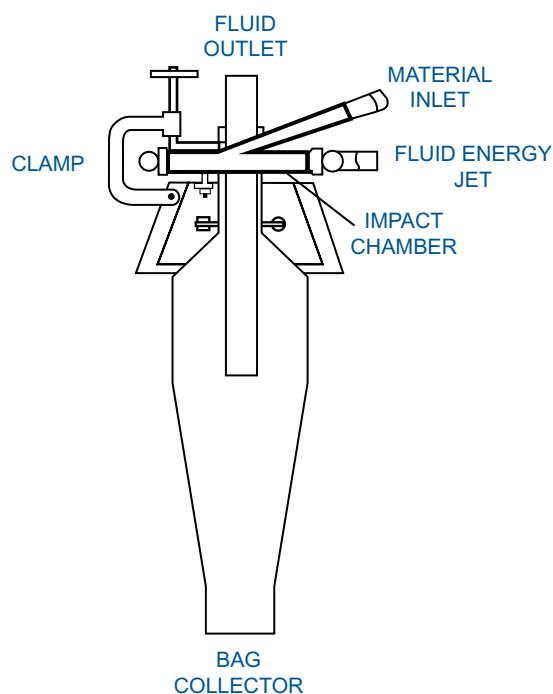


Figure 23-29. Fluid energy mill (jet mill). (From Parrott EL. Milling of pharmaceutical solids. *J Pharm Sci* 1974; 63: 813–829.)

even the cost of the size reduction method is considered, when accurate particle sizing is not required.⁴⁷

POWDER HANDLING

POWDER MIXING

Definition and Significance

Powder mixing is an operation that serves to make two or more components uniformly distributed in the powder bed. In most cases, solid dosage forms do not only contain one component. Several ingredients are combined together to contribute to either the manufacturing process or to the bioavailability. For instance, a tablet formulation often consists of the active ingredient (API) and many other excipients, such as diluents, binders, disintegrants, and lubricants. As a result, the mixing process is necessary to ensure that the API and other components are homogeneously distributed throughout the tablet. Powder mixing processes are involved in nearly every preparation process of pharmaceutical products, such as tablets, capsules, granules, pellets, bulk powders, and dry powder inhalers.⁵¹

Mechanisms

There are three main mechanisms involved in the mixing process, related to the different kinds of particle motion.

- **Convection:** The large group of particles are moving together in the mixer, driven by the rotational force provided by a mixer vessel, paddle, blade, ribbon, or gas flow. The rotating flow leads to convective motion of particles and mainly contributes to the macroscopic mixing, rather than microscopic mixing. The mixing rate of this process is relatively high, however, to achieve a more uniform mixing, prolonged times are required, because particles within the moving group are not able to mix well.
- **Shear:** Shear refers to the action that one layer of powder particles moves relative to another layer of particles. Shear mixing can be caused by particles from a different powder layer, having different velocities developed in a rotary vessel or due to the compression force and extension of the powder bed. Hence, shear mixing is the movement of a layer of particles, rather than the movement of individual particles. The rate of shear mixing is intermediate and mainly causes semimicroscopic mixing.
- **Diffusion:** Diffusion is known as the movement of individual particles, due to random motion of powder particles. When the powder bed is placed in a mixer, the volume of that powder bed increases, due to convection and shear forces, which drive the powder flowing around the mixer. Therefore, the voids between particles increase, leading to great potential for individual particles to mix with each other. The rate of diffusive mixing is relatively low, compared with convection and shear mixing; however, it contributes to the microscopic mixing.

For most of the mixing operations, all of these mechanisms are involved, rather than only one or two of them.⁵²

In a mixing process, the degree of mixedness is often defined as the uniformity of the powder components throughout the entire powder mixture, which is frequently calculated based on some statistic and expressed by variance of the composition from the mean value of the powder mixture. Imaging samples are randomly taken from the powder mixture, x_i is the composition of sample number i , the variance of samples can be calculated as follows:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n [(x_i) - \bar{x}]^2}{(n-1)}} \quad (23)$$

where $\bar{x}$ is the mean value of the composition.^{2,3}

Powder segregation (demixing) is the opposite process of mixing, which means the separation of the drug from the powder components. Under some circumstances, the mixing process provides the necessary situations for powder to segregate, because, during the mixing process, particles are moving relatively close to each other, and their movement may differ due to their different properties, such as size, shape, or density. Conversely, particles having similar properties tend to move together, causing that region of the powder bed to have a higher composition concentration. Some other handling processes may also cause powder segregation, such as transportation, packing, feeding, and storage. The smallest region that can measure the homogenous state in a powder mixture is called the scale of scrutiny and can be a length, area, or volume. The scale of scrutiny is a scale of the finished product. For instance, for a 100mg tablet, the scale of scrutiny is 100 mg. When the scale of scrutiny is large or the particle size is small, which indicates a large number of particles are present in the sample, the powder mixture are relatively uniform. Segregation can never be completely avoided, so powder sampling is preferred for the final product.^{52,53}

Mixing Equipment

- The mortar and pestle represent one of the most commonly used small scale mixing equipment. The mortar and pestle method combines comminution and mixing in a single operation. Hence, it is very useful when some degree of particle size reduction, as well as mixing, is required. Mortars can be made of different materials and shapes, and pestles are made of the same materials as the mortars. Conventional mortars include glass mortars, Wedgwood mortars, and porcelain mortars, and there is an increasing use of a newly developed mortar, called the electronic mortar and pestle (EMP), among pharmacists. Different types of mortars have specific utility in compounding different materials. For example, glass mortars are suitable for preparing solutions, suspensions, and ointments; Wedgwood mortars are designed primarily for size reduction for most of the materials in modern pharmacy practice; porcelain mortars are similar to Wedgwood mortars, but they are more preferable for comminution of soft aggregates, crystals, and mixing of powders with uniform particle size; EMP is specifically designed for blending of creams, ointments, and oral liquids. The EMP uses a spinning blade and a moving arm to mix products. The largest benefit of EMP is that the materials can be weighed, mixed, and dispensed all in the same container.
- Tumbling mixers/blenders are most commonly used for powders with similar densities. There are different types of tumbling mixers, including V-shaped, rotating cube, cylindrical, double-cone, oblique, and Y-cone mixers (Fig. 23-30a, b, c). Tumbling mixers are likely to mix powders with good flowability and granules, rather than cohesive or powders with poor flowability, because shear force provided by these mixers is not enough to separate the individual particles from agglomerates. The regular loading ability of tumbling mixers ranges from approximately 50 g to 100 kg. For this equipment, segregation tends to occur, if particle size differs significantly.
- In the ribbon mixer, spiral ribbons, paddles, or blades are mounted on a shaft, which is located in the center of the mixing equipment (Fig. 23-30d). Rotating of the shaft leads to the circular movement of the blades and subsequent mixing of the powder. The ribbon mixer can mix powders with poor flowability and avoid segregation as compared to tumbling mixers.
- A diagram of an industrial planetary mixer is shown in Fig. 23-30g. There is a mixing bowl located at the bottom of the equipment, which can move up and down. The

bowl is used for feeding and discharging of the product. The bowl can be raised to the position of the mixing blades for the actual mixing process. The mixing blades or paddles are located off the center of the mixer. They rotate around their center and simultaneously rotate around the center of the mixer. This double rotating process can cause a more complete mixing of the entire equipment, so “dead spots” are avoided. Planetary mixers can break agglomerates easily.

- The Nautamixer (Fig. 23-30e) is composed of a bottom discharger with a rotating screw fastened to the upper end of the rotating arm. The screw conveys the product to the top, where it can flow back to the powder feed. Hence, for this equipment, the vertical impeller and horizontal rotating arm are combined together to induce a combination of convection, shear, and diffusive mixing.
- The fluid-bed mixer is usually used for powder mixing, prior to granulation in the same bowl. Therefore, the fluid-bed mixer is part of a fluid-bed granulator system (Fig. 23-30f). Fluidization of the powder particles can be achieved by blowing heated and filtered air into the equipment. Efficient mixing is achieved by circulation of the fluidized powder. After mixing, granulation liquid is pumped from the liquid receiver/holder through a spray nozzle and on to the fluidized powder bed to facilitate granulation. When sufficient liquid has been sprayed to achieve appropriate granule size, the nozzle is turned off and the wet granules are dried by the fluidizing heated air.

POWDER SAMPLING

When powders are poured into a pile, segregation occurs. Particles with small size (fine particles) will stay in the center of the pile, and coarser particles tend to concentrate at the periphery. Size segregation is unavoidable, during powder transportation, storage, or packing. Therefore, powder sampling is important for the quality control of the final product. Certain

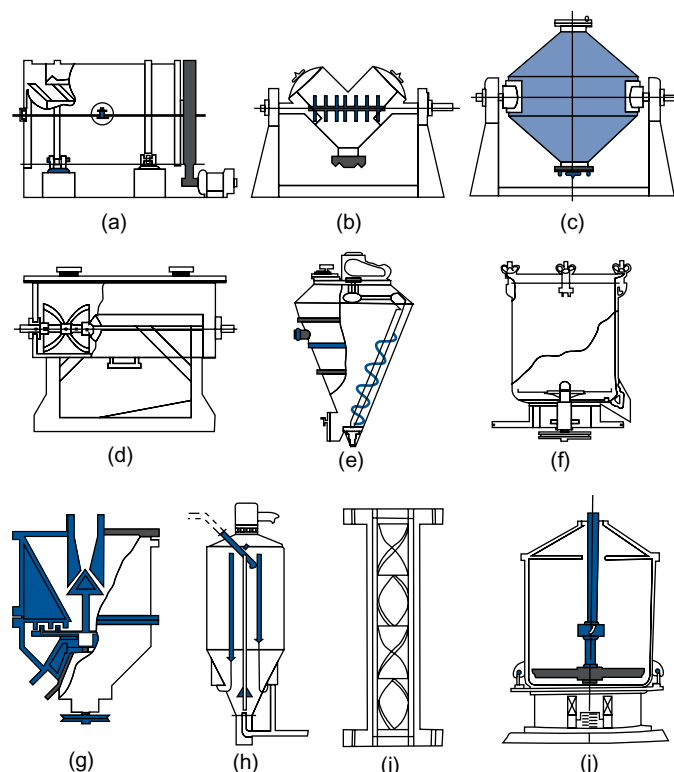


Figure 23-30. Typical powder mixers. (From Masuda H. *Powder Technology Handbook*, 3rd ed. Boca Raton, FL: Taylor & Francis, 2006.)

Table 23-4. Summary of Sampling Methods⁵³

Apparatus	Type of sampler	Method of sampling	Scale	Advantages	Disadvantages
Sample thief	Manual	Tubular steel retains a core sample when inserted into powder	Large and small	Good for free-flowing powders	Additional weight at bottom of bag may vary sample size; Can be hard to push into powder; Fines may lodge between tubes; Particles can fracture; Fines compact, impeding flow; Segregation may occur as fines percolate into sample more easily than coarse particles; A plug of powder can be pushed ahead of thief and surface material contaminates sample; Personal preference introduces bias for the area sample
Hand scoop	Manual	Cross-sectional sample from moving stream, bags, or barrels	Large	Simple and cheap	Thin layer may remain on belt, leading to bias; Overfilling can lead to an excess of fines.
Shovel	Manual	Pits are dug in the powder bed and a shovelful taken from bottoms and sides	Large, up to several tons	Simple and cheap	Cannot be used with particles of more than 5 cm diameter
Cross-cut sampling	Manual/semiautomatic/automatic	Material is shovelled from the conveyor belt	Large	Simple	Can leave a layer on the belt and result in bias
Pneumatic lance	Semiautomatic	Air flow used on entry and exit of lance from powder bed	Large	Disturbance of powder minimized over sample thief; Porous plate prevents too many fines due to strong air current	Personal preference may bias sample.
Vacuum probe sampler	Semiautomatic	Powder extracted by vacuum	Large	Simple	Difficult to sample below surface without contamination; Personal preference leads to bias; Fines more easily extracted than coarse particles.
Gravity-flow auger sampler	Semiautomatic	Slotted tube in flowing powder rotates and worm screw carries out material	Large and small	Easy to use	Bias is still a problem here; Difficult to sample all of powder stream, therefore bias.
Sampling from a moving stream	Manual/automatic	Powder is sampled as it falls off the conveyor	Large	If carried out properly can be very good sampling technique; easily designed into a new plant.	If overfilled, a greater number of fines than coarse particles are collected; Difficult and expensive to fit into an existing plant; Obtaining a fixed sample is difficult; Difficult to prevent dust escaping.
Full-stream trough sampler	Automatic	Powder is sampled as it falls off the conveyor	Large	Can be used to sample dusty material	Must not overfill; Difficult and expensive to install into existing plant.
Arc path cutter	Automatic	Chute moves through powder stream, and collects sample	Large	No operator bias	Difficult to take more than one sample size; Must cover the whole of the stream to avoid bias.
Straight path cutter	Automatic	Rectangular chute moves through powder stream	Large	Different sample sizes can be taken easily	Must cover the whole of the stream to avoid bias

Moving-flap sample divider	Automatic	A flap in the stream samples powder or allows it to be stored depending on its position	Large	Efficient as covers the whole of the stream when sampling	Resultant sample is large and subsampling needed. Bias may be introduced here; Bias is present due to one side being sampled more than the other.
Integrated automatic sampling plant	Automatic	Primary sample is selected and repeatedly screened, resulting in the final sample	Large	Quick	Variations in materials can cause problems
Chute splitter	Sub-sampler	A series of chutes split sample repeatedly	Large	Can be repeated until desired sample size is achieved	If segregation occurs the result can be misleading; Prone to operator bias.
Cone and quartering	Sub-sampler	Powder poured through cone and divided into four equal parts. This is repeated until the desired sample size is reached	Small	Simple	Prone to operator bias as fine particles remain in the center of the cone; Symmetry is difficult to achieve but essential for accuracy.
Spinning riffler	Sub-sampler	A steady stream of powder flows into a rotating basket of containers	Large/ small	Good for sub-sampling large samples; Good for powders with good flow properties; Minimal bias; More efficient than other samplers tested.	Air currents may displace fines but can be avoided with a slower rotation speed; Expensive; Time-consuming; Segregation may be a problem.
Free-fall tumbler mixer	Sub-sampler	A ladle in the lid of the mixer collects a good representative sample	Large/ small	Can be used with fines present; Representative sample produced in a short time period.	
Hopper sample divider	Sub-sampler	Hoppers oscillate and powder falls into two containers. Only one of the contents is kept.	Small	Sample size can be controlled by monitoring time over each container	Large number of increments needed for accuracy
Table sampler	Sub-sampler	Powder flows down inclined plane and prisms and holes split the powder	Small	Simple	Very low accuracy; After each separation should be a complete mix to avoid errors.

rules of sampling are typically applied: sample from a moving stream of powder; and sample the whole of the stream for many equal periods of time. This makes sure all parts of the powder have equal chances to be sampled, leading to a maximized accuracy and minimized error. However, sampling errors can be caused by the non-uniform distribution of the particles or the inherent differences in physicochemical properties of particles. Table 23-4 summarized a complete listing of sampling methods.⁵³

STORAGE

In most cases, powders are stored in bins (silos). Silos are classified by their height: shallow bins ($h \leq 1.5d$ or $1.5a$) and deep bins ($h < 1.5d$ or $1.5a$), where h is the height of the silo, d is the inside diameter of a circular silo, and a is the short side length of a rectangular silo. Silos can also be classified by bin numbers, such as single bins and group bins. Most of the silos are closed, such as vacuum silos; some are open, like ore silos. There are advantages to storing powders in the silos, such as large storage capacity, reduced cost of transportation or feeding, automatic loading, unloading and volume control, no damage or changing to the stored powders, and relative ease of incorporation to a manufacturing system.

POWDERS AS A DOSAGE FORM

When the term “powder” is referred to as a dosage form, it represents a formulation that is a mixture of powdered drug and excipients. Powders are one of the most conventional dosage forms. Nowadays, with the rapid development of formulations containing high potent compounds, the use of powders as a dosage form has declined. Most of the powders are replaced by tablets and capsules. However, under certain circumstances, powders still have some advantages over solid dosage forms on the market. Advantages of powders include: chemically stable; relatively convenient, when able to provide a large dose of drugs, rather than capsules or tablets; and the dissolution rate of oral powders containing water soluble drugs is generally faster than tablets or capsules, in which disintegration of the tablet or the capsule shell is required prior to dissolution. Disadvantages of powders as a dosage form include: they are not convenient for patients to carry, compared to capsules or tablets; masking unpleasant tastes is difficult; potent drugs requiring low doses may not be appropriate; and irritating drugs, which can cause damage to the stomach, are not suitable. There are a variety of powdered dosage forms commercially available, such as bulk powders, divided powders, dusting powders, insufflations, and dry powder inhalers.

BULK POWDERS

Bulk powders refer to a mixture of all the materials, packed into a properly designed bulk container, such as a glass or plastic bottle. The major problem of bulk powders is the inaccuracy of dose. The dose of bulk powders can be affected by many factors, including the measuring device (spoon), storage humidity, degree of settling, and patient factors. For example, the dose of bulk powder may vary for patients using differently sized spoons, or even those using the same spoon according to their technique. In addition, drugs present in the bulk powders are better suited, if they have a wider therapeutic window, a large dose, and pleasant taste. Effervescent powders are a special type of bulk powder. In addition to drugs and other excipients, effervescent powders contain an effervescent couple (i.e., sodium bicarbonate and citric acid), which react and effervesce when in contact with water. The effervescent dosage form is helpful to cover the unpleasant taste of salty or bitter drugs. In this chapter, bulk powders are described for internal use, including oral powders, as well as powders for injection. For drugs that are not stable when dissolved in an aqueous pharmaceutically acceptable diluent, such as water, sterile liquid can be added to sterile powders contained in ampoules to form the solution just prior to use.

DIVIDED POWDERS

Divided powders are bulk powders in which the individual dose has been packed separately. The traditional packing of divided powders is in wrapped paper. However, many problems are involved in this, when the materials are volatile, hygroscopic, or deliquescent. Therefore modern packing methods are developed to replace the use of paper wrapping, such as foil and plastic laminates. Effervescent powders can be packed into individual doses, because the plastic laminates can protect powders from moisture adsorption. The powdered product should always be protected from exposure to moisture.

DUSTING POWDERS

Dusting powders are designed for external use, acting as a therapeutic, lubricant, or protective. Dusting powders act locally and are intended to have no systemic absorption. Dusting powders are usually dispensed in a relatively fine state (micronized) to increase efficacy and decrease irritation. Dusting powders can be packed in glass or metal containers with a perforated lid to allow the powders to be dusted to the effective area. Excellent flowability is necessary for this dosage form. Pressure aerosols are another delivery form that can generate dusting powders. They are more expensive than the sifter-container, but several advantages are realized, such as convenient operation, and protection from moisture, air, and contamination.

INSUFFLATIONS AND DRY POWDER INHALERS (DPI)

Insufflations are fine powders of drugs, which are dosed into the nose, ear, or throat by the use of an insufflator. The use of conventional insufflators has declined, due to poor patient compliance and dose non-uniformity. Some newly developed devices have been introduced to replace the traditional insufflators. In these devices, drugs are usually dispensed with a carrier excipient, such as lactose, and placed into a hard gelatin capsule. When the device is operated, the capsule is broken and the fine powder is inhaled into the patient's body.⁴²

Pulmonary delivery of dry powder formulations is a popular approach to deliver the drug to the lung locally, for the treatment of such diseases as asthma and chronic obstructive pulmonary disease (COPD). Dry powder inhalers are similar to the new insufflators previously discussed above.⁵⁴

REFERENCES

1. Duran J. *Sands, Powders, and Grains: An Introduction to the Physics of Granular Materials*. New York: Springer, 2000.
2. Masuda KH, Hideto Y. *Powder Technology Handbook*, 3rd ed. Boca Raton, FL: Taylor & Francis, 2006.
3. Aulton ME. *Aulton's Pharmaceutics: The Design and Manufacture of Medicines*. Edinburgh/New York: Churchill Livingstone, 2007.
4. United States Pharmacopeial Convention. United States Pharmacopeia, 34th ed., and National Formulary, 29th ed. Baltimore, MD: United Book Press, 2001: 319.
5. United States Pharmacopeial Convention. United States Pharmacopeia, 34th ed., and National Formulary, 29th ed. Baltimore: United Book Press, 2001: 161.
6. Blott SJ, Pye K. *Sedimentology* 2008; 55: 31–63.
7. Sun Y, Xia Y. Shape-controlled synthesis of gold and silver nanoparticles. *Science* 2002; 298: 2176–2179.
8. Chatfield EJ. *J Sci Instrum* 1967; 44: 615–617.
9. Marple VA *et al.* Next generation pharmaceutical impactor (a new impactor for pharmaceutical inhaler testing). Part I: Design. *J Aerosol Med* 2003; 16: 283–299.
10. United States Pharmacopeial Convention. United States Pharmacopeia, 34th ed., and National Formulary, 29th ed. Baltimore, MD: United Book Press, 2001: 224–239.
11. Kasama SK. *ISIJ International* 2009; 49: 637–644.
12. Ito T *et al.* Simultaneous determination of the size and surface charge of individual nanoparticles using a carbon nanotube-based Coulter counter. *Anal Chem* 2003; 75: 2399–2406.

13. Aboofazeli R *et al.* Particle size analysis of concentrated phospholipid microemulsions I. Total intensity light scattering. *AAPS PharmSci* 2000; 2: E13.
14. Aboofazeli R *et al.* Particle size analysis of concentrated phospholipid microemulsions II. Photon correlation spectroscopy. *AAPS PharmSci* 2000; 2: E19.
15. Beaubien LJ, Vanderwielen AJ. Particle-size analysis of pharmaceutical powders. *J Pharm Sci* 1980; 69: 651–655.
16. Allen T. *Particle Size Measurement*, 5th ed. New York/London: Chapman & Hall, 1997.
17. Dia S *et al.* Air or helium comparison pycnometer. *J Pharm Belg* 1971; 26: 631–637.
18. Farmer WM. Measurement of particle size, number density, and velocity using a laser interferometer. *Appl Opt* 1972; 11: 2603–2612.
19. Gomi H. Multiple scattering correction in the measurement of particle size and number density by the diffraction method. *Appl Opt* 1986; 25: 3552.
20. Peter H *et al.* *Aerosol Science and Technology* 2002; 36: 227–238.
21. Giddings JC, Moon MH. Particle size distribution by sedimentation/steric field-flow fractionation: development of a calibration procedure based on density compensation. *Anal. Chem* 1991; 63.
22. Nichols G *et al.* A review of the terms agglomerate and aggregate with a recommendation for nomenclature used in powder and particle characterization. *J Pharm Sci* 2002; 91: 2103–2109.
23. Lowell S, Shields JE. *Powder Surface Area and Porosity*. London/New York: Chapman & Hall, 1991.
24. Mersmann A. *Crystallization Technology Handbook*, 2nd ed. New York: Marcel Dekker, 2001.
25. Tiwary AK. Modification of crystal habit and its role in dosage form performance. *Drug Dev Ind Pharm* 2001; 27: 699–709.
26. Hammond RB *et al.* Quantifying solubility enhancement due to particle size reduction and crystal habit modification: case study of acetyl salicylic acid. *J Pharm Sci* 2007; 96: 1967–1973.
27. Abaturon LV *et al.* Water as a stabilizer and plasticizer of the globular structure of lysozyme. *Dokl Akad Nauk* 1997; 355: 114–116.
28. Ogain ON *et al.* Particle engineering of materials for oral inhalation by dry powder inhalers. I-Particles of sugar excipients (trehalose and raffinose) for protein delivery. *Int J Pharm* 2011; 405: 23–35.
29. Yang W *et al.* High bioavailability from nebulized itraconazole nanoparticle dispersions with biocompatible stabilizers. *Int J Pharm* 2008; 361: 177–188.
30. Yang W *et al.* Comparison of bioavailability of amorphous versus crystalline itraconazole nanoparticles via pulmonary administration in rats. *Eur J Pharm Biopharm* 2010; 75: 33–41.
31. Brittain HG. *Polymorphism in Pharmaceutical Solids*, 2nd ed. New York: Informa Healthcare, 2009.
32. Rajjada DK *et al.* Influence of microenvironment pH, humidity, and temperature on the stability of polymorphic and amorphous forms of clopidoogrel bisulfate. *AAPS PharmSciTech* 2010; 11: 197–203.
33. Bruni G *et al.* Determination of the nateglinide polymorphic purity through DSC. *J Pharm Biomed Anal* 2011; 54: 1196–1199.
34. Lu GW *et al.* Characterization of a novel polymorphic form of celecoxib. *J Pharm Sci* 2006; 95: 305–317.
35. Lindberg NO *et al.* Flowability measurements of pharmaceutical powder mixtures with poor flow using five different techniques. *Drug Dev Ind Pharm* 2004; 30: 785–791.
36. Zhou YC *et al.* Numerical investigation of the angle of repose of monosized spheres. *Phys Rev E Stat Nonlin Soft Matter Phys* 2001; 64.
37. Javadzadeh Y *et al.* Effect of some commercial grades of microcrystalline cellulose on flowability, compressibility, and dissolution profile of piroxicam liquisolid compacts. *Drug Dev Ind Pharm* 2009; 35: 243–251.
38. Liu LX *et al.* Effect of particle properties on the flowability of ibuprofen powders. *Int J Pharm* 2008; 362: 109–117.
39. Kim EH *et al.* Effect of surface composition on the flowability of industrial spray-dried dairy powders. *Colloids Surf B Biointerfaces* 2005; 46: 182–187.
40. Kaerger JS *et al.* Influence of particle size and shape on flowability and compactibility of binary mixtures of paracetamol and microcrystalline cellulose. *Eur J Pharm Sci* 2004; 22: 173–179.
41. United States Pharmacopeial Convention. United States Pharmacopeia, 34th ed., and National Formulary, 29th ed. Baltimore, MD: United Book Press, 2001: p 319.
42. Pilcer G, Amighi K. Formulation strategy and use of excipients in pulmonary drug delivery. *Int J Pharm* 2010; 392: 1–19.
43. Patel RP *et al.* *Indian J Sci Technol* 2009; 2: 44–47.
44. Sollohub K, Cal K. Spray drying technique: II. Current applications in pharmaceutical technology. *J Pharm Sci* 2010; 99: 587–597.
45. Kawashima Y *et al.* Improvement of solubility and dissolution rate of poorly water-soluble salicylic acid by a spray-drying technique. *J Pharm Pharmacol* 1975; 27: 1–5.
46. Rogers TL *et al.* Micronized powders of a poorly water soluble drug produced by a spray-freezing into liquid-emulsion process. *Eur J Pharm Biopharm* 2003; 55: 161–172.
47. Parrott EL. Milling of pharmaceutical solids. *J Pharm Sci* (1974); 63: 813–829.
48. Niazi SK. *Handbook of Pharmaceutical Manufacturing Formulations*. Boca Raton, FL: CRC Press: Boca Raton, 2004.
49. Charpentier PA *et al.* Study of the RESS process for producing beclomethasone-17,21-dipropionate particles suitable for pulmonary delivery. *AAPS PharmSciTech* 2008; 9: 39–46.
50. Rhodes MJ. *Principles of Powder Technology*. New York: John Wiley and Sons, Inc., 1990.
51. Kaye BH. *Powder Mixing*. London: Chapman & Hall, 1997.
52. Venables HJ, Wells JI. Powder mixing. *Drug Dev Ind Pharm* 2001; 27: 599–612.
53. Venables HJ, Wells JI. Powder sampling. *Drug Dev Ind Pharm* 2002 28: 107–117.
54. Telko MJ, Hickey AJ. Dry powder inhaler formulation. *Respir Care* 2005; 50: 1209–1227.

Chapters 24

Solutions, Emulsions, Suspensions, and Extracts

Michael M. Crowley, PhD

SOLVENTS FOR LIQUID PHARMACEUTICAL PREPARATIONS 435

STABILITY CONSIDERATIONS 437

SOLUTIONS 439

NONAQUEOUS SOLUTIONS 446

EMULSIONS 448

SUSPENSIONS 458

EXTRACTS 465

The dosage forms described in this chapter are prepared by employing pharmaceutically and therapeutically acceptable vehicles. The active ingredient(s) may be dissolved in aqueous media, an organic solvent, or a combination of the two, by suspending the drug (if it is insoluble) in an appropriate medium or by incorporating the active pharmaceutical ingredient into one of the phases of an oil and water emulsion. Such solutions, suspensions, and emulsions are further defined in subsequent paragraphs, but some, with similar properties and applications, are considered in greater detail elsewhere in Remington.

These dosage forms are useful for a number of reasons. They can be formulated for different routes of administration: orally, introduction into body cavities, or external application. The dose can easily be adjusted by dilution, making the oral liquid form ready to be administered to children or people unable to swallow tablets or capsules. Extracts eliminate the need to isolate the drug in pure form, allow several ingredients to be administered from a single source (e.g., pancreatic extract), and permit the preliminary study of drugs from natural sources. Occasionally, solutions of drugs, such as potassium chloride, are used to minimize adverse effects in the gastrointestinal tract.

The preparation of these dosage forms involves several considerations on the part of the pharmacist, namely; the purpose of the drug, internal or external use, solubility and concentration of the drug, selection of the liquid vehicle(s), physical and chemical stability of the drug and any excipients, preservation of the preparation, and use of appropriate excipients, such as buffers, solubility enhancers, suspending agents, emulsifying agents, viscosity controlling agents, colors, and flavors. Oral preparations require consideration be given to improving patient compliance by making an acceptable product; consequently, color, odor, and taste must be considered. The viscosity of a product must also be considered, so it has the proper palatability for an oral preparation and has the appropriate suspending properties, if it is an emulsion or suspension. The theory of solutions, which involves solubility, ionization, pH control through the use of buffers, and solubilization, is discussed in Chapters 13 (Solutions and Phase Equilibria) and 15 (Ionic Solutions and Electrolyte Equilibria). Due to the complexity of some manufactured products, compounding may be carried out with the aid of linear programming models to obtain the optimal product. Chapters 26 through 28 should be consulted for information on the preparation and characteristics of those liquid preparations that are intended for parenteral and ophthalmic use.

Much has been written about the biopharmaceutical properties of solid dosage forms. Many researchers begin their absorption studies of drugs administered in solution to assess the

bioavailability relative to tablets and capsules. Absorption occurs when drugs are in a dissolved state, thus, it is frequently observed that the bioavailability of oral dosage forms decreases in the following order: aqueous solution > aqueous suspension > tablet or capsule. Formulation may influence the bioavailability and pharmacokinetics of drugs in solution, including drug concentration, volume of liquid administered, pH, ionic strength, buffer capacity, surface tension, specific gravity, viscosity, and excipients. Emulsions and suspensions are more complex systems; consequently, the bioavailability and pharmacokinetics of these systems may be affected by additional formulation factors, such as surfactants, type of viscosity agent, particle size and particle-size distribution, polymorphism, and solubility of drug in the oil phase.

Liquid preparations may be dispensed in one of three ways: 1) in its original container, 2) repackaging a bulk product at the time a prescription is presented by the patient, or 3) compounding the solution, suspension, or emulsion in the dispensary. Compounding may involve nothing more than mixing marketed products in the manner indicated on the prescription or, in specific instances, may require the incorporation of active ingredients and excipients in a logical and pharmaceutically acceptable manner into aqueous or organic solvents that will form the bulk of the product.

The pharmacist, in the first instance, depends on the pharmaceutical manufacturer to produce a product that is safe, efficacious, elegant, and stable until its expiration date, when stored at conditions described on its label. Manufacturers guarantee efficacy of their products, but, in some instances, consumer preference is variable. For example, cough syrups marketed by two different manufacturers may contain the same active ingredient(s), and the relative merits of the two products may appear interchangeable. In such instances, the commercial advantage may be based on factors, such as flavor, color, aroma, mouth feel, and packaging.

SOLVENTS FOR LIQUID PHARMACEUTICAL PREPARATIONS

The pharmacist's knowledge of the physical and chemical characteristics of a given drug dictates the selection of the appropriate solvent for a particular formulation. In addition to solubility, solvent selection is also based on clarity, toxicity, viscosity, compatibility with excipients, chemical inertness, palatability, odor, color, and economy. In most cases, especially solutions for oral, ophthalmic, or parenteral administration, water is the preferred solvent, because it meets the majority of the above criteria better than other available solvents. Often, an auxiliary solvent is also employed to augment the solvent action of water

or to contribute to a product's chemical or physical stability. Alcohol, glycerin, and propylene glycol have been frequently used for these purposes.

Solvents, such as acetone and isopropyl alcohol, are too toxic for use in oral pharmaceutical preparations, but they are useful as solvents in organic chemistry and in the preparatory stages of drug development. For purposes such as this, certain solvents are officially recognized in the compendia. A number of fixed oils, such as corn oil, cottonseed oil, peanut oil, and sesame oil, serve useful solvent functions, particularly in the preparation of oleaginous injections, and are recognized in the compendia for this purpose.

WATER

The major ingredient in most of the dosage forms described herein is water. It is used both as a vehicle and as a solvent for the desired flavoring or medicinal ingredients. Its tastelessness, freedom from irritating qualities, and lack of pharmacological activity make it ideal for such purposes. There is, however, a tendency to assume its purity is constant and it can be stored, handled, and used with a minimum of care. Although true that municipal supplies must comply with Environmental Protection Agency (EPA) regulations (or comparable regulations in other countries), drinking water must be purified before it can be used in pharmaceuticals. Water quality can have a significant impact on the stability of pharmaceutical dosage forms.¹ In manufacturing environments, the design of purified water systems must meet standards outlined in the United States Pharmacopeia (USP) and be validated.²⁻⁵

Five of the eight solvent waters described in the USP are used in the preparation of parenterals, irrigations, or inhalations. Purified Water must be used for all other pharmaceutical operations, dosage forms, and, as needed, in all USP tests and assays. It must meet rigid specifications for chemical purity. Purified Water is obtained by deionization, distillation, ion-exchange, reverse osmosis, filtration, or other suitable procedures. For parenteral administration, Water for Injection, Bacteriostatic Water for Injection, or Sterile Water for Injection must be used. Sterile Water may be sterile at the time of production but may lose this characteristic, if stored improperly.

The major impurities in water are calcium, iron, magnesium, manganese, silica, and sodium. These cations are combined with the bicarbonate, sulfate, or chloride anions. Hard waters are those that contain calcium and magnesium cations. Bicarbonates are the major impurity in alkaline waters. Deionization processes do not necessarily produce Purified Water that will comply with EPA requirements for drinking water. Resin columns retain phosphates and organic debris. Either alone or in combination, these substances can act as growth media for micro-organisms. Observations have shown that deionized water containing 90 organisms/mL contained 106 organisms/mL after 24-hour storage. Ultraviolet radiant energy (240–280 nm), heat, or filtration can be used to limit the growth of, kill, or remove micro-organisms in water. The latter method employs membrane filters and can be used to remove bacteria from heat-labile materials.

The phenomenon of osmosis involves the passage of water from a dilute solution across a semi-permeable membrane to a more concentrated solution. Flow of water can be stopped by applying pressure to the concentrated solution equal to the osmotic pressure. The flow of water can be reversed by applying a pressure greater than the osmotic pressure. The process of reverse osmosis uses the latter principle; by applying pressure greater than the osmotic pressure to the concentrated solution (e.g., tap water), pure water may be obtained. Organic molecules are rejected on the basis of a sieve mechanism related to their size and shape. Small organic molecules, with a molecular weight smaller than approximately 200, will pass through the membrane material. Because there are few organic molecules with a molecular weight of less than 200 in the municipal water

supply, reverse osmosis is sufficient for the removal of organic material. The pore sizes of the selectively permeable reverse-osmosis membranes are between 0.5 and 10 nm. Viruses and bacteria larger than 10 nm are rejected, if no imperfections exist in the membrane. The membranes may and do develop openings that permit the passage of micro-organisms. Due to the semi-static conditions, bacteria can grow both upstream and downstream of the membrane.

ALCOHOLS

Next to water, alcohol is the most commonly used solvent in pharmacy for many organic compounds. When mixed with water, a hydroalcoholic mixture is formed capable of dissolving both alcohol-soluble and water-soluble substances, a feature especially useful for extraction and purification of active constituents from crude drugs and synthetic procedures. Alcohol, USP, is 94.9–96.0% by volume, at 15.56°C of C₂H₅OH, and Dehydrated Alcohol, USP, contains not less than 99.5% C₂H₅OH by volume. Dehydrated alcohol is utilized when an essentially water-free alcohol is necessary. Alcohol is widely used for its miscibility with water and its ability to dissolve many water-insoluble ingredients, including drug substances, flavors, and antimicrobial preservatives. Alcohol is used in liquid products as an antimicrobial preservative or in conjunction with parabens, benzoates, sorbates, and other agents. Diluted Alcohol, NF, is prepared by mixing equal volumes of Alcohol, USP, and Purified Water, USP. Due to contraction upon mixing, the final volume of such mixtures is not the sum of the individual volumes of the two components, but is about 3% less.

The United States Food and Drug Administration (FDA) has expressed concern regarding undesired pharmacologic and potential toxic effects of alcohol when ingested by children. For this reason, manufacturers of over-the-counter (OTC) oral drug products have been asked to restrict, if possible, the use of alcohol and include appropriate warnings in the labeling. For OTC oral products intended for children under 6 years of age, the recommended alcohol content limit is 0.5%; for products intended for children 6–12 years of age, the recommended limit is 5%; and for products recommended for children older than 12 years of age and for adults, the recommended limit is 10%.

Rubbing Alcohol, USP, must be manufactured in accordance with the requirements of the US Treasury Department, Bureau of Alcohol, Tobacco, and Firearms, Formula 23-H (8 parts by volume of acetone, 1.5 parts by volume of methyl isobutyl ketone, and 100 parts by volume of ethyl alcohol). It contains not less than 68.5% and not more than 71.5% by volume of dehydrated alcohol, the remainder consisting of water and the denaturants with or without color additives and perfume oils. Rubbing Alcohol contains in each 100 mL not less than 355 mg of sucrose octaacetate or not less than 1.40 mg of denatonium benzoate. The preparation may be colored with one or more color additives listed by the FDA for use in drugs, and a suitable stabilizer may be added. The use of this denaturant mixture makes the separation of ethyl alcohol from the denaturants a virtually impossible task with ordinary distillation apparatus. This discourages the illegal removal and use of the alcoholic content of rubbing alcohol as a beverage. The product is volatile and extremely flammable and should be stored in tight containers remote from ignition sources. It is used externally as a soothing rub for bedridden patients, a germicide for instruments, and a skin cleanser prior to injection.

Isopropyl Rubbing Alcohol is about 70% by volume isopropyl alcohol, the remainder consisting of water with or without color additives, stabilizers, and perfume oils. It is used exclusively as a vehicle in topical products and applications. This preparation and a commercially available 91% isopropyl alcohol solution are commonly employed to disinfect needles and syringes for hypodermic injections of insulin and for disinfecting the skin.

Glycerin is a clear, syrupy liquid with a sweet taste and is miscible with water and alcohol. Glycerin is used in a wide variety of pharmaceutical formulations, including oral, otic, ophthalmic, topical, and parenteral preparations. In topical pharmaceutical formulations and cosmetics, glycerin is used primarily for its humectant and emollient properties. In parenteral formulations, glycerin is used mainly as a solvent. In oral solutions, glycerin is used as a solvent, sweetening agent, antimicrobial preservative, and viscosity-increasing agent.

Propylene glycol has become widely used as a solvent, extractant, and preservative in a variety of liquid pharmaceutical formulations, including parenterals. Propylene glycol is a viscous liquid and is miscible with water and alcohol. It is a useful solvent with a wide range of applications and is often used in place of glycerin. As an antiseptic, it is similar to ethanol, and against molds it is similar to glycerin and only slightly less effective than ethanol. Propylene glycol is also used as a carrier for emulsifiers and as a vehicle for flavors, as opposed to ethanol, due to its lack of volatility.

STABILITY CONSIDERATIONS

The stability of the active ingredient in the final product is a primary concern to the formulator, the pharmacist, and the patient. In general, drug substances are less stable in aqueous media than solid dosage forms, and it is important to properly stabilize and preserve solutions, suspensions, and emulsions that contain water. Acid-base reactions, acid or base catalysis, oxidation, and reduction can occur in these products. Reactions and interactions (adsorption) can arise from ingredient–ingredient interactions or container–product interactions. For pH sensitive compounds, any of these interactions may alter the pH and cause precipitation.

Vitamins, essential oils, and almost all fats and oils can be oxidized. Formulators use the word “auto-oxidation,” when the ingredient(s) reacts with oxygen but without drastic external interference. Such reactions can be initiated by heat, light, including ultraviolet radiant energy, peroxides, or other labile compounds or heavy metals, such as copper or iron. This initiation step results in the formation of a free radical that then reacts with oxygen. The free radical is regenerated and reacts with more oxygen (propagation). The reactions are terminated, when the free radicals react with one another.

The effect of trace metals can be minimized, using chelating agents, such as citric acid or EDTA. Antioxidants may retard or delay oxidation by rapidly reacting with free radicals as they are formed (quenching). Common antioxidants include propyl, octyl, and dodecyl esters of gallic acid, butylated hydroxyanisole (BHA), butylated hydroxytoluene (BHT), and the tocopherols or vitamin E. Connors and coworkers provide a detailed approach for the prevention of oxidative degradation of pharmaceuticals.⁶ Table 24-1 lists common antioxidants and chelating agents used in pharmaceutical preparations.

The USP states that, if a product must be repackaged, the container specified by the compendium must be used. For example, a suitable opaque plastic container should be used, if a light-resistant container is specified. If a product is diluted, or where two products are mixed, the pharmacist should use his or her knowledge to guard against incompatibility and instability. Oral antibiotic preparations constituted into liquid form should never be mixed with other products. If the chemical stability of extemporaneously prepared liquid preparations is unknown, their use should be minimized and every care taken to ensure product characteristics will not change while used by the patient.

Due to the number of excipients and additives in these preparations, it is recommended that all the ingredients be listed on the container to reduce the risks that confront hypersensitive patients when these products are administered. Finally, the pharmacist should inform the patient regarding the appropriate use of the product, the proper storage conditions, and the time after which it should be discarded.

Table 24-1. Common Antioxidants and Chelating Agents Used in Liquid Pharmaceutical Dosage Forms

Antioxidants	Alpha tocopherol Ascorbic acid Acorbyl palmitate Butylated hydroxyanisole Butylated hydroxytoluene Monothioglycerol Potassium metabisulfite Propionic acid Propyl gallate Sodium ascorbate Sodium bisulfite Sodium metabisulfite Sodium sulfite
Chelating Agents	Citric acid monohydrate Disodium edetate Dipotassium edetate Edetic acid Fumaric acid Malic acid Phosphoric acid Sodium edetate Tartaric acid Trisodium edetate

PRESERVATIVES

In addition to stabilization of pharmaceutical preparations against chemical and physical degradation, liquid and semisolid preparations must be protected against microbial contamination. Nearly all products described in this chapter contain water and, thus, with certain exceptions, such as aqueous acids, will support microbial growth. Aqueous solutions, syrups, emulsions, and suspensions often provide excellent growth media for micro-organisms, such as molds, yeast, and bacteria (typically *pseudomonas*, *E.coli*, *salmonella* and *staphylococcus*).

Kurup and Wan describe many preparations that are not preserved adequately and are not able to resist microbial contamination.⁷ Products, such as ophthalmic and injectable preparations, are sterilized by autoclaving (20 minutes at 15 pounds of pressure at 120°C, followed by dry heat at 180°C for 1 hour) or filtration. However, many require the presence of an antimicrobial preservative to maintain aseptic conditions throughout their stated shelf life.⁸ Certain hydroalcoholic and alcoholic preparations do not require addition of a chemical preservative, if the alcohol content is sufficient to prevent microbial growth. An alcohol content of 15% by weight in acid solutions and 18% by weight in alkaline solutions is sufficient to prevent microbial growth. Most alcohol containing preparations, such as elixirs, spirits, and tinctures, are self-preserving and do not require preservation. Indeed, the formulator should challenge any new preparation by procedures described in the General Tests and Assays, parts <51> and <61> of the USP, and other methods reported in the literature.^{9–12}

When a preservative is required, its selection is based on several considerations, in particular, the site of use, whether internal, external, or ophthalmic.¹³ Several researchers have described various interactions that must be considered when preservatives are selected.^{14–15} The major criteria that should be considered in selecting a preservative are as follows: its physicochemical properties, such as solubility and dissociation constant, and it should be effective against a wide spectrum of micro-organisms, stable for its shelf life, nontoxic, nonsensitizing, compatible with the ingredients in the dosage form, inexpensive, and relatively free of taste and odor.

The chosen preservative should be sufficiently stable and soluble to achieve adequate concentration to provide protection. This choice is more critical in two and three phase emulsion

systems in which the preservative may be more soluble in the oil phase than in the aqueous phase.^{12,16} The pH of the preparation must be considered to ensure the preservative does not dissociate, rendering it ineffective or degrade by acid or base catalyzed hydrolysis. The undissociated moiety or molecular form of a preservative possesses preservative capacity, because the ionized form is unable to penetrate micro-organisms. The preservative must be compatible with the formulation ingredients and the product container and closure. Finally, the preservative must not impact the safety or comfort of the patient when administered. For instance, preservatives used in ophthalmic preparations must be non-irritating. Chlorobutanol, benzalkonium chloride, and phenylmercuric nitrate are commonly used in these applications.

Although few micro-organisms are viable below a pH of 3 or above a pH of 9, most aqueous pharmaceutical preparations are manufactured within the favorable pH range. Acidic preservatives, such as benzoic acid, boric acid, and sorbic acid, are less dissociated and more effective in acidic formulations. Similarly, alkaline preservatives are less effective in acidic or neutral conditions and more effective in alkaline formulations. The scientific literature is rife with examples of incompatibilities between preservatives and other pharmaceutical adjuncts.^{17–19} Commonly used macromolecules, including cellulose derivatives, polyethylene glycol, and tragacanth gum, have been reported to cause preservative failure, due to binding and adsorption.^{20,21}

The mode of action by which preservatives interfere with microbial growth, multiplication, and metabolism occurs through one of several mechanisms. Preservatives often alter cell membrane permeability, causing leakage of cell constituents (partial lysis), complete lysis and cytoplasmic leakage, and/or coagulation of cytoplasmic constituents (protein precipitation). Other preservatives inhibit cellular metabolism by interference with enzyme systems or cell wall synthesis, oxidation of cellular constituents, or hydrolysis.

Table 24-2 lists preservatives commonly used in pharmaceutical products with typical concentration levels. Preservatives may be grouped into a number of classes, depending upon

their molecular structure. These basic groups are subsequently discussed.

Alcohols

Ethanol is useful as a preservative, when it is used as a solvent; however, it does need a relatively high concentration, somewhat greater than 15%, to be effective. Too high a concentration may result in incompatibilities in suspension and emulsion systems. Propylene glycol also is used as a solvent in oral solutions and topical preparations, and it can function as a preservative in the range of 15–30%. It is not volatile like ethanol and is used frequently not only in solutions, but also in suspensions and emulsions. Chlorobutanol and phenylethyl alcohol are other alcohols used in lower concentrations (approximately 1%) as preservatives.

Acids

Benzoic acid has a low solubility in water, about 0.34% at 25°C, but the apparent aqueous solubility of benzoic acid may be enhanced by the addition of citric acid or sodium acetate to the solution. The concentration range used for inhibitory action varies from 0.1 to 0.5%. Activity depends on the pH of the medium, because only the undissociated acid has antimicrobial properties. Optimum activity occurs at pH values below 4.5; at values above pH 5, benzoic acid is almost inactive.²² It has been reported that antimicrobial activity of benzoic acid is enhanced by the addition of the basic protein protamine.²³ Sorbic acid also has a low solubility in water, 0.3% at 30°C. Suitable concentrations for preservative action are in the range of 0.05–2%. Its preservative action is due to the nonionized form; consequently, it is only effective in acid media. The optimum antibacterial activity is obtained at pH 4.5, and practically no activity is observed above pH 6. Sorbic acid is subject to oxidation, particularly in the presence of light and in aqueous solutions. Activity against bacteria can be variable, due to its limited stability. Thus, sorbic acid is frequently used in combination with other antimicrobial preservatives or glycols in which synergistic effects occur.

Table 24-2. Common Preservatives Used in Liquid Pharmaceutical Dosage Forms and Their Typical Concentration Levels

Antimicrobial Preservatives	Typical Usage Level (% w/w)	Antifungal Preservatives	Typical Usage Level (% w/w)
Benzalkonium Chloride	0.002–0.02%	Butyl Paraben	0.1–0.4%
Benzethonium Chloride	0.01–0.02%	Methyl Paraben	0.1–0.25%
Benzyl Alcohol	3.0%	Ethyl Paraben	0.1–0.25%
Bronopol	0.01–0.1%	Propyl Paraben	0.1–0.25%
Cetrimide	0.005%	Benzoic Acid	0.1–0.5%
Cetylpyridinium chloride	0.0005–0.0007%	Potassium sorbate	0.1–0.2%
Chlorhexidine	0.002–0.5%	Sodium Benzoate	0.1–0.2%
Chlorobutanol	0.5%	Sodium Propionate	5–10%
Chlorocresol	0.2%	Sorbic Acid	0.05–0.2%
Chloroxylenol	0.1–0.8%		
Cresol	0.15–0.3%		
Ethyl Alcohol	15–20%		
Glycerin	20–30%		
Hexetidine	0.1%		
Imidurea	0.03–0.5%		
Phenol	0.1–0.5%		
Phenoxyethanol	0.5–1.0%		
Phenylethyl Alcohol	0.25–0.5%		
Phenylmercuric Nitrate	0.002–0.01%		
Propylene Glycol	15–30%		
Thimerosal	0.1%		

Esters

Parabens are esters of *p*-hydroxybenzoic acid and include the methyl, ethyl, propyl, and butyl derivatives. The water solubility of the parabens decreases as the molecular weight increases from 0.25% for the methyl ester to 0.02% for the butyl ester. These compounds are used widely in pharmaceutical products, stable over a pH range of 4–8, and have a broad spectrum of antimicrobial activity, although they are most effective against yeasts and molds. Antimicrobial activity increases as the chain length of the alkyl moiety is increased, but aqueous solubility decreases; therefore, a mixture of parabens is frequently used to provide effective preservation. Preservative efficacy is also improved by the addition of propylene glycol (2–5%) or by using parabens in combination with other antimicrobial agents, such as imidurea. Activity is reduced in the presence of nonionic surface active agents, due to binding. In alkaline solutions, ionization takes place, and this reduces their activity; in addition, hydrolytic decomposition of the ester group occurs with a loss of activity.

Quaternary Ammonium Compounds

Benzalkonium chloride is a mixture consisting principally of the homologs $C_{12}H_{25}$ and $C_{14}H_{29}$. This preservative is used at a relatively low concentration, 0.002–0.02%, depending on the nature of the pharmaceutical product. This class of compounds has an optimal activity over the pH range of 4–10 and is quite stable at room temperature. Due to the cationic nature of this type of preservative, it is incompatible with many anionic compounds and can bind to nonionic surfactants. It is used in preparations for external use or those solutions that come in contact with mucous membranes. In ophthalmic preparations, benzalkonium chloride is widely used at a concentration of 0.01–0.02% w/w. Often, it is used in combination with other preservatives or excipients, particularly 0.1% w/v disodium edetate, to enhance its antimicrobial activity against strains of *Pseudomonas*. A concentration of 0.002–0.02% is used in nasal and otic formulations, sometimes in combination with 0.002–0.005% thimerosal. Benzalkonium chloride 0.01% w/v is also employed as a preservative in small-volume parenteral products.

Clearly, when the pharmacist dispenses or compounds liquid preparations, responsibility is assumed, along with the manufacturer, for the maintenance of product stability. General chapter <1191> of the USP describes stability considerations for dispensing, which should be studied in detail.⁹ Stock should be rotated and replaced, if expiration dates on the label so indicate. Products should be stored in the manner indicated on the manufacturer's label or in the compendium. Further, products should be checked for evidence of instability. With respect to solutions, elixirs, and syrups, major signs of instability are color change, precipitation, and evidence of microbial or chemical gas formation. Emulsions may cream, but, if they break (i.e., there is a separation of an oil phase), the product is considered unstable. Sedimentation and caking are primary indications of instability in suspensions. The presence of large particles may mean excessive crystal growth has occurred (Ostwald Ripening). Additional details on these topics are provided in the pertinent sections of this chapter.

SOLUTIONS

A solution is a homogeneous mixture that is prepared by dissolving a solid, liquid, or gas in another liquid and represents a group of preparations in which the molecules of the solute or dissolved substance are dispersed among those of the solvent. Most solutions are unsaturated with the solute; in other words, the concentration of the solute in the solution is below its solubility limit. The strengths of pharmaceutical solutions are expressed in terms of percentage strength, although for very dilute preparations expressions of ratio strength are sometimes used. The term percentage when used without qualification

(as with w/v, v/v, or w/w) means percentage weight-in-volume for solutions or suspensions of solids in liquids; percentage weight-in-volume for solutions of gases in liquids; percentage volume-in-volume for solutions of liquids in liquids; and weight-in-weight for mixtures of solids and semisolids.

Solutions may also be classified on the basis of physical or chemical properties, method of preparation, use, physical state, number of ingredients, and particle size. For the pharmacist and pharmaceutical scientist, solutions are more defined by site of administration and composition than by physicochemical definitions. For instance, pharmaceutical solutions may be classified as an oral solution, otic solution, ophthalmic solution, or topical solution. These solutions may also be classified based upon their composition. Syrups are aqueous solutions containing a sugar; elixirs are sweetened hydroalcoholic (combinations of water and ethanol) solutions; spirits are solutions of aromatic materials, if the solvent is alcoholic, or aromatic waters, if the solvent is aqueous. Depending on their method of preparation and concentration, tinctures or fluid extracts are solutions prepared by extracting active constituents from crude drugs.

Many pharmaceutical chemicals are only slowly soluble in a given solvent and require an extended time for complete dissolution. To increase the dissolution rate, a pharmacist and pharmaceutical scientist may employ one or several techniques, such as applying heat, reducing the particle size of the solute, utilizing of a solubilizing agent, or subjecting the ingredients to rigorous agitation. In most cases, solutes are more soluble in solvents at elevated temperatures than at room temperature or below, due to the endothermic nature of the dissolution process. The pharmacist should ensure that the materials are heat stable and non-volatile, when using heat to facilitate the dissolution rate.

AQUEOUS SOLUTIONS

The narrower definition in this subsection limits the solvent to water and excludes those preparations that are sweet and/or viscid in character and nonaqueous solutions. This section includes those pharmaceutical forms that are designated as Aromatic Waters, Aqueous Acids, Solutions, Douches, Enemas, Gargles, Mouthwashes, Juices, Nasal Solutions, Otic Solutions, and Irrigation Solutions.

Aromatic Waters

The USP defines Aromatic Waters as clear, saturated aqueous solutions (unless otherwise specified) of volatile oils or other aromatic or volatile substances.⁹ Their odors and tastes are similar, respectively, to those of the drugs or volatile substances from which they are prepared, and they are free from empyreumatic and other foreign odors. Aromatic waters may be prepared by distillation or solution of the aromatic substance, with or without the use of a dispersing agent. They are used, principally, as flavored or perfumed vehicles.

Peppermint Water USP and Stronger Rose Water USP are examples of aromatic waters. Concentrated waters, such as peppermint, dill, cinnamon, and caraway, may be prepared as follows:

Dissolve 20 mL of the volatile oil in 600 mL of 90% ethanol. Add sufficient purified water in successive small portions to produce 1000 mL. Shake vigorously after each addition. Add 50 g of sterilized purified talc. Shake occasionally for several hours, and filter.

The aromatic water is prepared by diluting the concentrate with 39 times its volume of water.

The chemical composition of many of the volatile oils is known, and suitable synthetic substances may be used in preparing pharmaceuticals and cosmetics. Similarly, many synthetic aromatic substances have a characteristic odor; for example, geranyl phenyl acetate has a honey odor. Such substances, either alone or in combination, can be used in nonofficial preparations.

The principal difficulty experienced in compounding prescriptions containing aromatic waters is “salting out” certain

ingredients, such as very soluble salts. A replacement of part of the aromatic water with purified water is permissible, when no other function is being served than that of a vehicle. Aromatic waters will deteriorate with time and should, therefore, be made in small quantities, protected from intense light and excessive heat, and stored in airtight, light-resistant containers.

Aqueous Acids

Inorganic acids and certain organic acids, although of minor significance as therapeutic agents, are of great importance in pharmaceutical manufacturing and analysis. This is especially true of acetic, hydrochloric, and nitric acids. Many of the more important inorganic acids are available commercially in the form of concentrated aqueous solutions. The percentage strength varies from one acid to another and depends on the solubility and stability of the solute in water and on the manufacturing process. Thus, Hydrochloric Acid contains from 36.5 to 38.0% by weight of HCl, whereas Nitric Acid contains from 69 to 71% by weight of HNO₃.

Because the strengths of these concentrated acids are stated in terms of percent by weight, it is essential that specific gravities also be provided, if one is to calculate conveniently the amount of absolute acid contained in a unit volume of the solution as purchased. The mathematical relationship involved is given by the equation:

$$M = V \times S \times F,$$

where M is the mass in g of absolute acid contained in V mL of solution, having a specific gravity S and a fractional percentage strength F .

As an example, Hydrochloric Acid containing 36.93% by weight of HCl has a specific gravity of 1.1875. Therefore, the amount of pure HCl supplied by 100 mL of this solution is given by:

$$M = 100 \times 1.1875 \times 0.3693 = 43.85 \text{ g HCL.}$$

Although many of the reactions characteristic of acids offer opportunities for incompatibilities, only a few are of sufficient importance to require more than casual mention. Acids and acid salts decompose carbonates with liberation of carbon dioxide; in a closed container, sufficient pressure may be developed to produce an explosion. Inorganic acids react with salts of organic acids to produce the free organic acid and a salt of the inorganic acid. If insoluble, the organic acid will be precipitated. Thus, salicylic acid and benzoic acid are precipitated from solutions of salicylates and benzoates. Boric acid, likewise, is precipitated from concentrated solutions of borates. By a similar reaction, certain soluble organic compounds are converted into an insoluble form. Phenobarbital sodium, for example, is converted into phenobarbital that will precipitate in aqueous solution.

The ability of acids to combine with alkaloids and other organic compounds containing a basic nitrogen atom is used in preparing soluble salts of these substances. Certain solutions, syrups, elixirs, and other pharmaceutical preparations may contain free acid, which causes these preparations to exhibit the incompatibilities characteristic of the acid. Acids also possess the incompatibilities of the anions they contain, and, in the case of organic acids, these are frequently of prime importance. These are discussed under the specific anions.

Diluted Acids

The diluted acids in the USP are aqueous solutions of acids of a suitable strength (usually 10% w/v, but Diluted Acetic Acid is 6% w/v) for internal administration or for the manufacture of other preparations.

The strengths of the official undiluted acids are expressed as percentages in weight (w/w), whereas the strengths of the official diluted acids are expressed as percentages in volume (w/v). It, therefore, becomes necessary to consider the specific

gravities of the concentrated acids, when calculating the volume required to make a given quantity of diluted acid. The following equation will give the number of milliliters required to make 1000 mL of diluted acid:

$$\frac{\text{Strength of diluted acid} \times 1,000}{\text{Strength of undiluted acid} \times \text{Specific gravity of undiluted acid}}$$

Thus, if one wishes to make 1000 mL of Diluted Hydrochloric Acid USP (10% w/v) using Hydrochloric Acid that assays 37.5% HCl (sp gr 1.18), the amount required is

$$\frac{10 \times 1,000}{37.5 \times 1.18} = 226 \text{ mL}$$

Diluted Hydrochloric Acid, USP has been used in the treatment of achlorhydria. However, it may irritate the mucous membrane of the mouth and attack the enamel of the teeth. The usual dose is 2–4 mL, well-diluted with water. In the treatment of achlorhydria, no attempt is made to administer more than a relief-producing dose.

Douches

A douche is an aqueous solution directed against a part or into a cavity of the body. It functions as a cleansing or antiseptic agent. An eye douche, used to remove foreign particles and discharges from the eyes, is directed gently at an oblique angle and allowed to run from the inner to the outer corner of the eye. Pharyngeal douches are used to prepare the interior of the throat for an operation and cleanse it in suppurative conditions. Similarly, there are nasal douches and vaginal douches. Douches are directed to the appropriate body part by using bulb syringes.

Douches are often dispensed in the form of a powder with directions for dissolving in a specified quantity of water (usually warm). However, tablets for preparing solutions are available (e.g., Dobell's Solution Tablets), or the solution may be prepared by the pharmacist. If powders or tablets are supplied, they must be free from insoluble material to produce a clear solution. Tablets are produced by the usual processes, but any lubricants or diluents used must be readily soluble in water. Boric acid may be used as a lubricant, and sodium chloride is normally used as a diluent. Many tablets deteriorate on exposure to moist air and should be stored in airtight containers.

Douches are not official as a class of preparations, but several substances in the compendia are frequently employed as such in weak solutions. Vaginal douches are the most common type of douche and are used for cleansing the vagina and hygienic purposes. Liquid concentrates or powders, which may be prepared in bulk or as single-use packages, should be diluted or dissolved in the appropriate amount of warm water prior to use. The ingredients used in vaginal douches include antimicrobial agents, such as benzalkonium chloride, the parabens or chlorothymol, and anesthetics or antipruritics, such as phenol or menthol. Astringents, such as zinc sulfate or potassium alum, surface-active agents, such as sodium lauryl sulfate, and chemicals to alter the pH, such as sodium bicarbonate or citric acid, are also used.

Enemas

A number of solutions are administered rectally for the local effects of the medication (e.g., hydrocortisone) or for systemic absorption (e.g., aminophylline). In the case of aminophylline, the rectal route of administration minimizes the undesirable gastrointestinal reactions associated with oral therapy.²⁴ Clinically effective blood levels of the agents are obtained within 30 minutes following rectal instillation. Corticosteroids are administered as retention enemas or continuous drip as adjunctive treatment of some patients with ulcerative colitis.

Enema preparations are rectal injections employed to evacuate the bowel (evacuation enemas), influence the general

system by absorption, or to affect a local disease. The latter two are called retention enemas. They may possess anthelmintic, nutritive, sedative, or stimulating properties, or they may contain radiopaque substances for roentgenographic examination of the lower bowel.

Sodium chloride, sodium bicarbonate, sodium monohydrogen phosphate, sodium dihydrogen phosphate, glycerin, docosate potassium, and light mineral oil are used in enemas to evacuate the bowel. These substances may be used alone, in combination with each other, or in combination with irritants, such as soap. Evacuation enemas are given at body temperature in quantities of 1 to 2 pt injected slowly with a syringe.

An official retention enema used for systemic purposes is aminophylline. Retention enemas are to be retained in the intestine and should not be used in quantities larger than 150 mL for an adult. Usually, the volume is considerably smaller, such as a few milliliters. "Microenema" is a term used to describe these small-volume preparations. Vehicles for retention microenemas have been formulated with small quantities of ethanol and propylene glycol, and no significant difference in irritation, as compared with water, was found. A number of other drugs, such as valproic acid, indomethacin, and metronidazole, have been formulated as microenemas for the purpose of absorption.

Gargles

Gargles are aqueous solutions frequently containing antiseptics, antibiotics, and/or anesthetics used for treating the pharynx and nasopharynx by forcing air from the lungs through the gargle held in the throat; subsequently, the gargle is expectorated. Many gargles must be diluted with water prior to use. Although mouthwashes are considered as a separate class of pharmaceuticals, many are used as gargles either as is or diluted with water.

A gargle/mouthwash containing the antibiotic tyrothricin has been shown to provide higher levels of gramicidin, a component of tyrothricin, in saliva when used as a gargle, rather than a mouthwash²⁵. Higher saliva levels of gramicidin were obtained, when a lozenge formulation was employed. Rapid relief of pharyngeal and oral pain was obtained, when Cepacaine solution, which contains a topical anesthetic, was used as a gargle.²⁶

Nystatin is administered in both powder and liquid form to treat oral fungal infections.²⁷ The medication is taken by placing one-half of the dose in each side of the mouth, swishing it around as long as possible, gargling, and swallowing. Hydrogen peroxide is a source of nascent oxygen and a weak topical antibacterial agent. Hydrogen peroxide topical solution has been used as a mouthwash or gargle in the treatment of pharyngitis or Vincent's stomatitis.²⁸⁻²⁹ Hydrogen peroxide has also been applied in root canals of teeth or other dental pulp cavities. Although used topically as a 1.5-3% solution for cleansing wounds, hydrogen peroxide is diluted with an equal volume of water for use as a mouthwash or gargle.

Mouthwashes

Mouthwashes are aqueous solutions, often in concentrated form, containing one or more active ingredients and excipients. They are used by swishing the liquid in the oral cavity. Mouthwashes can be used for two purposes: therapeutic and cosmetic. Therapeutic rinses or washes can be formulated to reduce plaque, gingivitis, dental caries, and stomatitis. Cosmetic mouthwashes may be formulated to reduce bad breath through the use of antimicrobial and/or flavoring agents.

Recent information indicates that mouthwashes are being used as a dosage form for a number of specific problems in the oral cavity; for example, mouthwashes containing a combination of antihistamines, hydrocortisone, nystatin, and tetracycline have been prepared from commercially available suspensions, powders, syrups, or solutions for the treatment of stomatitis, a painful side effect of cancer chemotherapy. Other drugs include allopurinol, also used for the treatment of stomatitis,³⁰ pilocarpine for xerostoma (dry mouth),³¹ amphotericin

B for oral candidiasis,³² and chlorhexidine gluconate for plaque control.³³ Mouthwashes may be used for diagnostic purposes. For example, oral cancer and lesions are detected using toluidine blue mouth rinse.³⁴

Commercial products (e.g., Cepacol, Listerine, Micrin, or Scope) vary widely in composition. Tricca has described the excipients found in Mouthwashes as alcohols, surfactants, flavors, and coloring agents.³⁵ Alcohol is often present in the range of 10-20%. It enhances the flavor, provides sharpness to the taste, aids in masking the unpleasant taste of active ingredients, functions as a solubilizing agent for some flavoring agents, and may function as a preservative. Humectants, such as glycerin and sorbitol, may form 5-20% of the mouthwash. These agents increase the viscosity of the preparation and provide a certain body or mouth feel to the product. They enhance the sweetness of the product and, along with the ethanol, improve the preservative qualities of the product.

Surfactants of the nonionic class, such as polyoxyethylene/polyoxypropylene block copolymers or polyoxyethylene derivatives of sorbitol fatty acid esters, may be used. The concentration range is 0.1-0.5%. An anionic surfactant used occasionally is sodium lauryl sulfate. Surfactants are used, because they aid in the solubilization of flavors and in the removal of debris by providing foaming action. Cationic surfactants, such as cetylpyridinium chloride, are used for their antimicrobial properties, but these tend to impart a bitter taste.

Flavors are used in conjunction with alcohol and humectants to overcome disagreeable tastes; at the same time, flavors must be safe to use. The principle flavoring agents are peppermint, spearmint, cinnamon, wintergreen oils, menthol, or methyl salicylate. Other flavoring agents may be used singly or in combination. Finally, coloring agents are also used in these products.

Juices

A juice is prepared from fresh ripe fruit, is aqueous in character, and is used in making syrups employed as vehicles. The freshly expressed juice is preserved with benzoic acid and allowed to stand at room temperature for several days, until the naturally present pectins are destroyed by enzymatic action, as indicated by the filtered juice yielding a clear solution with alcohol. Pectins, if allowed to remain, would cause precipitation in the final syrup.

Cherry Juice and Tomato Juice are described in the USP. Artificial flavors now have replaced many of the natural fruit juices. Although they lack the flavor of the natural juice, they are more stable and easier to incorporate into the final pharmaceutical form. Commercial juices, such as orange, apple, grape, and mixed vegetables, have been used recently to prepare extemporaneous preparations of cholestyramine³⁶ and nizatidine.³⁷ Information on cranberry juice indicates it may be effective in controlling some urinary tract infections and urolithiasis.³⁸

Nasal Solutions

Nasal solutions are aqueous solutions designed to be administered to the nasal passages in drops or sprays. Other nasal preparations may be in the form of emulsions or suspensions. The adult nasal cavity has about a 20 mL capacity with a large surface area (about 180 cm²) for drug absorption afforded by the microvilli present along the pseudo-stratified columnar epithelial cells of the nasal mucosa.³⁹ The nasal tissue is highly vascularized, making it an attractive site for rapid and efficient systemic absorption. Another advantage of nasal delivery is it avoids first-pass metabolism by the liver. For some peptides and small molecular compounds, intranasal bioavailability has been comparable to that of injections. However, bioavailability decreases as the molecular weight of a compound increases, and, for proteins composed of more than 27 amino acids, bioavailability is quite low.⁴⁰ Various pharmaceutical techniques and functional excipients, such as surfactants, have been shown capable of enhancing the nasal absorption of large molecules.⁴¹⁻⁴²

Many drugs are administered for their local sympathomimetic effects to reduce nasal congestion, such as Ephedrine Sulfate Nasal Solution, USP, or Naphazoline Hydrochloride Nasal Solution, USP. A few other preparations, Lypressin Nasal Solution, USP, and Oxytocin Nasal Solution, USP, are administered in spray form for their systemic effect for the treatment of diabetes insipidus and milk letdown prior to breast feeding, respectively. Table 24-3 lists examples of commercial products for nasal use.

Nasal solutions are formulated to be similar to nasal secretions with regard to tonicity, pH, and viscosity, so normal ciliary action is maintained. Thus, aqueous nasal solutions are isotonic and slightly buffered to maintain a pH of 5.5–6.5. In addition, antimicrobial preservatives, similar to those used in ophthalmic preparations, and appropriate drug stabilizers, if required, are included in the formulation.

Current studies indicate that nasal sprays are deposited mainly in the atrium and cleared slowly into the pharynx with the patient in an upright position. Drops spread more extensively than the spray, and three drops cover most of the walls of the nasal cavity with the patient in a supine position and head tilted back and turned left and right.^{43,44} It is suggested that drop delivery, with appropriate movement by the patient, leads to extensive coverage of the walls of the nasal cavity.

Most nasal solutions are packaged in dropper or spray bottles, containing 15–30 mL of medication. The pharmaceutical scientist should ensure the product is stable in the containers, and the pharmacist should keep the packages tightly closed during

periods of nonuse. The patient should be advised that, should the solution become discolored or contain precipitated matter, it must be discarded.

Otic Solutions

These solutions are occasionally referred to as ear or aural preparations. Other otic preparations include suspensions and ointments for topical application in the ear. Ear preparations are placed in the ear canal by drops or in small amounts for the removal of excessive cerumen (ear wax) or for the treatment of ear infections, inflammation, or pain.

The main classes of drugs used for topical administration to the ear include analgesics, such as benzocaine; antibiotics, such as neomycin; and anti-inflammatory agents, such as cortisone (Table 24-4). The USP preparations include Antipyrine and Benzocaine Otic Solution. The Neomycin and Polymyxin B Sulfates and Hydrocortisone Otic Solutions may contain appropriate buffers, solvents, and dispersants in an aqueous solution. The main solvents used in these preparations include glycerin or water. The viscous glycerin vehicle permits the drug to remain in the ear for a long time. Anhydrous glycerin, being hygroscopic, tends to remove moisture from surrounding tissues, thus, reducing swelling. Viscous liquids like glycerin or propylene glycol are used either alone or in combination with a surfactant to aid in the removal of cerumen (ear wax). To provide sufficient time for aqueous preparations to act, it is necessary for patients to remain on their side for a few minutes, so the

Table 24-3. Examples of Commercial Nasal Preparations

Product Name	Manufacturer	Active Ingredient	Indication
Atrovent Nasal Spray	Boehringer Ingelheim	Ipratropium bromide 0.06%	Seasonal or Allergic Rhinitis
Beconase AQ Nasal Spray	GlaxoSmithKline	Beclomethasone dipropionate, monohydrate 42 mcg	Seasonal or Allergic Rhinitis
Miacalcin	Novartis	Calcitonin-salmon, 2200 I.U. per mL	Postmenopausal osteoporosis
Nasalcrom Nasal Spray	Pharmacia	Cromolyn sodium 5.2 mg	Seasonal or Allergic Rhinitis
Nasarel Nasal Spray	IVAX	Flunisolide	Seasonal or perennial rhinitis
Nicotrol Nasal Spray	Pfizer	Nicotine 0.5 mg	Smoking Cessation
Neo-Syneprhine	Bayer	Oxymetazoline hydrochloride 0.05%	Decongestion
Rhinocort Aqua Nasal Spray	Astra-Zeneca	Budesonide 32mcg	Seasonal or Allergic Rhinitis
Stadol Nasal Spray	Bristol-Myers Squibb	Butorphanol tartrate, 1 mg	Pain Relief, Migraines
Stimate Nasal Spray	Aventis	Desmopressin Acetate 1.5 mg/mL	Hemophilia A or von Willebrand disease
Synarel Nasal Solution	Searle	Nafarelin acetate 2 mg/mL	Endometriosis
Tyzine	Bradley Pharmaceuticals	Tetrahydrozoline hydrochloride	Decongestion

Table 24-4. Examples of Commercial Otic Preparations

Product Name	Manufacturer	Active Ingredient	Indication
Americaine-Otic	Celltech	Benzocaine	Local Anesthetics
Cerumenex Ear Drops	Purdue	Triethanolamine polypeptide oleate-condensate	Removal of earwax
Chloromycetin Otic	Pfizer	Chloramphenicol	Antiinfective
Cipro HC Otic	Alcon	Ciprofloxacin hydrochloride and hydrocortisone	Acute otitis externa
Cortisporin	GlaxoSmithKline	Neomycin and Polymyxin B Sulfates and Hydrocortisone	Antibacterial and anti-inflammatory
Debrox® Drops	GlaxoSmithKline	Carbamide peroxide	Removal of earwax
Floxin Otic	Daiichi	Ofloxacin	Antiinfective
Tympagesic	Savage	Antipyrine, Benzocaine, and Phenylephrine Hydrochloride	Topical anesthetic

drops do not run out of the ear. Otic preparations are dispensed in a container that permits the administration of drops.

Irrigation Solutions

Irrigation solutions are sterile, non-pyrogenic solutions used to wash or bathe surgical incisions, wounds, or body tissues. Because they come in contact with exposed tissue, they must meet stringent USP requirements for sterility, total solids, and bacterial endotoxins. These products may be prepared by dissolving the active ingredient in Water for Injection. They are packaged in single-dose containers, preferably Type I or Type II glass, or suitable plastic containers, and then sterilized. A number of irrigations are described in the USP, including Acetic Acid Irrigation for bladder irrigation, Dimethyl Sulfoxide Irrigation for relief of internal cystitis, Glycine Irrigation for transurethral prostatic resection, Ringer's Irrigation for general irrigation, Neomycin and Polymyxin B Sulfates Solution for Irrigation for infection, and Sodium Chloride Irrigation for washing wounds.

Extemporaneous formulations frequently are prepared using an isotonic solution of sodium chloride as the solvent. For example, cefazolin or gentamicin in 0.9% sodium chloride is used as anti-infective irrigations,⁴⁵ and 5-fluorouracil in 0.9% sodium chloride is employed for bladder irrigation.⁴⁶ Alum, either potassium or ammonium, in either sterile water or 0.9% sodium chloride for irrigation, has been used for bladder hemorrhage. Amphotericin in sterile water has been used for the treatment of localized infections on the dermis, the bladder, and the urinary tract.⁴⁷ All the extemporaneous preparations should meet the general requirements noted for USP irrigations.

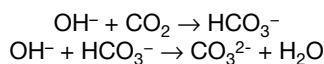
PREPARATION OF SOLUTIONS

The method of preparation for many solutions is given in the compendia. These procedures fall into three main categories: simple solutions, solution by chemical reaction, and solution by extraction.

Simple Solutions are prepared by dissolving the solute in most of the solvent, mixing until dissolved, then adding sufficient solvent to bring the solution up to the proper volume. The solvent may contain other ingredients that stabilize or solubilize the active ingredient. Calcium Hydroxide Topical Solution USP (Lime Water), Sodium Phosphate Oral Solution USP, and Strong Iodine Solution USP are examples.

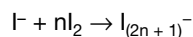
Calcium Hydroxide Topical Solution USP contains, in each 100 mL, not less than 140 mg of $\text{Ca}(\text{OH})_2$. The solution is prepared by agitating vigorously 3 g of calcium hydroxide with 1000 mL of cool, purified water. Excess calcium hydroxide is allowed to settle out, and the clear, supernatant liquid dispensed. An increase in solvent temperature implies an increase in solute solubility. This rule does not apply, however, to the solubility of calcium hydroxide in water, which decreases with increasing temperature. The official solution is prepared at 25°C.

Solutions containing hydroxides react with the carbon dioxide in the atmosphere.

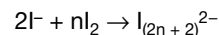


Calcium Hydroxide Topical Solution, therefore, should be preserved in well-filled, tight containers, at a temperature not exceeding 25°C.

Strong Iodine Solution USP contains, in each 100 mL, 4.5–5.5 g of iodine and 9.5–10.5 g of potassium iodide. It is prepared by dissolving 50 g of iodine in 100 mL of purified water containing 100 g of potassium iodide. Sufficient purified water is then added to make 1000 mL of solution. One gram of iodine dissolves in 2950 mL of water. However, solutions of iodides dissolve large quantities of iodine. Strong Iodine Solution is, therefore, a solution of polyiodides in excess iodide.



Doubly charged anions may be found also.

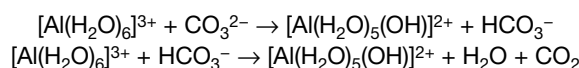


Strong Iodine Solution is used in the treatment of iodide deficiency disorders, such as endemic goiter.

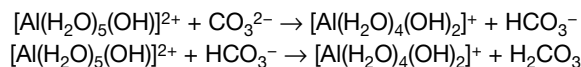
Several antibiotics (e.g., cloxacillin sodium, nafcillin sodium, and vancomycin), because they are relatively unstable in aqueous solution, are prepared by manufacturers as dry powders or granules in combination with suitable buffers, colors, diluents, dispersants, flavors, and/or preservatives. These preparations, Cloxacillin Sodium for Oral Solution, Nafcillin for Oral Solution, and Vancomycin Hydrochloride for Oral Solution, meet the requirements of the USP. Immediately prior to dispensing to the patient, the pharmacist adds the appropriate amount of water. The products are stable for up to 14 days when refrigerated.⁴⁸ This period provides sufficient time for the patient to complete the administration of all the medication.

Solutions by chemical reaction are prepared by reacting two or more solutes with each other in a suitable solvent. An example is Aluminum Subacetate Topical Solution USP. Aluminum sulfate (145 g) is dissolved in 600 mL of cold water. The solution is filtered, and precipitated calcium carbonate (70 g) is added, in several portions, with constant stirring. Acetic acid (160 mL) is added slowly, and the mixture is set aside for 24 hours. The product is filtered, and the magma on the Buchner filter is washed with cold water, until the total filtrate measures 1,000 mL.

The solution contains pentaquohydroxo- and tetraquodihydroxoaluminum(III) acetates and sulfates dissolved in an aqueous medium saturated with calcium sulfate. The solution contains a small amount of acetic acid. It may be stabilized by the addition of not more than 0.9% boric acid. The reactions involved in the preparation of the solution are subsequently given. The hexaquo aluminum cations are first converted to the nonirritating $[\text{Al}(\text{H}_2\text{O})_5(\text{OH})]^{2+}$ and $[\text{Al}(\text{H}_2\text{O})_4(\text{OH})_2]^+$ cations.



As the concentration of the hexaquo cations decreases, secondary reactions involving carbonate and bicarbonate occur.



The pH of the solution now favors the precipitation of dissolved calcium ions as the insoluble sulfate. Acetic acid is now added. The bicarbonate formed in the final stages of the procedure is removed as carbon dioxide.

Aluminum Subacetate Topical Solution is used in the preparation of Aluminum Acetate Topical Solution USP (Burow's Solution). The latter solution contains 15 mL of glacial acetic acid, 545 mL of Aluminum Subacetate Topical Solution, and sufficient water to make 1000 mL. It is defined as a solution of aluminum acetate in approximately 5%, by weight, of acetic acid in water. It may be stabilized by the addition of not more than 0.6% boric acid.

Often, drugs or pharmaceutical necessities of vegetable or animal origin are extracted with water or with water containing other substances. Preparations of this type may be classified as solutions but, more often, are classified as extracts and are described at the end of this chapter.

SWEET AND OTHER VISCID AQUEOUS SOLUTIONS

Solutions that are sweet or viscid include syrups, honeys, mucilages, and jellies. All of these are viscous liquids or semisolids. The basic sweet or viscid substances giving body to these preparations are sugars, polyols, and/or polysaccharides.

Syrups

Syrups are concentrated, viscous, aqueous solutions of sugar or a sugar substitute with or without flavors and medical

Table 24-5. Examples of Commercial Medicated Syrups

Product Name	Manufacturer	Active Ingredient & Dose	Indication
Chlor-Trimeton	Schering-Plough	2 mg chlorpheniramine maleate / 5 mL	Allergic rhinitis
Children's Benadryl	Pfizer	12.5 mg diphenhydramine HCl / 5 mL	Allergic rhinitis
Demerol Syrup	Sanofi	50 mg meperidine HCl / 5 mL	Narcotic analgesic
Ditropan Syrup	Ortho-McNeil	5 mg Oxybutynin chloride / 5 mL	Overactive bladder
Dramamine	Pfizer	12.5 mg dimenhydrinate / 5 mL	Antiemetic
Phenergan Syrup	Wyeth-Ayerst	25 mg promethazine HCl / 5 mL	Antiemetic
Symmetrel Syrup	Endo	50 mg amantadine HCl / 5 mL	Antiviral

substances. When Purified Water alone is used in making the solution of sucrose, the preparation is known as “syrup,” or “simple syrup,” if the sucrose concentration is 85%. Syrups are also used to apply sugar coatings to tablets, particularly those with disagreeable aromas or acrid taste. In addition to sucrose, certain other polyols, such as glycerin or sorbitol, may be added to retard crystallization of sucrose or to increase the solubility of added ingredients. Alcohol often is included as a preservative and also as a solvent for flavors; further resistance to microbial attack can be enhanced by incorporating antimicrobial agents. When the aqueous preparation contains some added medicinal substance, the syrup is called a medicated syrup. Flavored syrups are not medicated, but rather contain various aromatic or pleasantly flavored substances and are intended to be used as a vehicle or flavor for prescriptions, such as Acacia, Cherry, Cocoa, Orange, and Raspberry USP.

Flavored syrups offer unusual opportunities as vehicles in extemporaneous compounding and are accepted readily by both children and adults. Because they contain no, or very little, alcohol, they are vehicles of choice for many of the drugs prescribed by pediatricians. Their lack of alcohol makes them superior solvents for water soluble substances. However, sucrose based medicines continuously administered to children apparently cause an increase in dental caries and gingivitis; consequently, alternate formulations of the drug either unsweetened or sweetened with non-cariogenic substances should be considered. A knowledge of the sugar content of liquid medicines is useful for patients who are on a restricted calorie intake; a list has been prepared by Greenwood.⁴⁹

As noted, sucrose based syrups may be substituted in whole or in part by other agents in the preparation of medicated syrups. A solution of sorbitol, or a mixture of polyols, such as sorbitol and glycerin, is commonly used. Sorbitol Solution, USP, which contains 64% by weight of the polyhydric alcohol sorbitol, is often used in sugar free and children's preparations. However, reports of adverse reactions to sorbitol are largely due to its action as an osmotic laxative when ingested orally.⁵⁰ Ingestion of large quantities of sorbitol (>20 g/day in adults) should, therefore, be avoided.

Syrups possess remarkable taste masking properties for bitter or saline drugs. Syrups flavored with Glycyrrhizin, a triterpene glycoside extracted from licorice root, has been recommended for disguising the salty taste of bromides, iodides, and chlorides.⁵¹ This has been attributed to its colloidal character and its double sweetness—the immediate sweetness of the sugar and the lingering sweetness of the glycyrrhizin. This syrup is also of value in masking bitterness in preparations containing the B complex vitamins. Acacia Syrup USP is of particular value, as a vehicle for masking the disagreeable taste of many medications, due to its colloidal character. Raspberry Syrup, USP, is one of the most efficient flavoring agents and is especially useful in masking the taste of bitter drugs. Many factors, however, enter into the choice of a suitable flavoring agent. Literature reports are often contradictory, and there appears to be no substitute for the taste panel, when developing new formulations.⁵²

It is important that the concentration of sucrose approach, but not quite reach, the saturation point. In dilute solutions,

sucrose provides an excellent nutrient for molds, yeasts, and other micro-organisms. In concentrations of 65% by weight or more, the solution will retard the growth of such micro-organisms. However, a saturated solution may lead to crystallization of a part of the sucrose under conditions of changing temperature. Several commercial medicated syrups are available for a variety of indications (Table 24-5).

Preparation of Syrups

Syrups are prepared using one of four techniques: solution with heat, solution by agitation, addition of sucrose to a liquid medication or flavored liquid, and percolation. The method of choice depends on the physical and chemical characteristics of the substances entering into the preparation. In many cases, syrups may be successfully prepared by more than one method, and the selection may simply be a matter of preference on the part of the pharmacist. Many of the compendial syrups do not have a designated method for preparation, because most are commercially available and are not prepared extemporaneously by the pharmacist.

Solution with Heat—This method is a suitable preparation method, if the constituents are not volatile or degraded by heat, and when it is desirable to make the syrup rapidly. Purified water is heated to 80–85°C, and then removed from its heat source. Sucrose is added with vigorous agitation. Then, other required heat-stable components are added to the hot syrup, the mixture is allowed to cool, and its volume is adjusted to the proper level by the addition of purified water. In instances in which heat labile agents or volatile substances, such as flavors and alcohol, are added, they are incorporated into the syrup after cooling to room temperature.

When heat is used in the preparation of syrups, there is almost certain to be an inversion of a slight portion of the sucrose. Sucrose, a disaccharide, may be hydrolyzed into monosaccharides, dextrose (glucose), and fructose (levulose). This hydrolytic reaction is referred to as “inversion,” and the combination of the two monosaccharide products is “invert sugar.” Sucrose solutions are dextrorotary, but, as hydrolysis proceeds, the optical rotation decreases and becomes negative when the reaction is complete. The rate of inversion is increased greatly by the presence of acids; the hydrogen ion acts as a catalyst in this hydrolytic reaction. Invert sugar is more readily fermentable than sucrose and tends to be darker in color. Nevertheless, its two reducing sugars are of value in retarding the oxidation of other substances.

The fructose formed during inversion is sweeter than sucrose, thus, the resulting syrup is sweeter than the original syrup. The relative sweetness of fructose, sucrose, and dextrose is in the ratio of 173:100:74. Thus, invert sugar is $1/100 (173 + 74)1/2 = 1.23$ times as sweet as sucrose. Fructose is responsible for the darkening of syrup, as it is amber in color. If the syrup is significantly overheated, sucrose caramelizes and becomes darker. Excessive heating of syrups is undesirable, because inversion occurs with an increased tendency to ferment. Syrups cannot be sterilized in an autoclave without some caramelization.

Agitation without Heat—This method is used in cases in which heat would cause degradation or volatilize formulation constituents. On a small scale, sucrose and other formulation

ingredients may be dissolved in purified water by placing the ingredients in a vessel of greater capacity than the volume of syrup to be prepared, allowing intense agitation without spillage. This process is more time consuming than Solution with Heat, but the product has greater stability. Large glass lined and stainless steel tanks equipped with mechanical mixers are employed in the large scale preparation of syrups.

Often, simple syrup or some other non-medicated syrup, rather than sucrose, is employed as the sweetening agent and vehicle. When solid agents are added to a syrup, it is best to dissolve them in a minimal amount of purified water and then incorporate the resulting solution into the syrup. When solid substances are added directly to syrups, they dissolve slowly, because the viscous nature of the syrup does not permit the solid substance to distribute readily.

This method and that previously described are used for the preparation of a wide variety of preparations that are described popularly as syrups. Most cough syrups, for example, contain sucrose and one or more active ingredients. Many other active ingredients (e.g., ephedrine sulfate, dicyclomine hydrochloride, chloral hydrate, or chlorpromazine hydrochloride) are marketed as syrups. Like cough syrups, these preparations are flavored, colored, and recommended in those instances in which the patient cannot swallow the solid dosage form.

Addition of Sucrose to a Liquid Medication or Flavored Liquid—This method is often used with fluidextracts or tinctures. Syrups made in this way develop precipitates, because alcohol is often an ingredient of the liquids used, and the resinous and oily substances solubilized by the alcohol precipitate when water is added. A modification of this process entails mixing the fluidextract or tincture with the water, allowing the mixture to stand to permit the separation of insoluble constituents, filtering and then dissolving the sucrose in the filtrate. It is obvious that this procedure is not permissible when the precipitated ingredients are the valuable medicinal agents.

Percolation—In the percolation method, either purified water or the source of the medicinal component is passed slowly through a bed of crystalline sucrose, thus, dissolving it and forming a syrup. This latter method really involves two separate procedures: first, the preparation of the extractive of the drug, and then the preparation of the syrup. To be successful in using this process, technique is critical: 1) the percolator used should be cylindrical or semi-cylindrical and cone-shaped as it nears the lower orifice; and 2) a coarse granular sugar must be used, otherwise it will coalesce into a compact mass, which the liquid cannot permeate. The percolation method is applied on a commercial scale for the making of compendial syrups, as well as those for confectionary use.

Ipecac syrup is prepared by percolation by adding glycerin and syrup to an extractive of powdered ipecac obtained by percolation. The drug ipecac consists of the dried rhizome and roots of *Cephaelis ipecacuanha* and contains the medicinally active alkaloids: emetine, cephaeline, and psychotrine. These alkaloids are extracted from the powdered ipecac by percolation with a hydroalcoholic solvent. The syrup is categorized as an emetic with a usual dose of 15 mL. This amount of syrup is commonly used in the management of poisoning in children, when the evacuation of stomach contents is desirable. About 80% of children given this dose will vomit within 30 minutes. Bulimics have used ipecac to bring on attacks of vomiting in an attempt to lose more weight.⁵³ Pharmacists must be aware of this abuse and warn these individuals, because one of the active ingredients is emetine. With chronic abuse of the syrup, emetine builds up toxic levels within body tissues and, in 3–4 months, can do irreversible damage to heart muscles, resulting in symptoms mimicking a heart attack.

Syrups should be made in quantities that can be consumed within a few months, except in those cases where special facilities can be employed for their preservation; a low temperature is the best method. Concentration without super-saturation is also a condition favorable to preservation. The USP states that syrups

may contain preservatives. Glycerin, methylparaben, benzoic acid, and sodium benzoate may be used to prevent bacterial and mold growth. Combinations of alkyl esters of *p*-hydroxybenzoic acid are effective inhibitors of yeasts that have been implicated in the contamination of commercial syrups. Syrups should be preserved in well-dried bottles, preferably those that have been sterilized. These bottles should not hold more than is likely to be required during 4–6 weeks and should be filled completely, carefully closed, and stored in a cool, dark place.

Some examples of syrup formulations are noted:

Ferrous Sulfate Syrup

FERROUS SULFATE	40.0 G
CITRIC ACID	2.1 G
PEPPERMINT SPIRIT	2 ML
SUCROSE	825 G
PURIFIED WATER TO MAKE	1000.0 ML

Dissolve the Ferrous Sulfate, the Citric Acid, the Peppermint Spirit, and 200 g of the Sucrose in 450 mL of Purified Water, and filter the solution until clear. Dissolve the remainder of the Sucrose in the clear filtrate, and add Purified Water to make 1000 mL. Mix, and filter, if necessary, through a pledget of cotton.

Amantadine Hydrochloride Syrup

AMANTADINE HYDROCHLORIDE	10.0 G
CITRIC ACID	2.1 G
ARTIFICIAL RASPBERRY FLAVOR	2 ML
METHYL PARABEN	2 G
PROPYL PARABEN	0.5 G
SORBITOL SOLUTION TO MAKE	1000 ML

Dissolve the amantadine hydrochloride, the Citric Acid, flavor and preservatives in the sorbitol solution.

Syrups are useful for preparing liquid oral dosage forms from not only the pure drug, as described, but also injections, capsules, or tablets, if the pure drug is not readily available. If the drug and all the excipients in the preparation, such as injectables or capsules, are water soluble, a solution should result, if a syrup is prepared. Conversely, if the preparation used contains water-insoluble ingredients, as is the case with tablets and some capsules, a suspension will be formed. Several of these preparations have been described in the literature, in regard to their formulation, stability, and bioavailability. Some drugs that have been prepared from either the pure drug or an injectable form include midazolam, atropine, aminocaproic acid, terbutaline, procainamide, chloroquine, propranolol, and citrated caffeine.^{54,55} If the appropriate salt of the drug is used, a solution will result.

When tablets are introduced to a syrup formulation, a suspension is often formed, because there are water insoluble ingredients used in tablet preparations. Examples of medicated syrups prepared from tablets are clonidine hydrochloride, cefuroxime axetil, famotidine, terbutaline sulfate, spironolactone, ranitidine, and rifampin.^{56,57} The resulting suspensions should have a uniform distribution of particles, so a consistent dose is obtained. If the materials are not distributed uniformly, more appropriate suspending formulations should be considered. If pharmaceutical preparations contain a liquid insoluble in water, such as valproic acid or simethicone, to be incorporated into syrups, an emulsion will form, and it will be difficult to prepare a uniform product.

Honeys

Honeys are thick liquid preparations somewhat allied to the syrups, differing in that honey, instead of syrup, is used as a base. They are unimportant as a class of preparations today,

but, at one time, before sugar was available and honey was the most common sweetening agent, they were used widely. Honey and sugar pastes are used to a small extent and have been discussed in the pharmaceutical literature for topical application for the treatment of certain types of ulcers and abscesses.⁵⁸

Mucilages

Mucilages are thick, viscid, adhesive liquids, produced by dispersing gum in water or by extracting the mucilaginous principles from vegetable substances with water. The mucilages are prone to decomposition, showing appreciable decrease in viscosity on storage; they should never be made in quantities larger than can be used immediately, unless a preservative is added. Mucilages are used primarily to aid in suspending insoluble substances in liquids; their colloidal character and viscosity help prevent immediate sedimentation. Examples include sulfur in lotions, resin in mixtures, and oils in emulsions. Both tragacanth and acacia are either partially or completely insoluble in alcohol. Tragacanth is precipitated from solution by alcohol, but acacia, conversely, is soluble in diluted alcoholic solutions. A 60% solution of acacia may be prepared with 20% alcohol, and a 4% solution of acacia may be prepared even with 50% alcohol.

Recent research on mucilages includes the preparation of mucilage from plantain and the identification of its sugars, the preparation and suspending properties of cocoa gum, the preparation of glycerin ointments using flaxseed mucilage, and the consideration of various gums and mucilages obtained from several Indian plants for pharmaceutical purposes.

Several synthetic mucilage-like substances, such as polyvinyl alcohol, methylcellulose, carboxymethylcellulose, and related substances, are used at the appropriate concentration as mucilage substitutes and emulsifying and suspending agents. Methylcellulose is used widely as a bulk laxative, because it absorbs water and swells to a hydrogel in the intestine, in much the same manner as psyllium or karaya gum. Methylcellulose Oral Solution, USP, is a flavored solution of the agent. It may be prepared by slowly adding the methylcellulose to about one-third the amount of boiling water, with stirring, until it is thoroughly wetted. Cold water should then be added, and the wetted material allowed to dissolve while stirring. The viscosity of the solution will depend upon the concentration and the specifications of the methylcellulose. The synthetic gums are non-glycogenetic and may be used in the preparation of diabetic syrups. Sodium carboxymethyl cellulose of a medium grade in water (0.25 – 1%) is suitable for preparing a suspending vehicle. Several formulas for such syrups, based on sodium carboxymethylcellulose, have been proposed.

Uniformly smooth mucilages are sometimes difficult to prepare, due to the uneven wetting of the gums. It is best to use fine gum particles and disperse them with agitation in a small quantity of 95% alcohol or in cold water (except for methylcellulose). The appropriate amount of water can then be added with constant stirring. A review of the chemistry and properties of acacia and other gums has been prepared.⁵⁹

Jellies

Jellies are a class of gels in which the structural coherent matrix contains a high portion of liquid, usually water. They are similar to mucilages, in that they may be prepared from similar gums, but they differ from the latter in having a jelly-like consistency. A whole gum of the best quality, rather than a powdered gum, is desirable to obtain a clear preparation of uniform consistency. Although the specific thickening agent in the USP jellies is not indicated, reference is usually made in the monograph to a water-soluble, sterile, viscous base. These preparations may also be formulated with water from acacia, chondrus, gelatin, carboxymethylcellulose, hydroxyethylcellulose, and similar substances.

Jellies are used as lubricants for surgical gloves, catheters, and rectal thermometers. Lidocaine Hydrochloride Jelly, USP,

is used as a topical anesthetic. Therapeutic vaginal jellies are available, and certain jelly-like preparations are used for contraceptive purposes, which often contain surface-active agents to enhance the spermaticidal properties of the jelly. Aromatics, such as methyl salicylate and eucalyptol, are often added to give the preparation a desirable odor.

Jellies are prone to microbial contamination and, therefore, contain preservatives; for example, methyl *p*-hydroxybenzoate is used as a preservative in a base for medicated jellies. One base contains sodium alginate, glycerin, calcium gluconate, and water. The calcium ions cause a cross-linking with sodium alginate to form a gel of firmer consistency.

NONAQUEOUS SOLUTIONS

It is difficult to fairly evaluate the importance of nonaqueous solvents in pharmaceutical processes. That they are important in the manufacture of pharmaceuticals is an understatement. However, pharmaceutical preparations and, in particular, those intended for internal use rarely contain more than minor quantities of the organic solvents common to the manufacturing or analytical operation. Commercial products intended for internal use may contain solvents, such as ethanol, glycerin, propylene glycol, certain oils, and liquid paraffin. Preparations intended for external use may contain other solvents in addition to those mentioned, namely isopropyl alcohol, polyethylene glycols, various ethers, and certain esters.

Although the lines between aqueous and nonaqueous preparations tend to blur in those cases in which the solvent is water soluble, it is possible to categorize a number of products as nonaqueous. This section is, therefore, devoted to groups of nonaqueous solutions: thealecoholic or hydroalcoholic solutions (e.g., elixirs and spirits), ethereal solutions (e.g., collodions), glycerin solutions (e.g., glycerins), oleaginous solutions (e.g., liniments, oleovitamins, and toothache drops), inhalations, and inhalants.

Although this list is limited, a wide variety of solvents are used in various pharmaceutical preparations. Solvents, such as glycerol formal, dimethylacetamide, and glycerol dimethylketal, have been suggested for some products produced by the industry. However, the toxicity of many of these solvents is not well established, and, for this reason, careful clinical studies should be carried out on the formulated product, before it is released to the marketplace. It is essential that the toxicity of solvents be tested appropriately and approved to avoid problems; for example, lives were lost in 1937, when diethylene glycol was used in an elixir of sulfanilamide. The result of this tragedy was the 1938 Federal Food, Drug, and Cosmetic Act, which required that products be tested for both safety and effectiveness.

COLLODIONS

Collodions are liquid preparations containing pyroxylin, a partially nitrated cellulose, in a mixture of ethyl ether and ethanol. They are applied to the skin by means of a soft brush or other suitable applicator and, when the ether and ethanol have evaporated, leave a film of pyroxylin on the surface. Salicylic Acid Collodion, USP, contains 10% w/v of salicylic acid in Flexible Collodion, USP, and is used as a keratolytic agent in the treatment of corns and warts. Collodion, USP, and Flexible Collodion, USP, are water-repellent protectives for minor cuts, scratches, and chigger bites. Collodion is made flexible by the addition of castor oil and camphor. Collodion has been used to reduce or eliminate the side effects of fluorouracil treatment of solar keratoses.⁶⁰ Vehicles other than Flexible Collodion, such as a polyacrylic base, have been used to incorporate salicylic acid for the treatment of warts with less irritation.

ELIXIRS

Elixirs are clear, pleasantly flavored, sweetened hydroalcoholic liquids intended for oral use. The main ingredients in elixirs are ethanol and water but glycerin, sorbitol, propylene glycol, flavoring agents, preservatives, and syrups are often used in the

preparation of the final product. The solvents are often used to increase the solubility of the drug substance in the dosage form. Elixirs are more fluid than syrups, due to the use of less viscous ingredients, such as alcohol, and the minimal use of viscosity-improving agents, such as sucrose. They are used as flavors and vehicles, such as Aromatic Elixir, USP, for drug substances; when such substances are incorporated into the specified solvents, they are classified as medicated elixirs, such as Dexamethasone Elixir, USP, and Phenobarbital Elixir, USP.

The distinction between some of the medicated syrups and elixirs is not always clear. For example, Ephedrine Sulfate Syrup, USP, contains between 20 and 40 mL of alcohol in 1000 mL of product. Definitions are sometimes inconsistent and, in some instances, not too important with respect to the naming of the articles of commerce. To be designated an elixir, however, the solution must contain alcohol. The alcoholic content will vary greatly, from elixirs containing only a small quantity to those that contain a considerable portion as a necessary aid to solubility. For example, Aromatic Elixir, USP, contains 21–23% alcohol; Compound Benzaldehyde Elixir, USP, conversely, contains 3–5%.

Elixirs may also contain glycerin and syrup. These may be added to increase the solubility of the medicinal agent, for sweetening purposes, or to decrease the pharmacological effects of the alcohol. Some elixirs contain propylene glycol. Claims have been made for this solvent as a satisfactory substitute for both glycerin and alcohol.

Although alcohol is an excellent solvent for some drugs, it does accentuate the saline taste of bromides and similar salts. It is often desirable, therefore, to substitute some other solvent that is more effective in masking such tastes for part of the alcohol in the formula. If taste is a consideration, the formulator is more prone to use a syrup, rather than a hydroalcoholic vehicle.

Because only relatively small quantities of ingredients have to be dissolved, elixirs are more readily prepared and manufactured than syrups, which frequently contain considerable amounts of sugar. An elixir may contain both water and alcohol soluble ingredients. If such is the case, the following procedure is indicated:

Dissolve the water soluble ingredients in part of the water. Add and solubilize the sucrose in the aqueous solution. Prepare an alcoholic solution containing the other ingredients. Add the aqueous phase to the alcoholic solution, filter, and make to volume with water.

Sucrose increases viscosity and decreases the solubilizing properties of water and must be added after the primary solution has been effected. A high alcoholic content is maintained during preparation, by adding the aqueous phase to the alcoholic solution. Elixirs should always be brilliantly clear. They may be strained or filtered and, if necessary, subjected to the clarifying action of purified talc or siliceous earth.

Elixirs, and many other liquid preparations intended for internal use, such as the diabetic syrups thickened with sodium carboxymethylcellulose or similar substances, contain saccharin, aspartame, acesulfame potassium, and other sweeteners. Cyclamates and saccharin have been banned in some countries as ingredients in manufactured products. Much research has been done to find a safe synthetic substitute for sucrose.

Research concerning the preparation of a dry elixir has been conducted by Kim and coworkers.⁶¹ Dry Elixirs containing a non-steroidal anti-inflammatory drug and ethanol were encapsulated in a dextrin. The dissolution rate constant of the drug from the microcapsules increased considerably, compared to the drug alone, possibly due to the cosolvent ethanol. It is suggested that this type of dosage form may be useful to improve the solubility, dissolution rate, and bioavailability of the drug.

Because elixirs contain alcohol, incompatibilities of this solvent are an important consideration during formulation. Alcohol precipitates tragacanth, acacia, and agar from aqueous solutions. Similarly, it will precipitate many inorganic salts from similar solutions. The implication is that such substances should be absent from the aqueous phase or present in such concentrations that there is no danger of precipitation on standing.

If an aqueous solution is added to an elixir, a partial precipitation of alcohol soluble ingredients may occur. This is due to the reduced alcoholic content of the final preparation. However, the alcoholic content of the mixture is not sufficiently decreased to cause separation. As vehicles for tinctures and fluidextracts, the elixirs cause a separation of extractive matter from these products, due to a reduction of the alcoholic content. Many of the incompatibilities between elixirs, and the substances combined with them, are due to the chemical characteristics of the elixir per se, or of the ingredients in the final preparation. Thus, certain elixirs are acid in reaction, whereas others may be alkaline and will, therefore, behave accordingly.

Some example formulations of medicated elixirs are as follows:

Phenobarbital Elixir

PHENOBARBITAL	4.00 G
PROPYLENE GLYCOL	50 ML
ALCOHOL	200 ML
SORBITOL SOLUTION	600 ML
SACCHARIN SODIUM	5.0 G
FLAVOR	Q.S.
PURIFIED WATER, TO MAKE	1000 ML

Theophylline Elixir

THEOPHYLLINE	5.3 G
CITRIC ACID	10.0 G
SYRUP	132.0 ML
GLYCERIN	50.0 ML
SORBITOL SOLUTION	324.0 ML
ALCOHOL	200.00 ML
FLAVOR	Q.S.
PURIFIED WATER, TO MAKE	1000.0 ML

GLYCERINS

Glycerins or glycerites are solutions or mixtures of medicinal substances in not less than 50% by weight of glycerin. Most of the glycerins are extremely viscous, and some are of a jelly-like consistency. Few of them are used extensively. Glycerin is a valuable pharmaceutical solvent forming permanent and concentrated solutions not otherwise obtainable. Glycerin is used as the sole solvent for the preparation of Antipyrine and Benzocaine Otic Solution, USP. Glycerins are hygroscopic and should be stored in tightly closed containers.

INHALATIONS AND INHALANTS

Inhalation preparations are so used or designed that the drug is carried into the respiratory tree of the patient. The vapor or mist reaches the affected area and gives prompt relief from the symptoms of bronchial and nasal congestion. The USP defines Inhalations in the following way:

Inhalations are drugs or solutions or suspensions of one or more drug substances administered to the nasal or oral respiratory route for local or systemic effect. Solutions of drug substances in sterile water for inhalation or in sodium chloride inhalation solution may be nebulized by the use of inert gases. Nebulizers are suitable for the administration of inhalation solutions only if they give droplets sufficiently fine and uniform in size so that the mist reaches the bronchioles. Nebulized solutions may be breathed directly from the nebulizer, or the nebulizer may be attached to a plastic face mask, tent or intermittent positive pressure breathing (IPPB) machine.

Another group of products, also known as metered-dose inhalers (MDIs) are propellant-driven drug suspensions or solutions

in liquefied gas propellant (chlorofluorocarbons and hydrofluoroalkanes) with or without a cosolvent and are intended for delivering metered doses of the drug to the respiratory tract. An MDI contains multiple doses, often exceeding several hundred. The most common single dose volumes delivered are from 25 to 100 μL (also expressed as mg) per actuation. Examples of MDIs containing drug solutions are Epinephrine Inhalation Aerosol, USP, and Isoproterenol Hydrochloride and Phenylephrine Bitartrate Inhalation Aerosol, respectively. Both the solubility and stability of the drug in the propellant mixture must be investigated during formulation development. Ethanol is commonly used as a cosolvent for hydrofluoroalkane propellants, and was reported to significantly increase the solubility of steroids.⁶²

As stated in the USP, particle size is of major importance in the administration of this type of preparation. The various mechanical devices used in conjunction with inhalations are described in Chapter 33 (Aerosols). It has been reported that the optimum particle size for penetration into the pulmonary cavity is of the order of 0.5–7.0 μm .⁶³ Fine mists are produced by pressurized aerosols and, hence, possess basic advantages over the older nebulizers; in addition, metered aerosols deliver more uniform doses. A number of inhalations are described in the USP.

The USP defines “inhalants” as follows:

A special class of inhalations termed “inhalants” consists of drugs or combinations of drugs that, by virtue of their high vapor pressure, can be carried by an air current into the nasal passage where they exert their effect. The container from which the inhalant is administered is known as an inhaler.

Amyl nitrate, USP, and Propylhexedrine Inhalant, USP, are two examples. Amyl nitrite is a clear, yellowish, volatile liquid that acts as a vasodilator when inhaled. The drug is prepared in sealed glass vials covered with a protective gauze cloth. Upon use, the glass vial is broken in the fingertips, and the cloth soaks up the liquid, which is then inhaled. The vials contain 0.3 mL of the drug substance. The effects of the drug are rapid and are used in the treatment of anginal pain.

Propylhexedrine is the active ingredient in the widely used Benzedrex and Dristan Inhalers. Propylhexedrine is a liquid, vasoconstrictor agent that volatilizes slowly at room temperature. This quality enables it to be effectively used as an inhalant. The official inhalant consists of cylindrical rolls of suitable fibrous material impregnated with propylhexedrine, aromatized to mask its amine like odor, and contained in a suitable inhaler. The vapor of the drug is inhaled into the nostrils when needed to relieve nasal congestion due to colds and hay fever. It may also be employed to relieve ear block and the pressure pain in air travelers. Each plastic tube of the commercial product contains 250 mg of propylhexedrine with aromatics. The containers should be tightly closed after each opening to prevent loss of the drug vapors.

LINIMENTS

Liniments are alcoholic or oil based solutions or emulsions containing therapeutic agents intended for external application. These preparations may be liquids or semisolids that are rubbed onto the affected area; due to this, they were once called “embrocations.”

Liniments are applied with friction and rubbing of the skin, the oil or soap base providing for ease of application and massage. Alcoholic liniments are used for their rubefacient, counterirritant, mildly astringent, and penetrating effects. Such liniments penetrate the skin more readily than do those with an oil base. Oily liniments, therefore, are milder in their action but are more useful when massage is required. Depending on their ingredients, such liniments may function solely as protective coatings. Liniments should not be applied to skin that is bruised or broken.

Other liniments contain antipruritics, astringents, emollients, or analgesics and are classified on the basis of their active ingredient. Dermatologists prescribe products of this type, but only

those containing the rubefacients are advertised extensively and used by consumers for treating minor muscular aches and pains. It is essential that these applications be marked clearly “For External Use Only.” Liniments containing a capsaicin are being investigated for treatment of pruritus.⁶⁴

OLEOVITAMINS

Oleovitamins are fish liver oils diluted with edible vegetable oil or solutions of the indicated vitamins or vitamin concentrates (usually vitamin A and D) in fish-liver oil. The definition is broad enough to include a wide variety of marketed products.

In Oleovitamin A and D, USP, vitamin D may be present as ergocalciferol or cholecalciferol obtained by the activation of ergosterol or 7-dehydrocholesterol, or may be obtained from natural sources. Synthetic vitamin A, or a concentrate, may be used to prepare oleovitamin A. The starting material for the concentrate is fish-liver oil, the active ingredient being isolated by molecular distillation or by a saponification and extraction procedure. These vitamins are unstable in the presence of rancid oils; therefore, these preparations should be stored in small, tight containers, preferably under vacuum or under an atmosphere of an inert gas, protected from light and air.

SPIRITS

Spirits, sometimes known as essences, are alcoholic or hydroalcoholic solutions of volatile substances. Like the aromatic waters, the active ingredient in the spirit may be a solid, liquid, or gas. The genealogical tree for this class of preparations begins with a distinguished pair of products, Brandy (*Spiritus Vini Vitis*) and Whisky (*Spiritus Frumenti*), and ends with a wide variety of products that comply with the definition given. Physicians have debated the therapeutic value of the former products and these are no longer compendial.

The alcohol concentration of spirits is rather high, over 60%. Due to the greater solubility of aromatic or volatile substances in alcohol than in water, spirits can contain a greater concentration of these materials than the corresponding aromatic waters. When mixed with water or with an aqueous preparation, the volatile substances present in spirits separate from solution and form a milky preparation. Salts may be precipitated from their aqueous solutions by the addition of spirits, due to their lesser solubility in alcoholic liquids. Some spirits show incompatibilities characteristic of the ingredients they contain. For example, Aromatic Ammonia Spirit cannot be mixed with aqueous preparations containing alkaloids (e.g., codeine phosphate). An acid–base reaction (ammonia-phosphate) occurs, and, if the alcohol content of the final mixture is too low, codeine will precipitate. Spirits should be stored in tight, light-resistant containers in a cool place. This tends to prevent evaporation and volatilization of either the alcohol or the active principle and to limit oxidative changes.

Spirits may be used pharmaceutically as flavoring agents and medicinally for the therapeutic value of the aromatic solute. As flavoring agents, they are used to impart the flavor of their solute to other pharmaceutical preparations. For medicinal purposes, spirits may be taken orally, applied externally, or used by inhalation, depending upon the particular preparation. When taken orally, they are mixed with a portion of water to reduce the pungency of the spirit. Depending on the materials utilized, spirits may be prepared by simple solution, solution by maceration, or distillation. The spirits still listed in the USP/NF are aromatic ammonia spirit, camphor spirit, compound orange spirit, and peppermint spirit.

EMULSIONS

An emulsion is a two-phase system prepared by combining two immiscible liquids, in which small globules of one liquid are dispersed uniformly throughout the other liquid. The liquid dispersed into small droplets is called the dispersed, internal, or discontinuous phase. The other liquid is the dispersion

medium, external phase, or continuous phase. Where oil is the dispersed phase and an aqueous solution is the continuous phase, the system is designated as an oil-in-water (O/W) emulsion. Conversely, where water or an aqueous solution is the dispersed phase and oil or oleaginous material is the continuous phase, the system is designated as a water-in-oil (W/O) emulsion. Emulsions may be employed orally, topically, or parenterally, depending on the formulation ingredients and the intended application. Many pharmaceutical emulsions may not be classified as such, because they are described by another pharmaceutical category more appropriately. For instance, certain lotions, liniments, creams, ointments, and commercial vitamin drops may be emulsions but may be preferentially referred to in these terms.

Emulsions possess a number of important advantages over other liquid forms:

- Poorly water soluble drugs may be easily incorporated with improved dissolution rates and bioavailability.
- The unpleasant taste or odor of oils can be masked partially or wholly, by emulsification.
- The absorption rate and permeation of medicaments can be controlled.
- Absorption may be enhanced by the diminished size of the internal phase.
- Formulation and technology for organ targeted delivery is available.
- Various particle sizes of the internal phase can be achieved by preparation technique, from micro emulsions (micron sized particles) to nanoparticles.
- Water is an inexpensive diluent and a good solvent for the many drugs and flavors that are incorporated into an emulsion.

It is possible to prepare emulsions that are basically non-aqueous. For example, investigations of the emulsifying effects of anionic and cationic surfactants on the nonaqueous immiscible system, glycerin and olive oil, have shown that certain amines and three cationic agents produced stable emulsions. Although the USP definition is broad enough to encompass non-aqueous systems, emphasis is placed on those emulsions that contain water, as they are the most common in the pharmaceutical sciences.

When it is necessary to administer oils by the oral route, patient acceptance is enhanced when the oil is prepared in emulsion form. Thus, mineral oil (a laxative), valproic acid (an anticonvulsant), oil-soluble vitamins, vegetable oils, and preparations for enteral feeding are formulated frequently in an O/W emulsion form to enhance their palatability.

The bioavailability of oils for absorption may be enhanced when the oil is in the form of small droplets. Furthermore, the absorption of some drugs, such as griseofulvin, may be enhanced when they are prepared in the form of an O/W emulsion.⁶⁵ Emulsion formulations of drugs, such as erythromycin and physostigmine salicylate, have been considered to improve their stability.^{66,67} Finally, the greatest use of emulsions is for topical preparations. Both O/W and W/O emulsions are used widely, depending on the effect desired. Emulsion bases of the W/O type tend to be more occlusive and emollient than O/W emulsion bases, which tend to be removed more easily by water. The effects of viscosity, surface tension, solubility, particle size, complexation, and excipients on the bioavailability of emulsions have been reported.⁶⁸

Although this section on emulsions focuses primarily on those for oral use and, to a lesser degree, those for topical application, it should be noted that there are a number of emulsions used parenterally described in specialized books listed in the Bibliography at the end of this chapter. For example, emulsions of the O/W type are used for intravenous feeding of lipid nutrients. These are used to provide a source of calories and essential fatty acids. These emulsions must meet exacting standards in regard

to particle size, safety, and stability. Examples of commercial products include Diprivan Injectable Emulsion (AstraZeneca), EMLA Cream (AstraZeneca), Renova 0.02% Cream (OrthoNeutrogena), Bactroban Cream (GlaxoSmithKline), Cordran Lotion (Watson), Differin Cream (Galderma), and Renova 0.05% Cream (OrthoNeutrogena). Other specialized uses of emulsions include radiopaque emulsions used as diagnostic agents for X-ray examination.

THEORIES OF EMULSIFICATION

Several theories have been proposed to explain how emulsifying agents act in producing the multi-phase dispersion and in maintaining the stability of the resulting emulsion. Some of these theories apply to specific types of emulsifying agents and to certain conditions, such as pH of the system and the physicochemical nature and proportions of the internal and external phases. The most prevalent theories are the surface-tension theory, the oriented-wedge theory, and the interfacial film theory.

Liquids assume a shape to minimize their surface area, which is spherical for a small drop. In a spherical drop of liquid, there are attractive forces between the molecules of the liquid, resisting distortion into a less spherical form. If two or more drops of the same liquid come into contact with one another, it is more thermodynamically favorable for them to coalesce, making a larger drop with a decreased surface area, compared to the total surface area of the individual drops. The tendency of liquids to minimize their surface area can be measured quantitatively, and, when the liquid is surrounded by air, the measurement is called the surface tension.

When a liquid is in contact with another liquid in which it is insoluble and immiscible, the force causing each liquid to resist breaking into smaller particles is called interfacial tension. Surface active agents, or surfactants, are substances that reduce the resistance of a droplet to form smaller droplets. Surfactants are also called emulsifiers and wetting agents. According to the surface tension theory of emulsification, the use of surfactants results in a reduction in the interfacial tension of the two immiscible liquids, reducing the repellent force between the liquids and diminishing each liquid's attraction for its own molecules. Thus, surfactants enable large globules to break into smaller globules, and prevent small globules from coalescing into larger globules.

The oriented wedge theory proposes that the surfactant forms monomolecular layers around the droplets of the internal phase of the emulsion. The theory is based on the assumption that emulsifying agents orient themselves about and within a liquid relative to their solubility in that particular liquid. In a system containing two immiscible liquids, the emulsifying agent is preferentially soluble in one of the two liquids and becomes more embedded with that phase relative to the other. Many surfactants have a hydrophilic or water loving portion and a hydrophobic or water hating portion (but usually lipophilic or oil-loving), and the molecules position or orient themselves into each phase. Depending on the shape and size of the molecules, their solubility characteristics, and, thus, their orientation, the wedge shape theory proposes that emulsifiers surround either oil globules or water globules.

An emulsifying agent, having a greater hydrophilic character than hydrophobic character, will promote oil in water emulsions. Conversely, water in oil emulsions result with the use of an emulsifier that is more hydrophobic than hydrophilic. Putting it another way, the phase in which the emulsifying agent is more soluble will become the continuous or external phase of the emulsion. Although this theory does not represent a completely accurate depiction of the molecular arrangement of the emulsifier molecules, the concept that water soluble emulsifiers form oil in water emulsions is important.

The interfacial film theory proposes that the emulsifier forms an interface between the oil and water, surrounding the droplets of the internal phase as a thin layer of film adsorbed on the surface of the drops. The film prevents the contact and coalescing

of the dispersed phase; the tougher and more pliable the film, the greater the stability of the emulsion. Naturally, the surfactant must be available to coat the entire surface of each drop of internal phase. Similar to the oriented wedge theory, the formation of an oil in water or a water in oil emulsion depends on the degree of solubility of the emulsifier in the two phases, with water soluble agents encouraging oil in water emulsions and oil soluble emulsifiers promoting water in oil emulsions.

In reality, none of the emulsion theories can individually explain the mechanism by which the many and varied emulsifiers promote emulsion formation and stability. It is more than likely that, even within a given emulsion system, more than one of the theories of emulsification are applicable. For instance, reducing the interfacial tension is critical during initial formation of an emulsion, but the formation of a protective wedge of molecules or film of emulsifier is equally important for continued emulsion stability. Undoubtedly, many emulsifiers are capable of both tasks.

EMULSION FORMULATION INGREDIENTS

The first step in preparation of an emulsion is the selection of the emulsifier. The emulsifier must be compatible with the formulation ingredients and the active pharmaceutical ingredient. It should be stable, non-toxic, and promote emulsification to maintain the stability of the emulsion for the intended shelf life of the product. The selection of the oil phase for oral preparations depends on the purpose of the product. For example, mineral oil is used as a laxative, and corn oil is used for its nutrient properties. Vegetable oils can be used to dissolve or suspend pharmaceuticals, such as oil-soluble vitamins.

Emulsions are thermodynamically unstable, due to the large increase in surface energy that results from the combination of interfacial tension and large surface area of the dispersed phase and the different densities of the two phases. Thus, emulsions tend to cream—the less dense phase rises and the denser phase falls in the container. Subsequently, the droplets can coalesce with a considerable reduction in surface free energy.

Consequently, considerable research has been conducted on their preparation and stabilization. To prepare suitable emulsions that remain stable, a number of excipients are used in their preparation.

Emulsifiers often have a hydrophilic portion and a lipophilic portion with one or the other being more or less predominant. Griffin devised a method whereby emulsifying or surface-active agents may be categorized on the basis of their hydrophilic-lipophilic balance or HLB value.⁷⁰ By this method, each agent is assigned an HLB value or number that is indicative of the substance's polarity, which may vary from 40 for sodium lauryl sulfate to 1 for oleic acid. Although the numbers have been assigned up to 40, the usual range is between 1 and 20. Table 24-6 lists examples of HLB values for common emulsifiers used in pharmaceutical applications. HLB values have also been useful in describing the functional properties of materials. For example, HLB values from 1 to 3 exhibit antifoaming properties, values from 7 to 10 exhibit good wetting properties, values from 13 to 20 act as solubilizers, and values from 13 to 15 function as detergents. Oil in water emulsions have a weighted HLB value ranging from 8 to 16, whereas water in oil emulsions have weighted HLB values ranging from 3 to 8.

Materials highly polar or hydrophilic have been assigned higher numbers than materials that are less polar and more lipophilic. Lipophilic surfactants have an HLB value from 0 to 10 and are known for their antifoaming, water in oil emulsifying, or wetting properties. Hydrophilic surfactants have HLB values ranging from 10 to 20 and form oil in water emulsions. The HLB system also assigns values to oils and oil-like substances. In using the HLB concept in the preparation of an emulsion, one selects emulsifying agents having the same or nearly the same HLB value as the oleaginous phase of the intended emulsion. When needed, two or more emulsifiers may be combined to achieve the proper HLB value.

The ionic nature of a surfactant is an important consideration when selecting a surfactant for an emulsion. Nonionic surfactants are effective over pH range 3–10; cationic surfactants are

Table 24-6. HLB Values of Common Emulsifiers used in Pharmaceutical Systems

Agent	HLB	Class
Oleic Acid	1.0	Anionic
Ethylene glycol distearate	1.5	Nonionic
Sorbitan tristearate (Span 65)	2.1	Nonionic
Glyceryl monooleate	3.3	Nonionic
Propylene glycol monostearate	3.4	Nonionic
Glyceryl monostearate	3.8	Nonionic
Sorbitan monooleate (Span 80)	4.3	Nonionic
Sorbitan monostearate (Span 60)	4.7	Nonionic
Diethylene glycol monolaurate	6.1	Nonionic
Sorbitan monopalmitate (Span 40)	6.7	Nonionic
Acacia	8.0	Anionic
Polyoxyethylene lauryl ether (Brij 30)	9.7	Nonionic
Polyoxyethylene monostearate (Myrj 45)	11.1	Nonionic
Triethanolamine oleate	12.0	Anionic
Polyoxyethylene sorbitan monostearate (Tween 60)	14.9	Nonionic
Polyoxyethylene sorbitan monooleate (Tween 80)	15.0	Nonionic
Polyoxyethylene sorbitan monolaurate (Tween 20)	16.7	Nonionic
Pluronic F 68	17.0	Nonionic
Sodium oleate	18.0	Anionic
Potassium oleate	20.0	Anionic
Cetrimonium Bromide	23.3	Cationic
Cetylpyridinium chloride	26.0	Cationic
Poloxamer 188	29.0	Nonionic
Sodium lauryl sulfate	40.0	Anionic

effective over pH range 3–7; and, anionic surfactants require a pH of greater than 8.⁶⁹

Emulsifying agents may be divided into three classes: natural emulsifying agents, finely divided solids, and synthetic emulsifying agents.

Natural Emulsifying Agents are substances derived from vegetable sources and include acacia, tragacanth, alginates, chondrus, xanthan, and pectin. These materials form hydrophilic colloids, when added to water, and produce o/w emulsions. Although their surface activity is low, these materials achieve their emulsifying power by increasing the viscosity of the aqueous phase. Examples of emulsifying agents derived from animal sources include gelatin, egg yolk, casein, wool fat, cholesterol, wax, and lecithin. Due to the widely different chemical constitution of these compounds, they have a variety of uses, depending on the specific compound, in both oral and topical preparations. All naturally occurring agents show variations in their emulsifying properties from batch to batch.

Finely Divided Solids are the colloidal clays: bentonite (aluminum silicate) and Veegum (magnesium aluminum silicate). These compounds are good emulsifiers and tend to be adsorbed at the interface, increase the viscosity in the aqueous phase, and are often used in conjunction with a surfactant to prepare O/W emulsions. However, both O/W and W/O preparations can be prepared by adding the clay to the external phase. They are used frequently in external preparations, such as lotions or creams.

Synthetic Emulsifying Agents are very effective at lowering the interfacial tension between the oil and water phases, because the molecules possess both hydrophilic and hydrophobic properties. These emulsifying agents are available in different ionic types: anionic, such as sodium dodecyl sulfate; cationic, such as benzalkonium chloride; nonionic, such as polyethylene glycol 400 monostearate; and ampholytic, such as long-chain amino acid derivatives.

In addition to the emulsifying agents, viscosity agents are employed, namely the hydrophilic colloids, such as naturally occurring gums, and partially synthetic polymers, such as cellulose derivatives (e.g., methylcellulose, hydroxypropylmethylcellulose, sodium carboxymethylcellulose), or a number of synthetic polymers that may be used, such as carbomer polymers. These materials are hydrophilic in nature and dissolve or disperse in water to give viscous solutions and function as emulsion stabilizers.

Other functional excipients are often utilized in emulsions. High molecular weight alcohols, such as stearyl alcohol, cetyl alcohol, and glyceryl monostearate, are employed primarily as thickening agents and stabilizers for o/w emulsions of certain lotions and ointments used externally. Cholesterol and cholesterol derivatives may also be employed in externally used emulsions and to promote w/o emulsions.

The aqueous phase of the emulsion favors the growth of micro-organisms; due to this, a preservative is added to the product. Some of the preservatives that have been used include chlorocresol, chlorobutanol, mercurial preparations, salicylic acid, the esters of *p*-hydroxybenzoic acid, benzoic acid, sodium benzoate, or sorbic acid. The preservative should be selected with regard for the ultimate use of the preparation and possible incompatibilities between the preservative and the ingredients in the emulsion (e.g., binding between the surfactant and the preservative). Low pH values of 5–6 and low concentrations of water are characteristics also likely to inhibit microbiological growth in emulsions.

Emulsions consist of an oil or lipid phase and an aqueous phase, thus, the preservative may diffuse from the aqueous phase into the oil phase. It is in the aqueous phase that micro-organisms tend to grow. As a result, water soluble preservatives are more effective, because the concentration of the unbound preservative in the aqueous phase assumes a great deal of importance in inhibiting the microbial growth. Esters of *p*-hydroxybenzoic acid appear to be the most satisfactory preservatives for emulsions.

Many mathematical models have been used to determine the availability of preservatives in emulsified systems. One model takes into account the O/W partition coefficient of the preservative, interaction of the preservative with the surfactant, interfacial tension, and membrane permeability. However, due to the number of factors that reduce the effectiveness of the preservative, a final microbiological evaluation of the emulsion must be performed.

Although emphasis concerning preservation of emulsions deals with the aqueous phase, micro-organisms can also reside in the lipid phase. Consequently, it has been recommended that pairs of preservatives be used to ensure adequate concentration in both phases. A combination of esters of *p*-hydroxybenzoic acid can be used to ensure appropriate concentrations in both phases, due to their difference in oil and water solubilities.

The oxidative decomposition of certain excipients, the oil phase, and some pharmaceuticals is possible in emulsions, not only due to the usual amount of air dissolved in the liquid and the possible incorporation of air during the preparation of the product, but also the large interfacial area between the oil and water phase. The selection of the appropriate antioxidant, briefly described at the beginning of the chapter, depends on such factors as stability, compatibility with the ingredients of the emulsion, toxicity, effectiveness in emulsions, odor, taste, and distribution between the two phases.

PREPARATION OF EMULSIONS

After the purpose of the emulsions has been determined (e.g., oral or topical use), the type of emulsions (O/W or W/O) and appropriate ingredients selected, and the theory of emulsification considered, then experimental formulations may be prepared. One method is suggested by Griffin:⁷⁰

1. Group the ingredients on the basis of their solubilities in the aqueous and nonaqueous phases.
2. Determine the type of emulsion required and calculate an approximate HLB value.
3. Blend a low HLB emulsifier and a high HLB emulsifier to the calculated value. For experimental formulations, use a higher concentration of emulsifier (e.g., 10–30% of the oil phase) than required to produce a satisfactory product. Emulsifiers should be stable chemically, non-toxic, and suitably low in color, odor, and taste. The emulsifier is selected on the basis of these characteristics, as well as the type of equipment used to blend the ingredients and the stability characteristics of the final product. Emulsions should not coalesce at room temperature, when frozen and thawed repeatedly, or at elevated temperatures of up to 50°C. Mechanical energy input varies with the type of equipment used to prepare the emulsion. The more the energy input, the less the demand on the emulsifier. Both process and formulation variables can affect the stability of an emulsion.
4. Dissolve the oil-soluble ingredients and the emulsifiers in the oil. Heat, if necessary, to approximately 5–10°C over the melting point of the highest melting ingredient or to a maximum temperature of 70–80°C.
5. Dissolve the water-soluble ingredients (except acids and salts) in a sufficient quantity of water.
6. Heat the aqueous phase to a temperature that is 3–5°C higher than that of the oil phase.
7. Add the aqueous phase to the oily phase with suitable agitation.
8. If acids or salts are employed, dissolve them in water and add the solution to the cold emulsion.
9. Examine the emulsion and make adjustments in the formulation, if the product is unstable. It may be necessary to add more emulsifier, to change to an emulsifier with a slightly higher or lower HLB value, or to use an emulsifier with different chemical characteristics.

The technique of emulsification of pharmaceutical preparations has been described by Nielloud and Marti-Mestres.⁷¹ The preparation of an emulsion requires work to reduce the internal phase into small droplets and disperse them throughout the external phase. This can be accomplished by a mortar and pestle or a high-speed emulsifier. The addition of emulsifying agents not only reduces this work, but also stabilizes the final emulsion.

Emulsions are prepared by four principal methods.

Addition of Internal Phase to External Phase

This is the most satisfactory method for preparing emulsions, as there is always an excess of the external phase present that promotes the type of emulsion desired. If the external phase is water and the internal phase is oil, the water-soluble substances are dissolved in the water and the oil-soluble substances mixed thoroughly in the oil. The oil mixture is added in portions to the aqueous preparation with agitation. To give a better shearing action during the preparation, sometimes all of the water is not mixed with the emulsifying agent, until the primary emulsion with the oil is formed; subsequently, the remainder of the water is added. An example using gelatin Type A is given subsequently.

Addition of the External Phase to the Internal Phase, the Dry Gum Technique

Using an O/W emulsion as an example, the addition of the water (external phase) to the oil (internal phase) will promote the formation of a W/O emulsion, due to the preponderance of the oil phase. After further addition of the water, phase inversion to an O/W emulsion should take place. This method is especially useful and successful when hydrophilic agents, such as acacia, tragacanth, or methylcellulose, are first mixed with the oil, affecting dispersion without wetting. Water is added, and eventually an O/W emulsion is formed. This "dry gum" technique is a rapid method for preparing small quantities of emulsion. The ratio 4 parts of oil, 2 parts of water, and 1 part of gum provides maximum shearing action on the oil globules in the mortar. The emulsion can then be diluted and triturated with water to the appropriate concentration. The preparation of Mineral Oil Emulsion described subsequently is an example.

Mixing Both Phases after Heating

This method is used when waxes or other substances that require melting are used. The oil soluble emulsifying agents, oils, and waxes are melted and mixed thoroughly. The water soluble ingredients dissolved in the water are warmed to a temperature slightly higher than the oil phase. The two phases are then mixed and stirred until cold. For convenience, but not necessity, the aqueous solution is added to the oil mixture. This method is frequently used in the preparation of ointments and creams. An example of an oral preparation containing a poorly soluble drug is given subsequently.

Alternate Addition of the Two Phases to the Emulsifying Agent

A portion of the oil, if an O/W emulsion is being prepared, is added to all of the oil-soluble emulsifying agents with mixing, then an equal quantity of water containing all the water-soluble emulsifying agents is added with stirring, until the emulsion is formed. Further portions of the oil and water are added alternately, until the final product is formed. The high concentration of the emulsifying agent in the original emulsion makes the initial emulsification more likely, and the high viscosity provides effective shearing action, leading to small droplets in the emulsion. This method is often used successfully with soaps.

Examples of some emulsions are given subsequently.

Type A gelatin is prepared by acid-treated precursors and is used at a pH of about 3.2. It is incompatible with anionic emulsifying agents, such as the vegetable gums. The following formula was recommended.

Type A

Gelatin Emulsion

GELATIN	(TYPE A)8 G
TARTARIC ACID	0.6 G
FLAVOR	Q.S
ALCOHOL	60 ML
OIL	500 ML
PURIFIED WATER, TO MAKE	1000 ML

Add the gelatin and the tartaric acid to about 300 mL of purified water, allow to stand for a few minutes, heat until the gelatin is dissolved, then raise the temperature to about 98°C and maintain this temperature for about 20 minutes. Cool to 50°C, and add the flavor, the alcohol, and sufficient purified water to make 500 mL. Add the oil, agitate the mixture thoroughly, and pass it through a homogenizer or a colloid mill, until the oil is dispersed completely and uniformly. This emulsion cannot be prepared by trituration or by the use of the usual stirring devices.

Type B gelatin is prepared from alkali-treated precursors and is used at a pH of about 8. It may be used with other anionic emulsifying agents but is incompatible with cationic types. If the emulsion contains 50% oil, 5 g of Type B gelatin, 2.5 g of sodium bicarbonate, and sufficient tragacanth or agar should be incorporated into the aqueous phase to yield 1,000 mL of product of the required viscosity.

An emulsion that may be prepared by the mortar and pestle method is the following Mineral Oil Emulsion, USP.

Mineral Oil Emulsion, USP

MINERAL OIL	500 ML
ACACIA, IN VERY FINE POWDER	125 G
SYRUP	100 ML
VANILLIN	40 MG
ALCOHOL	L60 ML
PURIFIED WATER, TO MAKE	1000 ML

The mineral oil and acacia are mixed in a dry Wedgwood mortar. Purified water (250 mL) is added, and the mixture is triturated vigorously, until an emulsion is formed. A mixture of the syrup, 50 mL of purified water, and the vanillin dissolved in alcohol is added in divided portions with trituration; sufficient purified water is then added to the proper volume; the mixture is mixed well and homogenized.

An Oral Emulsion (O/W) Containing an Insoluble Drug⁷²

COTTONSEED OIL	460 G
SULFADIAZINE	200 G
SORBITAN MONOSTEARATE	84 G
POLYOXYETHYLENE 20 SORBITAN MONOSTEARATE	36 G
SODIUM BENZOATE	2 G
SWEETENER	QS
FLAVOR OIL	QS
PURIFIED WATER	1000 G

Heat the first three ingredients to 50°C and pass through colloid mill. Add the next four ingredients at 50°C to the first three ingredients at 65°C, and stir, while cooling to 45°C. Add the flavor oil and continue to stir, until room temperature is reached.

PROPERTIES AND STABILITY OF EMULSIONS

The type of emulsion (O/W or W/O) depends, to some extent, on the phase to volume ratio. The higher the fraction of one phase, the greater likelihood it will form the external phase.

Thus, O/W emulsions are favored, if water forms a greater fraction of the volume than the oil phase. However, it is possible for the internal phase of an emulsion to occupy up to 74% of the volume of the emulsion and still form a stable product.

The consistency of emulsions can be increased by increasing the viscosity of the continuous phase, increasing the fractional volume of the internal phase, reducing the particle size of the internal phase, increasing the proportion of the emulsifying agent, or adding hydrophobic emulsifying agents to the oil phase of the emulsion.

The physical stability of emulsions may be defined by a number of expressions. The first of these, which is called "creaming," is the movement of the droplets either upward or downward, depending upon their density. This gives a product that is not homogenous and can lead to poor content uniformity. Creaming is not a serious problem, because a moderate amount of shaking will redisperse the droplets uniformly. The rate of creaming may be decreased by considering the theory of creaming using Stokes' law. This equation relates the rate of creaming to the size of the droplets, the difference in densities, and the viscosity of the external phase. Thus, the rate of creaming may be decreased by decreasing the size of the droplets and increasing the viscosity of the external phases, both of which were previously discussed. Minimizing the difference between densities is more challenging, due to a number of practical difficulties.

When the droplets aggregate, they come together and act as a single unit, but do not fuse. As a result of the larger size, they tend to cream faster and further provoke physical instability. Aggregation is to some extent reversible and may be controlled by choosing a somewhat different surfactant system and controlling the electrical potential of the droplets. Coalescence of an emulsion is the fusion of the droplets, leading to a decrease in their numbers and, eventually, the complete separation of the two phases, yielding an unsatisfactory product that should be reformulated.

General methods are available for testing and challenging the stability of emulsions, including bulk changes, centrifugal and ultracentrifugal studies, dielectric measurement, surface area measurement, temperature cycling, preservative effectiveness, and accelerated motion studies. Low-shear rheological studies measuring viscoelasticity are suggested as the optimal method of stability testing.

MULTIPLE EMULSIONS

A recent innovation in emulsion technology is the development of multiple emulsions. The dispersed phase of these emulsions contains even smaller droplets that are miscible with the continuous phase. Thus, the multiple emulsion may be O/W/O, where the aqueous phase is between two oil phases, or W/O/W, where the internal and external aqueous phases are separated by an oil phase. In these systems both hydrophobic and hydrophilic emulsifiers are used, and both have an effect on the yield and stability.^{73,74}

It appears that O/W/O emulsions are formed better by lipophilic, nonionic surfactants, using acacia emulsified simple systems, whereas W/O/W multiple emulsions are formed better by nonionic surfactants in a two-stage emulsification procedure. A specific formulation for a W/O/W emulsion may be prepared by forming the primary (W/O) emulsion from isopropyl myristate (47.5%), sorbitan monooleate (2.5%), and distilled water to 100%. This primary emulsion (50%) is added to a polyoxyethylene sorbitan monooleate (2% w/v) solution in water.⁷⁴ Other formulations of multiple emulsions include carboxymethylcellulose sodium, microcrystalline cellulose, sorbitan monooleate, and sorbitan trioleate.

Although the technique of preparing these emulsions is more complicated, research indicates potential use of these emulsions for prolonged action, taste-masking, more effective dosage forms, improved stability, parenteral preparations, protection against the external environment, and enzyme entrapment. These emulsions may also be used to separate two

incompatible hydrophilic substances in the inner and outer aqueous phases by the middle oil phase. Some drugs that have been investigated in these types of emulsions are vancomycin⁷⁵ and prednisolone.⁷⁶

MICROEMULSIONS

Coarse pharmaceutical macroemulsions appear white and tend to separate on standing. Microemulsions are translucent or transparent, do not separate, and typically have a droplet diameter from 10 to 200 nanometers. The microemulsions are not always distinguishable from micellar solutions. Both O/W and W/O emulsions are possible, and one may be converted to the other by addition of more internal phase or by altering the type of emulsifier. As the internal phase is added, the emulsion will pass through a viscoelastic gel stage; with further addition, an emulsion of the opposite type will occur.

The most obvious benefit of microemulsions is their stability. The emulsifier should be 20–30% of the weight of the oil used. The W/O systems are prepared by blending the oil and emulsifier with a little heat, if required, and then adding the water. The order of mixing for O/W systems is more flexible. One of the simplest methods is to blend the oil and the emulsifier and pour this into water with a little stirring.

If the emulsifier has been selected properly, microemulsification will occur almost spontaneously, leading to a satisfactory and stable preparation. The details of various preparations and the relationship between microemulsions and micellar solutions have been reviewed by Bourrel.⁷⁷ Other authors suggest the preparation of microemulsions is considerably more difficult than the preparation of coarse emulsions. Rosano and colleagues discuss the use of a primary surfactant adsorbed at the interface that influences the curvature of the dispersed phase.⁷⁸ The amount of surfactant required may be estimated from the surface area of the droplets and the cross-sectional area of the surfactant molecule. The authors propose the use of a cosurfactant to form a duplex film and the order of mixing is important.

PROCESSING EQUIPMENT FOR EMULSIONS

The preparation of emulsions requires a certain amount of energy to form the interface between the two phases, and additional work must be done to stir the system to overcome resistance to flow. In addition, heat is often supplied to the system to melt waxy solids and/or reduce viscosity. Consequently, the preparation of emulsions on a large scale requires considerable amounts of energy for heating and mixing. Careful consideration of these processes has led to the development of low energy emulsification equipment that use an appropriate emulsification temperature and selective heating of the ingredients. This process, described by Lin, involves the preparation of an emulsion concentrate subsequently diluted with the external phase at room temperature.^{79,80}

Due to the variety of oils used, emulsifier agents, phase to volume ratios, and the desired physical properties of the product, a wide selection of equipment is available for preparing emulsions, and the main classes of equipment are discussed. Homogenization speed and time and rate of cooling may influence the viscosity of the product. Further information may be obtained from the **Bibliography**.

Special techniques and equipment in certain instances will produce superior emulsions, including rapid cooling, reduction in particle size, or ultrasonic devices. A wide selection of equipment for processing both emulsions and suspensions has been described.⁷¹ A number of improvements have been made to make the various processes more effective and energy efficient.

The mortar and pestle may be used to prepare small quantities of an emulsion in the pharmacy or laboratory, and it is one of the simplest and least expensive methods. It may be used for most of the different techniques of preparing emulsions. Final particle size is considerably larger than is achieved by the equipment subsequently described. In addition, it is necessary

for the ingredients to have a certain viscosity prior to trituration to achieve a satisfactory shear. Satisfactory emulsions of low viscosity ingredients and small volumes may be prepared using the appropriate equipment described subsequently.

Agitators

Ordinary agitation or shaking may be used to prepare the emulsion. This method is frequently employed by the pharmacist, particularly in the emulsification of easily dispersed, low-viscosity oils. Under certain conditions, intermittent shaking is considerably more effective than ordinary continuous shaking. Continuous shaking tends to break up not only the phase to be dispersed, but also the dispersion medium, thus, impairing the ease of emulsification. Laboratory shaking devices may be used for small-scale production.

Mechanical Mixers

Emulsions may be prepared using one of several mixers that are available. Propeller and impeller type mixers that have a propeller attached to a shaft driven by an electric motor are convenient and portable and can be used for both stirring and emulsification. This type operates best in mixtures that have low viscosity, that is, mixtures with a viscosity of glycerin or less. They are also useful for preparing emulsions. A turbine mixer has a number of blades that may be straight or curved, with or without a pitch, mounted on a shaft. The turbine tends to give a greater shear than propellers. The shear can be increased by using diffuser rings perforated and surrounding the turbine, so the liquid from the turbine must pass through holes. The turbines can be used for both low-viscosity mixtures and medium-viscosity liquids. The degree of stirring and shear by propeller or turbine mixers depends on several factors, such as the speed of rotation, pattern of liquid flow, position in the container, and baffles in the container.

Production sized mixers include high powered propeller, shaft stirrers immersed in a tank or self-contained units with propeller and paddle systems. The latter are constructed so the contents of the tank may be either heated or cooled during the production process. Baffles are often built into a tank to increase mixing efficiency. Examples of two production dispersion mixers are shown in Figures 24-1 and 24-2.

Small electric mixers may be used to prepare emulsions at the prescription counter. They save time and energy and produce

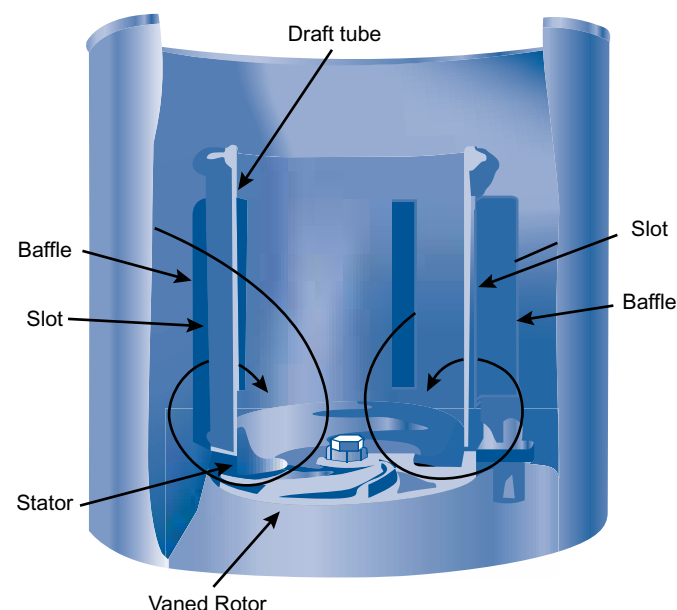


Figure 24-1. Standard slurry-type dispersal mixer with vaned-rotor mixing element and slotted draft-tube circulating element.

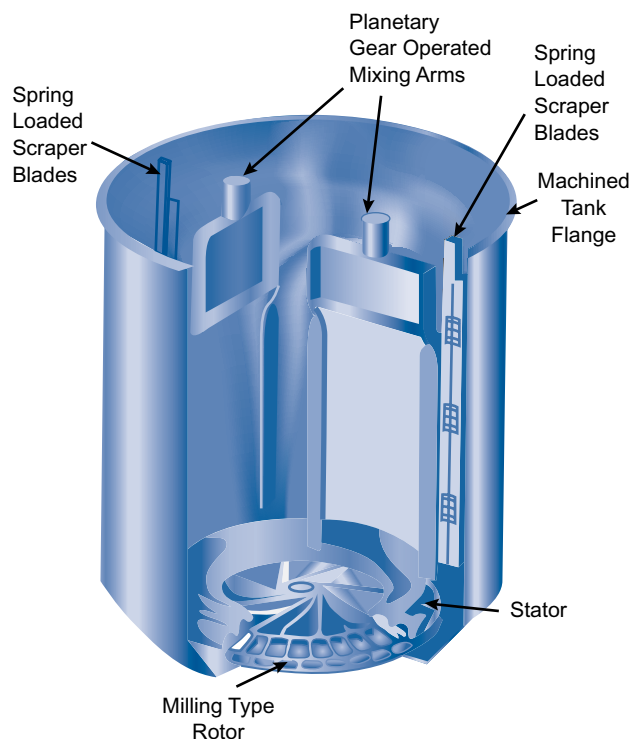


Figure 24-2. Standard paste-type dispersal mixer with cupped-rotor milling element and double-rotating mixing arm circulating element.

satisfactory emulsions, when the emulsifying agent is acacia or agar. The commercially available Waring Blender disperses efficiently by means of the shearing action of rapidly rotating blades. It transfers large amounts of energy and incorporates air into the emulsion. If an emulsion is first produced by using a blender of this type, the formulator must remember that the emulsion characteristics obtained in the laboratory will not necessarily be duplicated by the production-size equipment.

Rotor Stators

The principle of operation of the rotor stator is the passage of the mixed phases of an emulsion formula between a stator and a high speed rotor revolving at speeds of 2,000–18,000 rpm. The advantages of high shear rotor/stator mixers over simple conventional agitators and mechanical mixers stem from the multistage mixing and shearing action as materials are drawn through the specially designed stator workhead. As illustrated in Figure 24-3, in Stage 1, suction is created from the high-speed rotation of the rotor blades within the mixing workhead, drawing liquid and solid materials up from the bottom of the vessel and into the center of the workhead. Next, centrifugal force drives the materials toward the periphery of the workhead, where they are subjected to a milling action in the precision machined clearance between the rotor blades and the inner wall of the stator (Stage 2). The materials are then forced by intense hydraulic shear, at high velocity, out through the perforations in the stator and circulated back into the vessel (Stage 3). The materials expelled from the head are projected radially at high speed towards the sides of the mixing vessel (Stage 4). At the same time, fresh material is continually drawn into the workhead, maintaining the mixing cycle. The effect of the horizontal (radial) expulsion and suction into the head is to set up a circulation pattern that minimizes aeration caused by the disturbance of the liquid's surface. Figure 24-4 illustrate examples of rotor stator workheads.

Rotor stators are available from small, laboratory scale to very large, commercial scale. An example of a pilot plant rotor stator is shown in Figure 24-5. An inline rotor-stator, which can

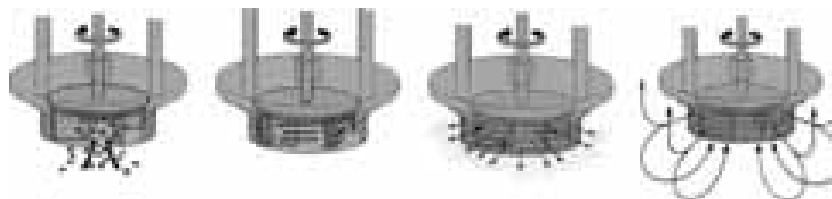


Figure 24-3. Solid and Liquid Flow through a Rotor Stator (Courtesy of Silverson Machines, Ltd.)



Figure 24-4. Types of workheads used in rotor stators. From top to bottom, these include a general purpose disintegrating head (versatile and useful for gels, thickeners, suspensions, solutions and slurries), a slotted disintegrating head (useful for fibrous materials), square hole high shear screen (provides high shear rates), and a standard emulsor head and screen (excellent for emulsions). (Courtesy of Silverson Machines, Ltd.)

be plumbed directly to a vessel or tank and has interchangeable screens and workheads, is shown in Figure 24-6.

The clearance between the rotor and the stator is often very small (from 0.001 inches and up) and requires precision machining and alignment. The emulsion mixture, in passing through and between the rotor and stator, is subjected to the tremendous forces that affect the chemical and physical properties of the emulsion.

The operating principle is the same for all, but each manufacturer incorporates specific features that result in changes in operating efficiency. The shearing forces applied in a rotor stator result in a temperature increase within the emulsion. It may be necessary, therefore, to use jacketed equipment to cool the emulsion during processing. Maa and Hsu have shown that droplet size of emulsions was mainly determined by shear force within the gap between the spinning rotor and stationary



Figure 24-5. Pilot Scale Rotor Stator. (Courtesy of Silverson Machines, Ltd.)

rotor.⁸¹ Droplet size decreased with homogenization intensity and duration, increasing viscosity of the continuous phase, and with decreasing viscosity of the dispersed phase.

Rotor stators are also used sometimes for the comminution of solids and for the preparation of suspensions, especially suspensions containing solids not wetted by the dispersion medium.

Homogenizers

Impeller types of equipment frequently produce a satisfactory emulsion; however, for further reduction in particle size, homogenizers may be employed.⁸² Homogenizers may be used in one of two ways:



Figure 24-6. Inline Rotor Stator for small production, pilot plant, and research laboratories up to 50 liters. (Courtesy of Silverson Machines, Ltd.)

1. The ingredients in the emulsion are mixed and then passed through the homogenizer to produce the final product.
2. A coarse emulsion is prepared in some other way and then passed through a homogenizer for the purpose of decreasing the particle size and obtaining a greater degree of uniformity and stability.

The mixed phases or the coarse emulsion are subjected to homogenization and are passed between a finely ground valve and seat under high pressure. This, in effect, produces an atomization enhanced by the impact received by the atomized mixture as it strikes the surrounding metal surfaces. They operate at pressures of 1000–5000 psi and produce some of the finest dispersions obtainable in an emulsion.

Figure 24-7 demonstrates the flow through the homogenizing valve, the heart of the high pressure, APV Gaulin homogenizer. The product enters the valve seat at high pressure, flows through the region between the valve and the seat at high velocity with a rapid pressure drop, causing cavitation; subsequently, the mixture hits the impact ring, causing further disruption, and then is discharged as a homogenized product. It is postulated that circulation and turbulence are responsible mainly for the homogenization that takes place. Different valve assemblies, two-stage valve assemblies, and equipment with a wide range of capacities are available.

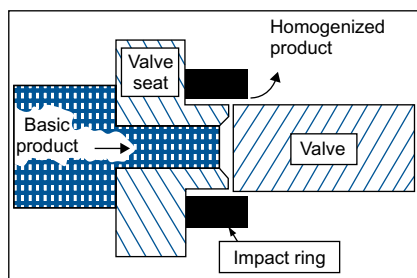


Figure 24-7. Material flow through a homogenizer. (Courtesy of SPX Corporation. Figure is copyrighted material owned by SPX Corporation and is used under license and permission from SPX Corporation. © 2011 SPX Corporation.)

Two-stage homogenizers are constructed so the emulsion, after treatment in the first valve system, is conducted directly to another, where it receives a second treatment. A single homogenization may produce an emulsion that, although its particle size is small, has a tendency to clump or form clusters. Emulsions of this type exhibit increased creaming tendencies. This is corrected by passing the emulsion through the first stage of homogenization at a high pressure (e.g., 3000–5000 psi) and then through the second stage at a greatly reduced pressure (e.g., 1,000 psi). This breaks down any clusters formed in the first step.

For small-scale extemporaneous preparation of emulsions, the inexpensive hand-operated homogenizer is particularly useful. It is probably the most efficient emulsifying apparatus available to the pharmacist and pharmaceutical scientist. The two phases, previously mixed in a bottle, are hand-pumped through the apparatus. Recirculation of the emulsion through the apparatus will improve its quality.

A homogenizer does not incorporate air into the final product. Air may ruin an emulsion, because the emulsifying agent is adsorbed preferentially at the air–water interface, followed by an irreversible precipitation termed “denaturization.” This is particularly prone to occur with protein emulsifying agents. Homogenization may spoil an emulsion, if the concentration of the emulsifying agent in the formulation is less than that required to accommodate the increase in surface area produced by the process.

The temperature rise during homogenization is not very large. However, temperature does play an important role in the emulsification process. An increase in temperature will reduce the viscosity and, in certain instances, the interfacial tension between the oil and the water. There are, however, many instances, particularly in the manufacturing of creams and ointments, where the ingredients will fail to emulsify properly, if they are processed at too high a temperature. Emulsions of this type are processed, first, at an elevated temperature, and then homogenized at a temperature not exceeding 40°C.

Homogenizers have been used most frequently with liquid emulsions, but now they may be used with suspensions, as the metal surfaces are formed from wear-resistant alloys that resist the wear of solid particles contained in suspensions.

Ultrasonic Devices

The preparation of emulsions by the use of ultrasonic vibrations also is possible. An oscillator of high frequency (100–500 kHz) is connected to two electrodes between which is placed a piezoelectric quartz plate. The quartz plate and electrodes are immersed in an oil bath and, when the oscillator is operating, high-frequency waves flow through the fluid. Emulsification is accomplished by simply immersing a tube containing the emulsion ingredients into this oil bath. Considerable research has been done on ultrasonic emulsification, particularly with regard to the mechanism of emulsion formation. The method has not been proven practical for large-scale production of emulsions, but evaluations are underway.⁸³

Microfluidizers

Microfluidizers have been used to produce very fine particles. The process subjects the emulsion to an extremely high velocity through micro-channels into an interaction chamber; as a result, particles are subjected to shear, turbulence, impact, and cavitation. Two advantages of this type of equipment are lack of contamination in the final product and ease of production scale up.

LIPOSOMES

Liposomes have been one of the most extensively studied drug delivery systems.^{84–87} Liposomes, meaning lipid body, may be broadly described as small vesicles of a bilayer of phospholipid

encapsulating an aqueous space ranging from about 0.03 to 10 μm in diameter. The lipid membrane of a liposome consists of a bilayer-forming amphiphile, cholesterol, and a charge generating molecule. The lipid membrane encloses a discrete aqueous compartment. This structure presents an overall hydrophilic membrane-like assembly in which the apolar or lipophilic portion of the amphiphilic molecule points inward, whereas the polar or hydrophilic portion points outward of the lamellar structure. These characteristics make liposomes useful as drug delivery systems. The enclosed vesicles can encapsulate water soluble drugs in the aqueous spaces or lipid soluble drugs in the membranes. Liposomes have been administered parenterally, topically, and by inhalation.

Liposomes offer several advantages as drug delivery systems: 1) they are biologically inert and completely biodegradable; 2) they can be prepared in various sizes, charge, compositions, and surface morphology; 3) they can encapsulate both water soluble and water insoluble drugs, including enzymes, hormones, and antibiotics; 4) encapsulated drugs are less susceptible to degradation; 5) organ targeted drug delivery is possible, since the entrapped drug is delivered intact to various tissues and cells after the liposome is destroyed; and 6) other tissues and cells of the body are protected from the drug, until they are released by the liposomes, thus, decreasing the drug's toxicity. The primary disadvantage of liposomes is their rapid removal from the blood following intravenous administration by cells of the reticuloendothelial system, particularly by Kupfer cells in the liver. Drug release is slowed by phagocytes through endocytosis, fusion, surface adsorption, or lipid exchange.

Several different amphiphiles have been investigated to create liposomal structures (vesicles). Only the bilayer forming lipid is the essential part of the lamellar structure and the other components are to impart specific characteristics. For example, cholesterol adds rigidity to the vesicular structure, rendering it less permeable. Phospholipids, such as phosphatidyl choline (lecithin), were the first amphiphiles used to produce bilayer structures to mimic cell membranes.

Liposomes can be prepared into several morphologies, which have been classified according to the vesicular shape. Multilamellar vesicles (MLV) were first prepared by Bangham and have multiple bilayer structures surrounding a relatively small internal core, much like an onion.⁸⁸ Oligolamellar vesicles (OLV) have large central aqueous cores surrounded by 2–10 bilayers. Unilamellar vesicles (ULV) have a single bilayer structure surrounding an internal aqueous core. Unilamellar vesicles can be prepared in a variety sizes: small unilamellar vesicles (20–40 nm), medium unilamellar vesicles (40–80 nm), large unilamellar vesicles (10–1,000 nm), and giant unilamellar vesicles (>1,000 nm).

Drug release in the blood following intravenous administration ranges from a few minutes to several hours, depending on the nature and composition of the lipids, surface properties, and size. In general, smaller unilamellar vesicles show much longer half-lives than multilamellar vesicles and large unilamellar vesicles. Negatively charged liposomes are cleared more rapidly from the circulation than neutral or positively charged liposomes. Circulation can be prolonged by blocking the reticuloendothelial system and allowing the liposomes to interact with vascular endothelial cells and blood cells. These "stealth liposomes" were developed by coating the liposomes with polymers, such as polyethylene glycol, enabling liposomes to evade detection by the body's immune system.

PREPARATION OF LIPOSOMES

Liposomes have been prepared using a number of techniques, including solvent evaporation, sonication, supercritical fluid techniques, spray drying, extrusion, and homogenization. A combination of these methods is often used, and the drug is added during the formation process. In this method, the lipid is dissolved in an organic solvent, such as acetone or chloroform. The solvent is evaporated, leaving a thin, lipid film on the walls

of the container. An aqueous solution of the drug is added and placed in an ultrasonic bath. The sound waves displace the lipid from the container walls, and they self-assemble into spheres or cylinders, entrapping the aqueous drug solution inside. If the drug is lipophilic, it is incorporated into the lipid phase and will reside within the lipophilic bilayers. Several advances have been made in liposome preparation to better control stability and size.

Liposomal products are now commercially available. Amphotec (distributed by InterMune, manufactured by Ben Venue Laboratories) is Amphotericin B Cholesteryl Sulfate Complex for Injection. It is a sterile, pyrogen-free, lyophilized powder for reconstitution and intravenous (IV) administration. Amphotec consists of a 1:1 (molar ratio) complex of amphotericin B and cholesteryl sulfate. Upon reconstitution, Amphotec forms a colloidal dispersion of microscopic disc-shaped particles. Each 50 mg single dose vial contains amphotericin B, 50 mg; disodium edetate dihydrate, 0.372 mg; lactose monohydrate, 950 mg; and hydrochloric acid, qs. Amphotec is indicated for the treatment of invasive aspergillosis in patients whose renal impairment or unacceptable toxicity precludes the use of amphotericin B deoxycholate in effective doses and in aspergillosis patients in whom prior amphotericin B deoxycholate therapy has failed. The drug is reconstituted with Sterile Water for Injection by rapidly adding the water to the vial; it is shaken gently by hand, rotating the vial until all the solids have dissolved. The fluid may be opalescent or clear.⁸⁹ For infusion, it is further diluted in 5% dextrose injection. The product should not be reconstituted with any fluid other than Sterile Water for Injection; do not reconstitute with dextrose or sodium chloride solutions. Also, for further dilution, it should not be admixed with sodium chloride or electrolytes. Solutions containing benzyl alcohol or any other bacteriostatic agent should not be used, as they may cause precipitation. An inline filter should not be used and the infusion admixture should not be mixed with other drugs. If infused using a y-injection site or similar device, flush the line with 5% dextrose injection before and after infusion of Amphotec. After reconstitution, the drug should be refrigerated and used within 24 hours; do not freeze. If further diluted with 5% dextrose injection, it should be refrigerated and used within 24 hours.

Doxil (Ortho Biotech) is doxorubicin hydrochloride encapsulated in stealth liposomes for intravenous administration. The product is provided as a sterile, translucent, red liposomal dispersion in a 10 mL glass, single use vial. Each vial contains 20 mg of doxorubicin HCl at a concentration of 2 mg/mL and a pH of 6.5. The stealth liposome carriers are composed of N-(carboxyl-methoxy polyethylene glycol 2000)-1,2-distearoyl-*sn*-glycerol-3-phosphoethanolamine sodium salt (MPEG-DSPE), 3.19 mg/mL; fully hydrogenated soy phosphatidylcholine (HSPC), 9.58 mg/mL; and cholesterol, 3.19 mg/mL. Each mL also contains ammonium sulfate, approximately 2 mg; histidine as a buffer; hydrochloric acid and/or sodium hydroxide for pH control; and sucrose to maintain isotonicity. Greater than 90% of the drug is encapsulated in the Stealth liposomes. The stealth liposomes are specially formulated to circulate in the body "undetected" by the mononuclear phagocyte system for a prolonged circulation time of about 55 hours. This is accomplished by pegylation, or binding methoxypolyethylene glycol on the surface of the liposomes. These liposomes are small, in the range of 100 nm in diameter. Doxil must be diluted in 250 mL of 5% dextrose injection prior to administration; once diluted, it should be refrigerated and administered within 24 hours. It should not be mixed with any other diluent or any preservative-containing solution. It should not be used with in-line filters. The product is not a clear solution, but a red, translucent liposomal dispersion. Unopened vials should be stored in a refrigerator, but freezing should be avoided, even though short term freezing (less than 1 month) does not appear to adversely affect the product.

SUSPENSIONS

The physical chemist defines the word “suspension” as a two-phase system consisting of an undissolved or immiscible material dispersed in a vehicle (i.e., solid, liquid, or gas). A variety of dosage forms fall within the scope of this definition, but emphasis is placed on solids dispersed in liquids. In more specific terms, the pharmaceutical scientist differentiates between such preparations as suspensions, mixtures, magmas, gels, and lotions. In these preparations, the substance distributed is referred to as the dispersed phase and the vehicle is termed the dispersing phase or dispersion medium. In a general sense, each of these preparations represents a suspension, but the state of subdivision of the insoluble solid varies from particles that settle gradually on standing to particles that are colloidal in nature.

The particles of the dispersed phase vary widely in size, from large, visible particles to colloidal dimensions, which fall between 1.0 nm and 0.5 μm in size. Course dispersions contain particles 10–50 μm in size and include suspensions and emulsions. Fine dispersions contain particles of smaller size, 0.5–10 μm . Magmas and gels represent such fine dispersions. Particles in a coarse dispersion have a greater tendency to separate from the dispersion medium than do the particles of a fine dispersion. Most solids in a dispersion tend to settle to the bottom of the container, because their density is higher than the dispersion medium.

Suspensions have a number of applications in pharmacy. They are used to supply drugs to the patient in liquid form. Many people have difficulty swallowing solid dosage forms; consequently, a liquid preparation has an advantage. In addition, the dose of a liquid form may be adjusted easily to meet the patient's requirements. Thus, if the drug is insoluble or poorly soluble, a suspension may be the most suitable dosage form. If a drug is unstable in an aqueous medium, a different form of the drug, such as an ester or insoluble salt that does not dissolve in water may be used in the preparation of a suspension. Drugs, such as antibiotics, that are unstable in the presence of an aqueous vehicle for extended periods of time are most frequently supplied as dry powder mixtures for reconstitution at the time of dispensing. This type of preparation is designated in the USP by the title “for Oral Suspension.” Suspensions that do not require reconstitution at the time of dispensing are simply designated as an “Oral Suspension.” Table 24-7 presents examples of commercial products.

To improve the stability of an antibiotic, such as ampicillin, formulations are made in such a way that the dispersion medium, water, is added upon dispensing to form a satisfactory suspension. The taste of pharmaceuticals can be improved, if they are supplied in suspension form, rather than solutions; thus, chloramphenicol palmitate is used instead of the more

soluble form, chloramphenicol. Another method to decrease the solubility of the drug is to replace part of the water with another appropriate liquid, such as alcohol or glycerin. Insoluble drugs may be formulated as suspensions for topical use, such as calamine lotion. Other preparations of suspensions, in addition to those noted previously, include parental preparations (Chapter 26), ophthalmic preparations (Chapter 28), aerosol suspensions (Chapter 33), and medicated topicals (Chapter 29).

PHYSICAL CHARACTERISTICS OF SUSPENSIONS

Formulation of suspensions involves more than mixing a solid in a liquid. Knowledge of the behavior of particles in liquids, suspending agents, wetting agents, polymers, buffers, preservatives, flavors, and colors is required to produce an acceptable and satisfactory suspension. Suspensions should possess several basic chemical and physical properties. The dispersed phase should settle slowly, if at all, and be re-dispersed readily upon shaking. The solid particles should have a narrow particle size distribution, which does not cake on settling, and the viscosity should be such that the preparation pours easily. In addition, the product should have an elegant appearance, be resistant to microbial growth and maintain its chemical stability.

Several factors influence the sedimentation rate of particles in a suspension. Stokes' law relates the diameter of the particles, the density of the particles and the medium, and the viscosity of the medium to the sedimentation rate:

$$\frac{ds}{dt} = \frac{d^2(\rho_p - \rho_m)g}{18\eta}$$

where ds/dt is the sedimentation rate, d is the diameter of the particles, ρ_p is the density of the particles, ρ_m is the density of the medium, g is the gravitational constant, and η is the viscosity of the medium.

Stokes' equation was derived for an ideal situation with perfectly spherical particles in a very dilute suspension. It assumes the spherical particles settle without causing turbulence, without particle to particle collision, and without chemical or physical attraction or affinity for the dispersion medium. Obviously, the typical pharmaceutical suspension contains particles that are irregularly shaped with a range of sizes, settling results in both turbulence and collision, and there is a reasonable affinity between the particles and suspension medium. However, the basic concepts of the equation offer an indication of the important variables for suspension of the particle and clues to formulation adjustments to decrease the rate of particle sedimentation.

Clearly, the sedimentation rate of large particles is greater than smaller particles, assuming all other factors remain constant. A slower rate of settling can be achieved by reducing particle size. Density also has a direct relationship with sedimentation rate: dense particles settle more rapidly than less

Table 24-7. Examples of Commercial Suspensions

Product Name	Manufacturer	Active Ingredient & Dose	Indication
Carafate	Aventis	1 g sucalfate / 10 mL	Antilulcer
Maalox	Novartis	225 mg aluminum hydroxide and 200 mg magnesium hydroxide / 5 mL	Antacid
Mepron	GlaxoSmithKline	750 mg atovaquone / 5 mL	Antiprotozoal
Mylanta Liquid	J&J-Merck	200 mg Aluminum Hydroxide, 200 mg Magnesium Hydroxide, and 20 mg simethicone / 5 mL	Antacid
Nystatin	Teva	100,000 units mycostatin / mL	Antifungal
Pepto-Bismol Liquid	Proctor & Gamble	262 mg bismuth subsalicylate / 15 mL	Antidiarrheal
Pred-G ophthalmic suspension	Allegan	0.3% gentamicin and 1.0% prednisolone acetate	Topical anti-inflammatory/anti-infective
Viramune	Boehringer Ingelheim	50 mg of nevirapine / 5 mL	Antiviral

dense particles. Most pharmaceutical suspensions are aqueous, and the density of the particles is greater than water, a desirable feature, since, if they were less dense, they would float, making a uniform product difficult to achieve.

The sedimentation rate is indirectly related to the medium viscosity, allowing the pharmaceutical scientist to manipulate settling by adjusting the viscosity of the medium. Settling is reduced by increasing the viscosity of the dispersion medium. One must keep in mind, a very high viscosity is not desirable, because it pours with difficulty and is equally difficult to re-disperse. The viscosity characteristics of a suspension may be altered not only by the vehicle used, but also by the solids content. As the proportion of solid particles is increased in a suspension, so is the viscosity. In most cases, the physical stability of a pharmaceutical suspension is adjusted by the dispersed phase, rather than through the dispersion medium. The dispersion medium supports the adjusted dispersed phase.

The most important consideration in formulation of suspensions is the size of the drug particles. In most pharmaceutical suspensions, the particle diameter is between 1 and 50 μm . The reduction in the particle size is beneficial to the stability of the suspension in that the rate of sedimentation is reduced as the particles are decreased in size. The reduction in particle size produces slow, more uniform rates of settling. However, reduction of the particle size to too great a degree of fineness should be avoided, since fine particles have a tendency to form a compact cake upon settling. The result may be that the cake resists breakup upon shaking, and forms rigid aggregates of particles. Particle shape can also affect caking and product stability.⁹⁰

Actions must be taken to prevent the agglomeration of particles into larger crystals or into masses, to avoid the formation of a cake. A common method to prevent rigid cohesion of small particles is through the intentional formation of a less rigid or loose aggregation of the particles by particle-to-particle bonding forces. An aggregation of this type is called a "floc" or "floccule" in which particles form a lattice structure that resists complete settling and compaction. Flocs form a higher sediment volume than unflocculated particles, and the loose structure permits the aggregates to break up easily and redistribute with agitation. There are several methods of preparing flocculated suspensions, the choice depending on the drug and type of product desired. For example, clays, such as bentonite, are commonly used as flocculating agents in oral suspensions. The structure of bentonite and of other clays assists the suspension by helping support the floc once formed. When clays are unsuitable, as in a parenteral suspension, a floc of the dispersed phase can be produced by an alteration in the pH of the preparation, to a region of low drug solubility. Electrolytes can also act as flocculating agents by reducing electrostatic interactions between the particles. Nonionic and ionic surfactants can also induce particle flocculation and increase the sedimentation volume.

Particle growth or Ostwald ripening is also a destabilizing process, resulting from temperature fluctuations during storage. Temperature fluctuations may change particle size distribution and polymorphic form of a drug, if the solubility of the drug is temperature dependent. For example, if the temperature is raised, drug crystals may dissolve and form a supersaturated solution, which favors crystal growth on cooling. As the dissolved drug crystallizes out of solution, it will preferentially occur on the surface of a crystal in the suspension.

SUSPENSION INGREDIENTS

The external phase is water for oral preparations; however, other polar liquids, such as glycerin or alcohol, may be considered to control solubility, stability, and taste. The selection of the external phase is based on taste, viscosity, density, and stability. Non-polar liquids, such as aliphatic hydrocarbons and fatty esters, may be considered, if the preparation is used for external purposes.

The main ingredients in a suspension are the drug and functional excipients that wet the drug, influence flocculation, control viscosity, adjust pH, and the external medium, usually water. In addition, flavoring, sweetening, and coloring agents and preservatives are employed. A wetting agent is a surfactant with an HLB value between 7 and 9. Surfactants with higher HLB values are sometimes recommended, such as polysorbates and poloxamers. They are employed at a low concentration (0.05–0.5%) to allow the displacement of air from hydrophobic material and permit the liquid, usually water, to surround the particles. If it is desirable to flocculate the particles, then flocculating agents are employed. Low concentrations, less than 1%, of electrolytes, such as sodium or potassium chloride, are employed to induce flocculation. Water soluble salts possessing divalent or trivalent ions may be considered, if the particles are highly charged.

Viscosity producing agents are polymers, including natural gums (acacia, xanthan) and cellulose derivatives, such as sodium carboxymethylcellulose and hydroxypropyl methylcellulose. These excipients are used at low concentrations to function as protective colloids, but, at higher concentrations, they function as viscosity increasing agents. At higher viscosity, the rate of settling of deflocculated particles is decreased, providing additional stability to the flocculated suspension. The choice of an appropriate viscosity agent depends on the use of the product (external or internal), the processing equipment, and the duration of storage. Suspension preparations for internal use exhibiting good flow and suspending properties often contain sodium carboxymethylcellulose 2.5%, tragacanth 1.25%, or guar gum 0.5%. For external applications, Carbopol polymers have been successfully used. Other common viscosity producing agents include acacia, methylcellulose, sodium alginate, or tragacanth.

Ideally, a suspension should be stable over a wide pH range. The chemical and/or physical stability of an active compound may occasionally require the pH of the medium be maintained within a specified range. Buffers must be carefully considered, so they produce their intended effects without interference with other ingredients in the formulation. Buffers can influence the solubility of the active, preservative ionization and its activity and ionic viscosity agents.

PREPARATION OF SUSPENSIONS

The preparation of suspensions involves several steps; the first is to obtain particles of the proper size, typically in the lower micrometer range. Oral preparations should not feel gritty; topical preparations should feel smooth to the touch; and injectables should not produce tissue irritation. Particle size and distribution should also be considered in terms of bioavailability or, from an *in vitro* perspective, the rate of release. Very small particles, less than 1 μm , will have a higher solubility than larger particles but also have a faster rate of dissolution. Thus, particle size of the dispersed solid in a suspension can influence the rate of sedimentation, flocculation, solubility, dissolution rate, and, ultimately, bioavailability.

Particle size reduction is accomplished by dry milling prior to the incorporation of the dispersed phase into the dispersion medium. Milling is the mechanical process of reducing particle size, which may be accomplished by a number of different types of machines. Impact mills and Hammer mills grind the powders by impact (Figs. 24-8, 24-9, and 24-10). Centrifugally rotating hammers, pins, or blades contact the particles and direct them against a screen, typically in the range of 4–325 mesh. The particles are forced through the screen, which regulates final particle size at the outlet of the milling chamber. The blade and screen act in conjunction to determine final product sizing, typically in the range of 10–50 μm .

Fluid energy or jet mills produce particles under 25 μm through violent turbulence in high velocity air (Fig. 24-11). Compressed air forms a high speed, jet stream, which passes the feed funnel and draws powders into grinding chamber.

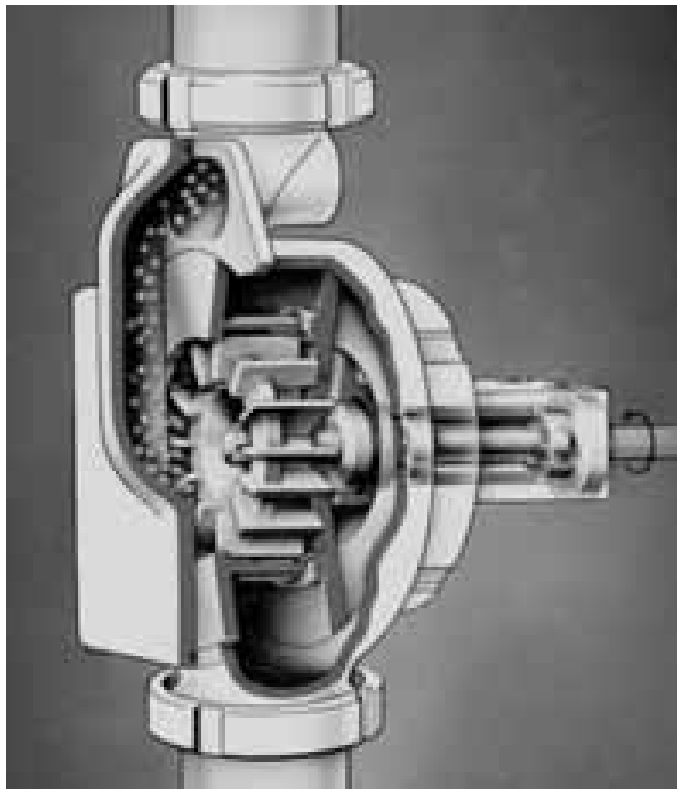


Figure 24-8. A cross section of an impact mill. (Courtesy of Hosokawa Micron Powder Systems.)

Pulverizing nozzles are installed around the grinding chamber and inject additional high speed air into the grinding chamber in a rotational direction. The centrifugal air flow accelerates particles and reduces particle size by particle-to-particle impaction and friction. The air flow drives large particles toward the perimeter, but small particles move toward the center, where they exit through the outlet.

A ball mill contains a number of steel or ceramic balls in a rotating drum. The balls reduce the particle size to a 20–200 mesh by both attrition and impact. Roller mills have two or more rollers that revolve at different speeds, and the particles are reduced to a mesh of 20–200 by means of compression and a shearing action. (See Chapter 23 (Powders) for a more detailed discussion on particle size reduction of solids.)

In the pharmacy, ceramic mortar and pestle are better at grinding and reducing particle size than glass. After reducing particle size, the drug powder is wetted thoroughly with a small quantity of water miscible solvent, such as glycerin or alcohol, which reduces the interfacial tension. The suspending agent in the aqueous medium is then added. Alternately, the suspending agent can be triturated with the drug particles, using a small quantity of glycerin or alcohol and then brought up to volume with the diluent water and triturated to a smooth uniform product.

On a large scale, the fine drug particles are treated with a small portion of water that contains the wetting agent and allowed to stand for several hours to release entrapped air. At the same time, the suspending agent should be dissolved or dispersed in the main portion of the external phase and allowed to stand until complete hydration takes place. Subsequently, wetted drug particles should be added slowly to the main portion of the dissolved suspending agent. Other excipients, such as electrolytes or buffers, should be carefully introduced. The preservatives, flavoring agents, and coloring agents are added last. Finally, the formulation is processed with homogenizers, ultrasonic devices, or colloid mills to produce a uniform product.

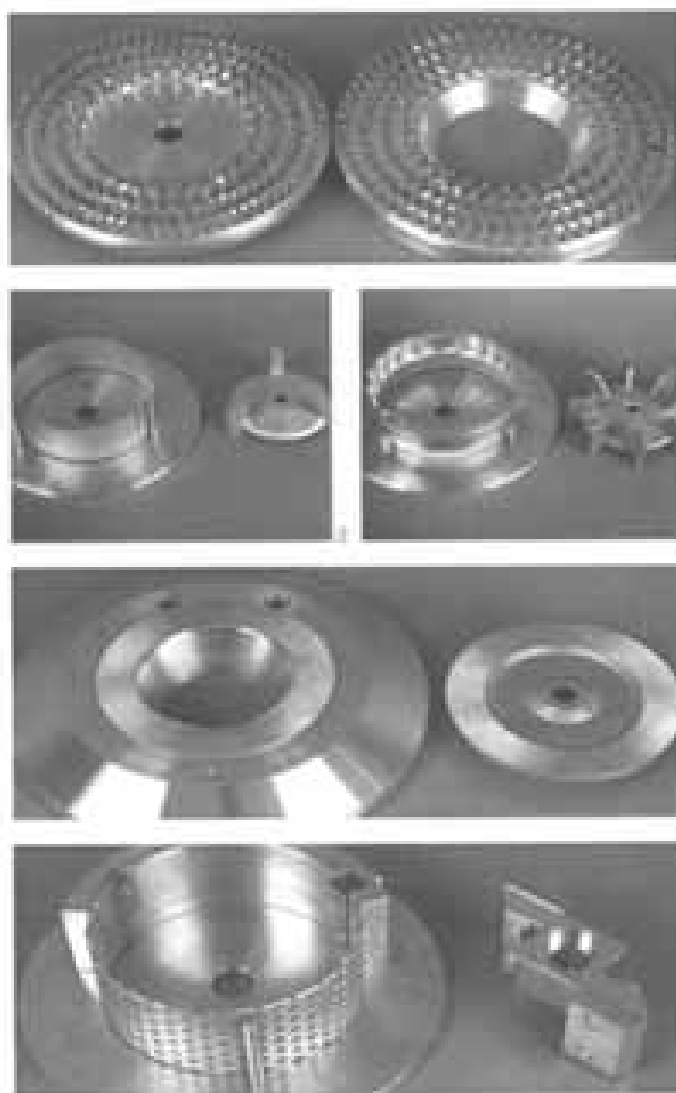


Figure 24-9. Multiple rotor options available for an impact mill. (Courtesy of Hosokawa Micron Powder Systems.)

A procedure for the preparation of Trisulfapyrimidines Oral Suspension is subsequently given.

Trisulfapyrimidines Oral Suspension

Veegum	1.00 g
Syrup USP	90.60 g
Sodium Citrate	0.78 g
Sulfadiazine	2.54 g
Sulfamerazine	2.54 g

Add the Veegum slowly and with continuous stirring to the syrup. Incorporate the sodium citrate into the Veegum–syrup mixture. Premix the sulfa drugs, add to the syrup, stir, and homogenize. Add sufficient 5% citric acid to adjust the pH of the product to 5.6. A preservative and a flavoring agent may be added to the product.

QUALITY CONSIDERATIONS

The quality of the suspension can be determined in a number of ways. Particle size, particle size distribution, and particle shape are often determined using photo microscopy or laser light diffraction techniques. Physical stability, the degree of settling or flocculation, may be determined using a device to measure the zeta potential. Viscosity may be determined by instruments,

THE FEED THROAT
Introduces material on a tangential path to the comminuting chamber

BLADE PROFILE
Helps determine degree of reduction based material being processed

SCREEN TYPE
Helps regulate particle output within a specified size range

ROTOR SPEED
Works with screen to regulate particle output within the size range

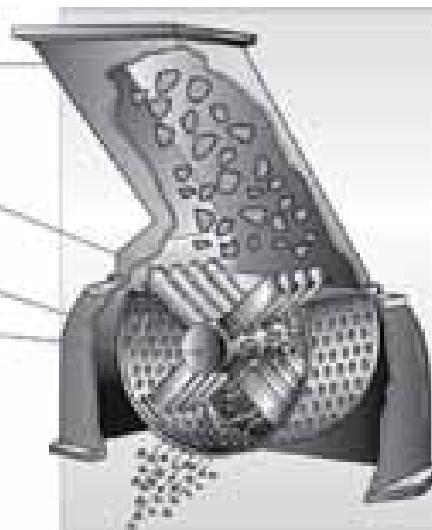


Figure 24-10. Material Flow through a Hammer Mill. Hammer mills operate by feeding material uniformly into a chamber in which a rotating blade assembly reduces the particles of the material by cutting or impacting them. The material discharges through a screen, which regulates final particle size at the outlet of the milling chamber. (Courtesy of The Fitzpatrick Company.)

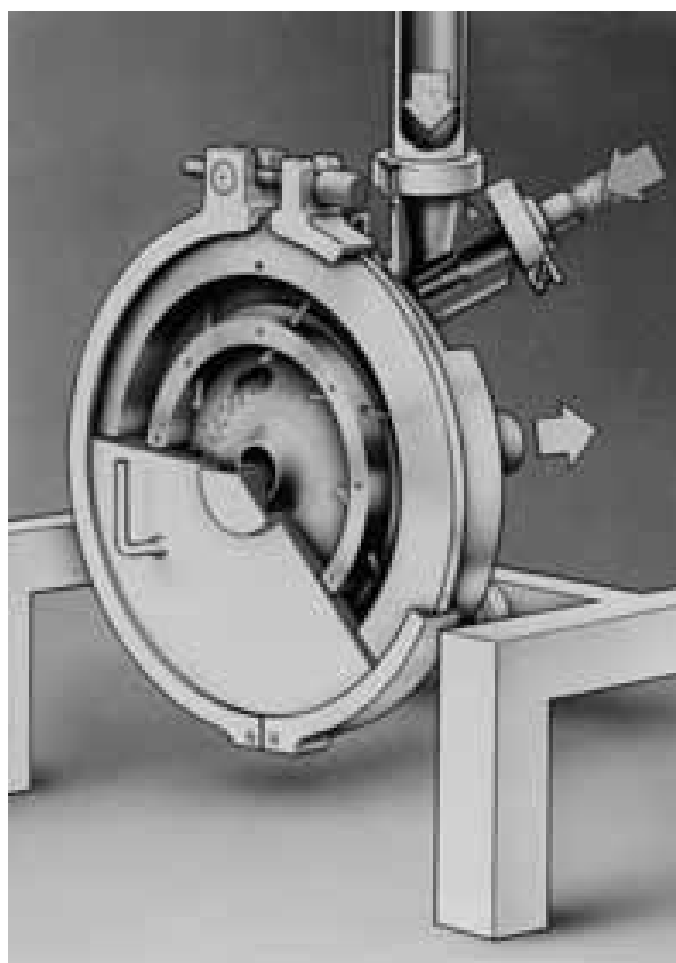


Figure 24-11. Schematic of particle size reduction in a fluid energy or jet mill. (Courtesy of Hosokawa Powder Systems.)

such as the Brookfield viscometer, or of the cone and plate configuration. Microbiological, as well as stability testing, according to ICH guidelines should be performed to determine the efficiency of the preservative and the appropriateness of the formulation with respect to time, temperature, and relative humidity.⁹¹

EXTEMPORANEOUS PREPARATIONS FROM TABLETS AND CAPSULES

Occasionally, it is necessary to prepare a liquid formulation of a drug to meet certain patient requirements. Consequently, patients who are unable to swallow solid medications and require a different route of administration or different dosing strength present a special need. Thus, the pharmacist may have to extemporaneously compound a liquid product. If the pure drug is available, it should be used to prepare the liquid dosage form. If it is necessary to prepare a liquid dosage form from tablets or capsules, a suspension is formed, if either the drug or one of the excipients in the tablets or capsules is insoluble. Insoluble excipients in these dosage forms include disintegrants, lubricants, glidants, colors, diluents, and coatings. Consequently, although the drug may be soluble in water, many excipients are not. It is preferable to use the contents of capsules, or tablets that are not coated. If coated, tablets with a water soluble coat are preferred to those with functional enteric coatings and the like. In any case, the contents of the capsules or the tablets should be ground finely with a ceramic mortar and pestle and then wetted using alcohol or glycerin.

Preservatives may be included in the liquid formulation to enhance the stability. However, preservatives have been found to cause serious adverse effects in infants. Benzyl alcohol should be omitted from neonatal formulations, because it can cause a gasping syndrome characterized by a deterioration of multiple organ systems and eventually death. Propylene glycol has also been implicated to cause seizures and stupor in some preterm infants. Thus, formulations for neonates should be purposely kept simple, not compounded to supply more than just a few days of medicine.

Finally, it may be desirable to use a hand homogenizer to prepare a more suitable product. Some drugs formulated in this manner include clonidine hydrochloride and simple syrup,⁹² cefuroxime axetil in an orange syrup vehicle,⁹³ and famotidine in cherry syrup.⁹⁴ Many other examples may be found in current hospital and community pharmacy journals, such as the *American Journal of Hospital Pharmacy*, *Canadian Journal of Hospital Pharmacy*, *U.S. Pharmacist*, *International Journal of Pharmaceutical Compounding*, and *Drug Development and Industrial Pharmacy*. Frequently, stability data and, occasionally, bioavailability and/or taste data are provided.

To minimize stability problems of the extemporaneously prepared product, it should be placed in air tight, light resistant containers and stored in the refrigerator by the patient.

Because it is a suspension, the patient should be counseled to shake it well prior to use and to be aware of any change that might indicate a stability problem with the formulation.

Tortorici reports an example of an extemporaneous suspension of cimetidine tablets that retained its potency at 40° over 14 days.⁹⁵ Twenty-four, 300 mg cimetidine tablets are compounded with 10 mL of glycerin and 120 mL of simple syrup. The tablets are triturated to a fine powder using a mortar, the mixture is levigated with the glycerin, and the simple syrup added. The suspension is mixed well, placed in a blender until smooth, and then refrigerated.

SUSTAINED RELEASE SUSPENSIONS

Sustained release suspensions represent a very specialized class of preparation. Sustained release, oral suspensions with morphine,⁹⁶ non-steroidal anti-inflammatory agents,⁹⁷ and other drugs⁹⁸ have been described in the literature. However, limited commercial success has been achieved, due to the difficulty in maintaining the stability. Formulation research for sustained release suspensions has focused on the similar technologies used in preparing sustained release tablets and capsules. Celltech licenses the Tussionex Pennkinetic system, which uses a combination of ion exchange resin and particle coating.⁹⁹ This novel system exploits the likelihood of complexation between ionic drugs and ion-exchange resins, which are then coated with ethyl cellulose. When administered orally, the coated particles with encapsulated drug adsorbed onto the resin are slowly released by an ion exchange process.

Durect markets the SABER system for sustained release suspension applications. SABER uses sucrose acetate isobutyrate (SAIB), a non-polymeric, non-water soluble high-viscosity liquid carrier material (>5000 cPs. at 37°C), to provide controlled release of active ingredients.¹⁰⁰ The drug is mixed with a small amount of a pharmaceutically acceptable solvent to form a low viscosity solution or suspension, which is then mixed with the high viscosity carrier. The resulting suspension can be administered via injection, orally, or as an aerosol, forming an adhesive, biodegradable depot upon contact with tissues. After administration of the SABER formulation, the solvent diffuses away, leaving a viscous, adhesive matrix of the three components—SAIB, drug, and any additives. The release rate can be easily modified by the ratio of non-polymeric, non-water soluble high-viscosity liquid carrier material present in the formulation. Extended systemic and local delivery for durations of 1 day to 3 months from a single injection has been demonstrated.

GELS AND MAGMAS

Gels are defined by the USP as:

Semisolid systems consisting of either suspensions made up of small inorganic particles or large organic molecules interpenetrated by a liquid. Where the gel mass consists of a network of small discrete particles, the gel is classified as a two-phase system. In a two-phase system, if the particle size of the dispersed phase is relatively large, the gel mass is sometimes referred to as a magma. Both gels and magmas may be thixotropic, forming semisolids on standing and becoming liquid on agitation.

Single-phase gels consist of organic macromolecules uniformly distributed throughout a liquid in such a manner that no apparent boundaries exist between the dispersed macromolecules and the liquid. Single-phase gels may be made from synthetic macromolecules or from natural gums. The latter preparations are also called mucilages. Although these gels are commonly aqueous, alcohols and oils may be used as the continuous phase. For example, mineral oil can be combined with a polyethylene resin to form an oleaginous ointment base.

Gels can be used to administer drugs topically or into body cavities.

Gels are also defined as semi-rigid systems in which the movement of the dispersing medium is restricted by an interlacing three dimensional network of particles or solvated

macromolecules in the dispersed phase. Physical and/or chemical cross linking may be involved. The interlacing and consequential internal friction is responsible for increased viscosity and the semisolid state.

Some gel systems are clear, and others are turbid, since the ingredients involved may not be completely soluble or insoluble or they may form aggregates, which disperse light. The concentration of the gelling agents is less than 10%, and in 0.5-2.0% range. Gels in which the macromolecules are distributed throughout the liquid in such a manner that no apparent boundaries exist between them and the liquid are called single-phase gels. In instances in which the gel mass consists of floccules of small distinct particles, the gel is classified as a two-phase system and frequently called a magma or a milk. Gels and magmas are considered colloidal dispersions, since they each contain particles of colloidal dimension.

Different types of colloidal dispersions have been given specific names. For instance, “sol” is a general term designating a dispersion of a solid substance in a liquid, a solid, or a gaseous dispersion medium. However, more often than not, it is used to describe the solid liquid dispersion system. A prefix, such as hydro- for water (hydrosol) or alco- for alcohol (alcosol), is used to specify the medium. Similarly, aerosol has similarly been developed to indicate a dispersion of a solid or a liquid in a gaseous phase.

The accepted size range for a substance “colloidal” is when particles fall between 1 nm and 0.5 μ m. One difference between colloidal dispersions and true solutions is the larger particle size of the dispersed phase in colloidal systems. The optical properties of the two systems are also different. True solutions do not scatter light and, therefore, appear clear, but colloidal dispersions contain discrete particles scatter light.

Gelling Agents

Several compendial materials function as gelling agents, including acacia, alginic acid, bentonite, carbomer, carboxymethylcellulose sodium, cetostearyl alcohol, colloidal silicon dioxide, ethylcellulose, gelatin, guar gum, hydroxyethylcellulose, hydroxypropyl cellulose, hydroxypropyl methylcellulose, magnesium aluminum silicate, maltodextrin, methylcellulose, polyvinyl alcohol, povidone, propylene carbonate, propylene glycol alginate, sodium alginate, sodium starch glycolate, starch, tragacanth, and xanthan gum.

Alginic acid is refined from seaweed. It is a tasteless, practically odorless, white to off-white colored, fibrous powder. It is used in concentrations between 1 and 5% as a thickening agent and swells in water to about 200 times its own weight without dissolving. Alginic acid can be cross linked by addition of calcium salts, resulting in substantially higher viscosity. Sodium alginate produces a gel at concentrations up to 10%. Aqueous preparations are most stable between pH values of 4–10; below pH 3, alginic acid is precipitated. Sodium alginate gels for external use should be preserved.

Carbomer resins are high molecular weight, acrylic acid-based polymers. The pH of 0.5% and 1.0% aqueous dispersions are 2.7–3.5 and 2.5–3.0, respectively. There are many carbomer resins, with viscosity ranges available from 0 to 80,000 cPs., depending on the pH to which it is neutralized. In addition to thickening, suspending and emulsifying in both oral and topical formulations, carbomers are also used to provide sustained release properties in both the stomach and intestinal tract for commercial products. Alcohol is often added to carbomer gels to decrease their viscosity. Carbomer gel viscosity is also dependent on the presence of electrolytes and the pH. A rubbery mass forms, if greater than 3% electrolytes are added. Carbomer preparations are primarily used in aqueous systems, although other liquids can be used. In water, a single particle of carbomer will wet very rapidly, but, like many other powders, carbomer polymers tend to form clumps of particles, when haphazardly dispersed in polar solvents. Rapid dispersion of carbomers can be achieved by adding the powder very slowly into the vortex

of the liquid that is very rapidly stirred. A neutralizer is added to thicken the gel after the carbomer is dispersed. Sodium hydroxide or potassium hydroxide can be used in carbomer dispersions containing less than 20% alcohol. Triethanolamine will neutralize carbomer resins containing up to 50% ethanol.

Carboxymethylcellulose (CMC) produces gels when used in concentrations of 4–6% of the medium viscosity grade. Glycerin may be added to prevent drying. Precipitation will occur at pH values less than 2; it is most stable at pH levels between 2 and 10, with maximum stability at pH 7 to 9. It is incompatible with ethanol. Sodium carboxymethylcellulose (NaCMC) is soluble in water and should be dispersed with high shear in cold water, before the particles hydrate and swell. Once the powder is well dispersed, the solution is heated with moderate shear to about 60°C for fastest dissolution. These colloidal dispersions are sensitive to pH, and the viscosity of the product decreases below pH 5 or above pH 10.

Tragacanth gum has been used to prepare gels that are stable at a pH range of 4–8. These gels must be preserved or sterilized by autoclaving. Tragacanth often lumps when added to water, thus, aqueous dispersions are prepared by adding the powder to rapidly mixed water. Also, lumps are also prevented by wetting the gum with ethanol, glycerin, or propylene glycol.

Colloidal silicon dioxide can be used to prepare transparent gels, when used with other ingredients of similar refractive index. Colloidal silicon dioxide adsorbs large quantities of water without liquefying, and its viscosity is largely independent of temperature. Changes in pH affect the viscosity: it is most effective at pH values up to about 7.5. Colloidal silicon dioxide (fumed silica) forms a hydrophobic gel, when combined with 1-dodecanol and n-dodecane. These are prepared by adding the silica to the vehicle and sonicating for about 1 minute to obtain a uniform dispersion, sealing, and storing at about 40°C overnight.

Gelatin gels are prepared by dispersing gelatin in hot water followed by cooling. Alternatively, gelatin can be wetted with an organic liquid, such as ethyl alcohol or propylene glycol, followed by the addition of the hot water and cooling. Magnesium aluminum silicate forms thixotropic gels at concentrations of about 10%. The material is inert and has few incompatibilities but is best used above pH 3.5. It may bind to some drugs and limit their availability.

Methylcellulose forms gels at concentrations up to about 5%. Since methylcellulose hydrates slowly in hot water, the powder is dispersed with high shear at 80–90°C in a portion of water. Once the powder is finely dispersed, the remaining water is added with moderate stirring. Alcohol or propylene glycol is often used to help wet the powders. High electrolyte concentrations will salt out the polymer, ultimately precipitating the polymer.

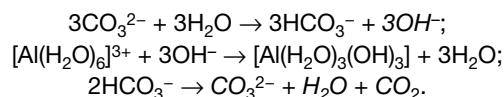
Poloxamer gels are made from selected forms of polyoxyethylene-polyoxypropylene copolymers in concentrations ranging from 15 to 50%. Poloxamers are white, waxy, free-flowing granules that are practically odorless and tasteless. Aqueous solutions of poloxamers are stable in the presence of acids, alkalis, and metal ions. Polyvinyl alcohol (PVA) is used at concentrations of about 2.5% in the preparation of various jellies, which dry rapidly when applied to the skin. Borax is often used to gel PVA solutions. For best results, disperse PVA in cold water, followed by hot water. It is less soluble in the cold water.

Povidone, in the higher molecular weight forms, can be used to prepare gels in concentrations up to about 10%. It has the advantage of being compatible in solution with a wide range of inorganic salts, natural and synthetic resins, and other chemicals. It has also been used to increase the solubility of a number of poorly soluble drugs.

Two-Phase Gels

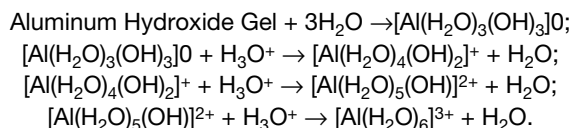
Two-phase gels containing bentonite may be used as a base for topical preparations, such as plaster and ointment. Aluminum

Hydroxide Gel, USP, is an example of a two phase gel. The USP states that “Aluminum Hydroxide Gel is a suspension of amorphous aluminum hydroxide in which there is a partial substitution of carbonate for hydroxide.” The gel is prepared by the interaction of a soluble aluminum salt, such as a chloride or sulfate, with ammonia solution, sodium carbonate, or bicarbonate. The reactions that occur during the preparation are:



The physical and chemical properties of the gel will be affected by the order of addition of reactants, pH of precipitation, temperature of precipitation, concentration of the reactants, the reactants used, and the conditions of aging of the precipitated gel.

Aluminum Hydroxide Gel is soluble in acidic (or very strongly basic) media. The mechanism in acidic media is:



It is unlikely that the last reaction proceeds to completion, because the activity of the gel is controlled by its insolubility. Further, because a certain quantity of insoluble gel is always available, the neutralizing capability of the gel extends over a considerable period of time.

Aluminum hydroxide gels may also contain peppermint oil, glycerin, sorbitol, sucrose, saccharin, and various preservatives. Sorbitol improves the acid consuming capacity by inhibiting a secondary polymerization that takes place on aging. In addition, polyols, such as mannitol, sorbitol, and inositol, have been shown to improve the stability of aluminum hydroxide and aluminum hydroxycarbonate gels.¹⁰¹

Single-Phase Gels

Single-phase gels are used more frequently in pharmacy for several reasons: semisolid state, high degree of clarity, ease of application, and ease of removal and use. The gels often provide a faster release of drug substance, independent of the water solubility of the drug, as compared to creams and ointments.

Some recent gel formulations include ophthalmic preparations of pilocarpine, carbachol, and betamethasone valerate; topical preparations for burn therapy, anti-inflammatory treatment, musculoskeletal disorders, and acne; peptic ulcer treatment with sucralfate gel; and bronchoscopy, using lidocaine. Gels may be used as lubricants for catheters and bases for patch testing, and sodium chloride gels are used for electrocardiography.

Some gel formulation examples are subsequently provided.

Methylcellulose and Carbomer Gel Base

METHYLCELLULOSE,	4000 CPS 1.0 %
CARBOMER	934 0.35 %
1 N SODIUM HYDROXIDE SOLUTION	QS TO PH 7
PROPYLENE GLYCOL	16.7 %
METHYL PARABEN	0.015%
PURIFIED WATER	QS 100%

Disperse the methylcellulose in a portion of hot (80–90°C) water. Cool to room temperature, and disperse the Carbomer 934 in the gel, using a bladed impeller. Adjust the pH of the dispersion to 7.0, by adding sufficient 1 N sodium hydroxide solution. Dissolve the methylparaben in the propylene glycol. Mix the methylcellulose, Carbopol 934, and propylene glycol fractions, using caution to avoid incorporating air.

Sodium Alginate Gel Base

SODIUM ALGINATE	10 G
GLYCERIN	10 G
METHYL HYDROXYBENZOATE	0.2 G
A SOLUBLE CALCIUM SALT (CALCIUM GLUCONATE)	0.5 G
PURIFIED WATER, TO MAKE	100 ML

Place a portion of water in a beaker and add the glycerin and preservative. Stir this solution with a high speed mixer and add the sodium alginate. The calcium salt is added next, which increases the viscosity. Continue mixing, until the preparation is homogeneous. The preparation should be stored in a tightly sealed, wide mouth jar or tube.

Carbomer Gel

CARBOMER	9342 G
TRIETHANOLAMINE	1.65 ML
METHYL PARABEN	0.2 G
PROPYL PARABEN	0.05 G
PURIFIED WATER, TO MAKE	100 ML

The parabens are dissolved in 95 mL of water with the aid of heat and allowed to cool. Carbomer 934 is added in small amounts to the solution using a high speed mixer, until a smooth dispersion is obtained. The preparation is allowed to stand, permitting entrapped air to separate. Then the neutralizing agent, triethanolamine, is added very slowly to avoid entrapping air. Finally, the remaining water is then incorporated.

LOTIONS

Lotions are not defined specifically in the USP, but a broad definition describes them as either liquid or semi-liquid preparations that contain one or more active ingredients in an appropriate vehicle. Lotions may contain antimicrobial preservatives and other appropriate excipients, such as stabilizers. Lotions are intended to be applied to the unbroken skin without friction. Lotions are suspensions of solids in an aqueous medium. Some lotions are, in fact, emulsions or solutions.

Even though lotions are applied without friction, the insoluble matter should be divided very finely. Particles approaching colloidal dimensions are more soothing to inflamed areas and effective in contact with infected surfaces. A wide variety of ingredients may be added to the preparation to produce better dispersions or to accentuate its cooling, soothing, drying, or protective properties. Bentonite is a good example of a suspending agent used in the preparation of lotions. Methylcellulose or sodium carboxymethylcellulose, for example, will localize and hold the active ingredient in contact with the affected site and be rinsed off easily with water. A formulation containing glycerin will keep the skin moist for a considerable period of time. The drying and cooling effect of a lotion may be accentuated by adding alcohol to the formula.

Dermatologists frequently prescribe lotions containing anesthetics, antipruritics, antiseptics, astringents, germicides, protectives, or screening agents to treat or prevent various types of skin diseases and dermatitis. Antihistamines, benzocaine, calamine, resorcin, steroids, sulfur, zinc oxide, betamethasone derivatives, salicylic acid, safflower oil, minoxidil, and zirconium oxide are ingredients common in lotions.

Lotions may be prepared by triturating the ingredients to a smooth paste and then adding the remaining liquid phase with trituration. High speed mixers or colloid mills produce better dispersions and, therefore, are used in the preparation of larger quantities of lotion. Calamine Lotion, USP, is the classic example of this type of preparation and consists of finely powdered,

insoluble solids held in more or less permanent suspension by the presence of suspending agents and/or surface-active agents. The formula and the method of preparation of Calamine Lotion, USP, follows.

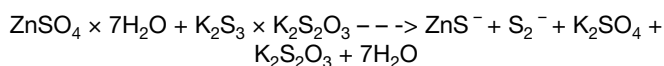
Calamine Lotion, USP

CALAMINE	80 G
ZINC OXIDE	80 G
GLYCERIN	20 ML
BENTONITE MAGMA	250 ML
CALCIUM HYDROXIDE TOPICAL SOLUTION	Q.S. 1,000 ML

Dilute the bentonite magma with an equal volume of calcium hydroxide topical solution. Mix the powder intimately with the glycerin and about 100 mL of the diluted magma, triturating until a smooth, uniform paste is formed. Gradually incorporate the remainder of the diluted magma. Finally, add enough calcium hydroxide topical solution to make 1000 mL, and shake well. If a more viscous consistency in the Lotion is desired, the quantity of bentonite magma may be increased to not more than 400 mL.

Many investigators have studied Calamine Lotion, and this has led to the publication of many formulations, each possessing certain advantages over the others, but none satisfying the collective needs of all dermatologists. Formulations containing hydrated microcrystalline cellulose and carboxymethylcellulose have a slower rate of sedimentation than the official preparation.

Although most lotions are prepared by trituration, some lotions are formed by chemical interaction in the liquid. White Lotion, USP, is an example, and the chemical reaction between Sulfurated Potash and Zinc Sulfate is as follows:

**White Lotion**

ZINC SULFATE	40 G
SULFURATED POTASH	40 G
PURIFIED WATER,	Q.S. 1,000 ML

Dissolve the zinc sulfate and the sulfurated potash separately, each in 450 mL of purified water, and filter each solution. Add slowly the sulfurated potash solution to the zinc sulfate solution with constant stirring. Then add the required amount of purified water, and mix.

Benzyl Benzoate Lotion, USP, is an example of a lotion that is also an emulsion. The formula and method of preparation are:

Benzyl Benzoate Lotion, USP

BENZYL BENZOATE	250 ML
TRIETHANOLAMINE	5 G
OLEIC ACID	20 G
PURIFIED WATERQ.S.	1,000 ML

Mix the triethanolamine with the oleic acid, add the benzyl benzoate, and mix. Transfer the mixture to a suitable container of about 2000 mL capacity, add 250 mL of purified water, and shake the mixture thoroughly. Finally, add the remaining purified water, and again shake thoroughly.

Triethanolamine forms a soap with the oleic acid and functions as the emulsifying agent to form a stable product. This type of emulsifying agent is almost neutral in water and gives a pH of about 8, thus, should not irritate the skin.

Certain lotions tend to separate or stratify on long standing, and they require a label directing they be shaken well before each use. All lotions should be labeled "For External Use Only."

Micro-organisms may grow in certain lotions, if no preservative is included. Care should be taken to avoid contaminating the lotion during preparation, even if a preservative is present.

Milk of Magnesia, USP, is a suspension of magnesium hydroxide containing approximately 80 mg of $\text{Mg}(\text{OH})_2$ per milliliter. The specifications for double strength or triple strength are that these products should contain approximately 160 mg or 240 mg of $\text{Mg}(\text{OH})_2$ per mL, respectively. It has an unpleasant, alkaline taste that can be masked with 0.1% citric acid (to reduce alkalinity) and 0.05% of a volatile oil or a blend of volatile oils. Magnesium hydroxide is prepared by the hydration of magnesium oxide.

For the most part, magmas are intended for internal use, although Bentonite Magma is used primarily as a suspending agent for insoluble substances for local application and occasionally for internal use. All magmas require "Shake Well" and "Avoid Freezing" labels.

EXTRACTS

Extraction, as the term is used pharmaceutically, involves the separation of medicinally active portions of plant or animal tissues from the inactive or inert components by using selective solvents in standard extraction procedures. The products obtained from plants are relatively impure liquids, semisolids, or powders intended only for oral or external use. These include classes of preparations known as decoctions, infusions, fluidextracts, tinctures, pilular (semisolid) extracts, and powdered extracts. Such preparations popularly have been called galenicals, after Galen, the 2nd century Greek physician.

Extraction continues to be of considerable interest to obtain improved yields of drugs derived from plant and animal sources. For example, extraction of digitalis glycosides has been carried out using super critical carbon dioxide.¹⁰² Other techniques include ultrasonics, rotary-film evaporators, hydro-distillation, liquid chromatography, multiple-solvent extraction, counter-current extraction, and gravitation dynamics.

This discussion is concerned primarily with basic extraction procedures for crude drugs to obtain the therapeutically desirable portion and eliminate the inert material by treatment with a selective solvent, known as the menstruum. Extraction differs from solution in that the presence of insoluble matter is implied in the former process. The principal methods of extraction are maceration, percolation, digestion, infusion, and decoction. The quality of the finished product can be enhanced by standardizing primary extracts and carrying out analytical assays during production on the raw materials, intermediate products, and manufacturing procedures.

The processes of particular importance, insofar as the USP is concerned, are those of maceration and percolation, as described specifically for Belladonna Extract, USP, and Cascara Sagrada Extract, USP. Most pharmacopeias refer to such processes for extraction of active principles from crude drugs. The USP provides general directions for both maceration and percolation under the heading of **Tinctures**.

Techniques of extraction continue to be investigated and applied to obtain higher yields of the active substance from natural sources. Some of these methods include the use of different grinding and shearing processes of plants, use of specific membranes for extraction, and different extraction procedures, such as distillation, digestion, percolation, and microwaves.

MACERATION

In this process the solid ingredients are placed in a stoppered container with 750 mL of the prescribed solvent and allowed to stand for a period of at least 3 days in a warm place with frequent agitation, until soluble matter is dissolved. The mixture is filtered and, after most of the liquid has drained, the residue on the filter is washed with sufficient quantity of the prescribed solvent or solvent mixture; the filtrates are combined to produce 1000 mL.

PERCOLATION

The ground solids are mixed with the appropriate quantity of the prescribed solvent to make it evenly and uniformly damp. It is allowed to stand for 15 minutes, then transferred to a percolator and packed. Sufficient prescribed solvent is added to saturate the solids. The top is placed on the percolator, and, when the liquid is about to drip from the apparatus, the lower opening is closed. The solids are allowed to macerate for 24 hours or for the specified time. If no assay is directed, the percolation is allowed to proceed slowly or at the specified rate, gradually adding sufficient solvent to produce 1000 mL of solution. If an assay is required, only 950 mL of percolate are collected and mixed and a portion assayed as directed. The rest of the percolate is diluted with the solvent to produce a solution that conforms to the required standard and then mixed.

DIGESTION

This is a form of maceration in which gentle heat is used during the process of extraction. It is used when moderately elevated temperature is not objectionable and the solvent efficiency of the menstruum is increased thereby.

INFUSION

An infusion is a dilute solution of the readily soluble constituents of crude drugs. Fresh infusions are prepared by macerating the solids for a short period of time with either cold or boiling water. The USP has not included infusions for some time.

DECOCTION

This once popular process extracts water soluble and heat stable constituents from crude drugs, by boiling in water for 15 minutes, cooling, straining, and passing sufficient cold water through the drug to produce the required volume.

EXTRACTIVE PREPARATIONS

After a solution of the active constituents of a crude drug is obtained by maceration or percolation, it may be ready for use as a medicinal agent, as with certain tinctures or fluidextracts, or it may be processed further to produce a solid or semisolid extract.

TINCTURES

Tinctures are defined in the USP as alcoholic or hydroalcoholic solutions prepared from vegetable materials or from chemical substances, an example of the latter being Iodine Tincture. Traditionally, tinctures of potent vegetable drugs essentially represent the activity of 10 g of the drug in each 100 mL of tincture, the potency being adjusted following assay. Most other tinctures of vegetable drugs represent the extractive from 20 g of the drug in 100 mL of tincture.

The USP, specifically, describes two general processes for preparing tinctures, one by percolation and the other by maceration. Percolation includes a modification, so tinctures that require assay for adjustment to specified potency, thus, may be tested before dilution to final volume. Belladonna Tincture, USP, is prepared in this manner. Compound Benzoin Tincture, USP, and Sweet Orange Peel Tincture, USP, are prepared by the maceration procedure.

FLUID EXTRACTS

The USP defines fluidextracts as liquid preparations of vegetable drugs, containing alcohol as a solvent or as a preservative, or both, so made that, unless otherwise specified in an individual monograph, each milliliter contains the therapeutic constituents of 1 g of the standard drug that it represents.

EXTRACTS

Extracts are defined in the USP as concentrated preparations of vegetable or animal drugs obtained by removal of the active constituents of the respective drugs with suitable menstrua,

evaporation of all or nearly all of the solvent, and adjustment of the residual masses or powders to the prescribed standards. There are three forms of extracts: semiliquids or liquids of syrupy consistency, plastic masses (known as pilular or solid extracts), and dry powders (known as powdered extracts). Extracts, as concentrated forms of the drugs from which they are prepared, are used in a variety of solid or semisolid dosage forms. The USP states that pilular extracts and powdered extracts of any one drug are interchangeable medicinally, but each has its own pharmaceutical advantages. Pilular extracts, called so because they are of a consistency to be used in pill masses and made into pills, are also suited for use in ointments and suppositories. Powdered extracts are better suited for incorporation into a dry formulation, as in capsules, powders, or tablets. Semiliquid extracts, or extracts of a syrupy consistency, may be used in the manufacture of some pharmaceutical preparations.

Most extracts are prepared by extracting the drug by percolation. The percolate is concentrated by distillation under reduced pressure. The use of heat is avoided where possible, due to potential injurious effect on active constituents. Powdered extracts made from drugs that contain inactive oily or fatty matter may have to be defatted or prepared from defatted drug. Pure Glycyrrhiza Extract, USP, is an example of a pilular extract and Belladonna Extract, USP, is an examples of a powdered extract.

FORTIFIED BEVERAGES

Our increasingly health conscious population has become more accustomed to dietary supplementation. Dietary supplements have, traditionally, been in the form of tablets and capsules that provide vitamins, minerals, herbs, and other naturally occurring ingredients. In the last decade, dietary supplementation has become available in many new forms, including beverages, shakes, powders, bars, and gels. These products are marketed as meal replacement shakes (Slim-Fast), vitamin fortified beverages (VitaminWater, LifeWater, etc.), protein and fiber powders for reconstitution (Muscle Milk), sports/electrolyte replacement drinks (Gatorade), and energy drinks (Redbull, Monster).

Currently, the Food and Drug Administration (FDA) regulates the sale and use of dietary supplements as a distinct subset of foods under the Federal Food, Drug and Cosmetic Act. Congress enacted the Dietary Supplement Health and Education Act of 1994, due to pressure from the dietary supplement industry to curtail FDA regulation and consumer advocacy groups desiring more quality standards. This legislation is intended to balance consumer access and the need for regulation to ensure product safety. The Act essentially confirms the status of all dietary supplements as foods and subjects such products to regulations accordingly.

Several of these products have excellent regulatory compliance records and quality standards, including confirmed expiration dates determined by stability testing. Other products have not had satisfactory compliance records and quality standards. Companies have been guilty of misleading labeling, unsubstantiated marketing claims, and excessive quantities of certain ingredients (e.g., sugar, caffeine). In one instance, a product marketed as a vitamin fortified beverage and a healthy alternative to a daily multivitamin was found to contain 85 g of sugar per serving, about 30% more sugar than a typical soft drink. Clinical nutritionists and consumer advocacy groups have expressed concern that some of these products may be more marketing than substance, suggesting scrutiny of the ingredients and claims associated with them.^{103,104} As a result, consumers should carefully review the labels of these products and exercise care in their selection.

REFERENCES

- Wang J *et al.* *Herald of Medicine* 2003; 22(Sept): 642–643.
- Schmidt-Nawrot J. *Pharm Ind* 2000; 62(6): 464–469.
- Eisinger HJ. *Pharm Ind* 2000; 62(6): 469–473.
- Pfafflin A. *Pharm Ind* 2000; 62(3): 223.
- Woiwode W, Huber S. *Pharm Ind* 2000; 62(5): 377–381.
- Connors KA *et al.* *Chemical Stability of Pharmaceuticals: A Handbook for Pharmacists*, 2nd edn. New York: Wiley-Interscience, 1986.
- Kurup T, Wan L. *Pharm J* 1986; 37: 761.
- Novack GD, Evans R. Commercially available ocular hypotensive products: preservative concentration, stability, storage, and in-life utilization. *J Glaucoma* 2001; 10(6): 483–486.
- US Pharmacopeial Convention. *USP27–NF22*. Rockville MD: US Pharmacopeial Convention, 2004.
- Sutton SVW, Porter D. Development of the antimicrobial effectiveness test as USP chapter <51>. *J Pharm Sci Technol* 2002; 56(6): 300–311.
- Clesceri LS *et al.* eds. *Standard Methods for Examination of Water & Wastewater*, 20th edn. Washington DC: American Public Health Association, 1998.
- Moll F *et al.* *Pharm Ind* 1997; 59(3): 258–264.
- Coates D. *Manuf Chem Aerosol News* 1974; 45(Jan): 19–20.
- Steinberg DC. Preserving foundations. *Cosmet Toiletries* 1995; 110(Feb): 71–73, 76.
- Koch CS. *Parfuem Kosmet* 1994; 75(Jan): 6, 18, 21.
- Bruch CW. *Drug Cosmet Ind* 1976; 118(June): 49–53, 161–162.
- Ma MH *et al.* Interaction of methylparaben preservative with selected sugars and sugar alcohols. *J Pharm Sci* 2002; 91(7): 1715–1723.
- Maa YF, Hsu CC. Aggregation of recombinant human growth hormone induced by phenolic compounds. *Int J Pharm* 1996; 140(2): 155–168.
- Sealzo M *et al.* Study of interaction effects of polyacrylic acid polymers (carbopol 940) on antimicrobial activity of methyl parahydroxybenzoate against some gram-negative, gram-positive bacteria and yeast. *J Pharm Pharmacol* 1996; 48(11): 1201–1205.
- Coates D. *Manuf Chem Aerosol News* 1973; 44(Oct): 34–37.
- Coates D. *Manuf Chem Aerosol News* 1973; 44(Aug): 41–42.
- Hurwitz SJ, McCarthy TJ. The effect of pH and concentration on the rates of kill of benzoic acid solutions against *E. coli*. *J Clin Pharm Ther* 1987; 12(2): 107–115.
- Boussard P *et al.* In vitro modification of antimicrobial efficacy by protamine. *Int J Pharm* 1991; 72(1): 51–55.
- Gionchetti P *et al.* Retrograde colonic spread of a new mesalazine rectal enema in patients with distal ulcerative colitis. *Aliment Pharmacol Ther* 1997; 11(4): 679–684.
- Matula C *et al.* Salivary levels of gramicidin after use of a tyrothricin-containing gargle/mouth-wash and tyrothricin lozenges. *Int J Clin Pharmacol Res* 1988; 8(4): 259–261.
- Breytenbach HS. *Curr Ther Res-Clin Exp* 1979; 26(5): 640–643.
- Allen LV. *US Pharmacist* 1990; 15(August): 88–90.
- Davis CC *et al.* Irritant contact stomatitis: a review of the condition. *J Periodont* 1998; 69(6): 620–631.
- Shibly O *et al.* Clinical evaluation of the effect of a hydrogen peroxide mouth rinse, sodium bicarbonate dentifrice, and mouth moisturizer on oral health. *J Clin Dent* 1997; 8(5): 145–149.
- Hanawa T *et al.* Development of patient-friendly preparations: preparation of a new allopurinol mouthwash containing polyethylene(oxide) and carrageenan. *Drug Dev Ind Pharm* 2004; 30(2): 151–161.
- Amerongen AVN, Veerman ECI. Current therapies for xerostomia and salivary gland hypofunction associated with cancer therapies. *Support Care Cancer* 2003; 11(4): 226–231.
- Ellis ME *et al.* Controlled study of fluconazole in the prevention of fungal infections in neutropenic patients with haematological malignancies and bone marrow transplant recipients. *Eur J Clin Microbiol Infect Dis* 1994; 13(1): 3–11.
- Ellepola ANB, Samaranayake LP. Adjunctive use of chlorhexidine in oral candidoses: a review. *Oral Dis* 2001; 7(1): 11–17.
- Joseph BK. Oral cancer: prevention and detection. *Med Princ Pract* 2002; 11: 32–35.
- Tricca RE. Formulating mouth rinses. *Drug Cosmet Ind* 1988; 142(5): 32.

36. Jungnickel PW *et al.* Blind comparison of patient preference for Flavored Colestid Granules and Questran Light. *Ann Pharmacother* 1993; 27(6): 700–703.
37. Abdel-Rahman SM *et al.* The bioequivalence of nizatidine (Axid) in two extemporaneously and one commercially prepared oral liquid formulations compared with capsule. *J Clin Pharmacol* 2003; 43(2): 148–153.
38. Krieger JN. Urinary tract infections: what's new? *J Urol* 2002; 168(6): 2351–2358.
39. Sarkar MA. Drug metabolism in the nasal mucosa. *Pharm Res* 1992; 9(1): 1–9.
40. Eppstein DA, Longenecker JP. Alternative delivery systems for peptides and proteins as drugs. *Crit Rev Ther Drug Carrier Syst* 1988; 5(2): 99–139.
41. Davis SS, Illum L. Absorption enhancers for nasal drug delivery. *Clin Pharmacokinet* 2003; 42(13): 1107–1128.
42. Arnold J *et al.* Nasal administration of low molecular weight heparin. *J Pharm Sci* 2002; 91(7): 1707–1714.
43. Bateman ND *et al.* A study of intranasal distribution of azelastine hydrochloride aqueous nasal spray with different spray techniques. *Clin Otolaryngol* 2002; 27(5): 327–330.
44. Tsikoudas A, Homer JJ. The delivery of topical nasal sprays and drops to the middle meatus: a semiquantitative analysis. *Clin Otolaryngol* 2001; 26(4): 294–297.
45. Adams WP *et al.* Optimizing breast-pocket irrigation: the post-betadine era. *Plast Reconstr Surg* 2001; 107(6): 1596–1601.
46. Connolly JG *et al.* Some newer approaches to the treatment of the mucosa of patients with superficial bladder cancer: 1. Continuous bladder irrigation with 5-fluorouracil. *Can Med Assoc J* 1979; 121: 318–320.
47. Abbas AAH *et al.* Aggressive cutaneous aspergillus infection: control by local irrigation with amphotericin B. *Med Pediatr Oncol* 2002; 39(2): 139–140.
48. Mallet L *et al.* Storage of vancomycin oral solution. *N Engl J Med* 1982; 307(7): 445.
49. Greenwood J. *Pharm J* 1989; 243(Oct 28): 553–557.
50. Jain NK *et al.* Sorbitol intolerance in adults. *Am J Gastroenterol* 1985; 80(9): 678–681.
51. Kim NC, Kinghorn AD. Highly sweet compounds of plant origin. *Arch Pharm Res* 2002; 25(6): 725–746.
52. Mitchell JC, Counselman FL. A taste comparison of three different liquid steroid preparations: prednisone, prednisolone, and dexamethasone. *Acad Emerg Med* 2003; 10(4): 400–403.
53. Cooper C *et al.* Ipecac—a substance of abuse. *Med J Aust* 1998; 168(2): 94–95.
54. Cote CJ *et al.* A comparison of three doses of a commercially prepared oral midazolam syrup in children. *Anesth Analg* 2002; 94(1): 37–43.
55. Allen Jnr, LV, Erickson MA. Stability of alprazolam, chloroquine phosphate, cisapride, enalapril maleate, and hydralazine hydrochloride in extemporaneously compounded oral liquids. *Am J Health-Syst Pharm* 1998; 55(18): 1915–1920.
56. Allen Jnr, LV, Erickson MA. Stability of bethanechol chloride, pyrazinamide, quinidine sulfate, rifampin, and tetracycline hydrochloride in extemporaneously compounded oral liquids. *Am J Health-Syst Pharm* 1998; 55(17): 1804–1809.
57. Horner RK, Johnson CE. Stability of an extemporaneously compounded terbutaline sulfate oral liquid. *Am J Hosp Pharm* 1991; 48(2): 293–295.
58. Al-Waili NS. Topical application of natural honey, beeswax and olive oil mixture for atopic dermatitis or psoriasis: partially controlled, single-blinded study. *Complement Ther Med* 2003; 11(4): 226–234.
59. Nussinovitch A. *Water-Soluble Polymer Applications in Foods*. Oxford: Blackwell Science, 2003.
60. Bedinghaus JM, Niedfeldt MW. Over-the-counter foot remedies. *Am Fam Physician* 2001; 64(5): 791–796.
61. Kim CK *et al.* Preparation and evaluation of flurbiprofen dry elixir as a novel dosage form using a spray-drying technique. *Int J Pharm* 1995; 120(1): 21–31.
62. Williams RO *et al.* Study of solubility of steroids in hydrofluoroalkane propellants. *Drug Dev Ind Pharm* 1999; 25(12): 1227–1234.
63. Smith KJ *et al.* Influence of flow rate on aerosol particle size distributions from pressurized and breath-actuated inhalers. *J Aerosol Med-Depo. Clear Eff Lung* 1998; 11(4): 231–245.
64. Weisshaar E *et al.* Topical capsaicin therapy in humans with hemodialysis-related pruritus. *Neurosci Lett* 2003; 345(3): 192–194.
65. Stozek T, Borysiewicz J. Bioavailability of griseofulvin in the form of liposomes. *Pharmazie* 1991; 46(1): 39–41.
66. Park SJ, Kim SH. Preparation and characterization of biodegradable poly(l-lactide)/poly(ethylene glycol) microcapsules containing erythromycin by emulsion solvent evaporation technique. *J Colloid Interface Sci* 2004; 271(2): 336–341.
67. Rubinstein A *et al.* In vitro release and intestinal absorption of physostigmine salicylate from submicron emulsions. *J Pharm Sci* 1991; 80(7): 643–647.
68. Constantinides PP. Lipid microemulsions for improving drug dissolution and oral absorption: physical and biopharmaceutical aspects. *Pharm Res* 1995; 12(11): 1561–1572.
69. Bhargava HN *et al.* *Pharm Tech* 1987; 11: 46.
70. Griffin WC *et al.* *Drug Cosmet Ind* 1967; 101(4): 41.
71. Nielloud F, Marti-Mestres G, eds. *Pharmaceutical Emulsions and Suspensions*. New York: Marcel Dekker Inc., 2000.
72. Lachman L *et al.* eds. *The Theory and Practice of Industrial Pharmacy*, 3rd edn. Philadelphia: Lea & Febiger, 1986.
73. Binks BP. Particles as surfactants—similarities and differences. *Curr Opin Colloid Interface Sci* 2002; 7(1–2): 21–41.
74. Florence AT, Whitehill D. The formulation and stability of multiple emulsions. *Int J Pharm* 1982; 11(4): 277–308.
75. Okochi H, Nakano M. Preparation and evaluation of w/o/w type emulsions containing vancomycin. *Adv Drug Deliv Rev* 2000; 45(1): 5–26.
76. Kassem MA *et al.* *STP Pharma Sci* 1995; 5(4): 309–315.
77. Bourrel M, Schechter RS, eds. *Microemulsions and Related Systems: Formulation, Solvency, and Physical Properties*. New York: Marcel Dekker, 1988.
78. Rosano HL *et al.* Microemulsions: A commentary on their preparation. *J Soc Cosmet Chem*. 1988; 39: 201–209.
79. Lin TJ, Shen YF. Low-energy emulsification. Part VI: Applications in high-internal phase emulsions. *J Soc Cosmet Chem* 1984; 35(7): 357–368.
80. Lin TJ. Low-energy emulsification—I—principles and applications. *J Soc Cosmet Chem* 1978; 29(3): 117–125.
81. Maa YF, Hsu C. Liquid-liquid emulsification by rotor/stator homogenization. *J Control Release* 1996; 38(2–3): 219–228.
82. Lieberman HA *et al.* eds. *Pharmaceutical Dosage Forms. Volume 3: Disperse Systems*, 2nd edn. New York: Marcel Dekker, 1998.
83. Maa YF, Hsu CC. Performance of sonication and microfluidization for liquid-liquid emulsification. *Pharm Dev Technol* 1999; 4(2): 233–240.
84. van Balen GP *et al.* Liposome/water lipophilicity: methods, information content, and pharmaceutical applications. *Med Res Rev* 2004; 24(3): 299–324.
85. Derycke ASL, de Witte PAM. Liposomes for photodynamic therapy. *Adv Drug Deliv Rev* 2004; 56(1): 17–30.
86. Gregoriadis G *et al.* eds. *Liposomes in Drug Delivery. Drug Targeting and Delivery v. 2*. Langhorne Pa: Harwood Academic Publishers, 1993.
87. Lasic DD, Papahadjopoulos D, eds. *Medical Applications of Liposomes*. Amsterdam: Elsevier, 1998.
88. Bangham AD *et al.* Diffusion of univalent ions across the lamellae of swollen phospholipids. *J Mol Biol* 1965; 13: 238–252.
89. Anon. *Physicians' Desk Reference*, 58th edn. Montvale NJ: Medical Economics, 2003.
90. Tsai SC *et al.* Effects of particle properties on the rheology of concentrated noncolloidal suspensions. *J Rheol* 1992; 36(7): 1291–1305.

91. FDA. *Guidance for Industry. Stability Testing of Drug Substances and Drug Products*. Rockville MD: Center for Drug Evaluation and Research (CDER), 1998: 1–114.
92. Levinson ML, Johnson CE. Stability of an extemporaneously compounded clonidine hydrochloride oral liquid. *Am J Hosp Pharm* 1992; 49(1): 122–125.
93. Harris AM, Rauch AM. Cefuroxime axetil oral suspension. *Pediatr Infect Dis J* 1994; 13(9): 838.
94. Echizen H, Ishizaki T. Clinical pharmacokinetics of famotidine. *Clin Pharmacokinet* 1991; 21(3): 178–194.
95. Tortorici MP. Formulation of a cimetidine oral suspension. *Am J Hosp Pharm* 1979; 36(1): 22.
96. Morales ME *et al.* Comparative study of morphine diffusion from sustained release polymeric suspensions. *J Control Release* 2004; 95(1): 75–81.
97. Shah KP, Chafetz L. Use of sparingly soluble salts to prepare oral sustained release suspensions. *Int J Pharm* 1994; 109(3): 271–281.
98. Sjoqvist R *et al.* In vivo validation of the release rate and palatability of remoxipride-modified release suspension. *Pharm Res* 1993; 10(7): 1020–1026.
99. Sheumaker JL. United States #4,762,709, 1988.
100. Tipton AJ, Holl RJ. United States #5,747,058, 1998.
101. Nail SL *et al.* Structure of aluminum hydroxide gel III: mechanism of stabilization by sorbitol. *J Pharm Sci* 1976; 65(8): 1195–1198.
102. Moore WN, Taylor LT. Extraction and quantitation of digoxin and acetyldigoxin from the *Digitalis lanata* leaf via near-supercritical methanol-modified carbon dioxide. *J Nat Prod* 1996; 59(7): 690–693.
103. Horovitz B. *Stress-busting drinks take off*. *USA Today* 2011; May 16.
104. O'Donnell J, Weise E. *Energy drink sales rise along with concerns*. *USA Today* 2011; March 17.

Lieberman HA *et al.* eds (1998). *Pharmaceutical Dosage Forms. Volume 3: Disperse Systems*, 2nd edn. New York: Marcel Dekker.

Nielloud F, Marti-Mestres G, eds (2000). *Pharmaceutical emulsions and suspensions*. New York: Marcel Dekker Inc.

SOLUTIONS, EMULSIONS AND SUSPENSIONS

Becher P (1983). *Encyclopedia of Emulsion Technology*. New York: Marcel Dekker.

Becher P (2001). *Emulsions: Theory & Practice*, 3rd edn. New York: Oxford University Press.

Kreuter, J (1994). *Colloidal Drug Delivery Systems*. New York: Marcel Dekker.

Osborne DW, Amann AH (1990). *Topical Drug Delivery Formulations*. New York: Marcel Dekker.

Yalkowsky SH (1981). *Techniques of Solubilization of Drugs*. New York: Marcel Dekker.

Yalkowsky SH (2003). *Handbook of Aqueous Solubility Data*. Boca Raton FL: CRC Press.

EQUIPMENT

Busse DJ (1990). *Mfg Chem* 61(7): 39.

Lagman B (1988). Change-Can Mixing Technology. *Drug Develop Ind Pharm* 14(18): 2705–16.

Oldshue JY (1983). *Fluid Mixing Technology*. New York: McGraw-Hill.

EXCIPIENT PROPERTIES

Rowe RC *et al.* eds (2009). *Handbook of Pharmaceutical Excipients*, 6th edn. Washington DC: American Pharmaceutical Association.

Reynolds J, ed (1996). *Martindale, The Extra Pharmacopoeia*, 31st edn. London: Pharmaceutical Press.

BIBLIOGRAPHY

GENERAL

Lachman L *et al.* eds (1986). *The Theory and Practice of Industrial Pharmacy*, 3rd edn. Philadelphia: Lea & Febiger.

Chapter 25

Sterilization Processes and Sterility Assurance

James Agalloco, BEChE, MSChE, MBA; William G. Lindboe, Jr, PhD and Russell E. Madsen, MS

INTRODUCTION 469

MICROBIAL GROWTH AND DEATH 469

EQUIPMENT QUALIFICATION 472

CONTAINER AND LOAD MAPPING 473

BIOLOGICAL INDICATORS 473

PHYSICAL MEASUREMENTS 474

EQUIVALENCE IN STERILIZATION 487

OTHER STERILIZATION METHODS 487

PARAMETRIC RELEASE OF STERILIZED MATERIALS 488

ASEPTIC PROCESSING 488

CONCLUSION 491

INTRODUCTION

Microorganisms represent a potential threat to human health in various settings. The presence of some microorganisms in foods, pharmaceuticals and even the environment can cause illness and even death. Louis Pasteur established the link between microbes and disease, and his seminal work influenced Joseph Lister to use phenol to reduce contamination risk during invasive surgery in the late 1800s.¹ This knowledge led to the introduction of multiple measures to mitigate the risks associated with microbial contamination and led to the development of sterilization methods, sanitization practices, and other microbial control measures intended to assure product and procedural safety.

In the pharmaceutical and medical device industries, sterile products are produced using two primary methods: terminal sterilization or aseptic processing.¹ Terminal sterilization provides for the clean assembly of the product as a final package, which is then subjected to a lethal process. Aseptic processing rearranges the steps: the product components are individually sterilized and then assembled in an extremely clean environment. This chapter will review the primary methods for the production of sterile products, using either approach, with emphasis on sterilization and sterility assurance related subject matter. Some of the processes described will also afford some measure of destruction/removal of pyrogens (endotoxin), viruses, and prions; however, their removal is not considered in detail with the exception of dry heat depyrogenation.

In addition to the therapeutic goal of all medical products there is a safety requirement; a major component of which in the context of sterile pharmaceuticals and medical devices is “sterility”. Sterility, which has been defined as the absence of life (or inability to reproduce), is an absolute concept. Items are “sterile,” or they are not; there is no middle ground. However, at the present time, there are no nondestructive means to demonstrate sterility. Instead we define processes to attain that goal, subject the materials to these processes, and then estimate the results. For the lethal sterilization processes, whether terminal or in-process, the “sterility” expectation is fulfilled by demonstrating a minimum Probability of a Non-Sterile Unit (PNSU) of not more than one positive unit in 1 000 000 units or 1×10^{-6} .^{2,3} It must be understood that the microorganism of concern in any process is the bioburden present prior to treatment. In actual practice this target is often exceeded by a substantial margin

as the industry is inherently conservative about such a critical quality attribute. The focus of the treatment is on the removal/destruction of any bioburden microorganisms that might be present on/in the materials prior to the process. Where liquids are sterilized by filtration, the filters utilized are shown capable of removing a minimum of seven logs of a challenge microorganism specifically cultured for its small size and potential penetration of the filter. In aseptic processing, which relies on prior sterilization of the materials, equipment, and other items, the goal is to prevent the recontamination of those items in the assembly process. Aseptic process simulations are used to demonstrate contamination rates of not more than one unit in 5000 units. The various estimations of performance from these processes are established during the validation efforts with the assumption that routine operations deliver comparable (or superior) results. In most instances everyday manufacturing performance is well in excess of the targeted and demonstrated minimum levels established by the validation efforts. Nevertheless, the level of sterility present after any of these processes is never known with any degree of precision.

MICROBIAL GROWTH AND DEATH

In considering means to control the microbial content of materials, a basic understanding of the typical microbial life cycle can be useful. In order to increase in number (or replicate), aerobic and facultative microorganisms require essentially what we need: food, water, and air.² While nutrient requirements across the spectrum of flora are diverse, the core needs are carbon and nitrogen, liquid moisture, and oxygen. Aseptic processing is performed in cleanrooms, in which the connotation of “clean” means absence of soil and moisture that might support microbial proliferation. The presence of human operators in many aseptic processing environments means that microbial contaminants will be continuously introduced. Appropriate cleanroom designs and operating practices are developed to provide for their prompt removal and adequate protection of the critical environments, such that those which do remain do not result in microbial contamination of sterilized items.

The other side of the microbial life cycle relates to their death. This can be accomplished by removing what they need to increase in number or, more expeditiously, by subjecting the microbes to some degree of stress. The stress can be thermal, chemical, or ionic in nature (filtration is not a lethal process

¹Where products are manufactured with aseptic processing followed by terminal sterilization, there is little difference in the manner in which either process is controlled compared to when they are performed individually.

²Obligate anaerobic microorganisms cannot survive in environments where the level of oxygen is greater than ~ 0.2% and thus are almost never isolated in manned environments.

and accomplishes microbial control by separative means) in which the death rate is accelerated. Sterilization processes deliver these conditions to extreme to rapidly kill the microbes.

While the lethal processes appear to be fundamentally different, the lethal aspect of each is at its core a chemical reaction that generally follows first order kinetics. Bigelow established that when the log number of the microorganisms is plotted against time, an essentially straight line results.⁴ This aspect of microbial destruction lies at the very core of all lethal processes as it enables predictive destruction of microorganisms. The inverse slope of this line is the D-value or the time required to reduce the population by one log (or 90%) (See Figure 25-1).

The D-value is influenced by a number of factors, most notably by the conditions that will ultimately result in the microorganism's destruction. When a D-value is reported, it should always indicate the exact conditions under which it was determined.

The semi-logarithmic plot of the Survivor Curve has the difficulty of being undefined mathematically at zero, which is the definition of "sterile". The concept of "Commercial Sterility" was initiated to address this difficulty and the practical nature of assessing sterilization processes that destroy microorganisms within the food industry. This concept expresses sterility as a probability in the region of the survivor curve below one surviving organism. This region has negative values for the log of concentration and is interpreted as non-sterile containers in a group being sterilized (e.g., one non-sterile unit in ten units for a survivor concentration of 10^{-1} , see Figure 25-1). This approach is called Probability of a Non-Sterile Unit (PNSU) or, as applied to all processes (not necessarily containers), Sterility Assurance Level (SAL). The D-value approach can be misused by applying it to curvilinear semi-log survivor curves, or extrapolating the linear portion from a limited range of experimental data. The latter misuse resulted in cases of paralytic polio and death when a vaccine manufacturer in the 1950s extrapolated the linear segment of a Survivor Curve for Poliomyelitis Virus treated with Formalin, published by Salk, *et al.*, to treat commercial quantities of vaccine.⁵

Thermal processes using moist heat have been extensively studied and the D-value has been determined for many important spoilage and resistant microorganisms to facilitate the control of lethal processes. The change in the D-value in response to changes in temperature is an important concern in thermal sterilization processes. The relationship between temperature

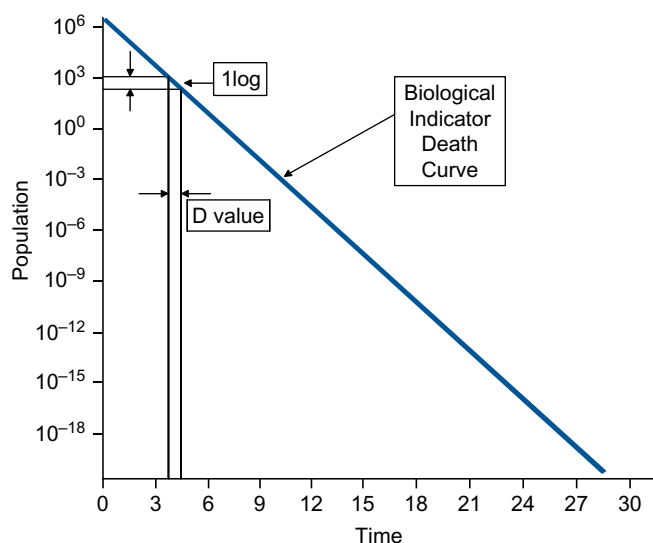


Figure 25-1. Graphical Representation of D-value from Survivor Curve.

and the D-value is established by the Thermal Death Time Curve and determination of the z-value (See Figure 25-2).

The z-value is the reciprocal slope of the semi-logarithmic plot of the Thermal Death Time Curve and allows for calculation of process lethality for processes where the conditions are not constant, and thus, microbial kill will occur at varying rates during the course of the treatment. A comparable approach is usable for dry heat sterilization albeit with less supportive data. There have been attempts to apply the D- and z-value approach to dry heat depyrogenation; however, these determinations are not universal because the endotoxin utilized in each varied.³

The sterilization process is modeled by a step functional isothermal process, in which the temperature is instantaneously raised to a reference temperature, held for "F" minutes, and instantaneously reduced to sublethal levels. The kinetic parameters of D- and z-values can be utilized to mathematically estimate the lethality delivered by an actual thermal process, according to the following calculation. The general form of this equation has been adapted for steam and dry heat processes.

F equation

$$L = \log^{-1} (T - T_b) / z = 10^{(T - T_b) / z} \quad (1)$$

Where:

T_0 = temperature in load

T_b = base (or reference) temperature

z = z - value [10°C] Note: units are a temperature difference, not a temperature

L = instantaneous process lethality

When the base temperature is 121.1°C and the z-value employed 10°C , the equation can be utilized in conjunction with thermal data to determine the F_0 , or the equivalent sterilization time for a particular process. The subscript "0" was initiated historically in the food industry, where 250°F and a z of 18° Fahrenheit degrees were more widely used. The use of a standard calculation method allows for the comparison of lethal conditions within load, between loads, and between processes.

F_0 equation

$$F_0 = 110^{(T - 121.1) / 10} dt \quad (2)$$

This integral is normally calculated as a summation of contributions over finite temperature measurements

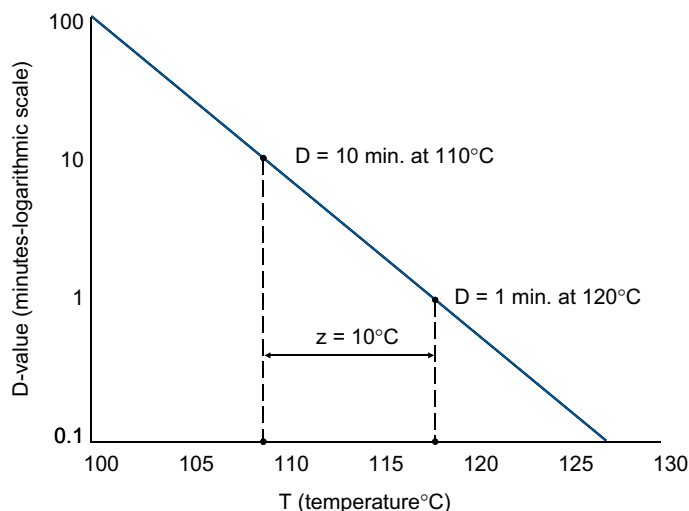


Figure 25-2. Graphical Representation of z-value from Thermal Death Time Curve.

³Dry heat depyrogenation is supported empirically by verification of endotoxin destruction. Mathematical determinations of dry heat depyrogenation are used for correlation to equipment parameters and not as definitive proof of process performance.

F_0 equation

$$F_0 = \sum 10^{(T-121.1)/10} \Delta t \quad (3)$$

Where:

Δt = time interval between measurements

T = temperature at each interval

A comparable but less accurate means for estimation of lethality has been developed for dry heat processes. Separate equations are used for sterilization and depyrogenation.

F_H equation sterilization

$$F_H = \sum 10^{(T-170)/20} \Delta t \quad (4)$$

F_H equation depyrogenation

$$F_H = \sum 10^{(T-250)/50} \Delta t \quad (5)$$

Where:

Δt = time interval between measurements

T = temperature at each interval

In the case of depyrogenation, differing base temperatures and z -values can be used.⁴ Gillis and Mosely have developed a mathematical model for ETO sterilization that considers gas concentration and temperature conditions, estimates the process lethality, and determines equivalent process time.⁶

Relative humidity was not considered as a contributing factor as the operation range evaluated showed no meaningful impact on lethality. The use of this model is not widespread although it affords a level of understanding for ETO sterilization.

$$U = \sum \left(10^{\left(\log t + \left(\frac{T_i - T_{REF}}{Z} \right) \right)} \right) \frac{C_i}{C_{REF}} \quad (6)$$

Where:

U = equivalent time

C = concentration

T = temperature

t = time

For the other sterilization processes, there are no current means to convert physical parameter measurements into lethality estimations. Their effectiveness is established using biological indicators, customarily Gram-positive spore formers, as surrogates for the bioburden. The delivered lethality should always be referenced to the bioburden present in order to support demonstration of process performance. This is accomplished by making a determination of the PNSU using equation 7.

PNSU calculation

$$\log N_u = \frac{-F}{D} + \log N_0 \quad (7)$$

Where:

N_u = PNSU or SAL

D = D-value of the microorganism

F = F-value of the process

N_0 = initial microbial population (extrapolated from linear segment of survivor curve)

In this calculation the D-value to be utilized is that of the bioburden as it is the target of the process in routine operation, and its destruction must be supported. The log reduction of a microorganism can be estimated for processes where the F value cannot be readily measured by substituting the process time in the equation (see equation 8). Note that the actual initial microbial concentration may be different if there is a significant curvilinear initial segment.

$$\text{Log reduction } \log N_u = \frac{-U}{D} + \log N_0 \quad (8)$$

Where:

N_u = PNSU or SAL

D = D-value of the microorganism

U = Duration of the process

N_0 = initial microbial population

This assumes the process is accomplished at the same conditions at which the D-value has been determined, which may not be precisely the case. Nevertheless, Equation 8 provides some means of assessment for these other processes.

While it may appear tempting to rely on the physical measurements to estimate process lethality and provide a more quantitative assessment of the process, the measurements must be acknowledged for their secondary nature. The mathematical models rely first on the response of microorganisms to the process (their D-value), as adjusted with respect to the specific conditions (their z -value). The physical measurements are assumed to be representative of the sterilizing conditions the microorganism experiences and responds to; however, those conditions are assumed and never precisely known. For this reason, sterilization microbiologists tend to place greater credence on the reaction of the biological challenge system to the process as proof of process efficacy.⁷

Sterilization by filtration processes differ from the other methods in that the mechanism is separative rather than lethal. As a consequence, the parametric measurements of the lethal methods are not applicable to sterilizing filtration. The important filtration process parameters relate to filter integrity, which can be used to substantiate microbial retention. An expanded discussion of sterilizing filtration is provided later in this chapter.

STERILIZATION CYCLE DEVELOPMENT

The determination of a suitable process (cycle) for sterilization must first consider the expected effect of the process on the materials. Certainly there is no value in defining a process that, while effective in destruction of microorganisms, has substantial adverse quality and/or reliability impact on the items being processed. A balance must often be struck between an extreme process that provides substantial microbial kill and less aggressive conditions that preserve the material's essential quality attributes. With the exception of stainless steel and glass (and even these can be impacted in some instances), the material effects have to be taken into account in cycle development. If the adverse material effects are excessive, a different sterilization process should be selected.

There are three fundamental approaches to the development, validation, and control of a sterilization process: overkill, bioburden/biological indicator (BB/BI), and bioburden.⁵ The default choice for cycle approach is the overkill method which is the simplest to establish and control but has the greatest potential adverse impact on the load items. At the other extreme is the bioburden method which has the least impact on materials but requires more rigorous evaluation initially and throughout its operational use. The bioburden/biological indicator approach falls between the other methods with respect to both material impact and operational ease (see Figure 25-3). The BB/BI method is used for the majority of terminally sterilized products and laboratory media, instances where the negative consequences of over-processing need to be avoided.⁷ Radiation sterilization is validated using a bioburden approach in part because the radiation resistance of many microorganisms has not been evaluated and non-sporeforming microorganisms have been isolated with substantial resistance to radiation.⁸

⁴See later section on dry heat depyrogenation for additional clarification.

⁵It is essential that the same process control methodology be utilized throughout.

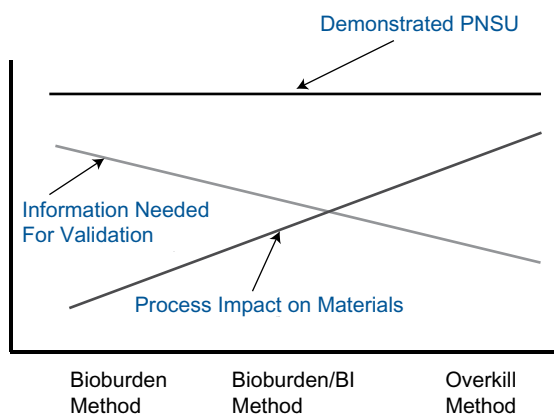


Figure 25-3. Comparison of Validation Approaches.

The methods can be briefly described as follows:

Overkill Approach

A process where the destruction of a high concentration of a resistant microorganism supports the elimination of bioburden that might be present in routine processing. That objective can be demonstrated by attaining any of the following: a defined minimum lethality, a defined set of process conditions, or confirmation of minimum log reduction of a biological indicator.

Bioburden / Biological Indicator Approach

A method where the incomplete destruction (or destruction of a modest population) of a resistant biological indicator can be used to demonstrate the capability of the process to reliably destroy any bioburden. This is accomplished using detailed knowledge of the bioburden/biological indicator populations and their relative resistance.

Bioburden Approach

A method where multiple bioburden isolates from the material are evaluated for resistance to the sterilization method and utilized as biological indicators to demonstrate the lethality of the process. Frequent monitoring of the bioburden population and resistance is mandatory for success.

The methods must be validated consistent with their initial development approach and maintained in the same way. Despite the difference in bioindicator performance with the different methods, each can provide materials that are adequately sterilized. This is because the PNSU for a sterilization process should be calculated using the population and resistance of the bioburden present (see Equation 7 above). The biological indicator serves as a measure of the process (as well as confirmation of lethality in locations and on surfaces where physical measurement is not possible), and not as definitive proof of performance in routine operation.⁶

The overkill method is intended primarily for items where the process's sensitivity to the treatment is such that the potential for over-processing is minimal. The process is demonstrated lethal under reduced conditions with both process monitors and biological indicators present. The routine use of somewhat more aggressive conditions provides an added measure of confidence in the outcome. The BB/BI and bioburden approaches are better suited where material impact places constraints on the conditions that can be used. Their validation is accomplished using the routine process, as it is the added controls on the bioburden that provides the desired process confidence.

There is an adaptation of the overkill method, called the half-cycle method, in which the process dwell period is doubled to

define the extent of the routine process. Originally developed for use with ETO before accurate in-process measures of concentration and relative humidity were available, the doubling of the dwell time was considered sufficient to support the full duration process if the biological indicators were inactivated in the half-cycle. Its use for other processes presupposes no adverse material impact with the extended duration, something that should not be assumed so readily. Lethal factors other than time – such as concentration, temperature, humidity, and others are never adjusted, making it more difficult to support process effectiveness when these parameters vary outside the demonstrated values. The half-cycle method is in widespread use, notwithstanding its limited utility in supporting variations in parameters other than time.⁹

EQUIPMENT QUALIFICATION

Qualification of sterilization equipment and its associated utility systems must be performed prior to beginning performance qualification. The reproducibility of sterilization cycles is largely dependent on the controls, utilities, and mechanics of the sterilizer and its ability to consistently operate according to expectations. The essential elements of the sterilizer and its supply services should be verified as properly installed and set to their specifications as part of the initial qualification and is often termed the installation qualification. This is followed by basic confirmation of equipment functionality in an operational qualification exercise. These can be performed sequentially or in a more integrated manner. The overall activity is often termed equipment qualification and precedes performance qualification, which entails rigorous evaluation of system performance under load, using independent parameter measurement and/or microbial challenges. An exact separation of the activities is non-critical; provided the appropriate measures are followed their designation as IQ, OQ, or PQ is arbitrary.

Qualification is performed to assure that the sterilizer has been installed and operates according to manufacturer's specifications. Its most important purpose is to establish a baseline for the installation and operation that can be used to assess the impact of subsequent changes or alterations. Any testing performed during the equipment qualification is commonly executed with the sterilizer empty.

Process equipment for sterilization, regardless of the process type, includes at least three major process stages: preparation/pre-treatment of the items to be sterilized; exposure to the sterilizing agent; and post-exposure steps. While these differ with respect to their execution in relation to each phase, there are common considerations:

- Preparation/pre-treatment steps (which can be performed prior to transfer to the sterilizer) encompass such activities as cleaning of the items, wrapping/placement into containers, pre-humidification/pre-conditioning of load items, and removal of air (to facilitate agent penetration).
- Exposure provides the regulation of sterilizing conditions for the required time period. This is variously called the dwell period or exposure time.
- Post-exposure activities include cooling, drying, sterilizing agent removal, and maintenance of sterility for unwrapped surfaces (essential in sterilization-in-place and chemical treatments).

The most important aspect of the equipment is its ability to deliver consistent performance over an extended period of use and the reliability/accuracy of the routine documentation it provides on each sterilization process. This mandates that the equipment be supported by a number of related practices:

⁶When comparing the height of two individuals to determine which is taller, the unit of measurement is not critical. Whether the measurement is made in meters, yards, rods, or fathoms, each measurement system will confirm the same result.

- A formal change control program addressing the sterilization equipment, utility supplies, materials, and procedures.
- Preventive maintenance as established by the equipment manufacturer for maintaining system performance.
- Calibration procedures for the installed instrumentation on sterilization equipment and critical utilities.
- Training of operating, maintenance, and other personnel in the procedures for operation maintenance and calibration.

CONTAINER AND LOAD MAPPING

In order to establish the process dwell time at the desired conditions, the load items and the load configuration itself must be studied to confirm attainment of the desired condition through the item (and load). This entails the execution of preliminary studies, in which the process conditions are evaluated within the item and the load to determine “worst case” locations for penetration. It is these locations that will define the expected process time to satisfy the sterilization objective.

COMPONENT MAPPING

Items to be evaluated are reviewed and physical monitoring devices introduced to assess penetration of the items. For thermal processes this is carried out with thermocouples placed in expected “worst case” locations, where, by virtue of mass, complexity, condensate/air retention, etc., heating will be slowest. Biological indicators can be used similarly in gas, liquid, and vapor processes. Radiation-sterilized items can be mapped using dosimeters to determine low-dose locations. Identification of these locations is essential to establish the desired dwell periods in these “worst case” locales. Measurements should be made in as non-obtrusive manner as possible so as to gain the most information with minimal alteration of the item and any external wrapping material and without impacting either positively or negatively the sterilization process. The items should be sealed in a manner comparable to that routinely used. Temperature measurements of load items should be made with the thermocouple junction in contact (or nearly so) with the item's surface. Biological indicator strips can be used; however, wires, dots, and strings with a surface dried spore suspension may be more appropriate in smaller items. The container mapping can be performed in other sterilized units, provided the process used is comparable to that expected in routine use. Once the appropriate “worst case” location has been identified, all subsequent monitoring should be performed at that location with the item wrapped as it was in the component mapping and oriented in the same way, e.g., open end down. Large items may only have their openings covered to facilitate handling.

Component mapping should be a part of all the destructive sterilization processes with the intent of locating the “most difficult to sterilize” location within the item. In gas sterilization this location would be the appropriate location for positioning of biological indicators on or within the device. The practice in radiation sterilization is comparable with dosimeters, positioned where the measured dose might be the lowest. These evaluations can be performed for a complete shipping carton in which a central or middle unit would be studied since these would represent the “most difficult to sterilize” location. These locations are roughly equivalent to the “cold spot” in thermal sterilization.

LOAD MAPPING

This activity provides for the assessment of the effect of load size, arrangement, and density on penetration across the sterilizing chamber. With load items probed as described above, the load is assembled and the effect of loading on the conditions assessed. Where the load consists of items potentially susceptible to the treatment, it is equally important to identify locations where maximum degradation might occur, in addition to those where minimum sterilization conditions are demonstrated. Those portions

of the load evidencing the most aggressive conditions would be evaluated for adverse effects on the materials, while those demonstrating the least aggressive are more appropriate for monitoring with biological challenges. Load mapping is less of a concern where items are largely unaffected by the sterilization process, such as steam sterilization of stainless steel and glass items. In these situations the differences in the item have substantially more influence on the process than its location in the sterilizer. It has added importance with dry heat processes where the ability of the sterilizing medium to penetrate / heat the items is limited.

Load mapping is expected for non-thermal destructive processes as well, with the effect of load density evaluated during cycle development and validation. In this exercise identification of maximum exposure conditions is valuable in identifying those locations where stability samples should be drawn from.

BIOLOGICAL INDICATORS

The use of biological indicators in sterilization has evolved. The first sterilization processes in the food industry were developed to eliminate *Clostridium botulinum*, which had been responsible for deaths in improperly processed canned foods. *Clostridium sporogenes*, a non-toxic anaerobic sporeformer with comparable resistance was identified as a suitable substitute for process development, validation, and monitoring. *Bacillus* spp. (and later *Geobacillus* spp.) spores with various levels of resistance (above and below that of the *Clostridia*) were introduced as biological indicators due to their ease of cultivation (they grow best in aerobic media), increased resistance to a variety of sterilization processes, and adaptability for use as suspensions and on various substrates.

The sporeforming *Bacillus* spp. and *Geobacillus* spp. have other characteristics that make them well suited for their use in sterilization process control. They are non-pathogenic; easy to produce, use and test; stable over extended periods; and substantially resistant to various treatments.

The commonly used biological indicators are indicated in Table 25-1. The use of other microbial strains is certainly acceptable. Bioburden microorganisms can be utilized as well for this purpose, and these are most commonly sporeforming microorganisms of the *Bacillus* species.

The purpose of the biological indicator is to directly assess the lethal conditions delivered by the process through their inactivation; however, it must be recognized that the BI serves as a measure of process performance, and while its destruction is generally sought in validation, that is not strictly required as the real concern is destruction of the bioburden during routine processing. Because the BI can be positioned on or in the load items at locations where physical measurements cannot always be made, it offers the most direct evidence of process conditions possible. The calculation of lethality, as discussed earlier, should always be considered as a secondary level assessment because it relies on numerous assumptions, including the linearity of the Survivor and Thermal Death Time Curves, accurate determination of D- and z-values, errors in physical measurement, and inaccuracies associated with spatial separation of physical measurement locations from the targeted surfaces, etc. Precedence should, therefore, be always given to the biological results when compared to physical data.⁷

The limitations of biological indicators in the evaluation and control of radiation sterilization procedures has been recognized for many years.⁸ The use of *Bacillus pumilus* for this purpose is no longer considered appropriate as “worst case”. Current practices for radiation sterilization process control rely heavily on actual bioburden monitoring without the necessity for challenge placement within load items.

Dry heat depyrogenation processes are commonly validated using endotoxin indicators; however, there are no standards for their resistance to processing. Suggestions as to D- and z-values for various endotoxin preparations have appeared in the literature; however, the results of these determinations cannot be

Table 25-1. Common Biological Indicator Microorganisms**Steam sterilization**

- *Geobacillus stearothermophilus* - ATCC 7953/12980*
- *Bacillus subtilis* - ATCC 5230
- *Bacillus smithii* (*Bacillus coagulans*) - ATCC 51232
- *Clostridium sporogenes* - ATCC 11437/9372

Dry heat sterilization

- *Bacillus atrophaeus* (previously *Bacillus subtilis* var. *niger*) - ATCC 49337/9372

Gas, liquid and vapor sterilization

- Ethylene Oxide - *Bacillus atrophaeus* - ATCC 49337/9372
- Hydrogen Peroxide - *Geobacillus stearothermophilus* - ATCC 7953/12980
- Chlorine Dioxide - *Bacillus atrophaeus* - ATCC 49337/9372
- Ozone - *Geobacillus stearothermophilus* - ATCC 7953/12908
- Peracetic Acid - *Bacillus atrophaeus* - ATCC 49337/9372

Radiation

Bacillus pumilis - ATCC 27142 (no longer in current use, see below)

* *Geobacillus stearothermophilus* is preferred for steam sterilization by direct contact, however its use in terminal sterilization is somewhat problematic because of its somewhat extreme resistance compared to possible bioburden microorganisms.

extended universally. Confirmation of endotoxin elimination must be established empirically, and correlation between F_{II} values and destruction of endotoxin is not currently possible.

PHYSICAL MEASUREMENTS

As sterilization processes rely on the exposure of items to extreme conditions intended to destroy microorganisms (and endotoxin) on them, physical measurements play an important role. The sterilization process equipment, depending upon its sophistication, will be equipped with instruments measuring (also recording) the essential process parameters. The most commonly measured variables are time, temperature, pressure, sterilizing medium concentration, and relative humidity, while others such as belt speed and agitation/recirculation rates may also be useful. Depending upon the equipment design, the measurements may be variously intended for control (the most critical application), confirmation/monitoring (useful in process evaluation), and informational (safety or convenience). During the validation exercise additional confirmation and information instruments may be added to provide greater information regarding the process conditions. Depending upon their use and type, the calibration frequency and accuracy/precision of these instruments will vary.^{10,11} Radiation processes utilize a unique measurement system, the dosimeter that is able to directly measure the delivered radiation dose.

MOIST HEAT STERILIZATION

The most widely used sterilization medium is moist heat, which is utilized for in-process sterilization of stainless steel, glass, and elastomeric materials (Dry Goods Sterilization); liquids in sealed containers (Liquid Good Sterilization); and for sterilization of large processing systems (Sterilization-in-Place). As each of these presents some unique challenges to execution, they will be addressed independently after discussion of the elements common to all.

Sterilization processes using moist heat entail the exposure of microorganisms to liquid water at elevated temperatures, which is believed to irreversibly denature cell wall proteins. In the absence of liquid water, the destructive process occurs much more slowly (see dry heat sterilization below). Steam sterilization is thus best accomplished in the presence of saturation conditions where both gas (steam) and liquid (condensate) are present. Saturated steam is defined as a state of H_2O that has both gas and liquid phases present simultaneously and

has a fixed relationship between temperature and pressure (see Figure 25-4).

The use of saturated steam for sterilization provides an advantage not present with other sterilization methods. When saturated steam comes into contact with a cooler surface, condensation of the gaseous steam phase occurs with concurrent release of energy, which serves to rapidly heat the surface to temperatures where microbial death is more rapid. The heat released is the latent heat of condensation, which must be added to boiling water to effect the phase change to the gaseous state. The heat of condensation is approximately 2240 kJ/kg. This amount of energy is substantially higher on a mass basis than is available when heating is accomplished without a phase change, using either superheated water or steam. Nevertheless, steam sterilization can be accomplished just as expeditiously when the microorganism is exposed to superheated liquid water at elevated temperatures, and this is used to considerable advantage in the sterilization of liquid-filled containers (see below).

STERILIZATION OF PARTS/HARD/DRY GOODS

The simplest of steam sterilization process is that utilized for stainless steel and glass items that are often termed parts loads or hard goods loads. Many of the items to be sterilized in parts

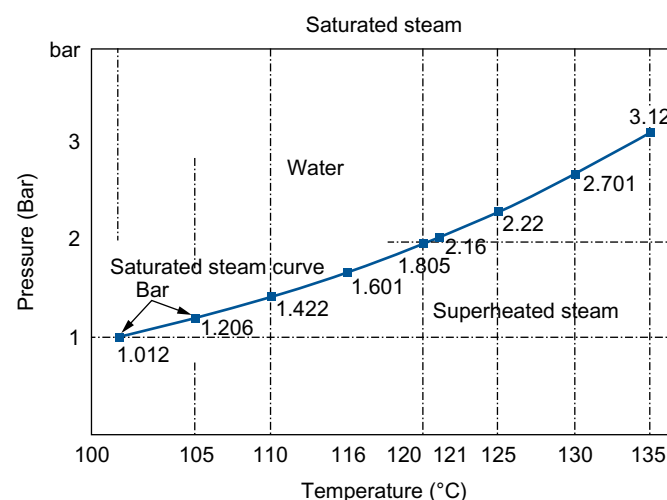


Figure 25-4. Saturated Steam Curve. Courtesy of Fedegari Autoclavi.

loads have little or no sensitivity to the steam process, and thus these processes are accomplished almost entirely using an overkill approach. Since the items to be sterilized may have an internal void volume, i.e., filters, tubing, fill sets, needle lumens, etc., air removal is necessary to allow for direct steam contact with the surfaces to be sterilized. Older sterilizers may use gravity displacement; however, this results in substantially longer cycle times and is frowned upon by some regulators. Newer sterilizers utilize multiple pre-vacuums (typically 3-6) for more efficient removal of the air and may have *in-situ* air indication devices. Items are placed in the sterilizer with the heavier, larger items (as confirmed to be “worst case,” based upon the component mapping) on the lower shelves. Where the load items are identical (e.g., elastomeric stopper loads), a discernable cold spot within the chamber may be identifiable and cycle dwell time derived from that location. Sterilization dwell periods are established to assure that the “worst case” items (or with identical items the “cold spot” in the sterilizer) attain the desired PNSU. Drying of the items may or may not be required. Items that are predominantly steel will require less drying, while elastomeric items and those with complex internal geometry may require drying under vacuum. International standards (ISO 17665) require specific steam quality and temperature distribution for the sterilizing medium.

A typical batch type steam sterilizer used for this purpose is shown in Figure 25-5. These sterilizers are commonly installed between areas of different classification and are used to supply sterilized items into the higher classified environment for use in subsequent processing.

Sterilization of these load items requires condensate removal throughout, and items within the load should be oriented to permit drainage from them. Items with complex internal geometries (e.g., filling assemblies) and substantial mass (e.g., stopper bowls) present the greatest difficulty in this regard.

Load items are customarily wrapped to reduce the contamination risk in post-sterilization, and this must be accomplished in a manner that does not restrict air/condensate removal and steam penetration. The wrapping need not enclose the item, merely those opening through which air/condensate and steam must pass to attain the sterilization process objective. Specifically designed rigid containers can also be used for this purpose.

The efficacy of hard goods sterilization is demonstrated in validation studies using penetration thermocouples (positioned as described earlier in component mapping) and biological indicators. The biological indicators are customarily either paper strips inoculated with spore suspension sealed in glassine envelopes or other materials inoculated with a *Geobacillus stearothermophilus* spore suspension. Complete destruction of the biological challenge is expected as outlined in the overkill approach to cycle development, validation, and control.

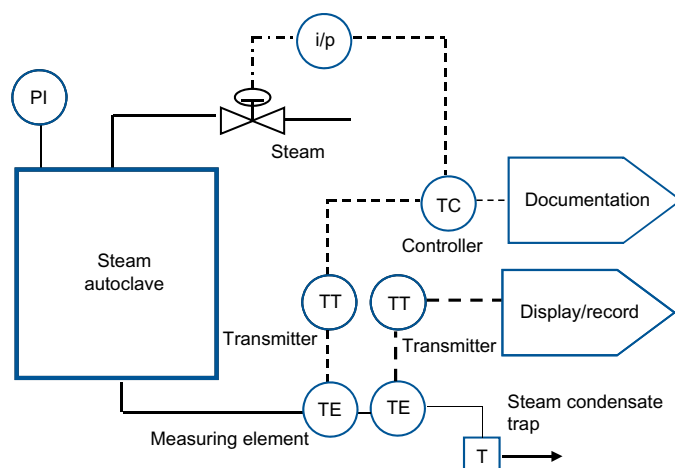


Figure 25-5. Typical Steam Sterilizer for Parts.

Sterilization cycle control is based upon confirmation of exceeding minimum time-temperature or F_0 targets. This is customarily derived from physical data obtained from the sterilizer drain, which is usually the coldest point in the system.

STERILIZATION OF LIQUIDS IN SEALED CONTAINERS

A seemingly similar process is used for the sterilization of containers filled with aqueous liquids. The outward similarity with parts sterilization is actually misleading. In these processes sterilization of the fluid is accomplished by heating of the container exterior, conduction through the container, and heating of the contents. The heat transferred superheats the aqueous phase and creates a moist heat condition in the container headspace. This type of process is utilized in a variety of situations including terminal sterilization of finished products, sterilization of in-process fluids, sterilization of laboratory media, and decontamination of liquid biological waste. Concerns that are generally common to all but the last of these are avoidance of adverse impact on items and maintaining container integrity throughout processing. Concerns relative to potential over-processing of these materials customarily adds an upper limit to the delivered lethality to ensure that essential product quality attributes are preserved. Heating closed containers to efficient sterilizing temperatures entails a commensurate increase in the internal pressure within the sealed containers, which presents a particular challenge in the post-cycle cooling. The sterilization of each container type is adapted to the specific needs of each and the available equipment.

Terminal Sterilization of Finished Products

The methods used for the sterilization of products in their final container are often specifically designed for that purpose, and this carries through to the equipment utilized. On the largest scale of operation, the sterilizers are substantial in size and customized for this purpose. The sterilization equipment includes steam-air sterilizers (see Figure 25-6), steam-air-water (see Figure 25-7), immersion sterilizer and continuous sterilizer in addition to the simpler designs used for sterilization by direct contact (see Figure 25-5). Because these units are utilized for sterilizing finished, sealed containers, they are commonly located in unclassified environments with defined flows for non-sterilized and sterilized materials to avoid potential mix-up. The efficient use of these batch sterilizers limits them to the sterilization of a single lot of a single size product container.

The preponderance of available literature on sterilization of aqueous containers (and actually all sterilization practices) is derived from practices originally developed for this process, and while appropriate for this specific process, their adoption across the spectrum of sterilization should be approached with some degree of caution. Many of these concepts were described in some detail in FDA's 1976 draft regulation on Large Volume Parenterals.¹²

In execution of these processes, provision for cooling of the containers at the conclusion of the exposure dwell cycle is often provided, which is intended to reduce the adverse effect of extended heating that might occur during an unassisted cooling. The microbial and chemical quality of the cooling water in direct contact with the containers is an essential consideration. An incident in the early 1970s, in which contaminated cooling water was aspirated into parenteral containers, resulted in the regulatory mandate for validation.¹³ This also entails maintenance of container integrity caused by stresses associated with increased differential pressure between the interior and exterior of container to prevent the ingress of cooling water. The use of air external over-pressure to assist in this is common place.

While the overkill sterilization method is preferred for items sterilized by direct contact, it is less commonly utilized in terminal sterilization, in which preservation of material properties is a primary consideration. The BB/BI and bioburden approaches are more appropriate in many applications. The validation of the process entails identification of both “cold” and “hot” spots

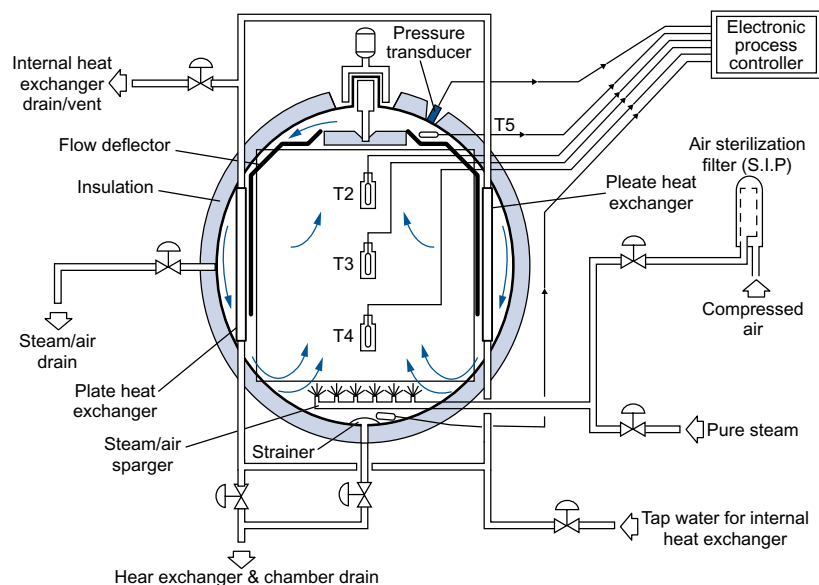


Figure 25-6. Steam-Air Sterilizer. Courtesy of Fedegari Autoclavi.

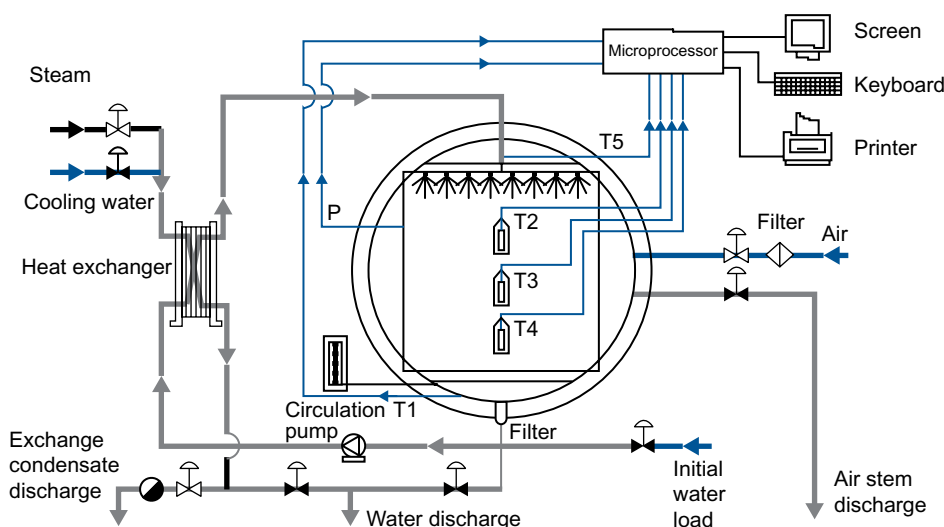


Figure 25-7. Steam-Air-Water Sterilizer. Courtesy of Fedegari Autoclavi.

to ensure that both minimum and maximum limits for sterilization are maintained.

Sterilization of In-Process Fluids

Formulation and filling processes often require small amounts of sterile aqueous materials for use in formulation, pH/volume adjustment, lubrication, and cooling of equipment. The sterilizers used for this process are typically of the same design as those used for direct sterilization. The only consideration given to their application for this process is the use of slow exhaust in an attempt to mitigate container integrity problems associated with sterilizing sealed containers. These containers are often sterilized in mixed loads with items sterilized by direct contact with the steam. The quantity of material processed in this manner is typically small, and the materials being sterilized generally have little heat sensitivity. The cycle dwell time is established to provide the appropriate exposure to the “worst case” item, regardless of whether it is a wrapped item or a sealed container. The overkill approach is commonly used.

Sterilization of fluids is sometimes necessary prior to the final filling of the distribution container. This may be necessitated by the desire to limit microbial growth during processing or

by the impracticality, or impossibility, of terminal sterilization or aseptic filtration. An example of this is influenza vaccine, in which low temperature sterilization by filtration is performed weeks, sometimes months, prior to final filling. In this case the sterile state is maintained by preservatives, such as formaldehyde and thimerisol.

Sterilization of Laboratory Media

The testing of many items requires a variety of sterile microbial growth media. Sterile media is available commercially in small containers; however, large users, especially those manufacturers having to validate aseptic filling operations using media fills and fermentation manufacturing, need to sterilize their own prepared media. Media for routine laboratory use may be prepared, filled into suitable bottles, and terminally sterilized in a batch-type autoclave. Larger quantities are prepared in stainless steel steam jacketed manufacturing tanks and sterilized in bulk. The sterilization processes are validated but have the advantage of being self-indicating after storage at ambient temperature. In laboratory settings these are processed in the same sterilizers used for direct sterilization in conjunction with the equipment and utensils necessary for the test. The “worst

case” item defines the process dwell period. In contrast with terminal sterilization processes, the sterilization of varying size containers with different media contents within them is commonplace. Laboratory sterilizers may provide some cooling capability (usually through the jacket) and air-overpressure because media represents their predominant use. The production of sterile media on a commercial scale resembles the terminal sterilization for finished products, with comparable equipment and controls. Over-sterilization of media can result in difficulty with preserving its growth promotion properties (and also makes the detection of growth more difficult), and thus BB/BI and bioburden methods are preferred.

Decontamination of Liquid Biological Waste

Many biotech manufacturing processes generate biohazards waste materials as a bi-product of the fermentation process. For example, large quantities of egg waste contaminated with influenza virus are generated during the manufacture of vaccine. Additionally, the ubiquitous microbiological laboratory produces biological waste materials as a course of normal operations. These waste materials have to be rendered safe prior to disposal. This is usually accomplished by using sterilization in batch type steam autoclaves with an extreme overkill process. The effectiveness of steam coupled with the disadvantage of chemical agents (replacing a bio-hazard with a chemical hazard) make steam the sterilizing medium of choice for this application. Incineration has been criticized as a decontamination method, especially for polyvinyl chloride (PVC) intravenous bags, where dioxin and acids are produced upon combustion. Knowledge of the materials to be decontaminated, specifically the decontamination kinetics, is necessary. Where possible, D and z values for the thermally vulnerable factor should be determined. This is not always easily obtained. In one instance, the authors recall having to decontaminate rattlesnake venom that had contaminated syringe equipment and bottles at an antivenom manufacturing facility. In this case it was necessary to contact the source producer of the venom to obtain process parameters.

Bio-waste decontamination is a common need in smaller facilities and laboratories where spent media and materials that are potentially contaminated with hazardous biological materials. Maintenance of container integrity is a consideration; however, over-processing is not a concern but rather an intention, and thus the overkill method is always utilized.¹⁴ In the most extreme installations, condensate from the sterilizer chamber is subject to overkill sterilization prior to discharge. Mixed items are commonly sterilized.

STERILIZATION-IN-PLACE (SIP)

The production of sterile products requires sterile containers to contain the items. On the smallest scales of operation, this entails the availability of sterile vials, bottles, and flexible bags in the compounding and holding of the materials. As the scale of operation increases, fixed vessels are more appropriate as intermediate processing and storage of sterile materials prior to aseptic filling operations. In the preparation of sterile bulk drugs, liposome formulation, and other complex formulations, the process often requires a substantial number of large sterile vessels. Another common need is for the sterilization of lyophilization equipment (chamber and condenser) prior to each batch.

The sterilization of these vessels and chambers is accomplished using direct contact with steam in an overkill process that is adapted from that used in steam autoclaves for individual smaller items. While identical in the core process, destruction of microorganisms by exposure to moist heat at elevated temperatures, there is one substantial difference that changes the

situation rather dramatically. Where a steam sterilizer is employed, the end user generally purchases the unit from an experienced manufacturer that has had considerable experience in the design and fabrication of the sterilizer, as well as in the establishment of process sequences and controls that assure its reliable performance. All purchasers of sterilizers utilize similar cycles and follow essentially similar methods in their cycle definition, control, and operation. The differences in sterilizer usage at each firm largely relate to the contents of the sterilizer and not to its design details.

The use of sterilization-in-place processes forces a major shift in the role of the end user. SIP systems are typically comprised of a vessel (or vessels) designed for use in a specified manner. As the usage changes, the details of the vessel and its inlet/outlet connections will change as well. As the intent with SIP is to sterilize the surfaces of the equipment rather than its contents, the details of the systems design are increasingly important. The variety of process applications requires the end user to develop a physical layout and process sequence amenable to SIP that may be unique unto it. As many systems subject to SIP involve liquid handling, the basics of the vessel are substantially different from a steam sterilizer. The scale of SIP systems can vary substantially as well. The smallest systems might be intended for portable tanks of 100 L or less with minimal additional piping. At the other extreme, a very large SIP system might include multiple large vessels (up to 5000 L) with numerous process filters and valves.⁷ This diversity of application entails unique approaches in design and execution.

The conventional starting point in SIP design is identification of the “sterile boundary,” that portion of the process system that is potentially exposed to sterile fluids. The boundaries of the system are commonly associated with filters and block valves. Once the boundary is defined, the design system requires many of the same considerations as sterilization in an autoclave. The most important of these include removal of air, removal of condensate, determination of a process sequence, and maintenance of system integrity post-cycle.

Air Removal

The presence of air within an SIP system prevents direct steam contact with the target surface. Removal of air from the system is most commonly managed using gravity displacement (there are some firms that endeavor to emulate steam sterilizers and use pre-vacuums). The most challenging areas of the system are process pipe runs that may extend some distance from the main vessel or chamber. Process piping that is remote from the vessel should be opened at the end to facilitate steam penetration. Steam is generally introduced at the top of the system to facilitate the air's downward movement. The steam may be introduced via a sterilizing grade filter, though larger systems may also utilize a direct steam inlet to ensure adequate supply.

Condensate Removal

In steam sterilizers condensate removal is accomplished by a single low point drain/steam trap, and the sterilizer is carefully leveled upon installation to be sure chamber is properly pitched to that location. The temperature control for the sterilizer chamber is commonly managed from a sensor in this drain line, which is colder than the chamber of the sterilizer. In an SIP system, especially those with a large number of inlet and outlet piping, there may be several low points where condensate could accumulate, and each must be opened to assure its removal. Process pipe should be sloped to facilitate the movement of condensate towards the low points. The SIP system may use steam traps, fixed orifices, or valves at each low point to restrict flow and build adequate steam pressure (and thus

⁷An example would be a bulk antibiotic process train that consisted of 5 major vessels, 10 sets of sterilizing grade filters for solvents and gases, the connecting piping and approximately 150 valves.

temperature). It is preferable to oversize these, as the loss of some steam is preferable to the adverse effect of condensate retention, which risks inadequate sterilization. The removal of air and condensate is accomplished by the same design elements with the only exceptions being vertical lines above the vessel chamber, where condensate flows downward into the vessel proper, while air might be removed at the top of the line farthest from the vessel.

Process Sequence

When operating a steam sterilizer, the operator merely has to press a button or make a key stroke to initiate a process sequence of multiple steps that will perform the sterilization. Sterilizer manufacturers have defined the steps, their sequence, and details for optimum performance. In an SIP application, it is the end user's responsibility to provide comparable certainty with respect to the execution of the process. Given the diversity of design and complexity, the process execution details may vary substantially.⁷ In larger facilities, SIP processes might be automated; however, the initial development of the software can be a trial and error exercise to define the proper timing and sequence of operation to assure reliable sterilization. Complete automation of SIP is uncommon; most automated SIP processes are supported by standard operating procedures requiring operator participation in the execution. The connection of spool pieces, hose connections, and adjustment of manual valves are the operator's responsibility. Fully manual SIP systems are also commonly used, in which the operator follows the SOP defined sequence.

Post-Sterilization Maintenance

The last portion of any sterilization process, and SIP is no exception, entails the removal of the sterilizing agent and returning the items to a usable state. With wrapped goods or sealed packages, this is rarely a major concern. In an SIP process the now-sterilized system must be protected until ready for use. This is accomplished by purging the system of steam at the conclusion of the dwell period with pressurized air (or other gas), delivered through a sterilizing grade filter. Initially the pressure of the gas must be higher than that of the vessel at sterilizing temperature. The filter is customarily sterilized *in-situ* with the rest of the system to eliminate the need for an aseptic connection. Attention must be paid to ensure that no portion of the system is allowed to become negative to the surrounding environment as that might result in the ingress of contamination. The internal pressure is maintained until the system has cooled to ambient condition, at which time it can be reduced to a level adequate to protect the system until it is used in the process.

SYSTEM DESIGN

The singular nature of many SIP systems makes the design of the system the most important concern of all with respect to their successful use. The design elements described below represent the more salient aspects of vessel and system design:

- The vessel should be rated for full vacuum and pressure to at least 130°C (and preferably higher).
- Inlet lines to the vessel should be kept short and installed vertically upwards to the extent possible.
- Where a line must extend some distance from the vessel, a bleed is usually required at the end of the line.
- Air and condensate bleeds should be placed at every low point in the system.
- Caution should be exercised in the use of jacket steam as it may result in a superheat condition within the vessel.
- Pipe runs should not be horizontal; the lines should be pitched (approximately 1/50-1/200) to provide for condensate drainage.
- Steam inlets to piping systems should be located at the high points in the system.
- Piping configurations shaped like the letters "W", "M", "V", "U" or "N" should be avoided as each of these will have a low point at which condensate can collect.
- Where flexibility in piping arrangement is required, permanent piping manifolds with appropriate valves or spool pieces to make temporary connections should be employed.
- Avoid flex hoses where possible; where they must be used, ensure they are sloped to drain at one or both ends.

FILTERS AND STERILIZATION-IN-PLACE

Within the context of SIP, the sterilizing grade filters used for steam, vent, process liquids, or gases require perhaps the greatest attention of all. They are a required part of virtually every installation to obviate the need for a post-sterilization aseptic connection. They are also prone to damage if subjected to excessive heat or pressure differential during the process, something that is rarely a concern with other physical components. The filter housing in larger systems is a small vessel in its own right, and the design considerations outlined above for process vessels apply. The porous nature of sterilizing grade filters makes them more difficult to sterilize as condensate within or adjacent to the membrane can have a deleterious impact on the sterilization efficacy.¹⁵

The outwardly similar process of sterilization-in-place may rely on the same lethal factors as steam sterilization in an autoclave; however, the uniqueness of system design where SIP is applied mandates an attention to design detail that transcends anything applied to autoclave designs. There are no universal solutions; each system must address the same fundamental concerns in a manner tailored for its unique configuration.

DRY HEAT PROCESSES

Dry heat processing for sterilization and depyrogenation has been a pharmaceutical process staple for many years. Items sterilized are typically glass containers, either Small Volume Parenteral (SVP) primary packages or laboratory glassware. Manufacturing equipment, consisting of glass or stainless steel items, is also sterilized. The sterilizing medium is hot air, and the predominant heat transfer mechanism is convection. The simplicity of process and abundant, inexpensive sterilizing medium has provided for its longevity. The disadvantages of dry heat relative to pressurized steam are the low thermal capacity of air, absence of latent heat (as provided in moist heat by condensation of the steam), and low surface coefficients for convective heat transfer. Whereas steam provides both sensible and latent heat upon condensation, dry heat has only the sensible heat. Consider how inserting an unprotected hand into boiling water versus into an oven at 100°C would compare, and the difference in heat transfer rates is obvious. Moreover, air heat transfer limitations also make air cooling inefficient. As a consequence, dry heat processes require longer process times to accommodate both the slow "come-up-time" and cooling of the load items. Dry heat equipment usually employs forced convection in which filtered hot air is blown by internal fans across the items to be processed to mitigate to the extent possible the poor heat transfer properties of air. Thermal process evaluation follows that used for autoclaves with the exception that the kinetic property of the thermally vulnerable factor (z) is 20°C for sterilization and approximately 50°C for depyrogenation. Note that this parameter, as in steam sterilization process evaluation, is a temperature difference and not a temperature. Also, the corresponding reference temperatures (T_b) are 170°C and 250°C, respectively, rather than the 121.1°C used in steam sterilization process evaluation. The corresponding sterilizing value to F_0 estimated for a dry heat process is called " F_H ." As described by Hitchens, bacteria enter a "self-preservation" state when deprived of conditions conducive to rapid growth.¹⁶ This "resistant capsule" or spore contains the vital part of the bacteria and, when entirely dry, can be heated to a temperature

far above the boiling point of water without killing them. The biological challenge of choice for dry heat processes is the heat resistant bacterial spores of *Bacillus subtilis* at a typical concentration of 10^6 spores per challenge location. Typical D-values are substantially different for sterilization and depyrogenation. Dry heat sterilization at 170°C uses either *Bacillus atrophaeus* or *Bacillus subtilis* that have D_{170} of approximately 1–2 minutes. Depyrogenation at 250°C presents a very different picture: vegetative cell D_{250} values estimated at NMT 0.00005 minutes, with *Bacillus atrophaeus* and *Bacillus subtilis* D_{250} estimated at 0.002–0.003 minutes, while purified endotoxin is in the range D_{250} of 0.5 to 2.0 minutes.⁸

The efficacy of dry heat sterilization is dependent upon three primary variables and three secondary variables.¹⁷ The primary variables are time, temperature and water content of the thermally vulnerable factor. The secondary variables are nature of adjacent support, nature of the system containing the thermally vulnerable factor (open or closed), and nature of the sterilizing medium. The secondary variables generally address whether the mode of heat transfer is convection or conduction. The moisture content of the cell appears to have an influence on mode of action of dry heat, which has been widely accepted as a denaturation of essential genetic material used for reproduction. There are two types of dry heat sterilizers, the batch type, consisting of ovens, and the continuous type, consisting of tunnels.

Dry Heat Ovens

Ovens can be single door or double door; the latter type is used when processing equipment to be sterilized into a classified area. Items to be processed are placed upon trays directly in smaller ovens or on wheel-in tray racks for the larger units. A diagram of the air flows and major components for a typical unit is shown in Figure 25-8. Note that there are no “standard” designs for ovens, and the arrangement of the major mechanical components of the oven may vary from that shown.

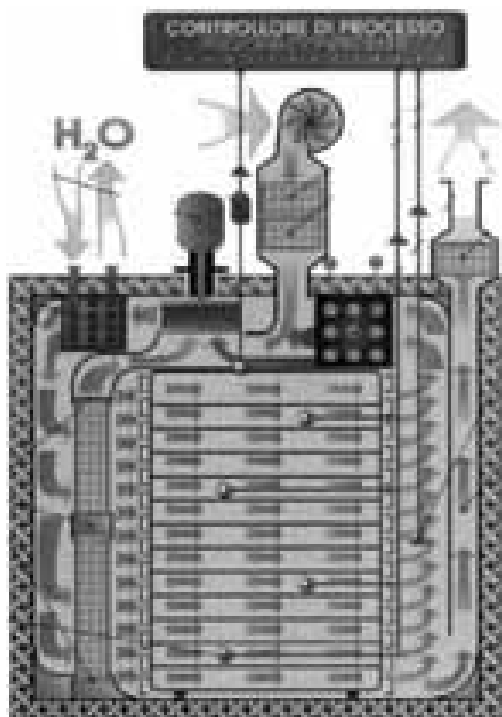


Figure 25-8. Typical Dry Heat Oven cross section. Courtesy of Fedegari Autoclavi.

Heated air is blown across the trays from one side of the oven to the other and recirculated. Filtered (HEPA) make-up air and filtered exhaust ducts are provided to ensure no contamination along with maintaining a positive air pressure with respect to non-classified areas. Materials of construction are selected to provide ease in cleaning and preclude particulate matter generation, with stainless steel a common choice. Non-stainless steel components, such as motors and conduits, are external to the oven chamber and recirculation ducts. A disadvantage of batch type units is that items are exposed to a different process, based upon their location within the oven chamber. The load items may have aluminum foil covers or wrappings to prevent contamination with particulate matter. Primary glass containers are typically loaded upside down. Older ovens may have broken heating elements, broken insulators, or residue from broken glass containers hidden behind plenum walls that produce particulate matter. These particles are blown around by the high velocity circulation air, which contaminates the load. Effective preventive maintenance programs, along with current monitors on heating elements, are necessary to prevent particle contamination of previously washed load items. Qualification protocols should ensure that effective cleaning and preventive maintenance programs are in place for these units. Typical temperature plots are shown in Figure 25-9.

Unlike some autoclaves, the dwell period start condition is based upon the heating medium temperatures and not load temperatures. The controlling temperature probe (thermocouple or resistance temperature detector) is typically located on the cool side recirculation plenum wall. Note in Figure 25-9 that the minimum load, consisting of a single slowest to heat item, shows the characteristic of having the process timer start well before the load is at temperature, considerably lower than the maximum load at the same point in time. It is essential that validation studies include the minimum load and that it not be an “empty chamber.” Maximum and minimum loads can be used to establish the process for intermediate size loads, provided all components are identical. In mixed loads the load size and configuration must be evaluated for each load. When ordering dry heat ovens, it is important to ensure that they are specified with thermocouple access ports for process development, determination of the cold spot(s), and process validation. Overkill processes (12D, 15D in the EU) are typically used, in which the D-value used is that of the resistant biochallenge rather than an actual naturally occurring bioburden. Glass containers are extruded at 1200°C , and both bioburden and pyroburden arises from point type contamination after processing, which is difficult to detect with sampling schemes.

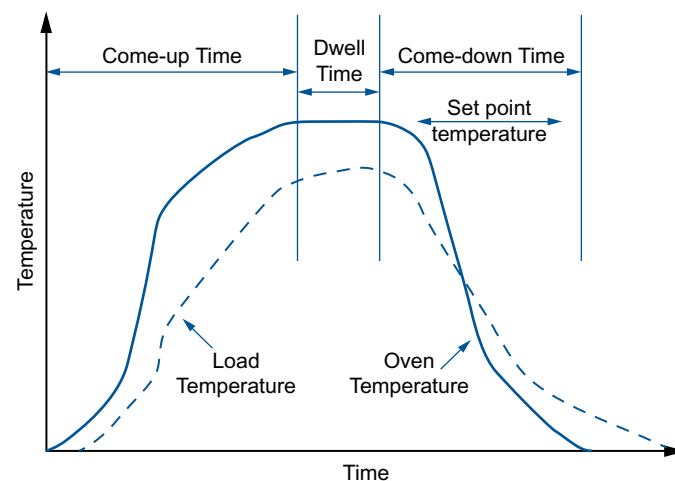


Figure 25-9. Typical Oven Cycle.

⁸Accurate determination of D_{250} values for these microorganisms is not possible because their death is so rapid at the elevated temperatures.

Depyrogenation Tunnels

Depyrogenation tunnels are continuous processing units used for use for primary glass containers, typically installed following a washing operation and providing in-feed directly to an aseptic filling operation. An advantage of the continuous process is that each container sees essentially the same process. Units using wide feed belts do have some variation across the belt, but this is small compared to the typically wider temperature distribution experience with batch ovens. Radiation is also a significant heat transfer mechanism in these units, in addition to laminar air flow. Typical units have three sections consisting of heat-up zone, process or dwell zone, and cooling zone (see Figure 25-10).

Components are slowly conveyed along on a stainless steel belt which exposes them to the three zones. Contamination is prevented by using filtered air and the use of positive pressures relative to adjacent non-classified spaces. The target thermally vulnerable factor in depyrogenation processes is lipopolysaccharide, a fever causing bacterial cell product. Commercially available biochallenge material is available for process validation studies.¹⁸ Unlike bacterial contaminants in steam sterilizers, first order reaction kinetics are not easily applied to these chemical entities and dry heat processes that have wide temperature distributions. A widely used empirical reaction kinetics equation¹⁹ is shown below:

$$\log Y_0 - \log Y = AB10^{ex} \quad (9)$$

Where Y_0 is the initial concentration, Y is the concentration after the depyrogenation process, and A , B , and C are constants for the selected temperature. The parameter x is the dry heat sterilizing value F_H described above. A disadvantage of this equation is that for a given temperature, a maximum log reduction of the initial concentration is observed, regardless of the level of the initial concentration or the magnitude of F_H . For example, for 250°C there is a maximum log reduction of slightly above six. As in ovens, overkill processes are employed; in this case, temperatures above 250°C are used to deliver a process equivalent to 30 minutes at 250°C within the process section. Belt speeds of 45 cm (1.5 ft) per minute are typical.

In summary, dry-heat processes are conducted to provide sterilization and/or depyrogenation. The latter typically

requires 250°C for 30 minutes, or equivalent. Sterilization can be accomplished at 170°C (T_b) with process times usually 20 minutes following a come-up time to achieve the process temperature (T_1), which usually exceeds the reference temperature. Equipment is selected, validated, and maintained to prevent contamination with particulates and/or microbes during and following processing.

GAS, LIQUID AND VAPOR STERILIZATION

Sterilization processes entail the exposure of microorganisms to conditions that destroy their viability. A variety of chemical agents are able to kill microbes by chemical reactions, many of which are oxidation related. The number of agents that can accomplish this in a rapid manner is extensive, and many of the agents are lethal in both the gaseous and liquid (water solution) state. Because the killing mechanism with these agents is based upon a chemical reaction between the chemical agent and the microorganism, first order kinetics as described in the Arrhenius equation (see Equation 10).

Arrhenius equation

$$-k = Ae^{-E_A/RT} \quad (10)$$

Where

k = Reaction rate

A = Prefactor

E_A = Activation energy

R = Gas constant

T = Absolute temperature

The rate of kill is predominantly a factor of the agent concentration, with temperature and humidity playing substantially less influential roles.¹⁰ The basis for all of the gaseous agents is ethylene oxide (ETO) where a substantial body of data has been developed, and the practices for ETO have largely defined the methods used for the other gases. Essentially identical approaches can be used for liquid sterilization as these operate in a single phase as well. The fundamentals of vapor processes differ somewhat because of the two-phase aspect of their delivery and action. Nevertheless some parallels to the base ETO process can be useful in those applications. These processes

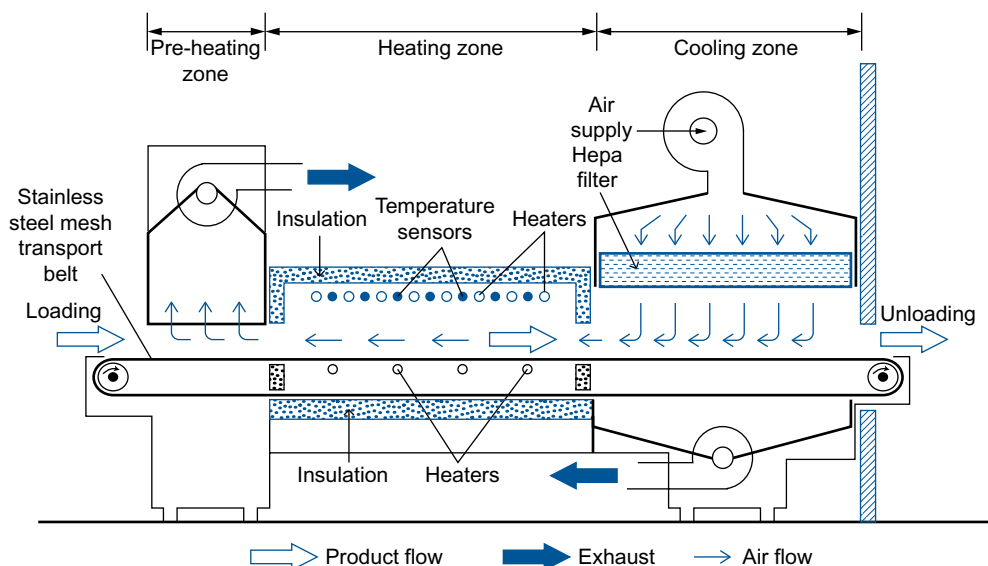


Figure 25-10. Typical Dry Heat Tunnel. Courtesy of Fedegari Autoclavi.

⁹Studies with ETO sterilization have demonstrated a z value of approximately 32° C and little difference in microbial resistance across relative humidities between 40-90% (Gillis & Mosely). There is a paucity of data on other agents, which neither supports nor repudiates the ETO experience.

can benefit from mixing or agitation in all phases of the process because mixing serves to improve uniformity.

Gas Sterilization

Ethylene Oxide

The most widely used gaseous sterilization agent is ethylene oxide (C_2H_4O), which kills by alkylation of the nucleic acid within the microorganism.²⁰ The range of lethal concentration begins near 300 mg/L and extends to approximately 1200 mg/L as higher levels have not been shown to result in substantially shorter process times.²¹ At one time ETO sterilization processes, whether in sterilizers or for lyophilizers, were commonly found in pharmaceutical plants. Recognition that ETO is both a carcinogen and mutagen led many pharmaceutical firms to rely on contract sterilizers for items requiring ETO sterilization. The violently explosive nature of ETO was another consideration in its displacement from pharmaceutical production sites. ETO is widely used for medical device sterilization, and the preferred methods are largely optimized for that application. Device sterilization is accomplished with the device in its final bulk package (ordinarily in pallet configuration), including multiple layers of packaging. Device packaging often entails Tyvek® lidded plastic trays with additional fiberboard/corrugate. In order to effect sterilization of the device, ETO must penetrate all of the layers. For effective kill adequate humidity must also be present. A typical ETO cycle (using 100% ETO) for medical devices would include the following steps:

- **Pre-Conditioning** – The load items are exposed to elevated humidity (90-95%) to increase moisture levels on the device surface prior to introduction to sterilizer. The pre-conditioning makes the microorganisms more susceptible to ETO. In temperate climates, this step is essential to ensure a consistent process despite seasonal variations in humidity that would otherwise alter process lethality. This step is performed in an area specifically designed (and qualified) for this purpose.
- **Re-Conditioning** – After transfer from the pre-conditioning area to the sterilizer, reconditioning in the sterilizer may be performed to re-humidify those peripheral portions of the load that may have dropped in humidity while in transit. This would be accomplished by low pressure steam injection. This step is optional.
- **Evacuation** – To mitigate the explosive potential of ETO sterilization, air is removed from the chamber. Unlike steam sterilization, in which air removal is sometimes carried to excess, the removal of air must be tempered with the desire to maintain adequate humidity within the load. Thus a single pre-vacuum is generally performed as desiccation of the load should be avoided.
- **Humidification** – The replacement of moisture within the load follows to return the load to the desired state.
- **Charge** – ETO gas and nitrogen are added to attain the desired concentration and pressure for the process dwell.
- **Exposure** – The lethal conditions are maintained for the pre-established process dwell period with ETO gas make-up to maintain the target concentration.
- **Exhaust** – The ETO /nitrogen mixture is removed from the chamber to make the materials safe for handling during the unloading process. The exhaust cycle may include multiple pulses of nitrogen and eventually air.
- **Unloading** – The sterilizer is unloaded and items transferred to the post-conditioning area. Operating personnel performing this task may need to wear personal protective equipment to minimize their exposure to trace amounts of ETO diffusing from the materials.
- **Post-Conditioning** – The load is aerated so that levels of ETO (as well as common process byproducts ethylene chlorohydrin and ethylene glycol) can be reduced to levels safe for the intended use. FDA established its expectations for these residuals in 1978.²² This remains the only

published expectation with respect to acceptable residual levels. The aeration area is typically supplied with 100% fresh air to aid in residue removal. The external aeration stage can last in excess of eight weeks.

Cycle Development, Validation, and Routine Process Control

The essentially linear death curves associated with ETO sterilization facilitates cycle development. There are two basic approaches utilized to establish the process D-value: fraction negative method and graduated spore strips. These can be applied in either a smaller test chamber or the production sterilizer. In the fraction negative method, a series of short duration test cycles are performed, in which some, but not all, of the biological indicators are inactivated. Graduated BIs exposed to an abbreviated cycle provides a similar assessment of lethality. The results of either of these studies can be utilized to determine the D-value for the microorganism within the load at the tested conditions.²³ The process dwell time that will be established is a multiple of the D-value.

The oldest approach to establishing and validating the process is the half-cycle, in which the D-value is multiplied by 12 to determine the half-cycle process dwell period. The load is challenged in triplicate with multiple biological indicators (from the same lot as evaluated in the D-value study) at locations across the full load. The complete inactivation of the indicators supports the process lethality. The routine sterilization cycle is defined as twice the cycle (thus the name half-cycle) in which the BIs were all killed, with the belief that this additional dwell time is more than adequate to support variations in gas concentration, relative humidity, and temperature during routine processing as compared to the conditions demonstrated in the half-cycle studies (see Figure 25-11).

An alternative approach to cycle development is the bracketing approach in which the process dwell as established by the half-cycle defines one side of the process bracket. The other side is defined by confirmation that materials exposed to increased gas concentration, higher humidity, higher temperature, and longer process dwell are acceptable for use. Useful where process parameters can be more narrowly controlled, this method results in generally shorter process times, which can be important to key material attributes and/or sterilizer capacity (see Figure 25-12).

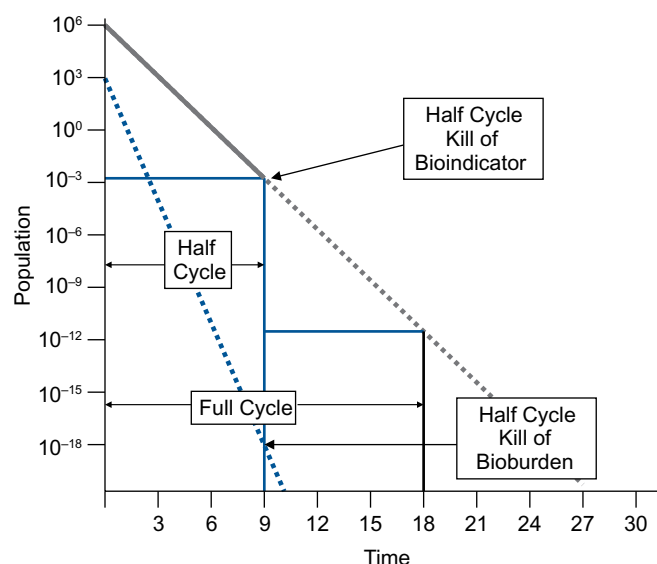


Figure 25-11. Visual Depiction of Half Cycle Sterilization.

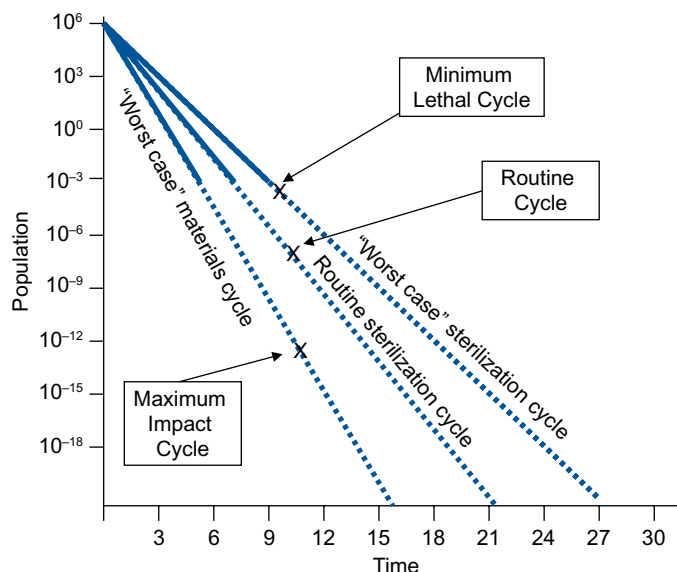


Figure 25-12. Visual Depiction of Bracketing Approach.

Routine process for ETO processes consists of process monitoring of the parameters along with each cycle biological indicator testing. Fully parametric release is also possible, and this is discussed later in this chapter.

Other Sterilizing Gases

There are a number of gases with demonstrated lethality against microorganisms other than ETO. Many of these gases, including hydrogen chloride (HCl), chlorine (Cl₂), sulfur dioxide (SO₂), and sulfur trioxide (SO₃), are unfortunately too corrosive and/or hazardous for easy application in a pharmaceutical plant.¹⁰ Less chemically aggressive but highly effective are such gases as chlorine dioxide (ClO₂), ozone (O₃), and nitrogen dioxide (NO₂).^{24,25,26} Each of these agents was developed and commercialized at least in part as an alternative to ETO, and thus the basic approaches to their use is similar. Pre-humidification prior to chamber loading is not a part of these processes, but each process does have an in chamber humidification step.

CHLORINE DIOXIDE

This moderately stable gas is prepared by chemical reaction in proximity to its point of use. It is less penetrating than ETO and appears to be less readily absorbed by materials and thus is easier to remove post-treatment. Chlorine dioxide is lethal in the range 5-30 mg/L at ambient temperature and humidity levels of 60-75% RH. Chlorine dioxide levels can be measured using relatively inexpensive UV sensors, facilitating routine process control. Exhaust from the sterilizer is chemically scrubbed to prevent atmospheric release.

OZONE

This unstable gas is manufactured just prior to introduction into the chamber by passing pure oxygen through an electrical field. It is also less penetrating than ETO, and because of its instability, comparatively easy to remove post-cycle. Ozone is effective at concentrations between 2-10% at room temperature and relative humidity levels 80%.

NITROGEN DIOXIDE

This gas is effective across the range of 8-21 mg/L at ambient temperature in the presence of humidity.²⁷

Liquid Sterilization

Historically, chemical sterilization was used as a direct additive to vaccine formulations as a super preservative, and formaldehyde was used for this purpose. Formaldehyde is still used, along with the mercury-containing preservative thimerisol in some current vaccine manufacturing, as a supplement to filter sterilization. Purified water systems can be sterilized by hypochlorite or ozone where high temperature sterilization is not possible. Contact time and concentration are the critical process parameters in chemical sterilization. Hydrogen peroxide solutions are used to in-line sterilize packaging components for form fill and seal aseptic filling operations. Sufficient dwell, or contact time, is needed to assure sterility. Following sterilization, the hydrogen peroxide solution must be neutralized and removed. This is accomplished by exposure to ultraviolet light, which may provide an additional lethal effect. Another common application is the use of dilute NaOH solutions for the cleaning (and effectively the sterilization) of chromatography columns.

Sterilization in the liquid phase obeys the same precepts as other sterilization processes and closely resembles gas sterilization. Agents effective as sterilants in aqueous solution include acids, bases, aldehydes, and various oxidizing agents.¹⁴ The more commonly used materials are hydrogen peroxide and glutaraldehyde in various concentrations. These are chosen because of their generally good compatibility with stainless steel, glass, and many plastics easing the selection of process equipment. The process can be as simple as immersion of materials in the tray or an automated system in a closed vessel. Regardless of the system, the microbial destruction occurs in the same linear fashion observed with gas phase sterilization. There is a paucity of kill rate data for liquid chemical sterilization, but the simplicity with which the process can be executed enables it to be generated rather easily. The most resistant microorganisms are believed to be bacterial spores. The only difficulty associated with the use of liquid chemical sterilization is the need to rinse and/or deactivate the chemical agent from the system or item and maintain its sterility before it can be used.

Vapor Sterilization

The gas and liquid sterilization processes described above operate similarly despite the difference in phase. When the same agent can be delivered as a gas or dissolved in a liquid, there are differences in the lethal rate. There are at least two major reasons for the lethal rate difference: the concentration of the chemical agent is generally higher in the liquid phase, and moisture levels in liquids are always higher. Where data are available on the same material, the kill rate in the liquid phase is typically more rapid. Vapors are loosely defined in the scientific community:

- Barely visible or cloudy diffused matter, such as mist, fumes, or smoke, suspended in the air.²⁸
- Any substance in the gaseous state, the condition of which is ordinarily that of a liquid or solid.²⁹
- Vapor is a condensable gas.³⁰

Clearly vapors are combinations of both gas and liquid phases. At their entry point they may resemble pure gases, but when exposed to lower temperatures, as is commonplace in ambient temperature applications, condensation of the less volatile components of that vapor can be expected. Most sterilizing vapors are solutions of a lethal chemical material in water. There are three vapors of interest to sterilization scientists: hydrogen peroxide, introduced as hot gas slightly above the boiling point of H₂O₂ (108°C for a 35% solution in water); peracetic acid, sprayed on as a room temperature liquid into closed chambers:

¹⁰Note all of these chemicals agent listed are effective as antimicrobials in aqueous solution as well; however, those that form strong acids in solution are rarely used exclusively for that purpose.

and formaldehyde, added as either as a heated gas or evaporated from its solid counterpart paraformaldehyde. Regardless of the vaporous agent, the potential for the presence of two phases in the system to be sterilized must be recognized. The inability to define the sterilizing conditions for 2 phase systems effectively eliminates a D-value determination and the attendant process control opportunities that affords. Vapor sterilization processes have to be established and validated empirically using biological indicators since parametric measurements cannot be relied upon to fully substantiate lethality. Achieving a consistent process with a vapor process mandates that the vapor be well mixed within the chamber to minimize stratification and the consequent concentration variation that the temperature differences between the surfaces and the vapor will create.

As vapors will have the lethal agent present in both phases, it can be useful to visualize the sterilization process in parallel (See Figure 25-13). There are no dimensions on this image since the slopes of the survivor curves are not well defined. It must be understood that the gas and liquid phases within the system are in equilibrium, and the concentration of the lethal material in each phase differs.

Line (A) in Figure 25-13 represents the survivor curve for microorganisms exposed to gas throughout the process with the slowest kill rate, as might be expected due to its lower concentration. Line (B) depicts the more rapid kill attainable in the liquid phase. Lines (C) and (D) represent locations where the microorganism sees both gas and liquid at some time during the process. Considering the entire system, an overall survivor curve considering a large number of microorganisms located throughout the chamber with varying phase exposure would reveal a "band" bounded by lines (A) and (B) rather than a single best-fit line (see Figure 25-14).

If vapor sterilization is to be adequately controlled, the inherent uncertainty of the rate of kill must be understood. The absence of metrics for the X and Y axis in these images is intentional; the rates of kill will vary with respect to several factors including concentration of the agent in the initial solution, rate of agent feed, temperature of the inlet vapor, temperature range across the chamber, and initial humidity level of the chamber. A suggested means for the empirical cycle design, validation, and routine control of vapor sterilization is depicted in Figure 25-15.

The execution of validation studies less lethal (as proof of sterilization efficacy) and more lethal (as proof of material compatibility) than the routine can be used to substantiate its acceptability. This mimics the bracketing approach described

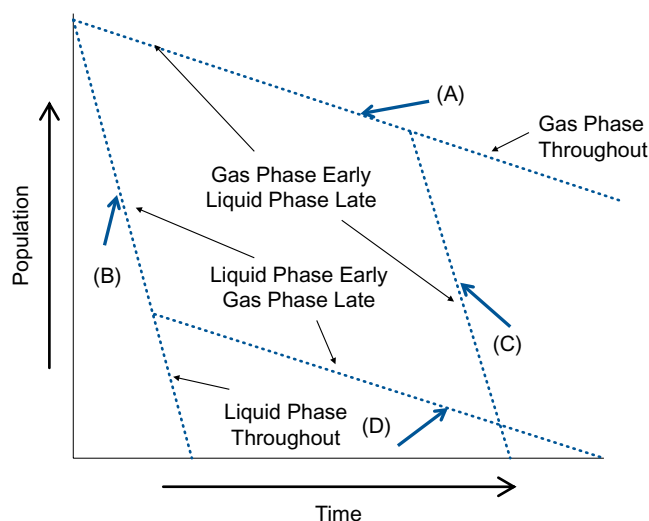


Figure 25-13. Bi-Phase Survivor Curves.

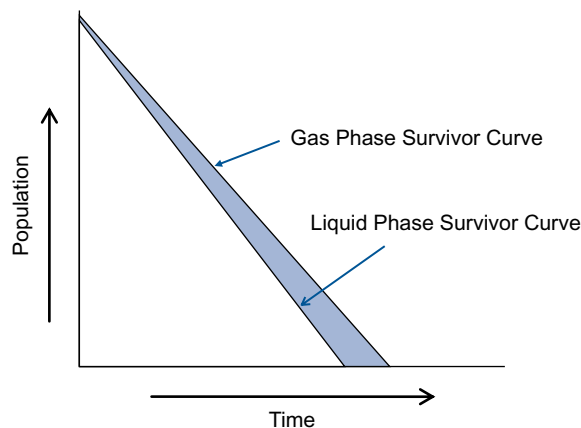


Figure 25-14. Vapor Phase Survivor Curves.

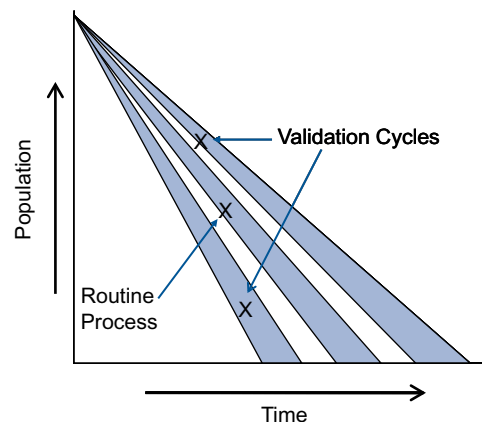


Figure 25-15. Validation Approach for Vapor Sterilization.

above for gas sterilization (see Figure 25-11) and is particularly appropriate for vapor sterilization where D-values are not available.

RADIATION STERILIZATION

Radiation is often chosen out of necessity and/or convenience as a sterilizing medium in pharmaceutical and medical device manufacturing and is increasingly used for plastic materials. There are three main types used: ionizing gamma (γ) irradiation from cobalt-60 (^{60}Co), ultraviolet (UV) radiation from lights, and electron beam from cathodes. There are other less used processes, such as microwave, X-ray, and high intensity visible light. The mechanism of sterilization involves the formation of free radicals and excitation of double covalent bonds, causing chemical reactions that degrade the reproductive systems of the target microorganisms. The different forms of radiation differ in the extent of their penetration through materials that will result in different process duration and control. Nevertheless, they achieve sterilization by the same basic principles and are validated in essentially the same way. Recall that it requires a three-inch thickness of lead to stop gamma radiation. Electron beam is much less penetrating, and ultraviolet light can be stopped by almost any thickness of an opaque material. Processing for each sterilizing medium varies to eliminate idiosyncratic disadvantages, such as shadow effects and dose variation, while maximizing penetration.

Out-sourced gamma ray terminal sterilization is attractive because often it can be performed for products already in tertiary packaging (corrugated cardboard shipping containers). Compare this to the costs and inconvenience of operating a sterile filling and packaging facility, and the choice appears obvious. The USP² and EU³¹ Pharmacopoeia originally recommended doses of 25 kGy and 32 kGy to assure a sterility

assurance level (SAL) of 10^{-6} .⁸ However, the major concern with the use of gamma ray sterilization is its effect on the product. The various forms of radiation degrade materials to varying levels dependent upon exposure levels. Dosimeters are made of plastic material (Harwell red Perspex 4034 dosimeters), and the dose is measured by a color change from dark red to darker red, which attests to the degradation effects of the radiation. The processing usually consists of palletized product containers moving into a chamber and remaining stationary or product in carriers, traveling around the source in a train-like conveying system to expose the product from all angles to ensure maximum penetration and increase delivered dose uniformity. Dosing can be highly variable, especially remote from the source. In a recent study Ameri *et al.*,³² evaluated and compared terminal sterilization of a parathyroid hormone transdermal micro-projection delivery system by gamma or e-beam versus aseptic processing.³² In developmental work they determined that a 21kGy dose of gamma radiation was necessary to have a SAL of 10^{-6} for and average bioburden of 88.7cfu. They found degradation after three months on stability testing for both irradiation and e-beam terminal sterilization and concluded that aseptic processing was the only feasible procedure that could provide two-year product stability. Even if the active is stable, the other ingredients may be adversely affected. In a recent study by Hegge *et al.*³³ on alginate foams containing curcumin, the active ingredient was exposed to a measured dose of 31kGy. The level of active was not affected; however, the absorption spectra differed from the pre-sterilized product, and a decrease in curcumin release was observed, indicating that the excipients were adversely affected by the treatment.

UV light is utilized primarily as an additional microbial control for water systems in pharmaceutical manufacturing. Some USP Purified Water systems and Water for Injection systems have supplemental UV light treatment units. Wavelengths of 253nm to 264nm are used to control bioburden levels in the recirculation loop. UV is also used to sterilize containers or as a treatment to remove a vaporized hydrogen peroxide sterilization treatment. The advantage of UV is that it can be used for continuous processing. The primary disadvantage of UV is the shadow effects and degradation of the UV source with time. The latter requires diligent preventive maintenance for production systems.

In summary, gamma sterilization is utilized when stability of all components is not an issue or where product cannot withstand heat or be aseptically sterilized and packaged. Electron beam does not have the penetration of gamma, but can be set up for semi-continuous processing. UV is applicable for continuous processing, primarily in water and packaging applications. Stability using any radiation process must be assessed, even if products and components are heat stable. Some materials experience a dramatic drop off in physical/chemical properties that cannot be predicted in accelerated stability studies; therefore, real-time data are much preferred.

STERILIZATION BY FILTRATION

Filtration removes particles from fluids. Unlike the other sterilization processes described in this chapter, sterilizing filtration relies on physical removal of microorganisms from the material to be sterilized; it does not kill or destroy the microorganisms. Sterilizing filtration is defined as the ability to remove a minimum of 1×10^7 per cm^2 of *P. diminuta* (currently classified as *B. diminuta*) when the filter is challenged under specified conditions.³⁴ Filtration is used where solutions may be degraded by other sterilization processes, such as autoclaving. Careful selection of the filter and filtration conditions is necessary to ensure that the process will yield a sterile filtrate.

History

Filtration as a mechanism for sterilizing pharmaceutical solutions was developed in the early part of the last century because of the need to sterilize heat-sensitive products. Three technologies were pursued: porcelain filter cartridges, asbestos-cellulose layers, and microporous membranes.³⁵ Filters made of porcelain and of asbestos-cellulose layers function as depth filters; that is, particles are retained by entrapment within the filter matrix. The primary retention mechanism of microporous membranes is sieve retention, wherein the particles are too large to fit through the pores of the filter. Porcelain filter cartridges were used to sterilize antibiotics; however, the filters were difficult to clean, and there were concerns about cross-contamination. Additionally, integrity testing that could be related to microbial retention was not possible. Seitz EK (*entkeimung* “germ removal”) asbestos-cellulose filters were used through the mid-1970s until the FDA’s Good Manufacturing Practice (GMP) regulations prohibited the use of asbestos and fiber-releasing filters in parenteral manufacture.^{36,37} As with porcelain filter cartridges, definitive integrity testing was not possible. Filters in the form of collodion membranes were manufactured by filter users in the early 1900s. This led to the development by Zsigmondy (Sartorius-Werke, Göttingen, Germany) of the first commercially successful membrane filter in 1929. Reliable integrity testing of these membrane filters was possible, allowing sterilizing filtration processes to be validated, another requirement of the GMP regulations. Flat-stock membranes have largely been replaced in sterilizing filtration applications by pleated cartridges, increasing throughput and minimizing handling concerns for fragile membrane materials. The use of single-use systems in the manufacture of high-value and highly potent pharmaceutical and biopharmaceutical products has facilitated the increasing use of pre-sterilized capsule filters (Figures 25-16 and 25-17).

Types of Filters and Membrane Materials

Filters for sterilizing pharmaceutical solutions are available in a variety of configurations and membrane materials. While



Figure 25-16. Cartridge filter used with a single-use production system. Courtesy of Sartorius Stedim Biotech.

¹¹This recommendation was removed from current USP 34 NF 29.



Figure 25-17. Capsule filter. Courtesy of Pall Corporation.



Figure 25-18. Cartridge filters. Courtesy of Pall Corporation.

flat-stock membranes continue to be used for small-scale filtration and bacterial retention studies, cartridge filters generally are used for production-scale sterilizing filtration because cartridge filters, either singly or in cartridge arrays, have the necessary flux and capacity to avoid the necessity of replacing them during the filtration process, potentially compromising the sterility of the filtration train (Figure 25-18).

Membranes with pore size ratings of 0.2 μm are typically used for sterilizing filtration, although 0.1 μm - and 0.45 μm -rated filters are sometimes used—0.1 μm -rated filters for long-duration filtration of products that may contain small, water-borne microorganisms such as *Pseudomonas* sp., while 0.45 μm -rated filters can be used where the filtration conditions require high flux and filtration capacity and where it has been demonstrated that the pre-filtration bioburden can be consistently removed under the validated filtration conditions.

Microporous membranes are composed of various polymers, including cellulose acetate, polysulfone, polyethersulfone,

polyethylene, polypropylene, polyvinylidene fluoride, and polyamide. Each of these membrane materials has properties that lend its use to a particular application.³⁸ For example, cellulose acetate exhibits low adsorption and high flow rates, but it has limited pH compatibility; polyamide (Nylon® 66) has good mechanical strength and broad pH compatibility, but it has high nonspecific protein adsorption and low hot water resistance; polyethersulfone exhibits high flow rates and throughput and broad pH compatibility, but it has limited solvent compatibility and moderate to low nonspecific adsorption, depending on surface modifications; and polyvinylidene fluoride has low nonspecific adsorption and good solvent compatibility, but only moderate flow rate and throughput, and because it is made hydrophilic by surface treatment, the membrane may lose hydrophilic properties during the filtration process. Studies should be conducted to ensure that the chosen membrane material is compatible with the product to be filtered, has the necessary flow rate and capacity, and is able to withstand the filtration conditions, including flow rate and pressure differential. The membrane and its housing should not adversely affect the product by introducing leachables or by adsorbing product components beyond established limits. Examples of filters and filter configurations are shown in Figures 25-19 and 25-20.

Retention Mechanisms

Filtration removes microorganisms from a liquid by two primary mechanisms: sieve retention and adsorption.

Sieve retention involves the capture of a microorganism that is too large to pass through a filter pore. It relies on physical blockage of particles that are larger than the pores they encounter. Although sieve retention is generally independent of filtration conditions and the microbial challenge level, factors such as fluid composition and filtration conditions can affect sieve retention. High ionic strength and osmolarity can draw water from



Figure 25-19. Large capacity, multiple cartridge, filter housing. Courtesy of Pall Corporation.



Figure 25-20. Assembling a cartridge filter and its housing. Courtesy of Sartorius Stedim Biotech.

microbial cells, reducing their size. High pressures and flow rates can “squeeze” microorganisms through the filter’s pores.

Adsorption retention is a charge-related phenomenon wherein microbial cells are captured by the surfaces they encounter without being physically restrained. Examples of the forces involved include electrical attractions, such as valence and ion-exchange, van der Waals forces, and hydrogen bonding. Adsorption is influenced by the composition of the membrane, the properties of the filtered fluid, the filtration conditions, and the number and type of microorganisms present in the fluid.

It is important to consider both mechanisms when developing and validating sterilizing filtration processes.

Validation

Good manufacturing practice regulations require validation of any sterilization process. Sterilizing filtration is validated in several stages, and it involves verification of microbial retention and determining that the filter does not contribute leachables to the product beyond established limits and that the filter does not unacceptably absorb formulation components.

Filter manufacturers ensure that membranes of a particular pore-size rating and composition are able to retain a microbial challenge of 1×10^7 per cm^2 of effective filtration area. This standardized testing is accomplished using ASTM F 838-05³⁹ which specifies the use of *B. diminuta* as the challenge microorganism and details the process conditions and apparatus to be employed. Successfully passing this test enables filter manufacturers to correlate integrity test results with microbial retention and to set integrity test specifications to be used for in-process production control and quality assurance purposes.

Filter manufacturers usually conduct a series of tests to determine the extractable and leachable elements of the microporous membranes and other filtration components, such as O-rings, gaskets, and housing materials.

Filter users must also validate the sterilizing filtration process to ensure it is capable of removing all microorganisms from the process stream under actual conditions of use and to establish a correlation between microbial retention and the integrity test value of the filter under actual conditions of use. Typically, *B. diminuta* at a level of 1×10^7 per cm^2 of effective filtration area in the product is used for this challenge. It is known that filtration efficacy can be affected by many factors, including the type and number of microorganisms to be removed, the properties of the liquid to be filtered, the membrane material, and the filtration process parameters. The surface tension, pH, and ionic strength of the liquid to be filtered can strongly influence microbial retention, while aspects of the filter that can impact the filtration include filter area, nominal pore size, pore-size distribution, filter membrane thickness, porosity, membrane polymer, filter pleat density, non-woven support layers, electrostatic charge, and hydrophilic nature of the filter membrane. The filtration process parameters influencing microbial retention include temperature, flow rate, differential pressure, and pressure pulsations. If the challenge organism is not viable in the product, simulated product (e.g., adjust pH, remove preservative) can be used for the validation testing. Another option is to use product indigenous bioburden microorganisms for the validation testing. In addition to microbial retention testing, leachable and adsorption studies are conducted with the product of interest to ensure the filtration process does not alter the product beyond acceptable established limits. PDA Technical Report No. 26 contains comprehensive information on appropriate validation strategies for sterilizing filtration.⁴⁰

Integrity Testing

The integrity test values established during validation of the filtration process are used prior to and after filtration to ensure the filter is integral and capable of retaining microorganisms under the intended conditions of use. Bubble point and diffusive flow are two commonly used integrity tests (Figures 25-21 and 25-22).



Figure 25-21. Automated integrity test unit. Courtesy of Pall Corporation.



Figure 25-22. Performing an automated integrity test on a cartridge filter. Courtesy of Sartorius Stedim Biotech.

The bubble point occurs when a gas displaces a wetting liquid from the largest membrane pores as gas pressure is slowly increased, resulting in bulk gas flow through those pores. The flow is evidenced by a steady stream of bubbles through a column of water on the downstream side of the membrane. There is a direct correlation between microbial retention and bubble point. Numerous studies have demonstrated that when the microbial reduction ratio is plotted against the bubble point a line of constant slope results.⁴¹ The bubble point is independent of the membrane surface area; however, diffusive flow through high surface area filters can mask the true bubble point. The bubble point test is easy to perform on small to medium-scale filters, the test time is short, and temperature effects are minor. Manual bubble point testing requires manipulation of downstream side of filter and is subjective.

Diffusive flow testing measures gas flow through a wetted filter at a constant pressure, usually about 80% of the bubble point. Diffusive flow is proportional to the differential pressure of the test gas, the diffusivity of the test gas in the wetting liquid, the thickness of the wetting liquid, the porosity, void volume, of the membrane, and the effective filtration area.

Unlike the bubble point, diffusive flow testing does not measure gas flow through only the largest set of wetted pores; therefore, diffusive flow does not correlate directly with microbial retention. However, bacterial challenge tests with a series of filters having decreasing diffusion rates show that a point exists where sterile filtrates occur. Diffusive flow testing is highly sensitive, especially for larger surface area membranes. Larger pores or flaws may be detected by thinning of the liquid layer. It is useful for membranes with small pore sizes (i.e., 0.1 μm and smaller), because of the high pressures required for bubble-point testing. Diffusive flow testing measures flow across the total pore volume, which may mask a flaw, especially in high surface-area multiple-cartridge arrays. It is also highly sensitive to temperature.

EQUIVALENCE IN STERILIZATION

The submission of an application for a pharmaceutical or biological product defines the methods by which that product is to be manufactured. There is a clear understanding that significant deviations from the approved process are to be avoided, and where substantial, submitted for regulatory review.⁴² In order to maintain compliance, firms must assure that sterilization processes are performed in a manner consistent with the originally filed procedure. Therapeutic equivalence is an expectation within ANDA regulation whereby changed products are compared to their prior state to determine whether the changes alter their effectiveness *in-vivo*. Within the ANDA regulatory framework, there is no detailed content on the impact of equipment changes on the product, nor is there any mention of sterilization processes or equipment.

Firms operating multiple pieces of identical equipment used for sterilization, or being sterilized as well as those operating even a single piece, must be able to support their (or its) operational equivalence. Where resources are not a constraint, firms would perform all of the initial qualification/validation work necessary for identical equipment during the initial effort. Further, they would repeat the entire exercise (not merely a portion of it) periodically to support that their sterilization processes are unchanged. This seems clearly excessive. The other extreme would be a firm that claimed that non-identical sterilization equipment can be considered equivalent and that validation in one piece supports the use of the other. A firm endeavored to utilize such an approach and received a warning letter from FDA that establishes the principles of equivalence in current use.⁴³ In essence the sterilizing equipment must be of the same size, control, fabrication, and operating principle. This expectation is tighter than is expected of non-sterilization processes in which substantially greater flexibility is afforded.⁴²

There is a simplifying aspect of sterilization for many processes in the F_0 or F_H determinations, commonly made during initial and ongoing validation studies. These terms are defined as equivalent sterilization time related to the base conditions for temperature and z-value of 10°C and as such is uniquely appropriate for use in equivalence studies for thermal sterilization processes.⁴⁴ Where the lethality values for thermal sterilization are comparable to those attended in identical equipment (or the same equipment at a different time), then the operational equivalence of the equipment has been established. In defining the acceptance criteria for equivalence determination, the measurement error and ordinary variation of operation must be considered, “given the material and procedural variation, uncertainties of sampling, and inherent variability in the analytical methods that even a non-varying piece of equipment will demonstrate variation in performance that are unavoidable because of limitations in the measurements applied to it.”⁴⁴

OTHER STERILIZATION METHODS

This chapter has reviewed the more commonly used sterilization methods, but there are other available processes that are microbially lethal. The decisions of which methods to include and which to omit was largely based upon their prevalence in the literature. Other lethal processes, some of which are derivatives of the method included, can be grouped into categories ranked by their market availability.

AVAILABLE NOW

- X-rays
- High Intensity Visible Light
- Microwave Sterilization
- Hydrogen Peroxide plasma
- HTST (High Temperature--Short Time)
- Semi-continuous H_2O_2 tunnels

EFFECTIVE WITH LIMITED OR NO COMMERCIAL SUPPORT

- Steam / Hydrogen Peroxide mixture
- Sulfur Dioxide
- High Pressures (>4,000-6,000 bar) (Batch or continuous process)
- Pulsed electrical field
- Cold plasma discharge
- Ultrasonification
- He/O₂ Plasma
- Hydrogen Bromide

POSSIBLY EFFECTIVE

- Supercritical fluid treatments (by itself or with a sanitizing agent)
- Pulsing magnetic fields

Some of these methods have close parallels in the processes described, and in those instances the described practices can serve as a guide. For those employing novel mechanisms, the path to implementation is more difficult. The general principles to be considered are constant regardless of the particular process, but certainly novel processes will need a substantial investment of time and expertise before they can be applied.

PARAMETRIC RELEASE OF STERILIZED MATERIALS

The sterility test was introduced in the 1930s as a means for confirming that sterile materials were safe for human use.⁴⁵ When the sterility test was first introduced, the process equipment was primitive by today's standards, lot sizes were much smaller and operations were heavily dependent on personnel. In this context, a laboratory test provided a measure of safety for what were potentially contaminated materials. Over the years this evolved to the present day sterility test as defined in the pharmacopeia and applied to both terminally sterilized and aseptically produced products.^{46,47} As technology changed, the capability of processes for sterile products improved substantially, and the sterility test became something of an anachronism, especially as it related to terminally sterilized products.⁴⁸ Parametric release can be defined as, "A sterility release procedure based upon effective control, monitoring, and documentation of a validated sterilization process cycle in lieu of release based upon end-product sterility testing."⁴⁹ Recognition that the sterility test did not afford adequate assurance of sterility led Baxter Healthcare to make the first application for parametric release to the FDA in 1991.⁵⁰ FDA was undoubtedly cautious, and it took nearly four years for approval of that submission. FDA issued a compliance policy guide, establishing its expectations for subsequent submissions.⁵¹ Later submissions by Hospira and B. Braun were approved with less delay.⁵² The EMEA (now EMA) published a position paper on parametric release in 2000 that addressed the subject at a very high level, followed by a definitive guidance document and Annex 17 of the EU CGMP's.^{53,54,55} In 2010 FDA issued submission guidance on parametric release summarizing its overall expectations.⁵⁶ Regrettably all of the regulatory documents on the subject lack clarifying detail that would enable practitioners to readily develop parametric release programs. Fortunately those firms that utilize parametric release for their terminally sterilized products have been generally open with respect to what they have provided. Essentially, the firms have individually established a list of critical factors derived from their

validation programs, which are used as the arbiters of sterility for each sterilizer load. The universal applicability of these lists is questionable, given the diversity of equipment designs. A detailed consensus document on the subject is lacking.

ASEPTIC PROCESSING

The production of sterile materials is divided into two major categories: terminally sterilized, using one or the other of the methods described earlier in this chapter, or aseptic processing, in which individually sterilized items are assembled in a pristine environment into the final product.¹⁰ Sterilization processes employ conditions that are lethal to microorganisms but also potentially detrimental to important properties of the product, container, or closure system. Aseptic processing allows the use of methods specifically suited for the individual component, minimizing the adverse impact on each (See Figure 25-23).

The difficulty lies in maintaining sterility of equipment and items from the point of sterilization through the sealing of the product container. The contamination risks associated with the aseptic assembly of sterile products mandate a strong regulatory (and weaker operational) preference that terminal sterilization be used wherever possible.^{57,58} The core of this expectation lies in the essentially universal belief that personnel are responsible for nearly all of the contamination isolated in an aseptic processing environment.⁵⁹ The central role of personnel in contamination, whether directly as the source or indirectly as the means for dissemination across the processing environment, has shaped the supportive element necessary for success: facilities, equipment, materials, procedures, and monitoring, as well as specific aspects associated directly with the presence of the operator (See Figure 25-24).

In the last 20 years, numerous technologies for aseptic processing have been introduced that endeavor to mitigate the contamination risk associated with personnel. These technological advances endeavor to alter the contamination vectors that result in loss of sterility. While the improvements may appear unrelated, their objective is singular – improvement in aseptic processing performance by reducing the impact of personnel. The core concepts behind this are common:⁶⁰

- Interventions are to be avoided at all times in aseptic processing.
- Interventions always mean increased risk to the patient.

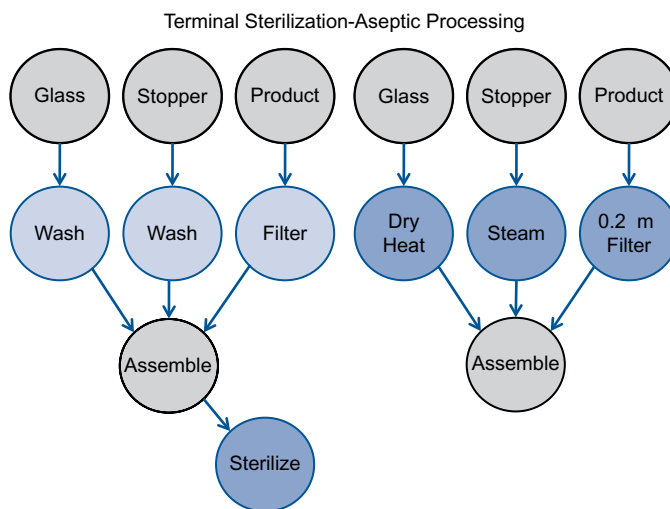


Figure 25-23. TS & AP Process Flow.

¹²The focus of the following discussion will be on the aseptic aspects; compounding and preparation areas must also be present, but aseptic processing is the focus of this discussion.

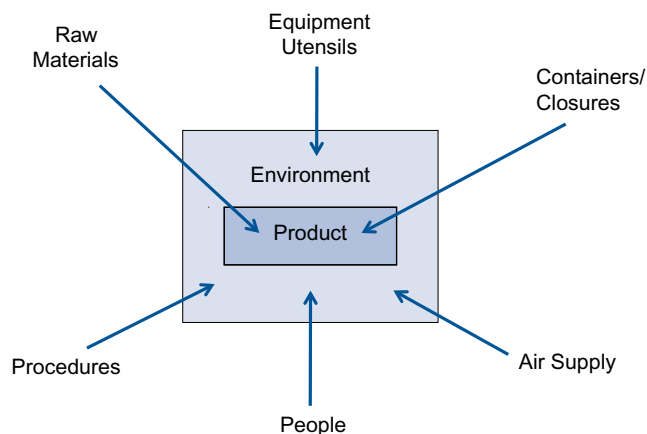


Figure 25-24. Contributions to Contamination in Aseptic Processing.

- There is no truly safe intervention.
- The “perfect” intervention is the one that doesn’t happen.

These ideas should remain central to the thinking of anyone involved with aseptic processing and are of substantially greater importance in older processing technologies that predate the newer approaches. Where technology is lacking, the operator is more intimately involved in the process, and the contamination risk will always be greater. The underlying goal for every aseptic process should be to minimize or eliminate personnel interaction with sterile items. The means to accomplish this is through careful design, selection, and operation of the aseptic process. An emerging term for these interrelated activities is “Sterility by Design”.⁶¹

FACILITY, ENVIRONMENT, AND AIR SYSTEMS

The facility provides the structure within which the aseptic process will be performed. At the building level it incorporates separate production, utility, laboratory, warehousing, office, and other areas. These should be arranged in a manner that provides for the flow of materials, personnel, and waste through the operating areas in a manner that minimizes the potential for cross-contamination.⁶¹ Some of the essential design considerations and operating practices for the core aseptic processing areas for sterile products are classified, beginning with ISO 8 in preparations areas and ranging to ISO 5 in the aseptic filling zone are the following:

- The classified environments should use materials of construction that are resistant to the intended agents, a minimum of horizontal surfaces and ledges, coved corners, flush windows, and other design details that facilitate cleaning and decontamination.
- The heating ventilating and air conditioning system should provide adequate air changes, air flows, temperature and relative humidity control, and pressurization to prevent the ingress of contamination.
- Air flow patterns within and between rooms and different classified zones should facilitate the removal and/or exclusion of contamination from critical environments.
- The classified areas should be supplied with HEPA filters (ceiling mounted in ISO 5 and 6) that are periodically tested for integrity.
- Differential pressures between different rooms and zones should be monitored and alarmed to support the integrity of the aseptic processing area.
- The entire aseptic area should be continuously maintained in a “clean” state and cleaned and decontaminated on a frequent basis. Isolators and RABS enclosures should

be treated with sporicidal agents on a periodic basis (these technologies lend themselves to longer intervals than ordinary manned environments).

- A minimum of equipment and materials should be retained in the aseptic portion of the facility to facilitate cleaning and decontamination.

Manned Aseptic Processing – Aseptic formulation and filling of aseptic formulations (and many terminally sterilized products) is performed in an ISO 5 environment, which is accessed from a manned ISO 6/7 background environment.⁸ A means of physical separation is provided between the ISO 5 and ISO 6/7 environments for product protection. The operator’s presence in the ISO 5 zone should be minimized. Partial barrier and air over-spill separates the ISO 5 environment from aseptically-garbed personnel. The non-product contact surfaces are manually cleaned and decontaminated.

Restricted Access Barrier System (RABS) Aseptic Filling – RABS are aseptic processing systems (ISO 5) intended to substantially reduce human borne contamination by the use of separative devices and defined mechanical features and operating procedures. Their capability varies with the operational practices employed. They are most effective when operated in a closed mode in which all activities are performed with the doors closed. Opening the door reduces the separation of personnel from sterilized items, and their performance may be only slightly better than a manned cleanroom (see above). Air over-spill protects the ISO 5 environment from the background environment (ISO 6-7), where aseptically-gowned personnel are present. The non-product contact surfaces are manually or automatically cleaned and decontaminated with the doors closed.

Isolator Based Aseptic Processing – Isolators are enclosed aseptic processing environments in which the filling process is performed. The ISO 5 isolator is never opened for access to the system by personnel during processing. All items enter and exit using systems and methods that prevent the ingress of contamination. The internal surfaces are decontaminated using an automated system. The background environment is ISO 8, and aseptically gowning of personnel is not necessary.

The majority of new, large-scale aseptic processing facilities are using either closed RABS or isolation technology since these have proven to be highly successful in preventing the ingress of contamination.

Equipment

Aseptic processing before the introduction of the HEPA filter relied heavily on operators. There would be some basic equipment, but personnel were responsible for material movement.⁶² The availability of HEPA filtration in the mid-1950s allowed for a large, protected area, in which equipment incorporating a modest amount of automation could be installed. Over time equipment improvements were gradually introduced, which increased reliability and sophistication, serving to reduce operator interaction with the equipment and components. The desirable features of filling equipment include the following:

- The product contact surfaces of the equipment can be readily sterilized by a validated method.
- Sterilization-in-place (see above) and clean-in-place should be utilized wherever possible.
- The product contact equipment should be assembled to the fullest extent possible prior to sterilization. Product contact equipment should be sterilized in sealed container and introduced in a manner that retains at least one layer of protection until entry into the ISO 5 zone.
- All equipment should be sterilized/depyrogenated using a just-in-time approach to reduce hold times and facilitate maintenance of the facility.

¹³The production of sterile suspensions, ointments, powders, and other complex formulations often requires the use of an ISO 5 environment.

- Filling equipment should be selected for high reliability, ease of change over, and adjustment. Remote adjustment of equipment parameters should be utilized.
- Tool-free changeover of the filler from one format to another.
- The filler and associated equipment should be largely tolerant of container-closure miss-feeds, jams, and other problems to minimize the need for operator interventions.
- The process equipment (and to some extent facility) should use PAT and other feedback systems for ease of control, operation, and improved documentation.
- Non-product contact portions of the equipment should be easily decontaminated and non-invasive of the critical zone. Equipment surfaces should be resistant to the potential corrosive action of sanitizing and decontamination agents.

More sophisticated equipment, including robotics, offers the means for dramatic reductions in the need for personnel intervention with sterile equipment, containers, closures, and other sterile items.

Containers and Closures

A core activity in aseptic processing is the assembly/filling of the product into its primary container. In the majority of instances this is a multiple component system in which the items are prepared and sterilized, using method specific for that type of item. In small-scale processing, much of the handling is performed by gowned personnel; however, commercial manufacturing generally entails equipment prepared as described in the previous section. The aseptic processing considerations relating to containers and closures include the following:

- The containers, closures and other primary components should be prepared and sterilized/depyrogenated using validated processes. The preparation processes should assure ease of handling, avoiding the need for interventions related to breakage, clumping, difficulties in feed, etc., that would require corrective interventions. Items should be sterilized/depyrogenated using a just-in-time approach. Inventories of materials in the aseptic environment should be minimized.
- Where possible, items should be introduced in a manner that retains at least one layer of sterilized, protective covering or wrap until entry into the critical zone.
- Containers, closures, and components should be selected for reliability of handling in the processing equipment.
- Containers /closures should be of appropriate quality. Higher AQLs for elimination of defects can result in a reduction in interventions.

Some of these same principles can be applied to utensils, environmental monitoring equipment, and other items to mitigate contamination concerns associated with them.

Excipients and Active Ingredients

Aseptic processing centers around the product being manipulated, and care must be taken to assure that the means for its introduction are properly established. The methods vary widely with the type of product being handled. Liquid solutions are the easiest to introduce, with sterilizing grade filters typically located close to the point of filling. Methods for introduction of sterile powders, whether for use in a suspension or as the final formulation, are more complex, usually entailing the removal of protective layers. Some of the concerns associated with their use are the following:

- Product materials must be prepared and sterilized using validated methods. The use of sterilization-in-place systems to the maximum extent possible is recommended.

- Product delivery in a closed container should be made directly into the ISO 5 critical zone. Any aseptic connections should be in that zone.
- Where product is supplied to the critical zone in sterile container (e.g., sterile powders), it should be introduced in a manner that retains at least one layer of sterilized, protective covering or wrap until entry into the critical zone.
- The package design for sterile materials must consider the aseptic processing risk of contamination at both the manufacturing site where bulk subdivision is performed and the fill-finish site where drug product containers are filled.⁶³

Operating Procedures

The operation and handling of the items described above must be defined in procedures, providing adequate detail to ensure that personnel can execute the task properly.^{64,65,66}

- The procedures central to the aseptic process should be critically reviewed to eliminate and/or simplify aseptic processes. The level of detail may need to be substantial as subtle variations can have a significant adverse impact. Interventions performed during aseptic processing must be recognized as increasing the risk of contamination dissemination.
- All interventions should be performed using sterilized utensils whenever possible.
- Intervention procedures should be established in detail for all inherent interventions and more broadly for corrective interventions (where some flexibility is necessary due to their greater diversity).
- Environmental monitoring is an inherent intervention, and means for its execution must be treated in the same manner as interventions involving product, containers, and closures.

Personnel

The personnel performing the aseptic process are the one element where the least amount of control is available. Their rate of movement, diligence in gowning, adherence to aseptic technique, and other factors can never be controlled precisely. They are, after all, humans and not machines or robots. If the other aspects of the aseptic process are well structured, the operator's role becomes less central to success. Ideally, the other elements are so well considered that the process will be successful even on their worst day. The reality is that operators play a central role in most aseptic processing operations, and their behavior has the greatest impact on contamination levels. These considerations are somewhat altered when isolators are used and aseptic gowning is eliminated; however, the requirements for aseptic technique apply in those systems without alteration.

- Personnel working in aseptic processing must receive initial and periodic formal training in CGMP, aseptic processing, microbiology, aseptic gowning, aseptic assembly/set-up (for those personnel that perform that tasks, aseptic technique, and job-specific tasks).
- Personnel entering aseptic environments should be initially and periodically evaluated and assessed for their proficiency in aseptic gowning. Less extensive monitoring on each exit from the aseptic core is commonplace as a means of assessing daily performance
- Personnel should be initially and periodically assessed for their proficiency in aseptic technique. Those individuals performing the set-up of the line must be highly capable. The proficiency checks of personnel can rely on off-line evaluations rather than actual media fill participation (see below). Gloves on enclosures should be monitored at the end of the batch or campaign.

- Personnel should conform to the highest standards of aseptic technique at all times. The presence of supervisory personnel with the environment is perhaps the best means to accomplish this.
- The gown materials should be cleaned and sterilized using validated methods. A fresh gown should be used for each entry into the aseptic area.
- Gloves on enclosures should be replaced periodically, decontaminated/sterilized and integrity-tested. Sleeves and half-suits tend to be more robust with replacement intervals somewhat longer than those defined for gloves.
- Manual filling should be recognized as a throwback to an earlier time and no longer used for aseptic processing. Isolators should be utilized for low-volume products requiring continuous, and thus excessive, operator intervention.⁶²

The focus of attention on the aseptic processing operator must never waiver. Their performance is all-important. Extraordinary measures with respect to any of the other influences on the process can be negated by inadequate personnel behavior. Isolators reduce some of this concern but do not eliminate it.

Monitoring, Testing and Capability Demonstration

The successful execution of a sterilization process is demonstrated on a daily basis by the review of records generated by the equipment that verifies that the process met the operational requirements.¹² In the context of aseptic processing, while there are some real-time measurements taken (i.e., pressure differentials, particle counts, temperature, etc.), what are considered the more explicit evidence of successful operation, microbial monitoring results, are not available until sometime after the process has been completed. Sterility testing is a compendial and regulatory requirement for aseptically manufactured sterile products, but the sample limitations are so severe that it should not be used as confirmation of aseptic process acceptability. Process simulations or media fills are means for addressing how well an aseptic process operates; however, there are numerous reasons that they are not considered definitive proof.

In-process microbiological samples are considered the primary evidence of aseptic processing acceptability. There are similar expectations established by both FDA and EMA with respect to the levels of microorganisms tolerated in aseptic environments.^{57,67} The sampling programs encompass air, surface, and personnel, with near zero limits in ISO 5 environments. While the goal of zero contamination is laudable, the reality is that that condition is not absolutely necessary for success in aseptic processing.⁶¹ No manned system, included those using closed RABS, can be truly be considered “sterile”; after all, the process is termed “aseptic”. Mandating rejection, extensive investigation, and corrective actions when a single CFU is detected suggests something that is not currently possible: that a manned aseptic environment can be devoid of microorganisms. A more rational approach has been incorporated into USP; however, its long term impact is unknown.⁶⁸ Nevertheless, monitoring must be performed, excursions from expected values looked into, and changes made to satisfy an unrealistic expectation and meet the regulatory levels. The design considerations outlined above can certainly assist in accomplishing that objective. The realities of microbial monitoring should be considered, regardless of the ultimate objective, and the following insights might prove useful in properly considering how monitoring, sterility testing, and process simulation results of aseptic processing operations can be considered:

- Despite any denigration of its utility, microbial monitoring of aseptic processing should be performed.
- The monitoring equipment and procedures, whether for viable or non-viables of any aseptic system, must not subject the product to increased risk of contamination. No monitoring is preferable to monitoring that risks contamination of sterile materials.
- All environmental monitoring activities, especially those performed in ISO 5 environments, must be recognized as interventional activity.
- Monitoring procedures are potential means for the introduction of adventitious contamination unrelated to the environment, material, or surface being sampled.
- Viable monitoring, even that of so-called “critical” areas and surfaces, is not an “in-process sterility test” and should never be used in that manner.
- The results of environmental monitoring, sterility testing, and process simulation should not be considered as “proof” of either sterility or non-sterility.
- Microbial monitoring can never recover all microorganisms present in an environment, nor on a surface.
- Significant excursions from the routine microbial monitoring results, with respect to number and type of microorganisms, should be investigated.
- Detection of low numbers of microorganisms in manned cleanrooms, including closed RABS, should be considered a rare but not unusual event.
- Sterility tests are not definitive proof of the acceptability of any batch; the sample size is too small for that purpose. It can only detect gross contamination.
- Process simulations are demonstrations of capability but cannot definitely establish the sterility of any material.
- Process simulations in excess of 5000–10 000 units are of relatively limited value; their greatest utility is in the evaluation of aseptic set-up practices.

Aseptic processing requires near absolute control bioburden in the processing environment. The goal is simple in concept; however, it is extremely difficult to attain. There are inherent risks related to aseptic processing since the sterile state must be maintained over an extended period of time with, in most cases, direct human intervention. Global regulatory agencies have properly expressed a clear preference for terminal sterilization wherever possible.^{57,58} There are numerous products that cannot be terminally sterilized, and thus aseptic processing will be necessary. The global industry has understood the inherent limitations associated with aseptic processing, and extensive improvements have been made to reduce the contamination potential.⁶⁹ The emerging technologies for aseptic processing suggest that future production methods will be substantially safer than those in current use.⁵⁹ The greater issue lies with the least capable aseptic processing facilities and operations; these are in far greater need of improvement than the more contemporary examples.

CONCLUSION

Sterilization and aseptic processing are essential practices for healthcare product manufacture and many healthcare services. The execution of these processes in an appropriate manner is essential for patient safety. Products that are sterile but non-stable, due to excessive conditions during sterilization, are unusable. Products that are stable but unsafe as a result of inadequate robust aseptic processing are equally inappropriate. Achieving suitably safe and stable products for administration as parenterals, ophthalmics, or by inhalation requires careful consideration of many factors, including the method of sterilization, its effects on the materials, and how those materials are handled post-sterilization.

¹⁴There are other less direct elements such as change control, calibration, preventive maintenance, training, etc., but *prima facie* are the sterilizer generated records.

REFERENCES

- Perkins J. *Principles and Methods of Sterilization in Health Sciences*, 2nd edn. Springfield, IL: Charles C. Thomas, 1969.
- USP <1211> *Sterilization & Sterility Assurance of Compensial Items*. USP 34, 2011.
- Food and Drug Administration. *Guidance for Industry for the Submission Documentation for Sterilization Process Validation in Applications for Human and Veterinary Drug Products*. FDA: 1994.
- Bigelow WD. The logarithmic nature of thermal death times, *J Infect Dis* 1921; 29: 528–536.
- Pflug IJ. *Microbiology and Engineering of Sterilization Processes*. 2nd edn. Philadelphia, PA, Parenteral Drug Association, Inc. 1978: 64–65.
- Gillis J, Mosely G. Calculating equivalent time for use in determining the lethality of EtO sterilization processes, *Med Device Diag Ind* 2002; 24(2): 54–63.
- Pflug IJ. *Microbiology & Engineering of Sterilization Processes*, 14th edn. Otterbein, IN: Environmental Sterilization Laboratory: 2010.
- ANSI/AAMI 11137 Sterilization of Health Care Products – Radiation Sterilization, 11137–2:2006.
- Agalloco J. Understanding overkill sterilization: putting an end to the confusion. *Pharmaceutical Technology* 2007; 30(5) (suppl.): 518–525.
- Bringert G. Calibration and Metrology. In Agalloco J, Carleton FJ, eds. *Validation of Pharmaceutical Processes*, 3rd edn. New York: Informa, US, 2007.
- Bringert G, Kemper C. Temperature Measurement. In Agalloco J, Carleton FJ, eds. *Validation of Pharmaceutical Processes*, 3rd edn. New York: Informa, US, 2007.
- FDA 21 CFR 212, Proposed June 1, 1976, withdrawn 1991.
- Chapman, K., A History of Validation in the United States - Part I, *Pharm Technol* 1991; 15(10): 82–96.
- CDC. *Guideline for Disinfection and Sterilization in Healthcare Facilities*, 2008.
- Agalloco J. Sterilization in Place Technology and Validation. In Agalloco J, Carleton FJ, eds. *Validation of Pharmaceutical Processes*, 3rd edn. New York: Informa, US, 2007.
- Hitchens AP. “Sterilization”. *J Am Pharm Associ* 2006; 3(6): 792.
- Pflug IJ. Sterilization of Space Hardware. *Environ Biol Med* 1971; 1: 63–81.
- Case L, Heffernan G. Dry Heat Sterilization and Depyrogenation Validation and Monitoring. In Agalloco J, Carleton FJ, eds. *Validation of Pharmaceutical Processes*, 3rd edn. New York: Informa, US, 2007.
- Tsuji K, Harrison S. Dry-heat destruction of lipopolysaccharide: dry-heat destruction kinetics. *Appl Environ Microbiol* 1978; 36: 710–714.
- Lawley PD, Brookes P. Further studies on the alkylation of nucleic acids and their constituent nucleotides. *Biochem J* 1963; 89(1): 127–138.
- Gillis J, Moseley G. Validation of Ethylene Oxide Sterilization Processes. In Agalloco J, Carleton FJ, eds. *Validation of Pharmaceutical Processes*, 3rd edn. New York, Informa, US, 2007.
- FDA. Ethylene oxide, ethylene chlorohydrin, and ethylene glycol – proposed maximum residue limits and maximum levels of exposure, *Federal Register* 43 FR: 27474–27483.
- Pflug IJ. *Syllabus for an Introductory Course in the Microbiology and Engineering of Sterilization Processes*, 4th edn. St. Paul, MN: Environmental Sterilization Services, 1980.
- www.chlordisys.com (accessed 4 June 2012).
- www.tso3.com (accessed 4 June 2012).
- www.noxyzilizer.com (accessed 4 June 2012).
- Kulla J *et al.* Sterilizing combination products using oxides of nitrogen, *Med Device Diagnostic Ind* 2009; 31(3): 46–54.
- www.thefreedictionary.com (accessed 4 June 2012).
- http://define.com (accessed 4 June 2012).
- http://chemistry.about.com (accessed 4 June 2012).
- Ph Eur. *Methods of preparation of sterile products*. 2009 (Suppl. 5.1.1).
- Ameri *et al.* Effect of irradiation on parathyroid hormone PTH(1–34) coated on a novel transdermal microparticle delivery system to produce a sterile product-adhesive compatibility. *J Pharm Sci* 2010; 99(4): 2123–2134.
- Hegge AB *et al.* Evaluation of novel alginate foams as drug delivery systems in antimicrobial photodynamic therapy (aPDT) of infected wounds—an in vitro study: studies on curcumin and curcuminoids XL. *J Pharm Sci* 2010; 99(8): 3499–3513.
- HIMA. *Microbiological Evaluation of Filters for Sterilizing Liquids*, HIMA Document No. 3, Vol. 4. Health Industry Manufacturers Association, Washington, DC, 1982.
- Wallhäuser KH. Foreword. In: Meltzer TH, ed. *Filtration in the Pharmaceutical Industry*. New York: Marcel Dekker, 1987: vii.
- Federal Register* 40: March 14, 1975: 11865.
- 21 CFR 211.72.
- Jornitz MW, Meltzer TH. *Sterile Filtration, A Practical Approach*. New York: Marcel Dekker, 2001: 75–76.
- ASTM. TASTM-F 838 *Standard Test Method for Determining Bacterial Retention of Membrane Filters Utilized for Liquid Filtration*. West Conshohocken, PA: ASTM International, 2005.
- PDA. Sterilizing filtration of liquids. *PDA J Pharm Sci Technol* 2008; (Suppl.)62: S-5.
- Johnston PR, Meltzer TH. Comments on organism challenge levels in sterilizing-filter efficiency testing. *Pharm Technol* 1979; 3(11):66–70, 110.
- FDA. Guidance for Industry: Changes to an Approved NDA or ANDA, September, 2004.
- FDA. Warning Letter to Sterile Recoveries, Inc., Nov, 4, 1994.
- Agalloco J. The importance of equivalence in the execution & maintenance of validation activities. *Pharm Technol* 2010; 34(12): 43–46.
- Sutton S. USP (1211): The compendial informational chapter on sterility assurance origins and future direction. *Pharm Technol* 2009; 33(Suppl): S16–S21.
- USP <71> Sterility Tests, USP 34, 2011.
- Ph Eur. Supplement 6.3, Sterility (reference 01/2009:20601), 2009.
- PDA. Parametric Release Task Force. Parametric Release of Sterile Pharmaceuticals, PDA Technical Report #30, *PDA J Pharm Sci Technol*, 1999: 53(4).
- Berger T. Developing an Approach for Parametric Release. In: Moldenhauer J, ed. *Steam Sterilization: A Practitioner's Guide*, Bethesda, MD: PDA, 2002.
- Sadowski M. Current Initiatives in Parametric Release. *PDA Sterilization Technology: Today and Tomorrow*, Chicago, March 2009.
- FDA. Parametric Release - Terminally Heat Sterilized Drug Products (CPG 7132a.13),
- Agalloco J. Personal communication, 2008.
- EMA. *Note for Guidance on Parametric Release*. CPMP/QWP/3015/99 Draft, March 2000.
- EMA. *Guidance on Real Time Release Testing*. EMA/CHMP/QWP/811210/2009 Rev 1, February 2010.
- EMA. Annex 17, Guide to EU CGMP, Parametric Release. ENTR/6270/00, July 2001.
- FDA. *Guidance for Industry Submission of Documentation in Applications for Parametric Release of Human and Veterinary Drug Products Terminally Sterilized by Moist Heat Processes*. FDA, February 2010.
- FDA. *Guidance on Sterile Drug Products produced by Aseptic Processing*. FDA, 2004.
- EMA. *Decision Trees for the Selection of Sterilization Methods*. (CPMP/QWP/155/96) 1999.
- Agalloco J, Akers J. The future of aseptic processing. *Pharm Technol* 2005; 29(3): S16–S23.
- Agalloco J, Akers, J. Simplified risk analysis for aseptic processing: the Akers-Agalloco method. *Pharm Technol* 2006; 30: 60–76.
- Agalloco J, Akers J. The Myth Called Sterility. *Pharm Technol* 2010; 34(3) (Suppl): 344–345. Continued online at Pharmtech.com.

62. Agalloco J. Risk management, CGMP, and the evolution of aseptic processing technology. *PDA J Pharm Sci Technol* 2009; 63(1): 8–10.
63. Agalloco J, DeSantis P. Validation of Active Pharmaceutical Ingredients. In: Nusim S, ed. *Bulk Pharmaceutical Chemical Production*, 2nd edn. New York: Informa, US, 2010.
64. Agalloco J. Management of aseptic interventions. *Pharm Technol* 2005; 29(3): 56–66.
65. Agalloco J, Akers J. The truth about interventions in aseptic processing. *Pharm Technol* 2007; 31(5): S8–S11.
66. Agalloco J, Akers J. Revisiting interventions in aseptic processing. *Pharm Technol* 2011; 35(4): 69–72.
67. EMA. Annex 1, Sterile Medicinal Products, 2009 revision.
68. USP. <1116> Microbiological Control and Monitoring of Aseptic Processing Environments, *Pharmacopeial Forum*, PF 36(6), Nov-Dec., 2010.
69. Agalloco J, Akers J. Innovations in Aseptic Processing Technology. In: Agalloco J, Akers J, eds. *Advanced Aseptic Processing Technology*, New York: Informa, US, 2010.

Chapter 26

Parenteral Preparations

Michael J. Akers, PhD

OVERVIEW OF UNIQUE CHARACTERISTICS OF PARENTERAL DOSAGE FORMS	495
FORMULATION PRINCIPLES	496
VEHICLES	496
SOLUTES	496
ADMINISTRATION	499
PARENTERAL COMBINATIONS	499
GENERAL CONSIDERATIONS	499
GENERAL MANUFACTURING PROCESS	500
COMPONENTS	501
WATER FOR INJECTION (WFI)	501
CONTAINERS AND CLOSURES	504
CONTAINER TYPES	504
RUBBER CLOSURES	506
NEEDLES	508
PYROGENS (ENDOTOXINS)	509
PRODUCTION FACILITIES	510
FUNCTIONAL AREAS	510
MAINTENANCE OF CLEAN ROOMS	514

PERSONNEL	514
ENVIRONMENTAL CONTROL EVALUATION	515
PRODUCTION PROCEDURES	517
CLEANING CONTAINERS AND EQUIPMENT	517
PRODUCT PREPARATION	519
FILTRATION	520
FILLING	521
SEALING	523
STERILIZATION	524
FREEZE-DRYING (LYOPHILIZATION)	525
QUALITY ASSURANCE AND CONTROL	528
STERILITY TEST	528
PYROGEN TEST	529
PARTICULATE MATTER EVALUATION	529
CONTAINER/CLOSURE INTEGRITY TEST	530
SAFETY TEST	530
PACKAGING AND LABELING	530

Parenteral (Gk, *para enteron*, beside the intestine) dosage forms differ from all other drug dosage forms, because they are injected directly into body tissue through the primary protective systems of the human body, the skin, and mucous membranes. They must be exceptionally pure and free from physical, chemical, and biological contaminants. These requirements place a heavy responsibility on the pharmaceutical industry to practice current good manufacturing practices (cGMPs) in the manufacture of parenteral dosage forms and on pharmacists and other health care professionals to practice good aseptic practices (GAPs) in dispensing parenteral dosage forms for administration to patients.

Certain pharmaceutical agents, particularly peptides, proteins, and many chemotherapeutic agents, can only be given parenterally, because they are inactivated in the gastrointestinal tract when given by mouth. Parenterally-administered drugs are relatively unstable and generally highly potent drugs that require strict control of administration to the patient. Due to the advent of biotechnology, parenteral products have grown in number and usage around the world.

This chapter focuses on the unique characteristics of parenteral dosage forms and the basic principles for formulating, packaging, manufacturing, and controlling the quality of these unique products. The references and bibliography at the end of this chapter contain the most up-to-date texts, book chapters, and review papers on parenteral product formulation, manufacture, and quality control.

OVERVIEW OF UNIQUE CHARACTERISTICS OF PARENTERAL DOSAGE FORMS

Parenteral products are unique from any other type of pharmaceutical dosage form for the following reasons:

- All products must be sterile.
- All products must be free from pyrogenic (endotoxin) contamination.
- Injectable solutions must be free from visible particulate matter. This includes reconstituted sterile powders.
- Products should be isotonic, although strictness of isotonicity depends on the route of administration. Products administered into the cerebrospinal fluid must be isotonic. Ophthalmic products, although not parenteral, must also be isotonic. Products to be administered by bolus injection by routes other than intravenous (IV) should be isotonic, or at least very close to isotonicity. IV infusions must be isotonic.
- All products must be stable, not only chemically and physically like all other dosage forms, but also 'stable' microbiologically (i.e., sterility, freedom from pyrogenic and visible particulate contamination must be maintained throughout the shelf life of the product).
- Products must be compatible, if applicable, with IV diluents, delivery systems, and other drug products co-administered.

FORMULATION PRINCIPLES

Parenteral drugs are formulated as solutions, suspensions, emulsions, liposomes, microspheres, nanosystems, and powders to be reconstituted as solutions. This section describes the components commonly used in parenteral formulations, focusing on solutions and freeze-dried products. General guidance is provided on appropriate selection of the finished sterile dosage form and initial approaches used to develop the optimal parenteral formulation.

VEHICLES

WATER

Since most liquid injections are quite dilute, the component present in the highest proportion is the vehicle. The vehicle of greatest importance for parenteral products is water. Water of suitable quality for compounding and rinsing product contact surfaces, to meet United States Pharmacopeia (USP) and other compendia specifications for Water for Injection (WFI), may be prepared either by distillation or by reverse osmosis. Only by these two methods is it possible to separate various liquid, gas, and solid contaminating substances from water. These two methods for preparation of WFI and specifications for WFI are also discussed in this chapter. With the possible exception of freeze-drying, there is no unit operation more important and none more costly to install and operate than that for the preparation of WFI.

WATER-MISCIBLE VEHICLES

A number of solvents that are miscible with water have been used as a portion of the vehicle in the formulation of parenterals. These solvents are used to solubilize certain drugs in an aqueous vehicle and to reduce hydrolysis. The most important solvents in this group are ethyl alcohol, liquid polyethylene glycol, and propylene glycol. Ethyl alcohol is used in the preparation of solutions of cardiac glycosides and the glycols in solutions of barbiturates, certain alkaloids, and certain antibiotics. Such preparations are given intramuscularly. There are limitations with the amount of these co-solvents that can be administered, due to toxicity concerns, greater potential for hemolysis, and potential for drug precipitation at the site of injection.¹ Formulation scientists needing to use one or more of these solvents must consult the literature (e.g., Mottu F *et al.* 2000)² and toxicologists to ascertain the maximum amount of co-solvents allowed for their particular product. Several references provide information on concentrations of co-solvents used in approved commercial parenteral products.^{3–8}

NON-AQUEOUS VEHICLES

The most important group of non-aqueous vehicles is the fixed oils. The USP provides specifications for such vehicles, indicating that the fixed oils must be of vegetable origin so they will metabolize, will be liquid at room temperature, and will not become rancid readily. The USP also specifies limits for the free fatty acid content, iodine value, and saponification value (oil heated with alkali to produce soap, i.e., alcohol plus acid salt). The oils most commonly used are corn oil, cottonseed oil, peanut oil, and sesame oil. Fixed oils are used as vehicles for certain hormone (e.g., progesterone, testosterone, deoxycorticosterone) and vitamin (e.g., Vitamin K, Vitamin E) preparations. The label must state the name of the vehicle, so the user may beware in case of known sensitivity or other reactions to it.

SOLUTES

Care must be taken in selecting active pharmaceutical ingredients and excipients to ensure their quality is suitable for parenteral administration. A low microbial level will enhance the effectiveness of either the aseptic or the terminal sterilization

process used for the drug product. Likewise, nonpyrogenic ingredients enhance the nonpyrogenicity of the finished injectable product. It is now a common GMP procedure to establish microbial and endotoxin limits on active pharmaceutical ingredients and most excipients. Chemical impurities should be virtually nonexistent in active pharmaceutical ingredients for parenterals, because impurities are not likely to be removed by the processing of the product. Depending on the chemical involved, even trace residues may be harmful to the patient or cause stability problems in the product. Therefore, manufacturers should use the best grade of chemicals obtainable and use its analytical profile to determine that each lot of chemical used in the formulation meets the required specifications.

Reputable chemical manufacturers accept the stringent quality requirements for parenteral products and, accordingly, apply good manufacturing practices to their chemical manufacturing. Examples of critical bulk manufacturing precautions include:

- Using dedicated equipment or properly validated cleaning to prevent cross-contamination and transfer of impurities;
- Using WFI for rinsing equipment;
- Using closed systems, wherever possible, for bulk manufacturing steps not followed by further purification; and
- Adhering to specified endotoxin and bioburden testing limits for the substance.

ADDED SUBSTANCES

The USP includes in this category all substances added to a preparation to improve or safeguard its quality. An added substance may:

- Increase and maintain drug solubility. Examples include complexing agents and surface active agents. The most commonly used complexing agents are the cyclodextrins, including Captisol. The most commonly used surface active agents are polyoxyethylene sorbitan monolaurate (Tween 20) and polyoxyethylene sorbitan monooleate (Tween 80).
- Provide patient comfort by reducing pain and tissue irritation, as do substances added to make a solution isotonic or near physiological pH. Common tonicity adjusters are sodium chloride, dextrose, and glycerin.
- Enhance the chemical stability of a solution, as do antioxidants, inert gases, chelating agents, and buffers.
- Enhance the chemical and physical stability of a freeze-dried product, as do cryoprotectants and lyoprotectants. Common protectants include sugars, such as sucrose and trehalose, and amino acids, such as glycine.
- Enhance the physical stability of proteins by minimizing self-aggregation or interfacial induced aggregation. Surface active agents serve nicely in this capacity.
- Minimize protein interaction with inert surfaces, such as glass and rubber and plastic. Competitive binders, such as albumin, and surface active agents are the best examples.
- Protect a preparation against the growth of microorganisms. The term 'preservative' is sometimes applied only to those substances that prevent the growth of microorganisms in a preparation. However, such limited use is inappropriate, being better used for all substances that act to retard or prevent the chemical, physical, or biological degradation of a preparation.
- Although not covered in this chapter, other reasons for adding solutes to parenteral formulations include sustaining and/or controlling drug release (polymers), maintaining the drug in a suspension dosage form (suspending agents, usually polymers and surface active agents), establishing emulsified dosage forms (emulsifying agents, usually amphiphilic polymers and surface active agents), and preparation of liposomes (hydrated phospholipids).

Although added substances may prevent a certain reaction from taking place, they may induce others. Not only may visible incompatibilities occur, but hydrolysis, complexation, oxidation, and other invisible reactions may decompose or otherwise inactivate the therapeutic agent or other added substances.⁹ Therefore, added substances must be selected with due consideration and investigation of their effect on the total formulation and the container-closure system.

ANTIMICROBIAL AGENTS

The USP states that antimicrobial agents in bacteriostatic or fungistatic concentrations must be added to preparations contained in multiple-dose containers. The European Pharmacopoeia requires multiple-dose products to be bacteriocidal and fungicidal. They must be present in adequate concentration at the time of use to prevent the multiplication of microorganisms inadvertently introduced into the preparation, while withdrawing a portion of the contents with a hypodermic needle and syringe. The USP provides a test for Antimicrobial Preservative Effectiveness to determine that an antimicrobial substance or combination adequately inhibits the growth of microorganisms in a parenteral product.¹⁰ Because antimicrobials may have inherent toxicity for the patient, the USP prescribes maximum volume and concentration limits for those commonly used in parenteral products (e.g., phenylmercuric nitrate and thimerosal 0.01%, benzethonium chloride and benzalkonium chloride 0.01%, phenol or cresol 0.5%, and chlorobutanol 0.5%).

The above limit is rarely used for phenylmercuric nitrate, most frequently employed in a concentration of 0.002%. Methyl *p*-hydroxybenzoate 0.18% and propyl *p*-hydroxybenzoate 0.02%, in combination, and benzyl alcohol 2% are also used frequently. Benzyl alcohol, phenol, and the parabens are the most widely used antimicrobial preservative agents used in injectable products. Although the mercurials are still allowed to be used in older products, they are not used for new products, due to concerns regarding mercury toxicity. In oleaginous preparations, no antibacterial agent commonly employed appears to be effective. However, it has been reported that hexylresorcinol 0.5% and phenylmercuric benzoate 0.1% are moderately bactericidal. A few therapeutic compounds have been shown to have antibacterial activity, thus, obviating the need for added agents.

Antimicrobial agents must be studied with respect to compatibility with all other components of the formula. In addition, their activity must be evaluated in the total formula. It is not uncommon to find a particular agent effective in one formulation but ineffective in another, possibly due to the effect of various components of the formula on the biological activity or availability of the compound; for example, the binding and inactivation of esters of *p*-hydroxybenzoic acid by macromolecules, such as Polysorbate 80, or the reduction of phenylmercuric nitrate by sulfide residues in rubber closures. A physical reaction encountered is that bacteriostatic agents are sometimes removed from solution by rubber closures.

Protein pharmaceuticals, because of their cost and/or frequency of use, are preferred to be available as multiple dose formulations (e.g., Human Insulin, Human Growth Hormone, Interferons, Vaccines, etc.). However, several proteins are reactive with antimicrobial preservative agents (e.g., Tissue Plasminogen Activator, Sargramostim, and Interleukins) and, therefore, are only available as single dosage units. Phenol and benzyl alcohol are the two most common antimicrobial preservatives used in peptide and protein products. Phenoxyethanol is the most frequently used preservative in vaccine products.

Single-dose containers and pharmacy bulk packs that do not contain antimicrobial agents are expected to be used promptly after opening or discarded. The ICH/CPMP guidelines (http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003476.pdf) require that products without preservatives be used immediately, although some package inserts define immediate use as within 3 hours after entering the primary package, or a longer usage period must be

justified. Large-volume, single-dose containers may not contain an added antimicrobial preservative. Therefore, special care must be exercised in storing such products after the containers have been opened to prepare an admixture, particularly those that support the growth of microorganisms, such as total parenteral nutrition (TPN) solutions and emulsions. It should be noted that, although refrigeration slows the growth of most microorganisms, it does not prevent their growth.

Buffers are used to stabilize a solution against chemical degradation or, especially for proteins, physical degradation (i.e., aggregation and precipitation) which might occur if the pH changes appreciably. Buffer systems should have as low a buffering capacity as feasible, so as not to significantly disturb the body's buffering systems when injected. In addition, the buffer type and concentration on the activity of the active ingredient must be evaluated carefully. Buffer components are known to catalyze degradation of drugs. The acid salts most frequently employed as buffers are citrates, acetates, and phosphates. Amino acid buffers, especially histidine, have become buffer systems of choice for controlling solution pH of monoclonal antibody solutions.

Anitoxidants are frequently required to preserve products, due to the ease with which many drugs, including proteins with methionine or cysteine amino acids conformationally exposed, are oxidized. Sodium bisulfite and other sulfurous acid salts are used most frequently. Ascorbic acid and its salts are also good antioxidants. The sodium salt of ethylenediaminetetraacetic acid (EDTA) has been found to enhance the activity of antioxidants, in some cases, by chelating metallic ions that would otherwise catalyze the oxidation reaction.

Displacing the air (oxygen) in and above the solution, by purging with an inert gas, such as nitrogen, can also be used as a means to control oxidation of a sensitive drug. Process control is required for assurance that every container is deaerated adequately and uniformly. However, conventional processes for removing oxygen from liquids and containers do not absolutely remove all oxygen. The only approach for completely removing oxygen is to employ isolator technology, where the entire atmosphere can be recirculating nitrogen or another non-oxygen gas.

Tonicity Agents are used in many parenteral and ophthalmic products to adjust the tonicity of the solution. Although it is the goal for every injectable product to be isotonic with physiologic fluids, this is not an essential requirement for small volume injectables administered intravenously. However, products administered by all other routes, especially into the eye or spinal fluid, must be isotonic. Injections into the subcutaneous tissue and muscles should also be isotonic to minimize pain and tissue irritation. The agents most commonly used are electrolytes and mono- or disaccharides.

Cryoprotectants and Lyoprotectants are additives that serve to protect biopharmaceuticals from adverse effects, due to freezing and/or drying of the product during freeze-dry processing. Sugars (non-reducing), such as sucrose or trehalose, amino acids, such as glycine or lysine, polymers, such as liquid polyethylene glycol or dextran, and polyols, such as mannitol or sorbitol, all are possible cryo- or lyoprotectants. Several theories exist to explain why these additives work to protect proteins against freezing and/or drying effects.^{11,12} Excipients that are preferentially excluded from the surface of the protein are the best cryoprotectants, and excipients that remain amorphous during and after freeze-drying serve best as lyoprotectants.

GENERAL GUIDANCE FOR DEVELOPING FORMULATIONS OF PARENTERAL DRUGS

The final formulation of a parenteral drug product depends on understanding the following factors that dictate the choice of formulation and dosage form.

1. Route of administration—Injections may be administered by such routes as intravenous, subcutaneous, intradermal, intramuscular, intraarticular, intralesional, and intrathecal. The type of dosage form (solution,

- suspension, etc.) determines the particular route of administration employed. Conversely, the desired route of administration places requirements on the formulation. For example, suspensions would not be administered directly into the bloodstream, due to the danger of insoluble particles blocking capillaries. Solutions administered subcutaneously require strict attention to tonicity adjustment, otherwise irritation of the plentiful supply of nerve endings in this anatomical area would give rise to pronounced pain. Injections intended for intraocular, intraspinal, intracisternal, and intrathecal administration require stricter standards of such properties as formulation tonicity, component purity, and limit of endotoxins, due to the sensitivity of tissues encountered to irritant and toxic substances.
- If the route of administration must be intravenous, then only solutions or microemulsions can be the dosage form. If the route of administration is subcutaneous or intramuscular, then the likely type of dosage form is a suspension or other microparticulate delivery system.
 - Pharmacokinetics of the drug—Rates of absorption (for routes of administration other than intravenous or intra-arterial), distribution, metabolism, and excretion for a drug have some effect on the selected route of administration and, accordingly, the type of formulation. For example, if the pharmacokinetic profile of a drug is very rapid, modified release dosage formulations may need developed. The dose of drug and the dosage regimen are affected by pharmacokinetics, so the size (i.e., concentration) of the dose will also influence the type of formulation and amounts of other ingredients in the formulation. If the dosage regimen requires frequent injections, then a multiple dose formulation must be developed, if feasible. If the drug is distributed quickly from the site injection, complexing agents or viscosity inducing agents may be added to the formulation to retard drug dissolution and transport.
 - Drug solubility—If the drug is insufficiently soluble in water at the required dosage, then the formulation must contain a co-solvent or a solute that sufficiently increases and maintains the drug in solution. If relatively simple formulation additives do not result in a solution, then a dispersed system dosage form must be developed. Solubility also dictates the concentration of drug in the dosage form.
 - Drug stability—If the drug has significant degradation problems in solution, then a freeze-dried or other sterile solid dosage form must be developed. Stability is sometimes affected by drug concentration that, in turn, might affect size and type of packaging system used. For example, if concentration must be low, due to stability and/or solubility limitations, then the size of primary container must be larger, and this might preclude the use of syringes, cartridges, and/or smaller vial sizes. Obviously, stability dictates the expiration date of the product that, in turn, determines the storage conditions. Storage conditions might dictate choice of container size, formulation components, and type of container. If a product must be refrigerated, then the container cannot be too large, and formulation components must be soluble and stable at colder conditions.
 - Compatibility of drug with potential formulation additives and packaging systems—It is well-known that drug-excipient incompatibilities frequently exist.⁹ Initial preformulation screening studies are essential to ensure that formulation additives, although possibly solving one problem, will not create another. Stabilizers, such as buffers and antioxidants, although chemically stabilizing the drug in one way, may also catalyze other chemical degradation reactions. Excipients and certain drugs can form insoluble complexes. Impurities in excipients can cause drug degradation reactions. Peroxide impurities in polymers

may catalyze oxidative degradation reactions with drugs, including proteins, which are oxygen sensitive.

- The use of silicone to lubricate vial rubber closures, syringe rubber plungers to coat the inner surface of glass syringes, and cartridges potentially can induce protein aggregation. Therefore, compatibility studies need to be designed to determine the potential for a new biopharmaceutical drug adversely affected by the presence of silicone applied to certain packaging surfaces. The increased popularity of laminated rubber closures and plungers has been due to the elimination of the need for applying silicone to these materials. Silicone coating is still required for glass syringes and cartridges, which provide new opportunities for the use of plastic syringes with biopharmaceuticals that further minimize the potential for incompatibilities between biopharmaceuticals and packaging systems.
- Desired type of packaging—Selection of packaging (i.e., type, size, shape, color of rubber closure, label, and aluminum cap) is often based on marketing preferences and competition. Knowing the type of final package early in the development process aids the formulation scientist in being sure the product formulation will be compatible and elegant in that packaging system.

Table 26-1 provides steps involved in the formulation of a new parenteral drug product. This can also be viewed as a list of questions, of which the answers will facilitate decisions on the final formulation that should be developed.

Table 26-1. Main Steps Involved in the Formulation of a New Parenteral Drug Product

1. Obtain physical properties of active drug substance
a. Structure, molecular weight
b. "Practical" solubility in water at room temperature
c. Effect of pH on solubility
d. Solubility in certain other solvents
e. Unusual solubility properties
f. Isoelectric point for a protein or peptide
g. Hygroscopicity
h. Potential for water or other solvent loss
i. Aggregation potential for protein or peptide
2. Obtain chemical properties of active drug substance
a. Must have a 'validatable' analytical method for potency and purity
b. Time for 10% degradation at room temperature in aqueous solution in the pH range of anticipated use
c. Time for 10% degradation at 5°C
d. pH stability profile
e. Sensitivity to oxygen
f. Sensitivity to light
g. Major routes of degradation and degradation products
3. Initial formulation approaches
a. Know timeline(s) for drug product
b. Know how drug product will be used in the clinic
i. Single dose vs multiple dose
ii. If multiple dose, will preservative agent be part of drug solution/powder or part of diluent?
iii. Shelf life goals
iv. Combination with other products, diluents
c. From knowledge of solubility and stability properties and information from anticipated clinical use, formulate drug with components and solution properties known to be successful at dealing with these issues, then perform accelerated stability studies.

i. High temperature storage
ii. Temperature cycling
iii. Light and/or oxygen exposure
iv. For powders, expose to high humidities
d. May need to perform several short-term stability studies as excipient types and combinations are eliminated
e. Selection of primary container and closure
i. Be aware of potential for tubing glass to be subject to glass delamination (glass lamellae); work with glass supplier to select type of glass
ii. Most rubber closure formulations are coated rubber to minimize leachables and do not require siliconization
f. Design and implement an initial manufacturing method of the product
g. Finalize formulation
i. Need for tonicity adjusting agent
ii. Need for antimicrobial preservative
h. Approach to obtain sterile product
i. Terminal sterilization
ii. Sterile filtration and aseptic processing

ADMINISTRATION

Injections may be classified in six general categories:

1. Solutions ready for injection
2. Dry, soluble products ready to be combined with a solvent just prior to use
3. Suspensions ready for injection
4. Dry, insoluble products ready to be combined with a vehicle just prior to use
5. Emulsions
6. Liquid concentrates ready for dilution prior to administration

When compared with other dosage forms, injections possess select advantages. If immediate physiological action is needed from a drug, it can usually be provided by the intravenous injection of an aqueous solution. Modification of the formulation or another route of injection can be used to slow the onset and prolong the action of the drug. The therapeutic response of a drug is controlled more readily by parenteral administration, since the irregularities of intestinal absorption are circumvented. Also, since the drug is administered by a professionally trained person, it confidently can be expected that the dose is accurately administered. Drugs can be administered parenterally, when they cannot be given orally, due to the unconscious or uncooperative state of the patient or due to inactivation or lack of absorption in the intestinal tract. Among the disadvantages of this dosage form are the requirement of asepsis at administration, the risk of tissue toxicity from local irritation, the real or psychological pain factor, and the difficulty in correcting an error, should one be made. In the latter situation, unless a direct pharmacological antagonist is immediately available, correction of an error may be impossible. One other disadvantage is that daily or frequent administration pose difficulties, as patients must either visit a professionally trained person or learn to self-inject. However, the advent of home health care as an alternative to extended institutional care and availability of new medications from biotechnology to treat chronic diseases have mandated the development of programs for training lay persons to administer these dosage forms.

PARENTERAL COMBINATIONS

Most dosage forms, when released to the marketplace by the manufacturer, are consumed by the patient without any significant manipulation of the product. For example, tablets and

capsules are ingested in the same form they were when released by the manufacturer. For many parenteral drug products, this is not the case. For example, products in vials must be withdrawn into a syringe prior to injection and often combined with other products in infusion solutions prior to administration. Freeze-dried products, first, have to be reconstituted with a specific or non-specific diluent prior to being withdrawn from the vial. Specifically, it is common practice for a physician to order the addition of a small-volume therapeutic injection (SVI), such as an antibiotic, to large-volume injections (LVIs), such as 1000 mL of 0.9% sodium chloride solution, to avoid the discomfort, for the patient, of a separate injection. Certain aqueous vehicles are recognized officially, due to their valid use in parenterals. Often, they are used as isotonic vehicles to which a drug may be added at the time of administration. The additional osmotic effect of the drug may not be enough to produce any discomfort when administered. These vehicles include Sodium Chloride Injection, Ringer's Injection, Dextrose Injection, Dextrose and Sodium Chloride Injection, and Lactated Ringer's Injection.

Although the pharmacist is the most qualified health professional to be responsible for preparing such combinations, as is clearly stated in the hospital accreditation manual of the Joint Commission on Accreditation of Healthcare Organizations,¹³ interactions among the combined products can be troublesome even for the pharmacist. In fact, incompatibilities can occur and cause inactivation of one or more ingredients or other undesired reactions. Patient deaths have been reported from the precipitate formed by two incompatible ingredients. In some instances, incompatibilities are visible as precipitation or color change, but, in other instances, there may be no visible effect.

The many potential combinations present a complex situation even for the pharmacist. To aid in making decisions concerning potential problems, a valuable compilation of relevant data has been assembled by Trissel¹⁴ and is updated regularly. Further, the advent of computerized data storage and retrieval systems has provided a means to organize and gain rapid access to such information. (Further information on this subject may be found in Chapter 27—Pharmaceutical Compounding – USP <797> Sterile Preparations.)

As studies have been undertaken and more information has been gained, it has been shown that knowledge of variable factors, such as pH and the ionic character of the active constituents, aids substantially in understanding and predicting potential incompatibilities. Kinetic studies of reaction rates may be used to describe or predict the extent of degradation. Ultimately, a thorough study should be undertaken of each therapeutic agent in combination with other drugs and IV fluids, not only of generic, but also of commercial preparations, from the physical, chemical, and therapeutic aspects.

Ideally, no parenteral combination should be administered, unless it has been studied thoroughly to determine its effect on the therapeutic value and safety of the combination. However, such an ideal situation may not exist. Nevertheless, it is the responsibility of the pharmacist to be as familiar as possible with the physical, chemical, and therapeutic aspects of parenteral combinations and to exercise the best possible judgment as to whether or not the specific combination extemporaneously prescribed is suitable for use in a patient.

GENERAL CONSIDERATIONS

An inherent requirement for parenteral preparations is that they be of the very best quality and provide the maximum safety for the patient. Constant adherence to high moral and professional ethics on the part of the responsible persons is most vital to achieving the desired quality in the products prepared.

TYPES OF PROCESSES

The preparation of parenteral products may be categorized as small-scale dispensing, usually one unit at a time, or large-scale manufacturing in which hundreds of thousands of units may

constitute one lot of product. The former category illustrates the type of processing done in early clinical phase manufacturing or in institutions, such as hospital pharmacies. The latter category is typical of the processing done in the later clinical phase and commercial manufacturing in the pharmaceutical industry. Wherever they are made, parenteral products must be subjected to the same basic practices of GMPs and good aseptic processing essential for the preparation of a safe and effective sterile product of highest quality, however, the methods used must be modified appropriately for the scale of operation.

The small-scale preparation and dispensing of parenteral products might use sterile components in their preparation. Therefore, the overall process focuses on maintaining, rather than achieving, sterility in the process steps. In the hospital setting, the final product might have a shelf life measured in hours. However, the extensive movement of patients out of the hospital to home care has modified hospital dispensing of parenteral products, wherein multiple units are made for a given patient and a shelf life of 30 days or more is required. Such products are sometimes made in hospital pharmacies, but increasingly in centers set up to provide this service. A discussion of such processing can be found in the USP general chapter <1206>.

This chapter emphasizes the preparation of parenteral products from non-sterile components in the highly technologically advanced plants of the pharmaceutical industry, using cGMP principles. In the pursuit of cGMP, consideration should be given to:

1. Ensuring that the personnel responsible for assigned duties are capable and qualified to perform them.
2. Ensuring that ingredients used in compounding the product have the required identity, quality, and purity.
3. Validating critical processes to be sure the equipment used and the processes followed ensure that the finished product has the qualities expected.
4. Maintaining a production environment suitable for performing the critical processes required, addressing such matters as orderliness, cleanliness, asepsis, and avoidance of cross contamination.
5. Confirming, through adequate quality-control procedures, that the finished products have the required potency, purity, and quality.
6. Establishing, through appropriate stability evaluation, that the drug products retain their intended potency, purity, and quality, until the established expiration date.
7. Ensuring that processes are always carried out in accord with established, written procedures.
8. Providing adequate conditions and procedures for the prevention of mix-ups.
9. Establishing adequate procedures, with supporting documentation, for investigating and correcting failures or problems in production or quality control.
10. Providing adequate separation of quality-control responsibilities from those of production to ensure independent decision making.

The pursuit of cGMP is an ongoing effort that must flex with new technological developments and new understanding of existing principles. Due to the extreme importance of quality in health care of the public, US Congress has given the responsibility of regulatory scrutiny over the manufacture and distribution of drug products to the FDA. Therefore, the operations of the pharmaceutical industry are subject to the oversight of the FDA and, with respect to manufacturing practices, to the application of the cGMPs. (These regulations are discussed more fully in Chapter 3—Quality Assurance and Control.)

In concert with the pursuit of cGMPs, the pharmaceutical industry has shown initiative and innovation in the extensive

technological development and improvement in quality, safety, and effectiveness of parenteral dosage forms in recent years. Examples include developments in:

- modular facility design and construction—smaller rooms, easier to clean, sanitize, and maintain;
- application of disposable technologies for compounding, mixing, and filling—reduce potential for cross-contamination;
- closure cleaning, siliconization (if applicable), and sterilization—all-in-one systems for rubber closures;
- sterilization technologies—well-defined sterilization validation principles, multiple approaches to sterilization cycles;
- filling technologies—greater speed, precision, and handling of viscous solutions;
- aseptic processing technology, including barrier isolator technology and restricted access barrier systems;
- freeze-drying technologies—automated loading and unloading, advances in process monitoring;
- control of particulate matter—greater diligence in cleaning methodologies, in-coming inspections, more experience with sources, causes, and minimization of particulate matter in facilities, on equipment and packaging, and personnel practices; and
- Automation—weight checking, inspection technologies, labeling and finishing operations.

GENERAL MANUFACTURING PROCESS

The preparation of a parenteral product may encompass four general areas:

1. Procurement and accumulation of all components in a warehouse area, until released to manufacturing;
2. Processing the dosage form in appropriately designed and operated facilities;
3. Packaging and labeling in a quarantine area, to ensure integrity and completion of the product; and
4. Controlling the quality of the product throughout the process.

Procurement encompasses selecting and testing according to specifications of the raw-material ingredients and the containers and closures for the primary and secondary packages. Microbiological purity, in the form of bioburden and endotoxin levels, has become standard requirements for raw materials.

Processing includes cleaning containers and equipment to validated specifications, compounding the solution (or other dosage form), filtering the solution, sanitizing or sterilizing the containers and equipment, filling measured quantities of product into the sterile containers, stoppering (either completely or partially for products to be freeze-dried), freeze-drying, terminal sterilization (if possible), and final sealing of the final primary container.

Packaging normally consists of the labeling and cartoning of filled and sealed primary containers. Control of quality begins with the incoming supplies, being sure that specifications are met. Careful control of labels is vitally important, as errors in labeling can be dangerous for the consumer. Each step of the process involves checks and tests to ensure the required specifications at the respective step are being met. Labeling and final packaging operations are becoming more automated.

The quality control unit is responsible for reviewing the batch history and performing the release testing required to clear the product for shipment to users. A common FDA citation for potential violation of cGMP is the lack of oversight by the quality control unit in batch testing and review and approval of results.

COMPONENTS

Components of parenteral products include the active ingredient, formulation additives, vehicle(s), and primary container and closure. Establishing specifications to ensure the quality of each of these components of an injection is essential. Secondary packaging is relevant more to marketing considerations, although some drug products might rely on secondary packaging for stability considerations, such as added protection from light exposure for light-sensitive drugs and antimicrobial preservatives.

The most stringent chemical-purity requirements will normally be encountered with aqueous solutions, particularly if the product is sterilized at an elevated temperature where reaction rates will be accelerated greatly. Dry preparations pose relatively few reaction problems but may require definitive physical specifications for ingredients that must have certain solution or dispersion characteristics when a vehicle is added.

Containers and closures are in prolonged, intimate contact with the product and may release substances into, or remove ingredients from, the product. Rubber closures are especially problematic (sorption, leachables, air and moisture transmission properties), if not properly evaluated for compatibility with the final product. Assessment and selection of containers and closures are essential for final product formulation, to ensure the product retains its purity, potency, and quality during the intimate contact with the container throughout its shelf life. Administration devices (e.g., syringes, tubing, transfer sets) that come in contact with the product should be assessed and selected with the same care as are containers and closures, even though the contact period is usually brief.

WATER FOR INJECTION (WFI)

PREPARATION

The source water can be expected to be contaminated with natural suspended mineral and organic substances, dissolved mineral salts, colloidal material, viable bacteria, bacterial endotoxins, industrial or agricultural chemicals, and other particulate matter. The degree of contamination varies with the source and will be markedly different, whether obtained from a well or from surface sources, such as a stream or lake. Hence, the source water must be pretreated by one or a combination of the following treatments: chemical softening, filtration, deionization, carbon adsorption, or reverse osmosis purification. Figure 26-1 shows a schematic of a typical process used to convert potable water to Water for Injection.

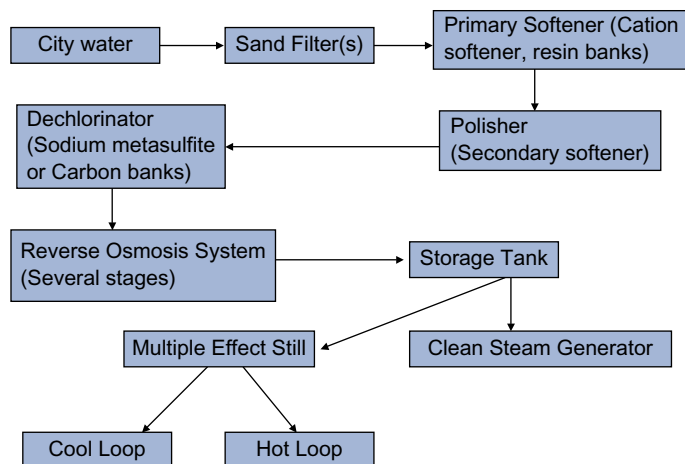


Figure 26-1. A schematic of a typical process used to convert potable water to Water for Injection.

Water for Injection can be prepared by distillation or by membrane technologies (i.e., reverse osmosis or ultrafiltration). The EP (*European Pharmacopeia*) only permits distillation as the process for producing WFI. The USP and JP (*Japanese Pharmacopeia*) allow all these technologies to be applied.

Distillation is a process of converting water from a liquid to its gaseous form (steam). Since steam is pure gaseous water, all other contaminants in the feedwater are removed. A conventional still consists of a boiler (evaporator), containing feed water (distilland); a source of heat to vaporize the water in the evaporator; a headspace above the level of distilland, with condensing surfaces for refluxing the vapor, thereby returning nonvolatile impurities to the distilland; a means for eliminating volatile impurities (demister/separation device) before the hot water vapor is condensed; and a condenser for removing the heat of vaporization, thereby converting the water vapor to a liquid distillate.

The specific construction features of a still and the process specifications have a marked effect on the quality of distillate obtained from a still. Several factors must be considered in selecting a still to produce WFI:

1. The quality of the feed water will affect the quality of the distillate. For example, chlorine in water, especially, can cause or exacerbate corrosion in distillation units, and silica causes scaling within. Controlling the quality of the feed water is essential for meeting the required specifications for the distillate.
2. The size of the evaporator will affect the efficiency. It should be large enough to provide a low vapor velocity, thus, reducing the entrainment of the distilland either as a film on vapor bubbles or as separate droplets.
3. The baffles (condensing surfaces) determine the effectiveness of refluxing. They should be designed for efficient removal of the entrainment at optimal vapor velocity, collecting and returning the heavier droplets contaminated with the distilland.
4. Redissolving volatile impurities in the distillate reduces its purity. Therefore, they should be separated efficiently from the hot water vapor and eliminated by aspirating them to the drain or venting them to the atmosphere.
5. Contamination of the vapor and distillate from the metal parts of the still can occur. Present standards for high-purity stills are that all parts contacted by the vapor or distillate should be constructed of metal coated with pure tin, 304 or 316 stainless-steel, or chemically resistant glass.

The design features of a still also influence its efficiency of operation, relative freedom from maintenance problems, or extent of automatic operation. Stills may be constructed of varying size, rated according to the volume of distillate that can be produced per hour of operation under optimum conditions. Only stills designed to produce high-purity water may be considered for use in the production of WFI. Conventional commercial stills designed for the production of high-purity water are available from several suppliers.

There are two basic types of WFI distillation units—the vapor compression still and the multiple effect still.

Compression Distillation

The vapor-compression still, primarily designed for the production of large volumes of high-purity distillate with low consumption of energy and water, is illustrated diagrammatically in Figure 26-2. To start, the feed water is heated from an external source in the evaporator to boiling. The vapor produced in the tubes is separated from the entrained distilland in the separator and conveyed to a compressor that compresses the vapor and raises its temperature to approximately 107°C. It then flows to the steam chest, where it condenses on the outer surfaces of the tubes containing the distilland; the vapor is, thus, condensed

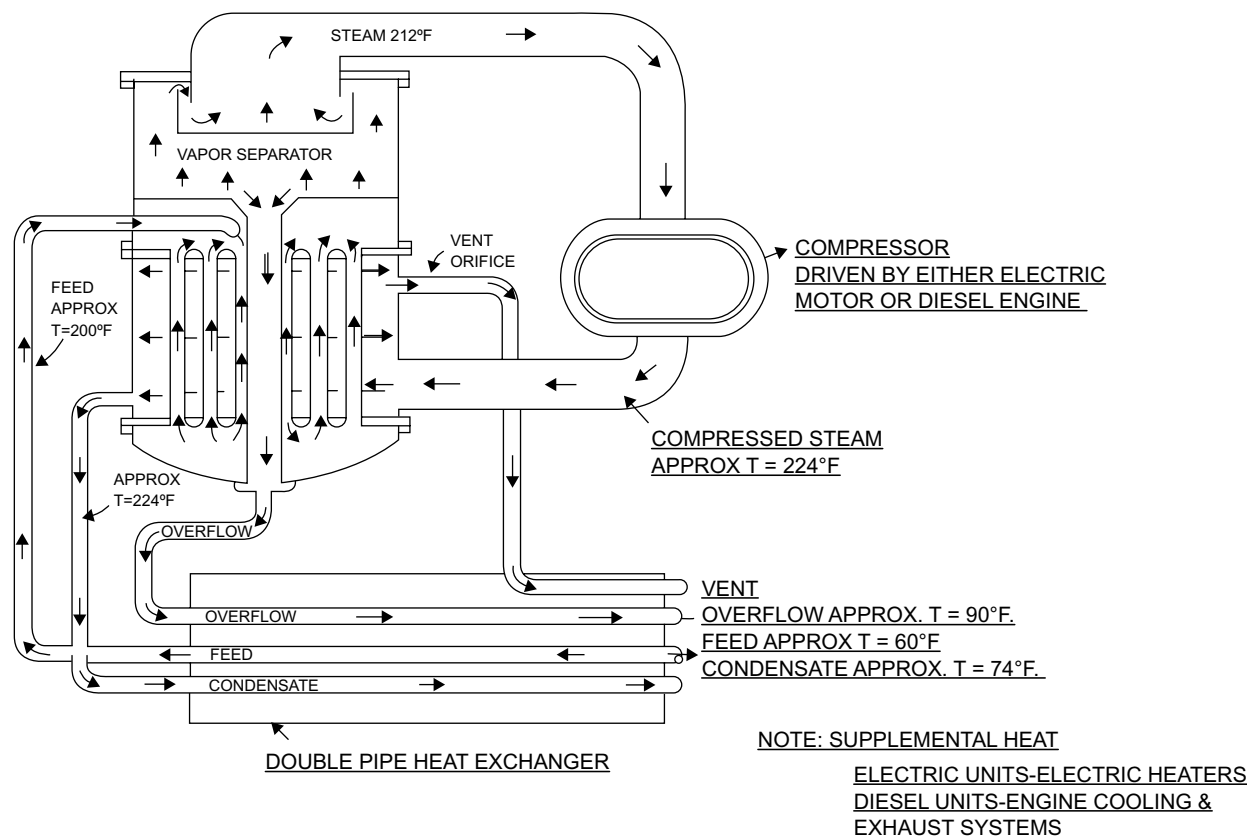


Figure 26-2. Vapor compressor still.

and drawn off as a distillate, while giving up its heat to bring the distilland in the tubes to the boiling point. Vapor-compression stills are available in capacities from 50 to 2800 gal/hr.

Multiple-Effect Stills

The multiple-effect still is also designed to conserve energy and water usage. In principle, it is simply a series of single-effect stills or columns running at differing pressures where phase changes of water take place. A series of up to seven effects may be used, with the first effect operated at the highest pressure and the last effect at atmospheric pressure. Figure 26-3 shows a schematic drawing of a multiple-effect still. Steam from an external source is used in the first effect to generate steam under pressure from feed water; it is used as the power source to drive the second effect. The steam used to drive the second effect condenses as it gives up its heat of vaporization and forms a distillate. This process continues until the last effect, when the steam is at atmospheric pressure and must be condensed in a heat exchanger.

The capacity of a multiple-effect still can be increased by adding effects. The quantity of the distillate will also be affected by the inlet steam pressure; thus, a 600-gal/hr unit designed to operate at 115 psig steam pressure could be run at approximately 55 psig and would deliver about 400 gal/hr. These stills have no moving parts and operate quietly. They are available in capacities from about 50 to 7000 gal/hr.

Reverse Osmosis (RO)

As the name suggests, the natural process of selective permeation of molecules through a semipermeable membrane separating two aqueous solutions of different concentrations is reversed. Pressure, usually between 200 and 400 psig, is applied to overcome osmotic pressure and force pure water to permeate through the membrane. Membranes, usually composed of

cellulose esters or polyamides, are selected to provide an efficient rejection of contaminant molecules in raw water. The molecules most difficult to remove are small inorganic molecules, such as sodium chloride. Passage through two membranes in series is sometimes used to increase the efficiency of removal of these small molecules and decrease the risk of structural failure of a membrane to remove other contaminants, such as bacteria and pyrogens.

Several WFI installations utilize both RO and distillation systems for generation of the highest quality water. Since feedwater to distillation units can be heavily contaminated and, thus, affect the operation of the still, water is first run through RO units to eliminate contaminants. (For additional information, see the book by Collentro.)¹⁵

Whichever system is used for the preparation of WFI, validation is required to be sure that the system, consistently and reliably, produces the chemical, physical, and microbiological quality of water required. Such validation should start with the determined characteristics of the source water and include the pretreatment, production, storage, and distribution systems. All of these systems together, including their proper operation and maintenance, determine the ultimate quality of the WFI.

Storage and Distribution

The rate of production of WFI is not sufficient to meet processing demands; therefore, it is collected in a holding tank for subsequent use. In large operations, the holding tanks may have a capacity of several thousand gallons and be a part of a continuously operating system. In such instances, the USP requires that the WFI be held at a temperature too high for microbial growth, normally a constant 80°C.

The USP also permits the WFI to be stored at room temperature but for a maximum of 24 hours. Under such conditions,

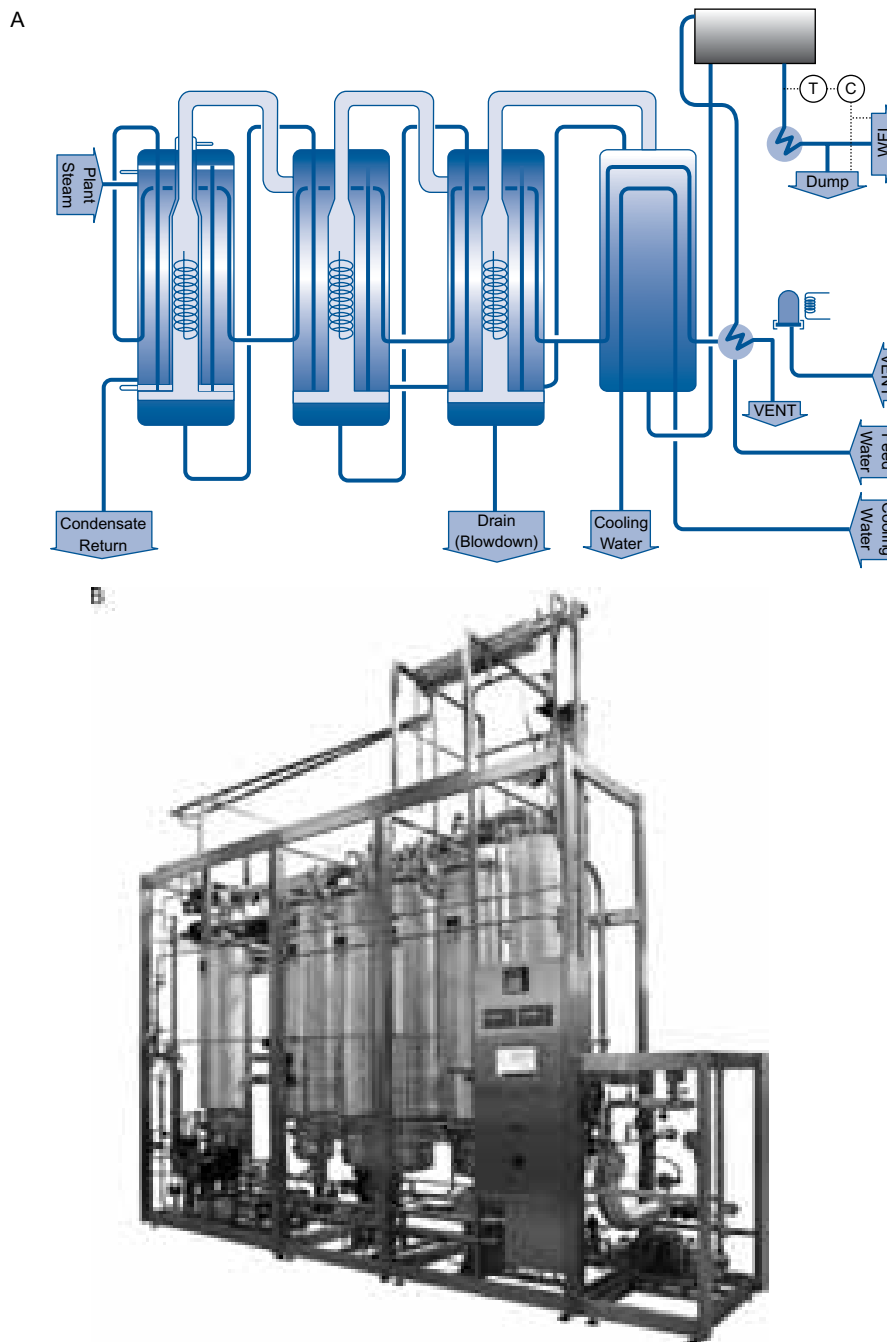


Figure 26-3. Multiple-effect water for injection distillation. Schematic (A) and (B) a space saving and energy efficient combination still and steam generator capable of delivering 750kg/hr (1650lb/hr) of pharmaceutical grade steam and 2200 liters/hr (580 gal/hr) of Water For Injection. (Courtesy of Getinge.)

the WFI is collected as a batch for a particular use with any unused water discarded within 24 hours. Such a system requires frequent sanitization to minimize the risk of viable microorganisms being present. The stainless-steel storage tanks in such systems are usually connected to a welded stainless-steel distribution loop, supplying the various use sites with a continuously circulating water supply. The tank is provided with a hydrophobic membrane vent filter capable of excluding bacteria and nonviable particulate matter. Such a vent filter is necessary to permit changes in pressure during filling and emptying. The construction material for the tank and

connecting lines is usually electropolished 316L stainless steel with welded pipe. The tanks also may be lined with glass or a coating of pure tin. Such systems are very carefully designed and constructed and often constitute the most costly installation within the plant.

When the water cannot be used at 80°C, heat exchangers must be installed to reduce the temperature at the point of use. Bacterial retentive filters should not be installed in such systems, due to the risk of bacterial buildup on the filters and the consequent release of pyrogenic substances.

Purity

Although certain purity requirements have been alluded to, the USP and EP monographs provide the official standards of purity for WFI and Sterile Water for Injection (SWFI).

The chemical and physical standards for WFI have changed in the past few years. The only physical/chemical tests remaining are the new total organic carbon (TOC), with a limit of 500 ppb (0.5 mg/L), and conductivity, with a limit of 1.3 $\mu\text{S}/\text{cm}$ at 25°C or 1.1 $\mu\text{S}/\text{cm}$ at 20°C. The former is an instrumental method capable of detecting all organic carbon present, and the latter is a three-tiered instrumental test measuring the conductivity contributed by ionized particles (in microSiemens or micromhos) relative to pH. Since conductivity is integrally related to pH, the pH requirement of 5 to 7 in previous revisions has been eliminated. The TOC and conductivity specifications are now considered adequate minimal predictors of the chemical/physical purity of WFI. However, the wet chemistry tests are still used when WFI is packaged for commercial distribution and for SWFI.

Biological requirements continue to be, for WFI, not more than 10 colony-forming units (CFUs)/100 mL and less than 0.25 USP endotoxin units/mL. The SWFI requirements differ in that, since it is a final product, it must pass the USP Sterility Test.

WFI and SWFI may not contain added substances. Bacteriostatic Water for Injection (BWFI) may contain one or more suitable antimicrobial agents in containers of 30 mL or less. This restriction is designed to prevent the administration of a large quantity of a bacteriostatic agent that would probably be toxic in the accumulated amount of a large volume of solution, even though the concentration was low.

The USP also provides monographs giving the specifications for Sterile Water for Inhalation and Sterile Water for Irrigation. The USP should be consulted for the minor differences between these specifications and those for SWFI.

CONTAINERS AND CLOSURES

Injectable formulations are packaged into containers made of glass or plastic. Container systems include ampoules, vials, syringes, cartridges, bottles, and bags (Fig. 26-4).

Ampoules are all glass, whereas bags are all plastic. The other containers can be composed of glass or plastic and must include



Figure 26-4. Various types of packaging for parenterals. (Courtesy of Gerresheimer.)

rubber materials, such as rubber stoppers for vials and bottles and rubber plungers and rubber seals for syringes and cartridges. Irrigation solutions are packaged in glass bottles with aluminum screw caps.

Table 26-2 provides a generalized comparison of the three compatibility properties—leaching, permeation, and adsorption—of container materials most likely involved in the formulation of aqueous parenterals. Further, the integrity of the container/closure system depends on several characteristics, including container opening finish, closure modulus, durometer and compression set, and aluminum seal application force. (Container-closure integrity testing is discussed in the Quality Assurance and Control section.)

CONTAINER TYPES

GLASS

Glass is employed as the container material of choice for most SVIs. It is composed, principally, of silicon dioxide, with varying amounts of other oxides, such as sodium, potassium, calcium, magnesium, aluminum, boron, and iron. The basic structural network of glass is formed by the silicon oxide tetrahedron. Boric oxide will enter into this structure, but most of the other oxides do not. The latter are only loosely bound, are present in the network interstices, and are relatively free to migrate. These migratory oxides may be leached into a solution in contact with the glass, particularly during the increased reactivity of thermal sterilization. The oxides dissolved may hydrolyze to raise the pH of the solution and catalyze or enter into reactions. Additionally, some glass compounds will be attacked by solutions and, in time, dislodge glass flakes into the solution. Such occurrences can be minimized by the proper selection of the glass composition.

Types

The USP provides a classification of glass:

- Type I, a borosilicate glass;
- Type II, a soda-lime treated glass;
- Type III, a soda-lime glass; and
- NP, a soda-lime glass not suitable for containers for parenterals.

Type I glass is composed, principally, of silicon dioxide (~81%) and boric oxide (~13%), with low levels of the non-network-forming oxides, such as sodium and aluminum oxides. It is a chemically resistant glass (low leachability), also having a low thermal coefficient of expansion (CoE) ($32.5 \times 10^{-7} \text{ cm}/\text{cm}^\circ\text{C}$ for 33 expansion glass; $51.0 \times 10^{-7} \text{ cm}/\text{cm}^\circ\text{C}$ for 51 expansion glass). In comparison, soda-lime glass has a thermal CoE of expansion of $8.36 \times 10^{-5} \text{ cm}/\text{cm}^\circ\text{C}$. The lower the thermal CoE, the more dimensionally stable the glass against thermal expansion stress that can result in cracking.

Types II and III glass compounds are composed of relatively high proportions of sodium oxide (~14%) and calcium oxide (~8%). This makes the glass chemically less resistant. Both types melt at a lower temperature, are easier to mold into various shapes, and have a higher thermal coefficient of expansion. Although there is no one standard formulation for glass among manufacturers of these USP type categories, Type II glass has a lower concentration of the migratory oxides than Type III. In addition, Type II has been treated under controlled temperature and humidity conditions, with sulfur dioxide or other dealkalizers to neutralize the interior surface of the container. Although it remains intact, this surface increases substantially the chemical resistance of the glass. However, repeated exposures to sterilization and alkaline detergents break down this dealkalized surface and expose the underlying soda-lime compound.

The glass types are determined from the results of two USP tests: the Powdered Glass Test and the Water Attack Test. A

Table 26-2. Comparative Compatibility Properties of Container Materials

	Leaching		Permeation		Adsorption (Selective) Extent ^a
	Extent ^a	Potential Leachables	Extent ^a	Potential Agents	
Glass					
Borosilicate	1	Alkaline earth and heavy metal oxides	0	N/A	2
Soda-lime	5	Alkaline earth and heavy metal oxides	0	N/A	2
Plastic polymers					
Low density	2	Plasticizers, antioxidants	5	Gases, water vapor, other molecules	2
High density	1	Antioxidants	3	Gases, water vapor, other molecules	2
PVC	4	HCl, especially plasticizers, antioxidants, other stabilizers	5	Gases, especially water vapor and other molecules	2
Polyolefins	2	Antioxidants	2	Gases, water vapor	2
Polypropylene	2	Antioxidants, lubricants	4	Gases, water vapor	1
Rubber polymers					
Natural and related synthetic	5	Heavy metal salts, lubricants, reducing agents	3	Gases, water vapor	3
Butyl	3	Heavy metal salts, lubricants, reducing agents	1	Gases, water vapor	2
Silicone	2	Minimal	5	Gases, water vapor	1

^a Approximate scale of 1 to 5, with 1 as the lowest.

proposal to revise USP <660> Containers-Glass had been published (*Pharmaceutical Forum*, 37(2), March-April, 2011) in which the USP would adopt the EP's methodology for the glass grains test to replace the USP powdered glass test and the water attack test would be deleted. The latter is used only for Type II glass and is performed on the whole container, due to the dealkalized surface; the former is performed on powdered glass, which exposes internal surfaces of the glass compound. The results are based on the amount of alkali titrated by 0.02 *N* sulfuric acid, after an autoclaving cycle with the glass sample in contact with a high-purity distilled water. Thus, the Powdered Glass Test challenges the leaching potential of the interior structure of the glass, whereas the Water Attack Test challenges only the intact surface of the container.

Selecting the appropriate glass composition is a critical facet of determining the overall specifications for each parenteral formulation. Glass can be the source/cause of leachables/extractables, particulates (glass delamination or glass lamellae formation), adsorption of formulation components, especially proteins, and cracks/scratches.

Leachables/Extractables—If the product is sensitive to the presence of ions, such as boron, sodium, potassium, calcium, iron, and magnesium, great care must be taken in selecting the appropriate glass container, as these ions may leach from the glass container and interact with the product, reducing chemical stability, inducing formation of particulate, or altering pH of solution. Drug formulations and thermal history can impact dissolution rates of glass and induce the leaching of significant quantities of glass metals. The presence of formulation components that can act as metal chelating agents, such as EDTA and citrate, is of special concern. Formulation studies involving accelerated aging (increased temperatures) should monitor changes in solution pH, metal content, and particulate formation, in addition to drug product stability (purity and potency). Leachables from the glass barrel, such as tungsten (used for the formation of the barrel) and silicone (used for lubrication of the syringe), were shown to have significant impact on the protein stability.¹⁶

The following rules apply with respect to glass leachables:

- Relatively low levels of leachables at pH 4-8.
- Relatively high levels of leachables at pH > 9.
- Major extractables are silicon and sodium.
- Minor extractables include potassium, barium, calcium, and aluminum.
- Trace extractables include iron, magnesium, and zinc.
- Treated glass gives less extractables, if pH < 8.

Delamination—Delamination, or glass particulate formation, is caused by chemical attack on the glass matrix by the formulation solution, resulting in weakening of the glass and eventual dislodgement of flakes from the glass surface.¹⁷ These fragments can be subvisible in size and, thus, difficult to detect. Delamination is of particular concern in tubing vials, and susceptibility may be driven by many of the same heat history factors that influence alkali leaching potential (as measured by compendial methods as a pH shift), although measures to reduce alkali leaching (ammonium sulfate treatment) may be completely ineffective in reducing delamination susceptibility.

Certain components or characteristics of the pharmaceutical product formulation may enhance the potential for delamination of the product container. Formulations with elevated pH (particularly >8), high sodium chloride content, or containing specific buffer components, known to attack or solubilize components, of the glass matrix (phosphate, citrate, tartrate, EDTA) are of potential concern. In some cases, delamination may be reduced or eliminated by careful control of process parameters during the glass forming processes; lower heat levels during the conversion process may be critical. Ammonium sulfate-treated glass containers are also known to be more susceptible to delamination.

Although compendial chemical resistance testing may be predictive of delamination potential, in some cases, experience indicates that additional testing with the actual formulation or a surrogate is required to properly evaluate container selection. For example, a product that undergoes terminal sterilization

could be evaluated by multiple sterilization cycles in the formulation matrix, followed by filtration and microscopic examination of the filters for subvisible particles of delaminated glass. For sterile-fill applications, it is recommended that the glass containers be filled with formulation placebo (all components except the unstable API) at the pH release limit(s) for the product and be challenged with a single autoclave cycle or an accelerated aging study at 55°C for at least 4 weeks, followed by filtration and microscopic examination for glass particles.

Adsorption—Adsorption of drug to solution contact surfaces and consequent loss of potency of delivered solution is a primary concern of container/solution compatibility and must be rigorously and formally evaluated during solution/container evaluation and stability studies. Glass containers are fairly inert surfaces, for most small drug products at relatively high concentrations, but pose a higher risk for therapeutic proteins and other smaller drug products formulated at low concentrations. Since adsorption is a surface phenomenon, increasing the surface area to volume ratio increases the risk of losses due to adsorption. Thus, small volume products carry higher risk for loss of potency due to adsorption and should be carefully evaluated for drug loss.¹⁸

Cracks and Scratches—Small cracks and scratches on glass containers can best be minimized by implementation of quality agreements between parenteral product manufacturers and glass container manufacturers. Not only does the glass container manufacturer need strict control procedures to minimize cracks and scratches from the time the container is formed until it reaches the finished product manufacturer, but there also needs to be high quality, 100% inspection practices by both glass and final product manufacturers. Also, local quality inspection procedures and practices need to have clearly understood definitions and a library of examples for what is defined as a crack and scratch. Cracks are considered unacceptable, whereas scratches are more of an esthetic indication of product elegance.

Type I glass will be suitable for all products, although sulfur dioxide treatment is sometimes used for even greater resistance to glass leachables. Because cost must be considered, one of the other, less-expensive types may be acceptable. Type II glass may be suitable, for example, for a solution that is buffered, has a pH below 7, or is not reactive with the glass. Type III glass is usually suitable for anhydrous liquids or dry substances. However, some manufacturer-to-manufacturer variation in glass composition should be anticipated within each glass type. Therefore, for highly chemically sensitive parenteral formulations, it may be necessary to specify both USP Type and specific manufacturer.

Schott developed a technology, called Plasma Impulse Chemical Vapor Deposition (PECVD), that coats the inner surface of Type I glass vials with an ultrathin film of silicon dioxide.¹⁹ This film forms a highly efficient diffusion barrier that practically eliminates glass leachables. Such treated glass is especially useful for drug products having high pH values, formulations with complexing agents, or products showing high sensitivity to pH shifts.

Physical Characteristics

Commercially available containers vary in size from 0.5 to 1000 mL. Sizes up to 100 mL may be obtained as ampoules and vials, and larger sizes as bottles. The latter are used for intravenous and irrigating solutions. Smaller sizes are also available as syringes and cartridges. Ampoules, syringes, and cartridges are drawn from glass tubing. The smaller vials may be made by molding or from tubing. Larger vials and bottles are made only by molding. Containers made by drawing tubing are optically clearer and have a thinner wall than molded containers (Fig. 26-4). Compared to molded glass, tubing glass also has better wall and finish dimensional consistency and no seams, is easier to label, weighs less, facilitates inspection, and has lower tooling costs. Tubing glass is preferable to molded glass for freeze-dried products, due to more efficient heat transfer from the shelf into the product. Molded containers are uniform in external dimensions, stronger,

and heavier. Also, molded glass is not as susceptible to leachables and delamination, because the glass formation temperatures to vaporize and condense the alkali components of the glass are not as high as for tubing container manufacture.²⁰ Easy-opening ampoules that permit the user to break off the tip at the neck constriction, without the use of a file, are weakened at the neck, by scoring or applying a ceramic paint with a different coefficient of thermal expansion. An example of a modification of container design to meet a particular need is the double-chambered vial, designed to contain a freeze-dried product in the lower, and solvent in the upper chamber. Other examples are wide-mouth ampoules with flat or rounded bottoms to facilitate filling with dry materials or suspensions and various modifications of the cartridge for use with disposable dosage units.

Glass containers must be strong enough to withstand the physical shocks of handling and shipping and the pressure differentials that develop, particularly during the autoclave sterilization cycle. They must be able to withstand the thermal shock resulting from large temperature changes during processing, for example, when the hot bottle and contents are exposed to room air at the end of the sterilization cycle. Therefore, a glass with a low coefficient of thermal expansion is necessary. The container must also be transparent to permit inspection of the contents.

Preparations that are light-sensitive must be protected, by placing them in amber glass containers or by enclosing flint glass containers in opaque cartons labeled to remain on the container during the period of use. It should be noted that the amber color of the glass is imparted by the incorporation of potentially leachable heavy metals, mostly iron and manganese, which may act as catalysts for oxidative degradation reactions. Silicone coatings are sometimes applied to containers to produce a hydrophobic surface, for example, as a means of reducing the friction of a rubber-tip of a syringe plunger.

The size of single-dose containers is limited to 1000 mL by the USP, and multiple-dose containers to 30 mL, unless stated otherwise in a particular monograph. Multiple-dose vials are limited in size to reduce the number of punctures for withdrawing doses and the accompanying risk of contamination of the contents. As the name implies, single-dose containers are opened or penetrated with aseptic care, and the contents used at one time. These may range in size from 1000-mL bottles to 1-mL or less ampoules, vials, or syringes. The integrity of the container is destroyed when opened, so that the container cannot be closed and reused.

A multiple-dose container is designed so that more than one dose can be withdrawn at different times, the container maintaining a seal between uses. It should be evident that, with full aseptic precautions, including sterile syringe and needle for withdrawing the dose and disinfection of the exposed surface of the closure, there is still a substantial risk of introducing contaminating micro-organisms and viruses into the contents of the vial. Due to this risk, the USP requires that all multiple-dose vials contain an antimicrobial agent or be inherently antimicrobial, as determined by the USP Antimicrobial Preservatives-Effectiveness tests. There are no comparable antiviral effectiveness tests, nor are antiviral agents available for such use. In spite of the advantageous flexibility of dosage provided by multiple-dose vials, single-dose, disposable container units provide the clear advantage of greater sterility assurance and patient safety.

Due to concerns for user safety and glass particulate matter occurring when glass is broken, glass sealed ampoules are no longer glass containers of choice for new SVIs in the United States.

RUBBER CLOSURES

To permit introduction of a needle from a hypodermic syringe into a multiple-dose vial and provide for resealing as soon as the needle is withdrawn, each vial is sealed with a rubber



Figure 26-5. Examples of rubber closures for vials and syringes. (Photo Copyright 2011, West Pharmaceutical Services, Inc., Used with Permission).

closure held in place by an aluminum cap (Fig. 26-5). This principle is also followed for single-dose containers of the cartridge type, except that there is only a single introduction of the needle to make possible the withdrawal or expulsion of the contents.

Rubber closures are composed of multiple ingredients plasticized and mixed together at an elevated temperature on milling machines. The elastomer primarily used in rubber closures, plungers, and other rubber items used in parenteral packaging and delivery systems is synthetic butyl or halobutyl rubber. Natural rubber is also used, but, if it is natural rubber latex, then the product label must include a warning statement, due to the potential for allergic reactions from latex exposure.

The plasticized mixture is placed in molds and vulcanized (cured) under high temperature and pressure. During vulcanization the polymer strands are cross-linked by the vulcanizing agent, assisted by the accelerator and activator, so that motion is restricted and the molded closure acquires the elastic, resilient character required for its use. Ingredients not involved in the cross-linking reactions remain dispersed within the

Table 26-3. Examples of Ingredients Found in Rubber Closures

Ingredient	Examples
Elastomer	Natural rubber (latex)
	Butyl rubber
	Neoprene
Vulcanizing (curing agent)	Sulfur
	Peroxides
Accelerator	Zinc dibutyldithiocarbamate
Activator	Zinc oxide
	Stearic acid
Antioxidant	Dilauryl thiodipropionate
Plasticizer/lubricant	Paraffinic oil
	Silicone oil
	Carbon black
Fillers	Clay
	Barium sulfate
	Inorganic oxides
Pigments	Carbon black

compound and, along with the degree of curing, affect the properties of the finished closure. Table 26-3 provides examples of rubber-closure ingredients.

The physical properties considered in the selection of a particular formulation include elasticity, hardness, tendency to fragment, and permeability to vapor transfer. The elasticity is critical in establishing a seal with the lip and neck of a vial or other opening and in resealing after withdrawal of a hypodermic needle from a vial closure. The hardness should provide firmness, but not excessive resistance to the insertion of a needle through the closure, and minimal fragmentation of pieces of rubber should occur as the hollow shaft of the needle is pushed through the closure. Although vapor transfer occurs to some degree with all rubber formulations, appropriate selection of ingredients makes it possible to control the degree of permeability. Physicochemical and toxicological tests for evaluating rubber closures are described in section <381> in the USP.

The ingredients dispersed throughout the rubber compound may be subject to leaching into the product contacting the closure. These ingredients (Table 26-3) pose potential compatibility interactions with product ingredients, if leached into the product solution, and these effects must be evaluated. Further, some ingredients must be evaluated for potential toxicity.

The example of pure red cell aplasia, an immunogenic reaction caused by leachables from a rubber closure in an erythropoietin prefilled syringe formulation, highlights the criticality of appropriate container-closure and the study of such leachables and extractables, even as a function of stability shelf life.²¹

To reduce the problem of leachables, laminates have been applied to the product contact surfaces of closures, with various polymers, the most successful being Teflon® (DuPont polytetrafluoroethylene [PTFE]) and Fluorotec® (West/Daikyo copolymer of tetrafluoroethylene and ethylene). Polymeric coatings have been developed that are claimed to have more integral binding with the rubber matrix, however, details of their function are trade secrets. Although rubber coatings do reduce the potential for extractables/leachables and eliminate the need for applied silicone treatment, they may have potential disadvantages of not flowing as easily during high speed filling operations and may not have the same container-closure integrity as uncoated stoppers with vial openings.

The physical shape of some typical closures may be seen in Figure 26-5. Most of them have a lip and a protruding flange that extends into the neck of the vial or bottle. Many disk closures

are being used now, particularly in the high-speed packaging of antibiotics. Slotted closures are used on freeze-dried products to permit the escape of water vapor, since they are inserted only partway into the neck of the vial until completion of the drying phase of the cycle. Also, the top design of the freeze-dry closure is important to minimize sticking of the closure to underneath the dryer shelf after stoppering the vial. Stoppers normally have a small protruding circle at the center of the top of the stopper. Gaps provided within the protruding circle minimize the tendency of the stopper to stick to the freeze-dryer shelf.

The plunger type of rubber is used to seal one end of a syringe or cartridge. At the time of use, the plunger expels the product by a needle inserted through the closure at the distal end of the package. Intravenous solution closures often have permanent holes for adapters of administration sets; irrigating solution closures are usually designed for pouring.

Rubber closures must be 'slippery' to move easily through a rubber closure hopper and other stainless steel passages, until they are fitted onto the filled vials. Traditionally, rubber materials are 'siliconized' (silicone oil or emulsion applied onto the rubber) to produce such lubrication. However, advances in rubber closure technologies have introduced closures that do not require siliconization, due to a special polymer coating applied to the outer surface of the closure. Examples are the Daikyo/West closures (Flurotec) and the Helvoet (Omniflex) closures.

PLASTIC

Thermoplastic polymers have been established as packaging materials for sterile preparations, such as large-volume parenterals, ophthalmic solutions, and, increasingly, small-volume parenterals. For such use to be acceptable, a thorough understanding of the characteristics, potential problems, and advantages for use must be developed. Three principal problem areas exist in using these materials:

1. Permeation of vapors and other molecules in either direction through the wall of the plastic container;
2. Leaching of constituents from the plastic into the product; and
3. Sorption (absorption and/or adsorption) of drug molecules or ions on the plastic material.

Permeation, the most extensive problem, may be troublesome by permitting volatile constituents, water, or specific drug molecules to migrate through the wall of the container to the outside and, thereby, be lost. This problem has been resolved, for example, by the use of an overwrap in the packaging of IV solutions in PVC bags to prevent loss of water during storage. Reverse permeation in which oxygen or other molecules may penetrate to the inside of the container and cause oxidative or other degradation of susceptible constituents may also occur. Leaching may be a problem, when certain constituents in the plastic formulation, such as plasticizers or antioxidants, migrate into the product. Thus, plastic polymer formulations should have as few additives as possible, an objective characteristically achievable for most plastics used for parenteral packaging. Sorption is a problem on a selective basis, that is, sorption of a few drug molecules occurs on specific polymers. For example, sorption of insulin and other proteins, vitamin A acetate, and warfarin sodium has been shown to occur on PVC bags and tubing, when these drugs were present as additives in IV admixtures. Table 26-2 gives a brief summary of some of these compatibility relationships.

One of the principle advantages of using plastic packaging materials is that they are not breakable, as is glass; also, there is a substantial weight reduction. The flexible bags of polyvinyl chloride or select polyolefins, currently in use for large-volume intravenous fluids, have the added advantage that no air interchange is required; the flexible wall simply collapses as the solution flows out of the bag.

Most plastic materials have the disadvantage of not being as clear as glass, and, therefore, inspection of the contents is

impeded. However, recent technologies have overcome this limitation, evidenced by plastic resins, such as CZ (polycyclopentane, Daikyo Seiko) and Topas COC (cyclic olefin copolymer, Ticona). In addition, many of these materials soften or melt under the conditions of thermal sterilization. However, careful selection of the plastic used and control of the autoclave cycle has made thermal sterilization of some products possible, large-volume injectables, in particular. Ethylene oxide or radiation sterilization may be employed for the empty container with subsequent aseptic filling. However, careful evaluation of the residues from ethylene oxide or its degradation products and their potential toxic effect must be undertaken. Investigation is required concerning potential interactions and other problems that may be encountered when a parenteral product is packaged in plastic.

Future trends in primary packaging for parenterals will continue to see significant growth in the application of plastic vials and syringes and the manufacturing of such packaging by form- (or blow-) fill-seal technologies. (For further details, see Chapter 35 (Pharmaceutical Packaging) and the review book chapter by Vilivalam and DeGrazio.²²

NEEDLES

Historically, stainless steel needles have been used to penetrate the skin and introduce a parenteral product inside the body. The advent of needleless injection systems has obviated the need for needles for some injections (e.g., vaccines) and is gaining in popularity over the conventional syringe and needle system. However, needleless injections are more expensive, can still produce pain on injection, are, potentially, a greater source of contamination (and cross-contamination from incessant use), and may not be as efficient in dose delivery.

Needles are hollow devices composed of stainless steel or plastic. Needles are available in a wide variety of lengths, sizes, and shapes. Needle lengths range from 1/4 inch to 6 inches. Needle size is referred to as its gauge (G), or the outside diameter (OD) of the needle shaft. Gauge ranges are 11 to 32 G, with the largest gauge for injection usually being no greater than 16 G. 16 G needles have an OD of 0.065 inches (1.65 mm), whereas 32 G have an OD of 0.009 inches (0.20 mm). Needle shape includes regular, short bevel, intradermal, and winged. Needle shape is defined by one end of a needle enlarged to form a hub with a delivery device, such as a syringe, or other administration device. The other end of the needle is beveled, meaning it forms a sharp tip to maximize ease of insertion.

The route of administration, type of therapy, and whether the patient is a child or adult dictate the length and size of needle used. Intravenous injections use 1–2 inch 15–25 G needles. Intramuscular injections use 1–2 inch 19–22 G needles. Subcutaneous injections use 1/4–5/8 inch 24–25 G needles. Needle gauge for children rarely is larger than 22 G, usually 25–27G. Winged needles are used for intermittent heparin therapy. Many different types of therapies (e.g., radiology, anesthesia, biopsy, cardiovascular, ophthalmic, transfusions, tracheotomy, etc.) have their own peculiar types of needle preferences.

Needles are purchased either alone (e.g., Luer-Lok) to be attached to syringes, cartridges, and other delivery systems, or, for syringes, can be part of the syringe set (stake needle). Syringes with needles may also have needle protectors (for example, see http://www.bd.com/vacutainer/pdfs/blood_transfer_device_with_safetyglide_needle_VS5985.pdf) to avoid potential dangers of accidental needle sticks post-administration (for more detail regarding the 2000 Needlestick Safety Act, see http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=106_cong_public_laws&docid=f:publ430.106). Such protectors can either be part of the assembly or be assembled during the finishing process. Needlestick prevention can be manual (shield activated manually by the user, although there is still the risk of accidental sticking), active (automated needle shielding activated by user), or passive (automated needle shielding without action by the user).

PYROGENS (ENDOTOXINS)

Water and packaging materials are the greatest sources of pyrogens (pyrogenic contamination). Pyrogens are products of metabolism of microorganisms. The most potent pyrogenic substances (endotoxins) are constituents (lipopolysaccharides, LPS) of the cell wall of gram-negative bacteria (e.g., *Pseudomonas* sp, *Salmonella* sp, *Escherichia coli*). Gram-positive bacteria and fungi also produce pyrogens but of lower potency and of different chemical nature. Gram positive bacteria produce peptidoglycans, whereas fungi produce β -glucans, both of which can cause non-endotoxin pyrogenic responses. Endotoxins are lipopolysaccharides that exist in high molecular weight aggregate forms. However, the monomer unit of LPS is less than 10,000 daltons, enabling endotoxin to easily pass through sterilizing 0.2 micron filters. The lipid portion of the molecule is responsible for the biological activity. Since endotoxins are the most potent pyrogens and gram-negative bacteria are ubiquitous in the environment, especially water, this discussion focuses on endotoxins and the risk of their presence as contaminants in sterile products.

Pyrogens, when present in parenteral drug products and injected into patients, can cause fever, chills, pain in the back and legs, and malaise. Although pyrogenic reactions are rarely fatal, they can cause serious discomfort and, in the seriously ill patient, shock-like symptoms that can be fatal. The intensity of the pyrogenic response and its degree of hazard are affected by the medical condition of the patient, the potency of the pyrogen, the amount of the pyrogen, and the route of administration (intrathecal is most hazardous followed by intravenous, intramuscular, and subcutaneous). When bacterial (exogenous) pyrogens are introduced into the body, LPS targets circulating mononuclear cells (monocytes and macrophages) that, in turn, produce pro-inflammatory cytokines, such as interleukin 2, interleukin 6, and tissue necrosis factor. Besides LPS, gram-negative bacteria also release many peptides (e.g., exotoxin A, peptidoglycan, and muramyl peptides) that can mimic the activity of LPS and induce cytokine release. The Limulus Amebocyte Lysate (LAL) test can only detect the presence of LPS. It has been suggested that the Monocyte Activation Test, replace LAL as the official pyrogen test, due to its greater sensitivity to all agents that induce the release of cytokines that cause fever and a potential cascade of other adverse physiological effects.²³

CONTROL OF PYROGENS

It is impractical, if not impossible, to remove pyrogens, once present, without adversely affecting the drug product. Therefore, the emphasis should be on preventing the introduction or development of pyrogens in all aspects of the compounding and processing of the product.

Pyrogens may enter a preparation through any means that will introduce living or dead micro-organisms. However, current technology permits the control of such contamination, and the presence of pyrogens in a finished product indicates processing under inadequately controlled conditions. It also should be noted that time for microbial growth to occur increases the risk for elevated levels of pyrogens. Therefore, compounding and manufacturing processes should be carried out as expeditiously as possible, preferably planning completion of the process, including sterilization, within the maximum allowed time, according to process validation studies. Aseptic processing guidelines require establishment of time limitations throughout processing for the primary purpose of preventing the increase of endotoxin (and microbial) contamination that, subsequently, cannot be destroyed or removed.

Pyrogens can be destroyed by heating at high temperatures. A typical procedure for depyrogenation of glassware and equipment is maintaining a dry heat temperature of 250°C for 45 min. Exposure of 650°C for 1 min or 180°C for 4 hours, likewise, will destroy pyrogens. The usual autoclaving cycle will not

do so. Heating with strong alkali or oxidizing solutions destroys pyrogens. It has been claimed that thorough washing with detergent will render glassware pyrogen-free, if subsequently rinsed thoroughly with pyrogen-free water. Rubber stoppers cannot withstand pyrogen-destructive temperatures, so reliance must be on an effective sequence of washing, thorough rinsing with WFI, prompt sterilization, and protective storage to ensure adequate pyrogen control. Similarly, plastic containers and devices must be protected from pyrogenic contamination during manufacture and storage, since known ways of destroying pyrogens affect the plastic adversely. It has been reported that anion-exchange resins and positively-charged membrane filters remove pyrogens from water. Also, although reverse osmosis membranes will eliminate them, the most reliable method for their elimination from water is distillation.

A method that has been used for the removal of pyrogens from solutions is adsorption on adsorptive agents. However, since the adsorption phenomenon may also cause selective removal of chemical substances from the solution, this method has limited application. Other in-process methods for their destruction or elimination include selective extraction procedures and careful heating with dilute alkali, dilute acid, or mild oxidizing agents. In each instance, the method must be studied thoroughly to be sure it will not have an adverse effect on the constituents of the product. Although ultrafiltration now makes pyrogen separation on a molecular-weight basis possible and the process of tangential flow is making large-scale processing more practical, use of this technology is limited, except in biotechnological processing.

SOURCES OF PYROGENS

Through understanding the means by which pyrogens may contaminate parenteral products, their control becomes more achievable. Therefore, it is important to know that water is probably the greatest potential source of pyrogenic contamination, since water is essential for the growth of micro-organisms and frequently contaminated with gram-negative organisms. When micro-organisms metabolize, pyrogens will be produced. Therefore, raw water can be expected to be pyrogenic and only when it is appropriately treated to render it free from pyrogens, such as WFI, should it be used for compounding the product or rinsing product contact surfaces, such as tubing, mixing vessels, and rubber closures. Even when such rinsed equipment and supplies are left wet and improperly exposed to the environment, there is a high risk they will become pyrogenic. Although proper distillation will provide pyrogen-free water, storage conditions must be such that micro-organisms are not introduced and subsequent growth is prevented.

Other potential sources of contamination are containers and equipment. Pyrogenic materials adhere strongly to glass and other surfaces, especially rubber closures. Residues of solutions in used equipment often become bacterial cultures, with subsequent pyrogenic contamination. Since drying does not destroy pyrogens, they may remain in equipment for long periods. Adequate washing reduces contamination, and subsequent dry-heat treatment can render contaminated equipment suitable for use. However, all such processes must be validated to ensure effectiveness. Aseptic processing guidelines require validation of the depyrogenation process by demonstrating at least 3-log reduction in an applied endotoxin challenge.

Solutes may be a source of pyrogens. For example, the manufacturing of bulk chemicals may involve the use of pyrogenic water for process steps, such as crystallization, precipitation, or washing. Bulk drug substances derived from cell culture fermentation will almost certainly be heavily pyrogenic. Therefore, all lots of solutes used to prepare parenteral products should be tested to ensure they will not contribute unacceptable quantities of endotoxin to the finished product. It is standard practice, today, to establish valid endotoxin limits on active pharmaceutical ingredients and most solute additives.

The manufacturing process must be carried out with great care and as rapidly as possible, to minimize the risk of microbial contamination. Preferably, no more product should be prepared than can be processed completely within one working day, including sterilization.

PRODUCTION FACILITIES

The production facility and its associated equipment must be designed, constructed, and operated properly for the manufacture of a sterile product to be achieved at the quality level required for safety and effectiveness. Materials of construction for sterile product production facilities must be 'smooth, cleanable, and impervious to moisture and other damage'. Further, the processes used must meet cGMP standards. Since the majority of SVIs are aseptically processed (finished product not terminally sterilized), strict adherence to cGMP standards with respect to sterility assurance (particularly, the FDA and EU aseptic processing guidance documents, which can be found at <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm070342.pdf> and Eudra. ex Vol 4, Annex I. Manufacture of Sterile Medicinal Products, http://ec.europa.eu/health/files/eudralex/vol-4/2008_11_25_gmp-an1_en.pdf) is essential.

FUNCTIONAL AREAS

To achieve the goal of a manufactured sterile product of exceptionally high quality, many functional production areas are involved: warehousing or procurement; compounding (formulation); materials (containers, closures, equipment) preparation; filtration and sterile receiving; aseptic filling; stoppering; lyophilization (if warranted); and packaging, labeling, and quarantine. The extra requirements for the aseptic area are designed to provide an environment where a sterile fluid may be exposed to the environment for a brief period during subdivision from a bulk container to individual-dose containers, without becoming contaminated. Contaminants, such as dust, lint, and other particles and micro-organisms, are found floating in the air, lying on counters and other surfaces, attached to clothing and body surfaces of personnel, concentrated in the exhaled breath of personnel, and deposited on the floor. The design and control of an aseptic area is directed toward reducing the presence of these contaminants, so they are no longer a hazard to aseptic filling.

Although the aseptic area must be adjacent to support areas, so an efficient flow of components may be achieved, barriers must be provided to minimize ingress of contaminants to the critical aseptic area. Such barriers may consist of a variety of forms, including sealed walls, manual or automatic doors, air-lock pass-throughs, ports of various types, or plastic curtains. Figure 26-6 shows an example of a floor plan for a clinical supply production facility (selected as an example of a small-scale, noncomplex facility) in which the two fill rooms and the staging area constitute the walled critical aseptic area, access to which is only by means of pass-through airlocks. Adjacent support areas (rooms) consist of glass preparation, equipment wash, capping, manufacturing (compounding), and various storage areas. Figure 26-7 shows an example of a Class 100/Grade A small scale filling room with operators properly gowned and practicing good aseptic techniques.

FLOW PLAN

In general, the components for a parenteral product flow from the warehouse, after release, to either the compounding area, as for ingredients of the formula, or the materials support area, as for containers and equipment. After proper processing in these areas, the components flow into the security of the aseptic area for filling of the product in appropriate containers. From there the product passes into the quarantine and packaging area, where it is held until all necessary tests have been performed. If the product is to be sterilized in its final container, its passage

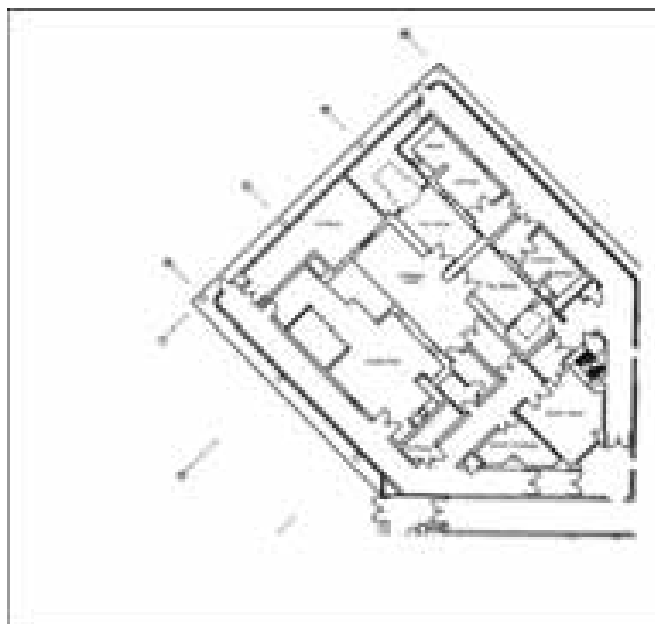


Figure 26-6. Floor plan of aseptic filling rooms and staging room with adjacent support areas.



Figure 26-7. Properly gowned and trained aseptic processing operators in a Class 100/Grade A Clean Room. (Courtesy of Dr. Laura Thoma, The University of Tennessee College of Pharmacy.)

is interrupted after leaving the aseptic area for subjection to the sterilization process. After the results from all tests are known, the batch records have been reviewed, and the product has been found to comply with its release specifications, it passes to the finishing area for final release for shipment. There, sometimes, are variations from this flow plan to meet the specific needs of an individual product or to conform to existing facilities. Automated operations have much larger capacity and convey the components from one area to another with little or no handling by operators.

Clean Room Classified Areas

Due to the extremely high standards of cleanliness and purity that must be met by parenteral products, it has become standard practice to prescribe specifications for the environments (clean rooms) in which these products are manufactured (Table 26-4). Table 26-4A compares US and European classifications and

Table 26-4A. Clean Room Classifications

European	United States	International Society of	Max No. of Particles	Max No. of Particles
Grade	Classification	Pharm. Eng. Description	per m ³ ≥ 0.5 μm	per m ³ ≥ 5 μm
A	100	Critical	3,500	0
B	100	Clean	3,500	0
C	10,000	Controlled	350,000	2,000
D	100,000	Pharmaceutical	3,500,000	20,000

Table 26-4B. ISO 14644 Classification of Cleanroom Particle Limits

ISO Classification	Maximum Concentration Limits (Particles per Cubic Meter of Air) for Particles ≥ the Sizes per Each Column				
	0.1 μm	0.3 μm	0.5 μm	1 μm	5 μm
1	10	-	-	-	-
2	100	10	4	-	-
3	1,000	102	35	8	-
4	10,000	1,020	352	83	-
5	100,000	10,200	3,520	832	29
6	1,000,000	102,000	35,200	8320	290
7	-	-	352,000	83,200	2,930
8	-	-	3,520,000	832,000	29,300
9	-	-	-	8,320,000	293,000

clean room designations assigned by the International Society of Pharmaceutical Engineers. Table 26-4B provides the International Standards Organization (ISO) 14644 Classification of Cleanroom Particle Limits adhered to by the parenteral manufacturing industry. Table 26-4A numbers are based on the maximum allowed number of airborne particles/ft³ or particles/m³ of 0.5 μm or larger size and, for Europe, 5.0 μm or larger size. The classifications used in pharmaceutical practice normally range from Class 100,000 (Grade D) for materials support areas to Class 100 (Grade A) for aseptic areas. To achieve Class 100 conditions, HEPA filters are required for the incoming air, with the effluent air sweeping the downstream environment at a uniform velocity, 100 ft/min ± 20%, along parallel lines (laminar air flow). HEPA filters are defined as 99.99% or more efficient in removing, from the air, 0.3 μm particles generated by vaporization of the hydrocarbon Emory 3004.

Because so many parenteral products are manufactured at one site for global distribution, air quality standards in aseptic processing areas must meet both US and European requirements. European standards differ from US standards, as European standards:

- use Grades A, B, C, and D classifications, rather than Class X (100, 1,000, etc);
- use particle and microbial limits per cubic meter, rather than per cubic foot;
- require particle measurements at 5 microns in addition to 0.5 microns in Grade A and B areas; and
- differentiate area cleanliness dynamically and 'at rest.'

For the sake of convenience, the remainder of this chapter uses Class X (e.g., 100, 1,000, 10,000, 100,000) designations, although it is recognized that the use of Grades or ISO numbers are more contemporary.

Air Cleaning—Since air is one of the greatest potential sources of contaminants in clean rooms, special attention must be given to air drawn into clean rooms by the heating, ventilating, and air conditioning (HVAC) systems. This may be done by a series of treatments that vary somewhat from one installation to another.

In one such series, air from the outside, first, is passed through a prefilter, usually of glass wool, cloth, or shredded plastic, to remove large particles. Then, it may be treated by passage through an electrostatic precipitator. Such a unit induces an electrical charge on particles in the air and removes them by attraction to oppositely charged plates. The air then passes through the most efficient cleaning device, a HEPA filter.

For personnel comfort, air conditioning and humidity control should be incorporated into the system. The latter is also important for certain products, such as those that must be lyophilized, and for the processing of plastic medical devices. The clean, aseptic air is introduced into the Class 100 area and maintained under positive pressure, which prevents outside air from rushing into the aseptic area through cracks, temporarily open doors, or other openings.

Laminar-Flow Enclosures—The required environmental control of aseptic areas has been made possible by the use of laminar airflow, originating through a HEPA filter, occupying one entire side of the confined space. Therefore, it bathes the total space with very clean air, sweeping away contaminants. The orientation for the direction of airflow can be horizontal (Figure 26-8A) or vertical (Figure 26-8B) and may involve a limited area, such as a workbench, or an entire room. Figure 26-9 shows a syringe-filling line in a Grade A/Class 100 area using vertical laminar airflow. The machine is placed in a conventional clean room with vertical LF provided through either the ceiling or a LF hood on top of the machine. The machine guarding is a stainless steel frame that can hold the LF hood. The panes are safety glass. This could be an example of a Restricted Access Barrier System (RABS), although there are no gloves installed, thus, requiring doors to be open to access the equipment, which is contrary to the requirements of an authentic RABS. The area outside the RAB can be maintained at a slightly lower level of cleanliness than that inside, perhaps Class 10,000 down to Class 1,000.

Critical areas of processing, wherein the product or product contact surfaces may be exposed to the environment, even for a brief period of time, must meet Class 100 clean room standards.

Any contamination introduced upstream by equipment, arms of the operator, or leaks in the filter will be blown downstream.

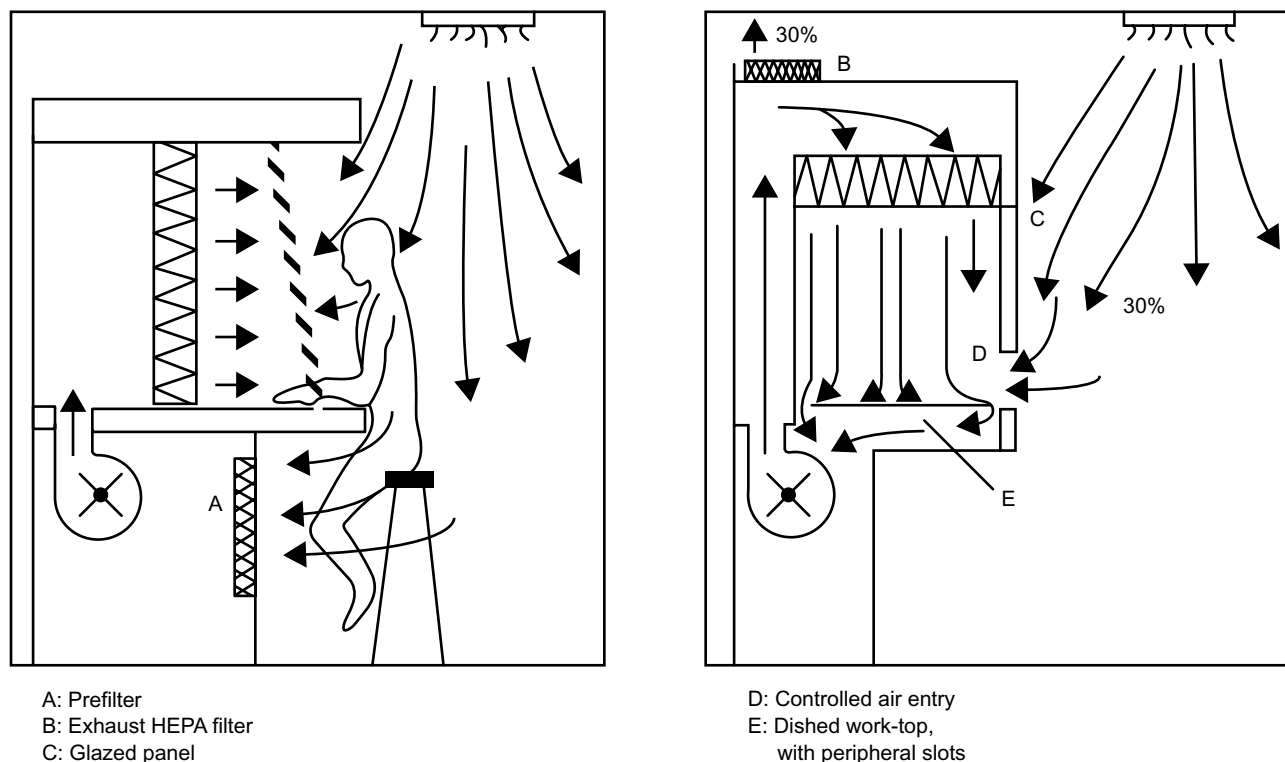


Figure 26-8. Horizontal and vertical laminar air flow. (Courtesy of Cosslett AG, The design of controlled environments. In: Denyer SP, Baird RM, eds. *Guide to Microbiological Control in Pharmaceuticals and Medical Devices*, 2nd ed. London: CRC Press, Taylor & Francis, 2007: 69–88.)

In the instance of horizontal flow, this may be toward the critical working site, the face of the operator, or across the room. Should the contaminant be, for example, penicillin powder, a biohazard material, or viable micro-organisms, the danger to the operator is apparent.

Further, great care must be exercised to prevent cross-contamination from one operation to another, especially with horizontal laminar air flow. For most large-scale operations, as shown in Figure 26-8B and Figure 26-9, a vertical system is much more desirable, with the air flowing through perforations in the countertop or through return louvers at floor level, where it can be directed for decontamination. Laminar-flow environments provide well-controlled work areas, only if proper precautions are observed. Any reverse air currents or movements exceeding the velocity of the HEPA-filtered airflow may introduce contamination, as may coughing, reaching, or other manipulations of the operator. Therefore, laminar-flow work areas should be protected by being located within controlled environments. Personnel should be attired for aseptic processing, as subsequently described. All movements and processes should be planned carefully, to avoid the introduction of contamination upstream of the critical work area. Checks of the air stream should be performed initially and at regular intervals (usually every six months), to be sure no leaks have developed through or around the HEPA filters.

Clean room design, traditionally, has Class 100 rooms adjacent to Class 100,000 rooms. Regulatory authorities have raised great concerns about this significant change in air quality from critical to controlled areas. It is now preferable to have an area classified from Class 1,000 to Class 10,000 in a buffer area between a Class 100 and Class 100,000 area.

Materials Support Area—This area is constructed to withstand moisture, steam, and detergents and is, usually, a Class 100,000 clean room. The ceiling, walls, and floor should be constructed of impervious materials, so moisture runs off and is not held. One of the finishes with a vinyl or epoxy-sealing coat



Figure 26-9. High speed syringe filling machine for pre-sterilized syringes. (Courtesy of Robert Bosch GmbH.)

provides a continuous surface free from all holes or crevices. All such surfaces can be washed at regular intervals to keep them thoroughly clean. These areas should be exhausted adequately, so the heat and humidity are removed for the comfort of personnel. Precautions must be taken to prevent the accumulation of dirt and the growth of micro-organisms due to the high humidity and heat. In this area, preparation for the filling operation, such as cleaning and assembling equipment, is undertaken. Adequate sink and counter space must be provided. This area must be cleanable, and the microbial load must be monitored and controlled. Precautions must also be taken to prevent deposition of particles or other contaminants on clean containers and equipment, until they have been properly boxed or wrapped preparatory to sterilization and depyrogenation.

Compounding Area—The formula is compounded in this area. Although it is not essential that this area be aseptic, control of micro-organisms and particulates should be more stringent than in the materials support area. For example, means may need provided to control dust generated from weighing and compounding operations. Cabinets and counters should, preferably, be constructed of stainless steel. They should fit snugly to walls and other furniture, so there are no catch areas dirt can accumulate. The ceiling, walls, and floor should be similar to those for the materials support area.

Aseptic Area—The aseptic area requires construction features designed for maximum microbial and particulate control. The ceiling, walls, and floor must be sealed, so they may be washed and sanitized with a disinfectant, as needed. All counters should be constructed of stainless steel and hung from the wall, so there are no legs to accumulate dirt, where they rest on the floor. All light fixtures, utility service lines, and ventilation fixtures should be recessed in the walls or ceiling to eliminate ledges, joints, and other locations for the accumulation of dust and dirt. As much as possible, tanks containing the compounded product should remain outside the aseptic filling area, with the product fed into the area through hose lines. Proper sanitization is required, if the tanks must be moved in. Large mechanical equipment located in the aseptic area should be housed as completely as possible within a stainless steel cabinet, to seal the operating parts and their dirt-producing tendencies from the aseptic environment. Further, all such equipment parts should be located below the filling line. Mechanical parts that will contact the parenteral product should be demountable, so they can be cleaned and sterilized.

Personnel entering the aseptic area should enter only through an airlock. They should be attired in sterile coveralls with sterile hats, masks, goggles, foot covers, and double gloves. Movement within the room should be minimal, and in-and-out movement rigidly restricted during a filling procedure. The requirements for room preparation and the personnel may be relaxed, if the product is to be sterilized terminally in a sealed container. Some are convinced, however, it is better to have one standard procedure meeting the most rigid requirements.

Isolation (Barrier) Technology—Isolator (or barrier) technology has long been used in the pharmaceutical industry and ranges from simple screens to restricted access barriers (RABS) to full isolation systems, all designed to isolate aseptic operations from personnel and the surrounding environment. Sterility tests are now almost exclusively conducted within isolators. A false-positive sterility test is practically unheard of these days, such that, if a positive test does occur, it likely is a true contamination, not as a result of contamination introduced during the test. Isolation technology in various formats has been adapted to automated, large-scale, aseptic filling operations.²⁴ An example of a sterility test isolator is shown in Figure 26-10, and an example of a filling operation within an isolator is shown in Figure 26-11. The sealed enclosures are presterilized, usually with peracetic acid, hydrogen peroxide vapor, or steam. Sterile supplies are introduced from sterilizable movable modules through uniquely engineered transfer ports or directly from attached sterilizers, including autoclaves and hot-air sterilizing tunnels.



Figure 26-10. Example of sterility test isolator with vapor phase hydrogen peroxide generator. (Courtesy of Baxter Healthcare Corporation.)



Figure 26-11. Example of a large-scale filling line within an isolator. (Courtesy of Robert Bosch GmbH.)

Results have been very promising, giving expectation of significantly enhanced control of the aseptic processing environment.

Isolators are enclosed, usually positively pressurized units with high efficiency particulate air (HEPA) filters, supplying ISO 5 airflow in a unidirectional manner to the interior. Air recirculates by returning it to the air handlers through sealed ductwork. Cleaning can be manual or automated (clean-in-place). Access to an isolator is through glove ports and sterile transfer systems. Isolators can be located in an ISO 8 or better environment.

The operations are performed within windowed, sealed walls, with operators working through glove ports. The sealed enclosures are presterilized, usually with peracetic acid, hydrogen peroxide vapor, or steam. Sterile supplies are introduced from sterilizable, movable modules, through uniquely engineered transfer ports or directly from attached sterilizers, including autoclaves and hot-air sterilizing tunnels.

Recent isolator development has been significant, and such systems are now better specified than ever. In advanced aseptic processing facilities, it has been proven that isolators can provide zero colony forming unit (0 cfu) contamination in process operations, whereas the background environment is only at ISO 8 - EC grade D level. However, cost-savings in cleanroom construction and operation may be offset by the construction and validation costs of the isolator system.

For existing production lines, where conversion from conventional filling to filling within an isolator is time and cost

prohibitive, modifications of isolation systems, RABS, have been applied. RABS offer a combined physical (e.g., plexiglas partitions) and aerodynamic barrier, ideally controlled by positive pressure with clean-air filtration, providing air exchanges and particulate clean-up for an ISO 5 critical process zone. RABS come in two types—‘passive,’ where there is no in-process open door access; and ‘active,’ where, under certain validated system configurations and control conditions, access may be included.

Several factors have contributed to the increased importance and utilization of barrier isolator technology:

1. The high level of concern from manufacturers and regulatory agencies over the level of sterility assurance in aseptic processing.
2. Continued, relatively high level of product recalls, due to concerns—proven or suspected—over contamination potential.
3. The surge of potential heat-labile products from biotechnology and the inability to terminally sterilize these molecules. There are needs to control the environment not only from contamination, but also with respect to stability considerations—temperature, humidity, and, if necessary, anaerobics.
4. Many new drug compounds are cytotoxic or otherwise highly potent, where safety considerations demand separation of these drugs from human operators.
5. Because so many biopharmaceutical drugs are so expensive, there is a trend toward smaller batch production. Smaller batch production makes construction of large manufacturing facilities unnecessary, yet there is still the need to manufacture in Class 100/Grade A/ISO 5 clean rooms. Isolators are ideal for smaller facilities, plus are much more economical from the standpoint of capital, labor and maintenance, and operator (e.g., number of employees, gowning) costs.

The main features of barrier/isolator technology are the ability to sterilize (more than sanitize) the environment to which sterile solution is exposed during filling and stoppering and the removal of direct human contact with the exposed sterile product. Isolators not only protect the product from potential human contamination, but also protect the human from potential toxic effects of direct exposure to the drug product, especially important for cytotoxic drugs.

MAINTENANCE OF CLEAN ROOMS

Maintaining the clean and sanitized conditions of clean rooms, particularly the aseptic areas, requires diligence and dedication of expertly trained custodians. Assuming the design of the facilities is cleanable and sanitizable, a carefully planned cleaning schedule should be developed, ranging from daily to monthly, depending on the location and its relation to the most critical Class 100 areas. Tools used should be non-linting, designed for clean room use, held captive to the area, and, preferably, sterilizable.

Liquid disinfectants (sanitizing agents) should be selected carefully, due to data showing their reliable activity against inherent environmental micro-organisms. They should be recognized as supplements to good housekeeping, never as substitutes. They should be rotated with sufficient frequency to avoid the development of resistant strains of micro-organisms. Figure 26-12 features an example of the ‘three bucket’ system used to sanitize facilities. One bucket is to remove as much of the remnant of the ‘dirty’ mop or sponge, the second bucket contains a rinse solution to help clean the mop/sponge, and the third bucket contains the sanitizing solution. The sanitizing solution should be rendered sterile prior to use, although, once in use, it will no longer be sterile.

It should be noted that ultraviolet (UV) light rays of 237.5 nm wavelength, as radiated by germicidal lamps, are an effective



Figure 26-12. Three bucket sanitizing system. (Courtesy of Contec, Inc. and Steve Fincher Photography.)

surface disinfectant. However, it must also be noted that they are only effective, if they contact the target micro-organisms at a sufficient intensity for a sufficient time. The limitations of their use must be recognized, including no effect in shadow areas, reduction of intensity by the square of the distance from the source, reduction by particulates in the ray path, and the toxic effect on epithelium of human eyes. It is stated that an irradiation intensity of 20 $\mu\text{w}/\text{cm}^2$ is required for effective antibacterial activity.

PERSONNEL

Personnel selected to work on the preparation of a parenteral product must be neat, orderly, and reliable. They should be in good health and free from dermatological conditions that might increase the microbial load. If personnel show symptoms of a head cold, allergies, or similar illness, they should not be permitted in the aseptic area, until recovery is complete. However, a healthy person with the best personal hygiene will still shed large numbers of viable and nonviable particles from body surfaces.²⁵ This natural phenomenon creates continuing problems, when personnel are present in clean rooms; effective training and proper gowning can reduce, but not eliminate, the problem of particle shedding from personnel.

Aseptic-area operators should be given thorough, formal training in the principles of aseptic processing and the techniques to be employed.²⁶ Subsequently, the acquired knowledge and skills should be evaluated, to assure that training has been effective, before personnel are allowed to participate in the preparation of sterile products. Retraining should be performed on a regular schedule to enhance the maintenance of the required level of expertise. An effort should be made to imbue operators with an awareness of the vital role they play in determining the reliability and safety of the final product. This is especially true of supervisors, since they should be individuals who not only understand the unique requirements of aseptic procedures, but are able to obtain the full participation of other employees in fulfilling these exacting requirements.

The uniform worn is designed to confine the contaminants discharged from the body of the operator, thereby preventing their entry into the production environment. For use in the



Figure 26-13. Fully gowned personnel in Class 100/Grade A/B Cleanrooms. (Courtesy of Baxter Healthcare Corporation.)

aseptic area, uniforms should be sterile. Fresh, sterile uniforms should be used after every break period or whenever the individual returns to the aseptic area. In some plants, this is not required, if the product is to be sterilized in its final container. Uniforms, usually, consist of coveralls for both men and women, hoods to cover the hair completely, face masks, and Dacron or plastic boots (Fig. 26-7 and Fig. 26-13). Sterile rubber or latex-free gloves are also required for aseptic operations, preceded by thorough scrubbing of the hands with a disinfectant soap. Two pairs of gloves are put on, one pair at the beginning of the gowning procedure, the other pair after all other apparel has been donned. In addition, goggles are required to complete the coverage of all skin areas.

Dacron or Tyvek uniforms are usually worn, are effective barriers to discharged body particles (viable and nonviable), are essentially lint-free, and are reasonably comfortable. Air showers are sometimes directed on personnel entering the processing area to blow loose lint from the uniforms.

Gowning rooms should be designed to enhance pregowning and gowning procedures by trained operators, so it is possible to ensure the continued sterility of the exterior surfaces of the sterile gowning components. Degowning should be performed in a separate exit room.

ENVIRONMENTAL CONTROL EVALUATION

Manufacturers of sterile products use extensive means to control the environment, so these critical injectable products can be prepared free from contamination. Nevertheless, tests should be performed to determine the level of control actually achieved. Normally, the tests consist of counting viable and non-viable particles suspended in the air or settled on surfaces in the workspace. A baseline count, determined by averaging multiple counts, when the facility is operating under controlled conditions, is used to establish the optimal test results expected. During the subsequent monitoring program, the test results are followed carefully for high individual counts, a rising trend, or other abnormalities. If the results exceed selected alert or action levels, a plan of action must be put into operation to determine if or what corrective and follow-up measures are required.

The tests used measure either the particles in a volume of sampled air or the particles settling or present on surfaces. To measure the total particle content in an air sample, electronic particle counters are available, operating on the principle of the measurement of light scattered from particles as they pass through the cell of the optical system. These instruments not only count particles, but also provide a size distribution, based on the magnitude of the light scattered from the particle.

Although a volume of air measured by an electronic particle counter will detect all particles instantly, these instruments cannot differentiate between viable (e.g., bacterial and fungal) and non-viable particles. Due to the need to control the level of micro-organisms in the environment in which sterile products are processed, it is also necessary to detect viable particles. These are usually fewer in number than non-viable particles and are only detectable as colony-forming units (CFUs) after a suitable incubation period at, for example, 30°C to 35°C for up to 48 hours. Thus, test results will not be known until 48 hours after the samples are taken, unless more rapid microbial test procedures become dependable and acceptable.

Locations for sampling should be planned to reveal potential contamination levels that may be critical in the control of the environment. For example, the most critical process step is usually the filling of dispensing containers, a site obviously requiring monitoring. In fact, the FDA aseptic processing guidelines require air particle counts be measured during actual filling and closing operations and not more than one foot from the actual work site.²⁷ Other examples include the gowning room, high-traffic sites in and out of the filling area, the penetration of conveyor lines through walls, and sites near the inlet and exit of the air system.

The sample should be large enough to obtain a meaningful particle count. At sites where the count is expected to be low, the size of the sample may need increased; for example, in Class 100 areas, Whyte and Niven²⁸ suggest that the sample be at least 30 ft³ and, probably, much more. Many firms employ continuous particle monitoring in Class 100 areas to study trends and/or to identify equipment malfunction.

The slit-to-agar (STA) sampler draws, by vacuum, a measured volume of air through an engineered slit, causing the air to impact the surface of a slowly rotating nutrient agar plate (Fig. 26-14). Micro-organisms adhere to the surface of the agar and grow into visible colonies counted as CFUs, since it is not known whether the colonies arise from a single micro-organism or a cluster.

A widely used method for microbiological sampling consists of the exposure of nutrient agar culture plates to the settling of micro-organisms from the air. This method is very simple and inexpensive to perform, but will detect only those organisms that have settled on the plate; therefore, it does not measure the number of micro-organisms in a measured volume of air (a non-quantitative test). Nevertheless, if the conditions of exposure are repeated consistently, a comparison of CFUs at one sampling site, from one time to another, can be meaningful.

Whyte and Niven suggested that settling plates should be exposed in Class 100 areas for an entire fill (up to 7 to 8 hours), rather than the more common 1 hour. However, excessive dehydration of the medium must be avoided, particularly in the path of laminar-flow air. The European Union GMP guidelines for sterile manufacture of medicinal products suggest an exposure period of not more than four hours that has been adopted by the FDA aseptic processing guidelines.

The number of micro-organisms on surfaces can be determined with nutrient agar plates having a convex surface (Rodac Plates; Fig. 26-15). With these, it is possible to roll the raised agar surface over flat or irregular surfaces to be tested. Organisms are picked up on the agar and grow during subsequent incubation. This method can also be used to assess the number of micro-organisms present on the surface of the uniforms of operators, either as an evaluation of gowning technique, immediately after gowning, or as a measure of the accumulation of micro-organisms during processing. Whenever used, care must be taken to remove any agar residue left on the surface tested.

(Further discussion of proposed, viable particle test methods and the counts accepted can be found in Section <1116> 'Microbial Evaluation and Classification of Clean Rooms and Other Controlled Environments' in the USP.)

Results from these tests, although not available until 2 days after sampling, are valuable to keep cleaning, production, and



Figure 26-14. Examples of a Slit-To-Agar (STA) Quantitative Air Sampler.



Figure 26-15. Example of a Rodac plate. (Courtesy of Baxter Healthcare Corporation.)

quality-control personnel apprised of the level of contamination in a given area and, by comparison with baseline counts, will indicate when more-extensive cleaning and sanitizing is needed. The results may also serve to detect environmental control defects, such as failure in air-cleaning equipment or the presence of personnel who may be disseminating large numbers of bacteria without apparent physical ill effects.

Issues regarding environmental monitoring remain among the most controversial aspects of cGMP regulatory inspections of parenteral manufacturing and testing environments. Regulatory trends include requiring an increase in the number and frequency of locations monitored in the clean room and on clean room personnel, enforcing numerical alert and action limits, and linking environmental monitoring data to the decision to release or reject the batch. It has been pointed out that fully gowned personnel still release a finite number of micro-organisms (typically 10 to 100 cfu per hour), so it is unreasonable to impose the requirement of zero microbial contamination limits at any location in the clean room.²⁹

MEDIA FILL (PROCESS SIMULATION TESTING)

FDA inspections have increasingly focused on media fill studies that truly simulate the production process. The media fill or process simulation test involves preparation and sterilization (often by filtration) of sterile trypticase soy broth and filling sterile containers with this broth, under conditions simulating, as closely as possible, those characteristics of a filling process for a product. The key is designing these studies to simulate all factors that occur during the normal production of a lot (Table 26-5).²⁷

The media fill provides a 'one-time' representation of the capabilities of an aseptic processing operation. Media fills are conducted, when a new filling line or new product container is introduced. For initial qualification of a line or product, three consecutive, separate, and successful media fill runs must take place. The FDA stresses that three is a minimum number of runs. Today, the term 'successful' means there is no growth in any of the units filled with sterile broth. All activities and interventions representative of each shift on each line must be simulated during the media fill. All personnel involved in the aseptic filling of a product (i.e., operators, maintenance personnel, microbiology support personnel) must participate in at least one media fill run per year. Typically, for each filling line

Table 26-5. Factors to Consider in the Design of Media Fill Studies

- Duration of longest run
- Worst-case environmental conditions
- Number and type of interventions, stoppages, adjustments, transfers
- Aseptic assembly of equipment
- Number and activities of personnel
- Number of aseptic additions
- Shift breaks, changes, multiple gownings
- Number/type of aseptic equipment disconnections and connections
- Aseptic samples
- Line speed/configuration
- Manual weight checks
- Operator fatigue
- Container/closure types run on the line
- Temperature/relative humidity extremes
- Conditions permitted before line clearance
- Container/closure surfaces which contact formulation during aseptic process

and process, the filling operation is validated for the smallest and largest container size that will be used.

After initial qualification, media fills are then conducted on a periodic basis, usually twice a year on the same filling line, to ensure that conditions that existed during the initial qualification have been maintained. For periodic qualification, only one successful media fill run is required. If any media fill run fails or significant changes occur with the line, facility, or personnel, then the initial qualification media fill (three consecutive successful runs) must be conducted. Any changes in the process must be evaluated for its level of significance (change control quality system) that would necessitate a media fill validation run. Any media fill failure must be thoroughly investigated and followed by multiple repeat media fill runs. It is considered inappropriate to 'invalidate' a media fill run.

The number of containers filled with media, ideally, should be the same as the actual number filled, according to the batch record for the product being validated. Of course, this is unrealistic for large batch sizes. Therefore, the number of units filled must be sufficient to reflect the effects of all worst case filling rates. For example, operator fatigue and the maximum number of interventions and stoppages must be incorporated into the media fill protocol. When media filling first started, the acceptable rate of positives (number of containers that showed contamination after incubating the culture media) was 1 out of 1000 (0.1%). Later that number became 1 out of 3000 to account for 95% confidence of a contamination rate of 0.1%. Today, 1 positive out of 3000 is no longer acceptable. Table 26-6 presents that ISO standard used to determine the minimum number of containers filled with media and the acceptable number of positives. The most common number of containers filled with media in the industry is 4750 with three consecutive runs of 4750 used for initial performance qualification of a new product and/or filling/closing line. This same number of units filled—4750—is also used for the routine semi-annual requalification media fills. The expected number of positive media fills (growth seen upon incubation) is zero. One or more failures likely means there is a significant breach in the aseptic manufacturing process, and the ensuing investigation must do everything possible to find the assignable cause.

After filling with culture media, but prior to incubation, all units should be inverted or swirled to enable the media to make contact with all internal surfaces of the container/closure system.

The culture media used for each media fill exercise must be tested to ensure it will support the growth of micro-organisms, if they are present. Challenge organisms used in the media challenge pretesting should include those isolated from

environmental/personnel monitoring, those isolated from positive sterility test results and USP growth promotion micro-organisms. The positive control units inoculated with approximately 100 CFUs of these challenge organisms are incubated at temperatures and times validated to show microbial growth if present. After the 14 day incubation period of the media fill containers, negative control units should then be inoculated with challenge organisms, to prove the media will still support growth, if present.

Inspection of media filled units, before and after incubation, is conducted by individuals trained as qualified inspectors and certified by the quality control unit. It is permissible that any unit, after filling, that is found to lack integrity be rejected from being part of the media fill incubation, just as a product vial would be rejected if a critical defect were found. However, if a media fill unit is found damaged after incubation is underway, it must remain incubated and counted in the data for the media fill batch. Procedures must be very clear and specific regarding samples taken during the media fill that simulate the actual sampling process and why these units are not part of those incubated.

Other requirements of a valid media fill experiment include:

- Must have the appropriate criteria for batch yield and accountability, just like a product batch.
- Must identify any contaminant to the species level and perform complete investigations of failed media fills.
- FDA advocates videotape media fills to identify personnel practices that could negatively impact the aseptic process.
- Media fill duration, according to FDA, EU, ISO, CEN (European Committee of Standardization), and PIC, must be sufficiently long to include all required manipulations and cover the same length of time normally consumed by the commercial process. Most media fills are a minimum of 3 hours; some may be as long as 24 hours.

PRODUCTION PROCEDURES

The processes required for preparing sterile products constitute a series of events initiated with the procurement of approved raw materials (drugs, excipients, vehicles, etc.) and primary packaging components (containers, closures, etc.) and ending with the sterile product sealed in its dispensing package. Each step in the process must be controlled very carefully, so the product has its required quality. To ensure the latter, each process should be validated to ensure it is accomplishing what it is intended to do. For example, an autoclave sterilization process must be validated by producing data showing it effectively kills resistant forms of micro-organisms; or, a cleaning process for rubber closures should provide evidence that it is cleaning closures to the required level of cleanliness; or, a filling process should provide evidence that it repeatedly delivers the correct fill volume per container. The validation of processes requires extensive and intensive effort to be successful and is an integral part of cGMP requirements.

CLEANING CONTAINERS AND EQUIPMENT

Containers and equipment coming in contact with parenteral preparations must be cleaned meticulously. It should be obvious that even new, unused containers and equipment are contaminated with such debris as dust, fibers, chemical films, and other materials arising from such sources as the atmosphere, cartons, the manufacturing process, and human hands. Residues from previous use must be removed from used equipment, before it is suitable for reuse. Equipment should be reserved exclusively for use only with parenteral preparations and, where conditions dictate, only for one product to reduce the risk of contamination. For many operations, particularly with biologic and biotechnology products, equipment is dedicated for only one product.

Table 26-6. ISO 13408-1 Standards for Minimum Number of Containers Filled with Media and the Acceptable Number of Positives

Number of Media Fill Units	Allowable Number of Failed Units (95% C.L.) by ISO	Allowable Number of Failed Units by Simple Math
3,000	1	3
4,750	2	4
6,300	3	6
7,760	4	7
9,160	5	9
10,520	6	10
11,850	7	11
13,150	8	13
14,440	9	13
15,710	10	15
16,970	11	17

A variety of machines are available for cleaning new containers for parenteral products. These vary in complexity from a small, hand loaded, rotary rinsers to large, automatic washers capable of processing several thousand containers per hour. The selection of the particular type is determined largely by the physical type of containers, the type of contamination, and the number to be processed in a given period of time.

Validation of cleaning procedures for equipment is another 'hot topic,' with respect to cGMP regulatory inspections. Inadequate cleaning processes have been a frequent citing by FDA and other regulatory inspectors, when inspecting both active ingredient and final product manufacturing facilities. It is incumbent upon parenteral manufacturers to establish scientifically justified acceptance criteria for cleaning validation. If specific analytical limits for target residues are arbitrarily set, this will cause concern for quality auditors. Validation of cleaning procedures can be relatively complicated, due to issues with sample methods (e.g., swab, final rinse, and testing of subsequent batch), sample locations, sensitivity of analytical methods, and calculations used to establish cleaning limits. Cleaning validation involves challenging the 'hardest to clean' surfaces with a well-defined sample, typically an active pharmaceutical ingredient (API) or a known 'hard to remove' substance like a sparingly soluble pharmaceutical or 'sticky' protein. The cleaning procedure is applied, using either swab samples or rinse samples obtained from the 'hardest to clean' surface locations. These samples are analyzed for residual API, using either a specific analytical technique, such as high performance liquid chromatography, or non-specific method, such as total organic carbon. Acceptance limits must be justified that the cleaning procedure accomplishes repeatedly before the cleaning method can be considered valid. (Additional discussion of this topic is found in Chapter 25 – Sterilization Processes and Sterility Assurance.)

CHARACTERISTICS OF MACHINERY

Regardless of the type of cleaning machine selected, certain fundamental characteristics are usually required:

1. The liquid or air treatment must be introduced in such a manner that it will strike the bottom of the inside of the inverted container, spread in all directions, and smoothly flow down the walls and out the opening with a sweeping action. The pressure of the jet stream should be such that there is minimal splashing and turbulence inside. Splashing may prevent cleaning all areas, and turbulence may redeposit loosened debris. Therefore, direct introduction of the jet stream within the container with control of its flow is required.
2. The container must receive a concurrent outside rinse.
3. The cycle of treatment should provide a planned sequence, alternating very hot and cool treatments. The final treatment should be an effective rinse with WFI.
4. All metal parts coming in contact with the containers and with the treatments should be constructed of stainless steel or some other non-corroding and non-contaminating material.

TREATMENT CYCLE

The cycle of treatments to be employed varies with the condition of the containers to be cleaned. In general, loose debris can be removed by vigorous rinsing with water. Detergents are rarely used for new containers, due to the risk of leaving detergent residues. However, a thermal-shock sequence in the cycle is usually employed to aid, by expansion and contraction, loosening of debris that may be adhering to the container wall. Sometimes, only an air rinse is used for new containers, if only loose debris is present. In all instances, the final rinse, whether air or WFI, must be ultraclean, so no particulate residues are left by the rinsing agent.



Figure 26-16. Loading end of large conveyer vial washer for vials from a rotary rinser through a sterilizing tunnel with vertical laminar-airflow protection of clean vials. (Courtesy of Baxter Healthcare Corporation.)

Only new containers are used for parenterals. Improvements have been made in maintaining their cleanliness, during shipment from the manufacturer through tight, low-shedding packaging, including plastic blister packs.

MACHINERY FOR CONTAINERS

The machinery available for cleaning containers embodies the previously mentioned principles, but varies in the mechanics by which it is accomplished. In one manual loading type, the jet tubes are arranged on arms like the spokes of a wheel, which rotate around a center post through which the treatments are introduced. An operator places the unclean containers on the jet tubes, as they pass the loading point, and removes the clean containers as they complete one rotation. A continuous automated line operation, capable of cleaning hundreds of containers per hour, is shown in Figure 26-16. The vials are fed into the rotary rinser in the foreground, transferred automatically to the covered sterilizing tunnel in the center, conveyed through the wall in the background, and discharged into the filling clean room.

HANDLING AFTER CLEANING

The wet, clean containers must be handled in such a way that contamination is not reintroduced. A wet surface will collect contaminants much more readily than a dry surface will. For this reason, wet, rinsed containers must be protected (e.g., by a laminar flow of clean air until covered, within a stainless steel box, or within a sterilizing tunnel). In addition, microorganisms are more likely to grow in the presence of moisture. Therefore, wet, clean containers should be dry-heat sterilized, as soon as possible after washing. Doubling the heating period is also adequate to destroy pyrogens; for example, increasing the dwell time at 250° from 1 to 2 hours, however, the actual time-temperature conditions required must be validated.

Increases in process rates have necessitated the development of continuous, automated line processing, with a minimum of individual handling, still maintaining adequate control of the cleaning and handling of the containers.¹⁷ The clean, wet containers are protected by filtered, laminar-flow air from the rinser, through the tunnel, and until they are delivered to the filling line (Fig. 26-16).

CLOSURES

The rough, elastic, and convoluted surface of rubber closures renders them difficult to clean. In addition, any residue of lubricant from molding or surface 'bloom' of inorganic constituents must be removed. The normal procedure calls for gentle agitation in a hot solution of a mild water softener or

detergent. The closures are removed from the solution and rinsed several times, or continuously for a prolonged period, with filtered WFI. The rinsing is done in a manner that flushes away loosened debris. The wet closures are carefully protected from environmental contamination, sterilized, usually by steam sterilization (autoclaving), and stored in closed containers, until ready for use. This cleaning and sterilizing process must also be validated with respect to rendering the closures free from pyrogens. Actually, it is the cleaning and final, thorough rinsing with WFI that must remove pyrogens, since autoclaving does not destroy pyrogens. If the closures were immersed during autoclaving, the solution is drained off, before storage, to reduce hydration of the rubber compound. If the closures must be dry for use, they may be subjected to vacuum drying at a temperature in the vicinity of 100°C. Some freeze-dried products require extremely dry closures to avoid desorption of moisture from the closure into the moisture-sensitive powder during storage. This may require drying times of hours, following steam sterilization.

The equipment used for washing large numbers of closures is usually an agitator or horizontal basket-type automatic washing machine. Due to the risk of particulate generation from the abrading action of these machines, some procedures simply call for heating the closures in kettles in detergent solution, followed by prolonged flush rinsing. The final rinse should always be with low-particulate WFI. Figure 26-17 shows an example of a modern closure processor that washes, siliconizes, sterilizes, and transports closures directly to the filling line.

It is also possible to purchase rubber closures already cleaned and lubricated in sterilizable bags supplied by the rubber closure manufacturer.

EQUIPMENT

All equipment should be disassembled as much as possible to provide access to internal structures. Surfaces should be scrubbed thoroughly with a stiff brush, using an effective detergent and paying particular attention to joints, crevices, screw threads, and other structures where debris is apt to collect. Exposure to a stream of clean steam aids in dislodging residues from the walls of stationary tanks, spigots, pipes, and similar

structures. Thorough rinsing with distilled water should follow the cleaning steps.

Due to the inherent variation in manual cleaning, the difficult accessibility of large stationary tanks, and the need to validate the process, computer-controlled systems, usually automated, known as clean-in-place (CIP) systems, have been developed. Such an approach involves designing the system, normally of stainless steel, with smooth, rounded internal surfaces and without crevices. That is, for example, with welded, rather than threaded, connections. Cleaning is accomplished with the scrubbing action of high-pressure spray balls or nozzles delivering hot detergent solution from tanks captive to the system, followed by thorough rinsing with WFI. The system is often extended to allow sterilizing-in-place (SIP), to accomplish sanitizing or sterilizing as well.

Rubber tubing, rubber gaskets, and other rubber parts may be washed in a manner as described for rubber closures. Thorough rinsing of tubing must be done by passing WFI through the tubing lumen. However, due to the relatively porous nature of rubber compounds and the difficulty in removing all traces of chemicals from previous use, it is considered by some inadvisable to reuse rubber or polymeric tubing. Rubber tubing must be left wet when preparing for sterilization by autoclaving.

PRODUCT PREPARATION

The basic principles employed in the compounding of the product are essentially the same as those used, historically, by pharmacists. However, large-scale production requires appropriate adjustments in the processes and their control.

A master formula should be developed and be on file. Each batch formula sheet should be prepared from the master and confirmed for accuracy. All measurements of quantities should be made as accurately as possible and checked by a second qualified person. Frequently, formula documents are generated by computer, and the measurements of quantities of ingredients are computer controlled. Although most liquid preparations are dispensed by volume, they are prepared by weight, since weighing can be performed more accurately than volume measurements, and no consideration needs to be given to the temperature.



Figure 26-17. Examples of Rubber Stopper Preparation Systems that clean, siliconize, and depyrogenate stoppers. (Courtesy of Getinge.)

Care must be taken that equipment is not wet enough to dilute the product significantly or, in the case of anhydrous products, to cause a physical incompatibility. The order of mixing of ingredients may affect the product significantly, particularly those of large volume, where attaining homogeneity requires considerable mixing time. For example, the adjustment of pH by the addition of an acid, even though diluted, may cause excessive local reduction in the pH of the product, so adverse effects are produced before the acid can be dispersed throughout the entire volume of product.

Parenteral dispersions, including colloids, emulsions, and suspensions, provide particular problems. In addition to the problems of achieving and maintaining proper reduction in particle size under aseptic conditions, the dispersion must be kept in a uniform state of suspension throughout the preparative, transfer, and subdividing operations.

Biopharmaceuticals are usually extremely sensitive to many environmental and processing conditions exposed to during production, such as temperature, mixing time and speed, order of addition of formulation components, pH adjustment and control, and contact time with various surfaces, such as filters and tubing. Development studies must include evaluation of manufacturing conditions to minimize adverse effects of the process on the activity of the protein.

Among many causes for protein aggregation are protein particles, resulting from heterogeneous nucleation on foreign micro- or nanoparticles, originating from the manufacturing process (i.e., mixing tanks, process tubing, filter systems, filling machines³⁰ or any other stainless steel, rubber, glass, or plastic surface³¹) and from the container/closure system.³⁰ It is well known that silicone oil, used as a lubricant for rubber closures, on vials, on rubber plungers, in prefilled syringes, and to coat the inner surface of glass syringes and cartridges can also induce protein aggregation.^{32,33} Although switching from silicone-coated containers to plastic containers and using coated rubber closures, rather than siliconized rubber closures, may minimize or eliminate the problems of protein aggregation, other challenges surface, such as unknown leachate potential, sterilization of components, vendor reliability, and cost.

The formulation of a stable product is of paramount importance. Certain aspects of this are mentioned in the discussion of components of the product. (Exhaustive coverage of the topic is not possible within the limits of this text, but further coverage is provided in Chapters 24 [Solutions, Emulsions, Suspensions and Extracts] and 4 [Stability of Pharmaceutical Products]). It should be mentioned here, however, that the thermal sterilization of parenteral products increases the possibility of chemical reactions. Such reactions may progress to completion, during the period of elevated temperature in the autoclave, or be initiated at this time but continue during subsequent storage. The assurance of attaining product stability requires a high order of pharmaceutical knowledge and responsibility.

FILTRATION

After a product has been compounded, it must be filtered, if it is a solution. The primary objective of filtration is to clarify a solution. A further step, removing particulate matter down to 0.2 μm in size, would eliminate micro-organisms and would accomplish cold sterilization. A solution with a high degree of clarity conveys the impression of high quality and purity, desirable characteristics for a parenteral solution.

Filters are thought to function by one or, usually, a combination of: 1) sieving or screening, 2) entrapment or impaction, and 3) electrostatic attraction (Fig. 26-18). When a filter retains particles by sieving, they are retained on the surface of the filter. Entrapment occurs when a particle smaller than the dimensions of the passageway (pore) becomes lodged in a turn or impacted on the surface of the passageway. Electrostatic attraction causes particles opposite in charge to that of the surface of the filter pore to be held or adsorbed to the surface. It should be noted that

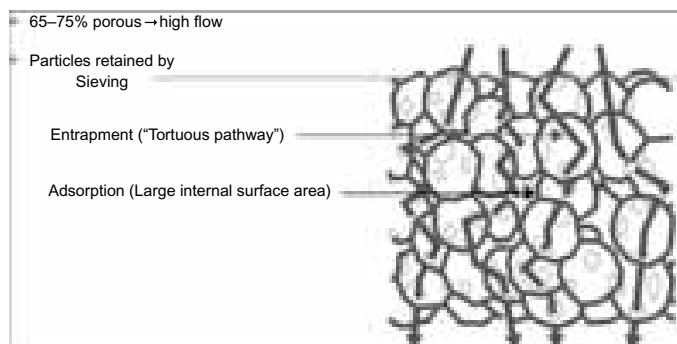


Figure 26-18. Membrane filter characteristics. (Courtesy of EMD Millipore Corporation.)

increasing, prolonging, or varying the force behind the solution may tend to sweep particles initially held by entrapment or electrostatic charge through the pores and into the filtrate.

Membrane filters are used exclusively for parenteral solutions, due to their particle-retention effectiveness, non-shedding property, non-reactivity, and disposable characteristics. However, it should be noted that non-reactivity does not apply in all cases. For example, polypeptide products may show considerable adsorption through some membrane filters, but those composed of polysulfone and polyvinylidene difluoride (PVDF) have been developed to be essentially non-adsorptive for these products. The most common membranes are composed of Cellulose esters, Nylon, Polysulfone, Polycarbonate, PVDF, or Polytetrafluoroethylene (Teflon).

Filters are available as flat membranes or pleated into cylinders (Fig. 26-19) to increase surface area and, thus, flow rate. Each filter in its holder should be tested for integrity before and after use, particularly, if it is being used to eliminate micro-organisms. This integrity test is performed either as the 'bubble-point test' or as the 'diffusion or forward flow' test. The bubble point test is commonly used on smaller filters. As the surface area of filters becomes large, diffusion of air through the water-filled pores tends to obscure the bubble point. Therefore, the diffusion test has been developed as an integrity test for filters with large surface areas. A 'pressure hold test' can also be applied to large surface area filters. The filter manufacturer will recommend the best integrity test for the filter system in question.

These are tests to detect the largest pore or other opening through the membrane. The basic test is performed by gradually raising air pressure on the upstream side of a water-wet filter. The bubble point is the pressure obtained when air bubbles first appear downstream of the filter. The diffusion or forward flow test raises pressure to some point below the known bubble point pressure, then diffusion flow (usually in mL/min) is measured. These pressures are characteristic for each pore size of a filter and are provided by the filter manufacturer. For example, a 0.2- μm cellulose ester filter will bubble at about 50 psig or a diffusive flow rating of no greater than 13 mL/min at a pressure of 40 psig. If the filter is wetted with other liquids, such as a product, the bubble point will differ and must be determined experimentally. If the bubble point is lower than the rated

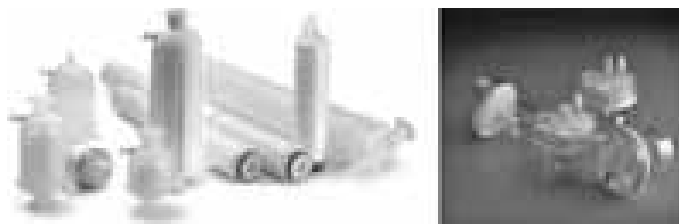


Figure 26-19. Cartridge and disc filters. Courtesy of EMD Millipore Corporation.

pressure, the filter is defective, probably due to a puncture or tear, and should not be used.

Although membrane filters are disposable and, thus, discarded after use, the holders must be cleaned thoroughly between uses. Today, clean, sterile, pretested, disposable assemblies for small, as well as large, volumes of solutions are available commercially. (Other characteristics of these filters, important for a full understanding of their use, are given in Chapter 14 – Separation Methods.)

There have been some reports that that 0.2 μm filters do not remove all possible microbial contamination, necessitating the need to use certain types of 0.1 μm membrane filters. However, most of the parenteral pharmaceutical industry continues to use 0.2 μm filters, although now employing redundant (two 0.2 μm filters side-by-side) filtration systems.

FILLING

During the filling of containers with a product, the most stringent requirements must be exercised to prevent contamination, particularly if the product has been sterilized by filtration and will not be sterilized in the final container. Under the latter conditions, the process is called an ‘aseptic fill’ and is validated with media fills. During the filling operation, the product must be transferred from a bulk container or tank and subdivided into dose containers. This operation exposes the sterile product to the environment, equipment, and manipulative technique of the operators, until it can be sealed in the dose container. Therefore, this operation is carried out with a minimum exposure time, even though maximum protection is provided by filling under a blanket of HEPA-filtered laminar-flow air within the aseptic area.

Most frequently, the compounded product is in the form of a liquid. However, products are also compounded as dispersed systems (e.g., suspensions and emulsions) and as powders. A liquid is more readily subdivided uniformly and introduced into a container having a narrow mouth than is a solid. Mobile liquids are considerably easier to transfer and subdivide than viscous, sticky liquids, which require heavy-duty machinery for rapid production filling.

Although many devices are available for filling containers with liquids, certain characteristics are fundamental to them all. A means is provided for repetitively forcing a measured volume of the liquid through the orifice of a delivery tube introduced into the container. The size of the delivery tube will vary from that of about a 20-gauge hypodermic needle to a tube 1/2 in or more in diameter. The size required is determined by the physical characteristics of the liquid, the desired delivery speed, and the inside diameter of the neck of the container. The tube must enter the neck and deliver the liquid well into the neck to eliminate spillage, allowing sufficient clearance to allow air to leave the container as the liquid enters. The delivery tube should be as large in diameter as possible to reduce resistance and decrease velocity of flow of the liquid. For smaller volumes of liquids, delivery usually is obtained from the stroke of the plunger of a syringe, forcing the liquid through a two-way valve, providing for alternate filling of the syringe and delivery of mobile liquids. For heavy, viscous liquids, a sliding piston valve, the turn of an auger in the neck of a funnel, or the oscillation of a rubber diaphragm may be used. Also, stainless steel syringes are required with viscous liquids, because glass syringes are not strong enough to withstand the high pressures developed during delivery. For large volumes, the quantity delivered is measured in the container by the level of fill in the container, the force required to transfer the liquid being provided by gravity, a pressure pump, or a vacuum pump.

The narrow neck of an ampoule limits the clearance possible between the delivery tube and the inside of the neck. Since a drop of liquid normally hangs at the tip of the delivery tube after a delivery, the neck of an ampoule will be wet as the delivery tube is withdrawn, unless the drop is retracted. Therefore, filling machines should have a mechanism by which this drop can be

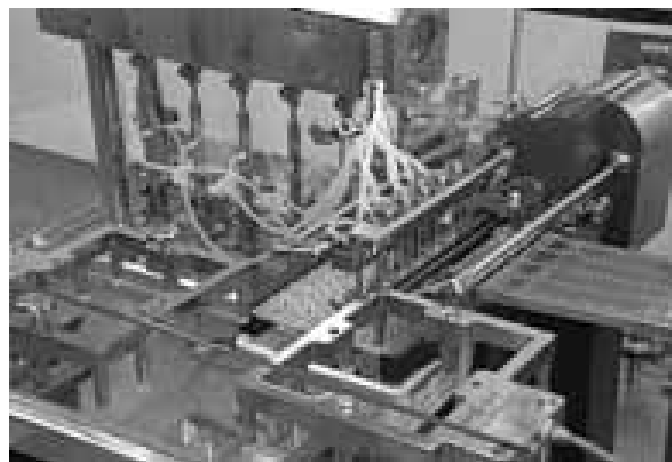


Figure 26-20. Syringe filling machine. (Courtesy Baxter Healthcare Corporation.)

drawn back into the lumen of the tube. Since the liquid will be in intimate contact with the parts of the machine through which it flows, these must be constructed of non-reactive materials, such as borosilicate glass or stainless steel. In addition, they should easily be demountable for cleaning and sterilization.

LIQUIDS

There are three main methods for filling liquids into containers with high accuracy: volumetric filling, time/pressure dosing, and net weight filling. Volumetric filling machines, employing pistons or peristaltic pumps, are most commonly used.

When high-speed filling rates are desired but accuracy and precision must be maintained, multiple filling units are often joined in an electronically coordinated machine (Fig. 26-20 and Fig. 26-21). When the product is sensitive to metals, a peristaltic-pump filler may be used, because the product comes in contact only with silicone rubber tubing. However, this sacrifices filling accuracy.

Time-pressure (or time-gravity) filling machines are gaining popularity in filling sterile liquids. A product tank is connected to the filling system equipped with a pressure sensor. The sensor continuously measures pressure and transmits values to the PLC system controlling the flow of product from tank to filling manifold. Product flow occurs when tubing is mechanically unpinched and stops when tubing is mechanically pinched. The main advantage of time/pressure filling operations is that these filling apparatuses do not contain mechanical moving parts in the product stream. The product is driven by pressure (usually nitrogen) with no pumping mechanism involved. Thus, especially for proteins that are quite sensitive to shear forces, time/pressure filling is preferable.

Most high-speed fillers for large-volume solutions use the bottle as the measuring device, transferring the liquid either by vacuum or by positive pressure from the bulk reservoir to the individual unit containers. Therefore, a high accuracy of fill is not achievable.

The USP requires that each container be filled with a sufficient volume in excess of the labeled volume to ensure withdrawal of the labeled volume and provides a table of suggested fill volumes.

The filling of a small number of containers may be accomplished with a hypodermic syringe and needle, the liquid drawn into the syringe and forced through the needle into the container. An example of such a device that provides greater speed of filling is the Cornwall Pipet (Becton Dickinson). The device has a two-way valve between the syringe and the needle and a means for setting the stroke of the syringe, so the same volume is delivered each time. Clean, sterile, disposable assemblies operating on the same principle have particular usefulness in hospital pharmacy or experimental operations.

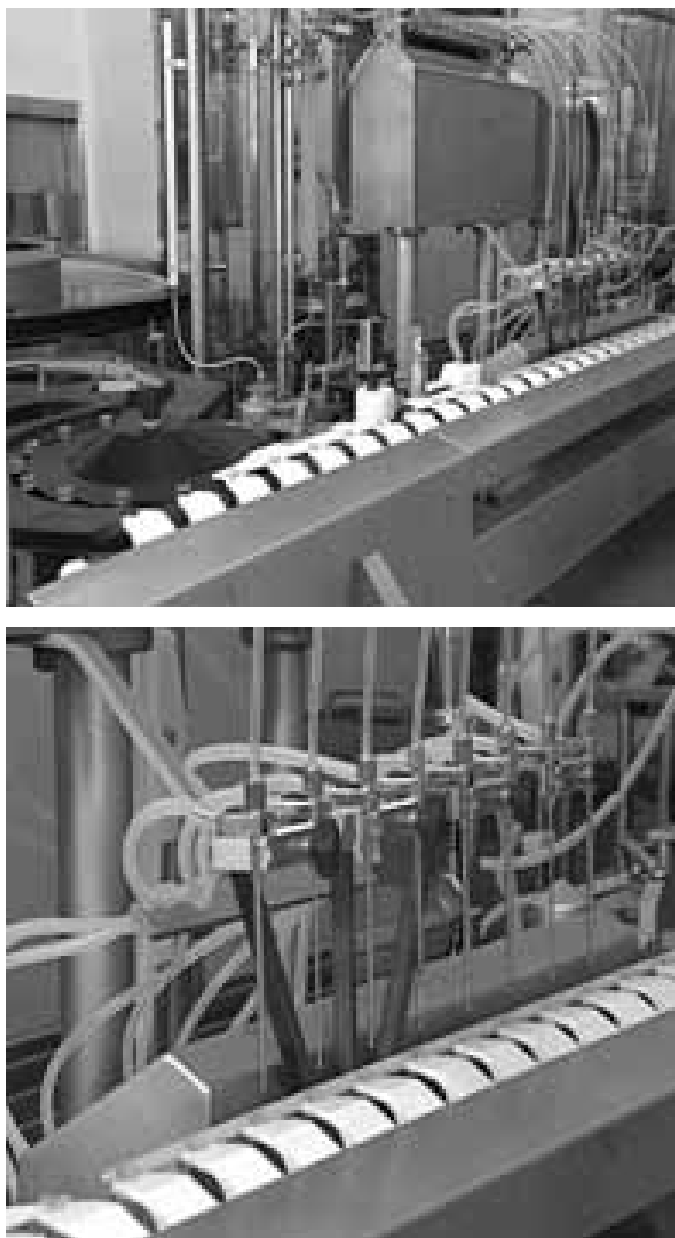


Figure 26-21. Vial filling machine, distant and close-up views. (Courtesy, Baxter Healthcare Corporation.)

SOLIDS

Sterile solids, such as antibiotics, are more difficult to subdivide evenly into containers than are liquids. The rate of flow of solid material is slow and often irregular. Even though a container with a larger-diameter opening is used to facilitate filling, it is difficult to introduce the solid particles, and the risk of spillage is ever-present. The accuracy of the quantity delivered cannot be controlled, as well as with liquids. Due to these factors, the tolerances permitted for the content of such containers must be relatively large.

Some sterile solids are subdivided into containers by individual weighing. A scoop is usually provided to aid in approximating the quantity required, but the quantity filled into the container is finally weighed on a balance. This is a slow process. When the solid is obtainable in a granular form, so it will flow more freely, other methods of filling may be employed. In general, these involve the measurement and delivery of a volume of the granular material that has been calibrated in terms of the weight desired. In the machine shown in Figure 26-22, an adjustable cavity in the rim of a wheel is filled by vacuum and

A

Principle of operation

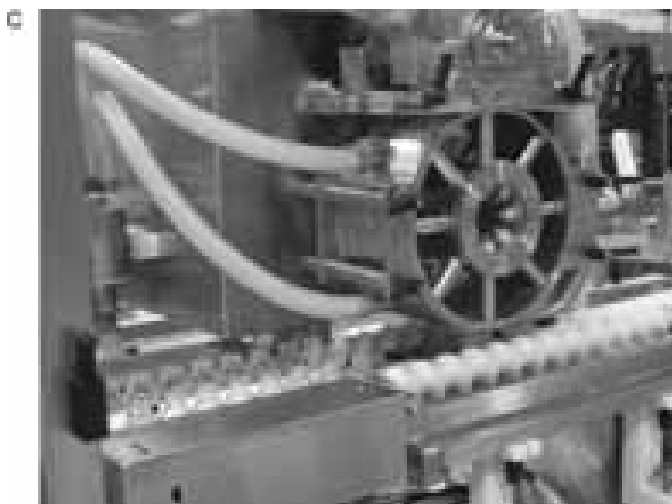
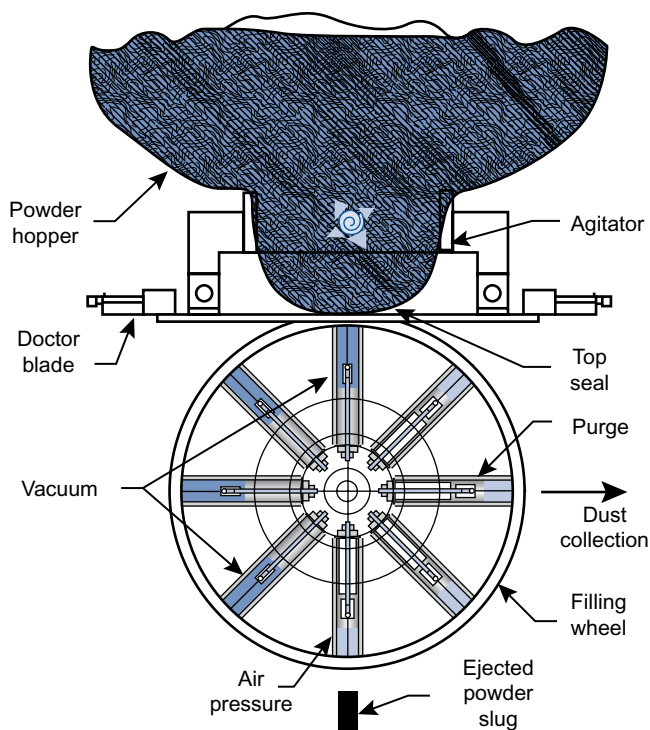


Figure 26-22. Perry Accofil Sterile Powder Filling Machine. A. Principle of operation. B. Filler inside a barrier system. C. Close-up view of filler. (Courtesy of M&O Perry Industries, Inc.)

the contents held by vacuum, until the cavity is inverted over the container. The solid material is then discharged into the container by a puff of sterile air.

SEALING

AMPOULES

Filled containers should be sealed as soon as possible, to prevent the contents from being contaminated by the environment. Ampoules are sealed by melting a portion of the glass neck. Two types of seals are employed normally: tip-seals (bead-seals) or pull-seals.

Tip-seals are made by melting enough glass at the tip of the neck of an ampoule to form a bead and close the opening. These can be made rapidly in a high-temperature gas-oxygen flame. To produce a uniform bead, the ampoule neck must be heated evenly on all sides, such as by burners on opposite sides of stationary ampoules or by rotating the ampoule in a single flame. Care must be taken to properly adjust the flame temperature and the interval of heating to completely close the opening with a bead of glass. Excessive heating results in the expansion of the gases within the ampoule against the soft bead seal, which causes a bubble to form. If the bubble bursts, the ampoule is no longer sealed; if it does not, the wall of the bubble will be thin and fragile. Insufficient heating will leave an open capillary through the center of the bead. An incompletely sealed ampoule is called a 'leaker'.

Pull-seals are made by heating the neck of the ampoule below the tip, leaving enough of the tip for grasping with forceps or other mechanical devices. The ampoule is rotated in the flame from a single burner. When the glass has softened, the tip is grasped firmly and pulled quickly away from the body of the ampoule, which continues to rotate. The small capillary tube, thus, formed is twisted closed. Pull-sealing is slower, but the seals are more secure than tip-sealing. Figure 26-23 shows a machine combining the steps of filling and pull-sealing ampoules.

Ampoules having a wide opening must be sealed by pull-sealing. Fracture of the neck of ampoules during sealing may occur, if wetting of the necks occurs at the time of filling. Also, wet necks increase the frequency of bubble formation and unsightly carbon deposits, if the product is organic.

To prevent decomposition of a product, it is sometimes necessary to displace the air in the space above the product in the ampoule with an inert gas, by introducing a stream of the gas, such as nitrogen or carbon dioxide, during or after filling with the product. Immediately thereafter, the ampoule is sealed, before the gas can diffuse to the outside. This process should be validated to ensure adequate displacement of air by the gas in each container.

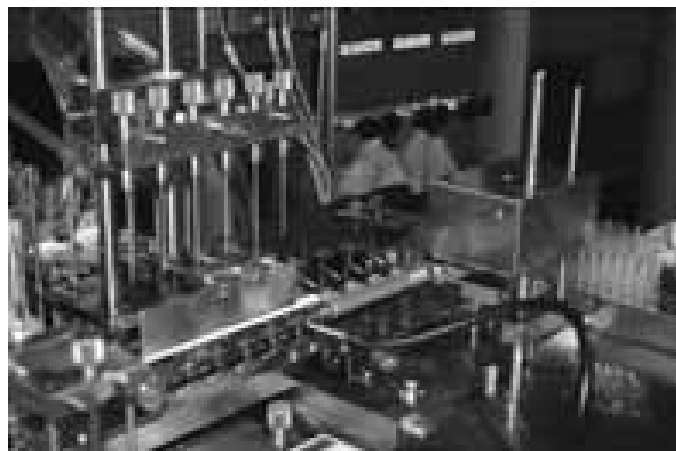


Figure 26-23. FPS2CA Automatic Monoblock Closed Ampule Filling and Sealing Machine. (Courtesy of and with permission from Cozzoli Machine Company.)

VIALS AND BOTTLES

Glass or plastic vials and bottles are sealed by closing the opening with a rubber closure (stopper). This must be accomplished as rapidly as possible after filling and with reasoned care, to prevent contamination of the contents. The large opening makes the introduction of contamination much easier than with ampoules. Therefore, during the critical exposure, the open containers should be protected from the ingress of contamination, preferably with a blanket of HEPA-filtered laminar airflow.

The closure must fit the mouth of the container snugly enough, so its elasticity seals rigid to slight irregularities in the lip and neck of the container. However, it must not fit so snugly that it is difficult to introduce into the neck of the container. Preferably, closures are inserted mechanically, using an automated process, especially with high-speed processing. To reduce friction, so the closure may slide more easily through a chute and into the container opening, the closure surfaces are halogenated or treated with silicone. When the closure is positioned at the insertion site, it is pushed mechanically into the container opening (Fig. 26-24). When small lots are encountered, manual stoppering with forceps may be used, but such a process poses greater risk of introducing contamination than automated processes. This is a good test for evaluation of operator aseptic techniques, but not recommended for any product filling and stoppering.

Container-closure integrity testing has become a major focus for the industry, due to emphasis by regulatory agencies. Container-closure integrity measures the ability of the seal between the glass or plastic container opening and the rubber closure to remain tight and fit and to resist any ingress of microbial contamination during product shelf life. Container-closure integrity test requirements are covered in USP <1207>, and the various test methods are described by Guazzo.³⁴

Rubber closures are held in place by means of aluminum caps. The caps cover the closure, crimped under the lip of the vial or bottle to hold them in place. The closure cannot be removed without destroying the aluminum cap; it is tamperproof. Therefore, an intact aluminum cap is proof that the closure has not been removed intentionally or unintentionally. Such confirmation is necessary to ensure the integrity of the contents, as to sterility and other aspects of quality.

The aluminum caps are designed so the outer layer of double-layered caps, or the center of single-layered caps, can be removed to expose the center of the rubber closure, without disturbing the band that holds the closure in the container. Rubber closures for use with intravenous administration sets often have a permanent hole through the closure. In such cases, a thin rubber disk, overlaid with a solid aluminum disk, is

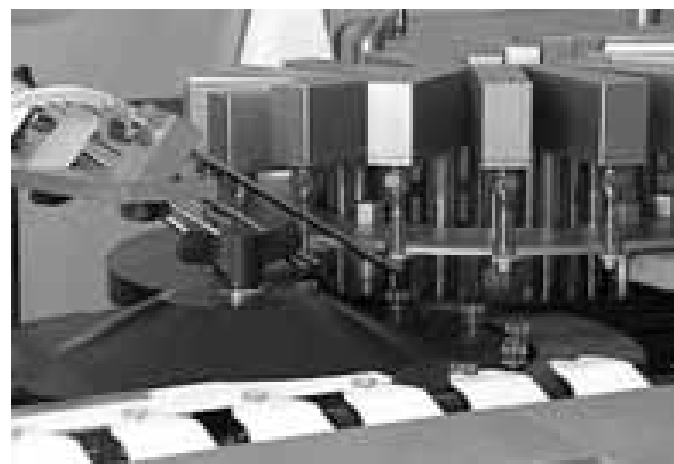


Figure 26-24. Mechanical device for inserting rubber closures in vials. (Courtesy of Baxter Healthcare Corporations.)



Figure 26-25. Examples of aluminum-plastic seals. A. Flip-Off Seals: Aluminum shell with a removable plastic button in order to access stopper surface. B. Flip-Tear Seals: Aluminum shell is completely removed from container by flipping off the plastic button—allows stopper removal. (Courtesy of West Pharmaceutical Services.) (Flip-Off is a registered trademark of West Pharmaceutical Services in the United States and other jurisdictions.)



Figure 26-26. CM200 Continuous Motion Crimping Machine. (Courtesy of and with permission from Cozzoli Machine Company.)

placed between an inner and outer aluminum cap, thereby providing a seal of the hole through the closure.

Single-layered aluminum caps may be applied using a hand crimper. Double- or triple-layered caps (Fig. 26-25) require greater force for crimping; therefore, heavy-duty mechanical crimpers (Fig. 26-26) are required.

STERILIZATION

Whenever possible, the parenteral product should be sterilized, after being sealed in its final container (terminal sterilization) and within as short a time as possible after filling and sealing are completed. Since this usually involves a thermal process, although there is a trend in applying radiation sterilization to finished products, due consideration must be given to the effect of the elevated temperature, upon the stability of the product. Many products, both pharmaceutical and biological, are affected adversely by elevated temperatures required for thermal sterilization. Heat-labile products must, therefore, be sterilized by a non-thermal method, usually by filtration through bacteria-retaining filters. Subsequently, all operations must be carried out in an aseptic manner, so contamination is not introduced into the filtrate. Colloids, oleaginous solutions, suspensions, and emulsions that are thermolabile may require a process in which each component is sterilized separately and the product is formulated and processed under aseptic conditions.

The performance of an aseptic process is challenging, but technical advances in aseptic processing, including improved automation, use of isolator systems, formulations to include antimicrobial effects, and combinations of limited sterilization with aseptic processing, have decreased the risk of contamination. Therefore, the successes realized should encourage continued efforts to improve the assurance of sterility achievable with aseptic processing. The importance of this is that, for many drug solutions and essentially all biopharmaceutical products, aseptic processing is the only method that can be considered for preparing a sterile product.

Interaction among environmental conditions, the constituents in the closure, and the product may result in undesirable closure changes, such as increased brittleness or stickiness, which may cause loss of container-closure seal integrity. Thus, shelf life integrity is an important consideration in closure selection and evaluation.

The assessment of aseptic-processing performance is based on the contamination rate resulting from periodic process simulations using media-filling, instead of product-filling of containers. A contamination rate no greater than 0.1% at 95% confidence has been considered indicative of satisfactory performance in the industry. However, with current advances in aseptic processing capabilities, lower contamination rates may be achievable.

Radiation sterilization, as mentioned, is gaining momentum as an alternative terminal sterilization method. There has been limited understanding of the molecular transformations that may occur in drug molecules and excipients under exposure to the high-energy gamma radiation levels of the process. However, lower energy beta particle (electron beam) radiation has seen some success. Significant research must still be accomplished, before radiation sterilization is used as a terminal sterilization process. The use of radiation for the sterilization of materials, such as plastic medical devices, is well established.

Dry-heat sterilization may be employed for a few dry solids not affected adversely by the high temperatures and for the relatively long heating period required. This method is applied most effectively to the sterilization of glassware and metalware. After sterilization, the equipment will be sterile, dry, and, if the sterilization period is long enough, pyrogen-free.

Saturated steam under pressure (autoclaving) is the most commonly used and the most effective method for the sterilization of aqueous liquids or substances that can be reached or penetrated by steam. A survival probability of at least 10^{-6} is readily achievable with terminal autoclaving of a thermally stable product.



Figure 26-27. An example of a modern autoclave for sterilization. (Courtesy of Getinge.)

However, it needs noted that, for terminal sterilization, the assurance of sterility is based on an evaluation of the lethality of the process (i.e., of the probable number of viable microorganisms remaining in product units). However, for aseptic processing, where the components used have been sterilized separately by validated processes and aseptically put together, the level of sterility assurance is based on an evaluation of the probable number of product units contaminated during the process.

Figure 26-27 shows an example of a modern autoclave for sterilization. Since the temperature employed in an autoclave is lower than that for dry-heat sterilization, equipment made of materials, such as rubber and polypropylene, may be sterilized, if the time and temperature are carefully controlled. As mentioned previously, some injections are affected adversely by the elevated temperature required for autoclaving. For some products, such as Dextrose Injection, a shortened cycle, using an autoclave designed to permit a rapid temperature rise and rapid cooling with water spray or other cooling methods, makes it possible to use this method. It is ineffective in anhydrous conditions, such as within a sealed ampoule containing a dry solid or an anhydrous oil. Other products that will not withstand autoclaving temperatures may withstand marginal thermal methods, such as tyndallization or pasteurization (e.g., 10 to 12 hours at 60°C). These methods may be rendered more effective for some injections, by the inclusion of a bacteriostatic agent in the product.

Articles to be sterilized must be properly wrapped or placed in suitable containers to permit penetration of sterilants and provide protection from contamination after sterilization. Sheets or bags made of special steam-penetrating paper or polymeric materials are available for this purpose. Further, containers or bags impervious to steam can be equipped with a microbe-excluding vent filter to permit adequate steam penetration and air exit. Multiple wrapping permits sequential removal of outer layers, as articles are transferred from zones of lower to higher environmental quality. The openings of equipment subjected to dry-heat sterilization are often covered with metal or glass covers. Laboratories often used silver-aluminum foil for covering glassware used for endotoxin testing. Wrapping materials commonly used for steam sterilization may be combustible or otherwise become degraded under dry-heat sterilization conditions.

The effectiveness of any sterilization technique must be proved (validated), before it is employed in practice. Since the goal of sterilization is to kill micro-organisms, the ideal indicator to prove the effectiveness of the process is a resistant form of an appropriate micro-organism, normally resistant spores (a biological indicator, or BI). Therefore, during validation of a sterilization process, BIs of known resistance and numbers are used in association with

physical-parameter indicators, such as recording thermocouples. Once the lethality of the process is established in association with the physical measurements, the physical measurements can be used for subsequent monitoring of in-use processes without the BIs. Eliminating the use of BIs in direct association with human-use products is appropriate, due to the ever-present risk of an undetected, inadvertent contamination of a product or the environment. The number of spores and their resistance in BIs used for validation studies must be accurately known or determined. Additionally, the manner in which BIs are used in validation is critical and must be controlled carefully.

In addition to the data printout from thermocouples, sometimes other physical indicators are used, such as color-change and melting indicators, to give visual indication that a package or truckload has been subjected to a sterilization process. Such evidence can become a part of the batch record to confirm that sterilization was accomplished.

Further details concerning methods of sterilization and their application can be found in Chapter 25 (Sterilization Processes and Sterility Assurance). In addition, the USP provides suggestions concerning the sterilization of injections and related materials.

FREEZE-DRYING (LYOPHILIZATION)

Many parenteral drugs, particularly biopharmaceuticals, are too unstable in solution to be available as ready-to-use liquid dosage forms. Such drugs can still be filled as solutions, placed in a chamber, where the combined effects of freezing and drying under low pressure will remove the solvent and residual moisture from the solute components, resulting in a dry powder that has sufficient long term stability. The process of freeze-drying has taken on greater prominence in the parenteral industry, due to the advent of recombinant DNA technology. Proteins and peptides must be freeze-dried for clinical and commercial use. There are other technologies available to produce sterile dry powder drug products besides freeze-drying, such as sterile crystallization or spray-drying and powder filling. However, freeze-drying is the most common unit process for manufacturing drug products too unstable to be marketed as solutions.

The term 'lyophilization' describes a process to produce a product that 'loves the dry state.' However, this term does not include the freezing process. Therefore, although lyophilization and freeze-drying are used interchangeably, freeze-drying is a more descriptive term. Equipment used to freeze-dry products are called freeze-dryers or lyophilizers.

Table 26-7 lists the advantages, features, and disadvantages of freeze-drying.

Freeze-drying, essentially, consists of:

- Freezing stage— Freezing the product solution at a temperature below its eutectic (crystalline) or glass transition temperature.
- Primary drying stage— Removing the solvent (ice) from the product, by evacuating the chamber, usually below 0.1 torr (100 μ m Hg), and subliming the ice onto a cold, condensing surface at a temperature below that of the product, the condensing surface being within the chamber or in a connecting chamber. During primary drying, the temperature of the product must remain slightly below its critical temperature, called 'collapse temperature.' Collapse temperature is best measured by visual observation using a freeze-dry microscope that simulates the freeze-drying process. Collapse temperature is similar to the eutectic or glass transition temperature of the product.
- Secondary drying stage— Removing bound water from solute(s) to a level that assures long term stability of the product. This is accomplished by introducing heat to the product under controlled conditions, thereby providing additional energy to the product to remove adsorbed water. The temperature for secondary drying should be as high as possible, without causing any chemical degradation of the active ingredient. For small molecules,

Table 26-7. Advantages and Disadvantages of Freeze-Drying and Desirable Characteristics of the Finished Freeze-Dried Dosage Form

Advantages of Freeze-Dried Products	
1.	Product is stored in dry state-few stability problems
2.	Product is dried without elevated temperatures
3.	Good for oxygen and/or air-sensitive drugs
4.	Rapid reconstitution time
5.	Constituents of the dried material remain homogeneously dispersed
6.	Product is process in the liquid form
7.	Sterility of product can be achieved and maintained
Disadvantages of Freeze-Dried Products	
1.	Volatile compounds may be removed by high vacuum
2.	Single most expensive unit operation
3.	Stability problems associated with individual drugs
4.	Some issues associated with sterilization and sterility assurance of the dryer chamber and aseptic loading of vials into the chamber
Desired Characteristics of Freeze-Dried Products	
•	Intact cake
•	Sufficient strength
•	Uniform color
•	Sufficiently dry
•	Sufficiently porous
•	Sterile
•	Free of pyrogens
•	Free of particulates
•	Chemically stable

the highest secondary drying temperature used is 40°C, whereas for proteins it is no more than 30°C.

A small-scale lyophilization system and its functional components is shown in Figure 26-28. The product may be frozen on the shelf in the chamber by circulating refrigerant (usually silicone) from the compressor through pipes within the shelf. After freezing is complete, which may require several hours, the chamber and condenser are evacuated by the vacuum pump, the condenser surface having been chilled previously by circulating refrigerant from the large compressor.

Heat then is introduced from the shelf to the product under graded control by electric resistance coils or by circulating

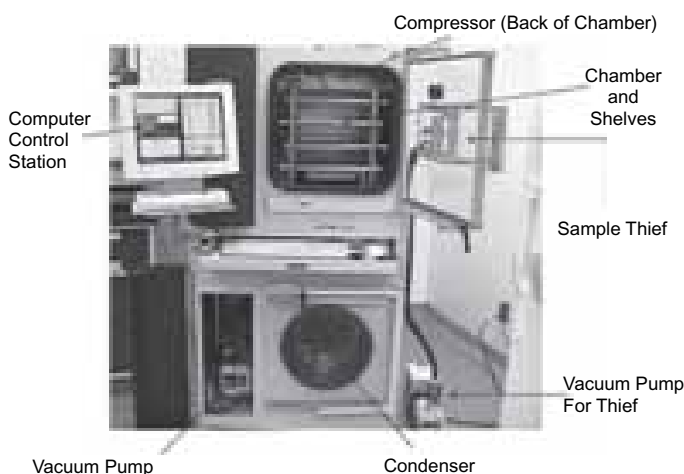


Figure 26-28. Example of a laboratory freeze-dryer. (Courtesy of Baxter Healthcare Corporation.)

silicone or glycol. Heat transfer proceeds from the shelf into the product vial and mass transfer (ice) proceeds from the product vial by sublimation through the chamber and onto the condenser. The process continues, until the product is dry (usually 1% or less moisture, except for some proteins that require a minimum amount of water for conformational stability), leaving a sponge-like matrix of the solids originally present in the product, the input of heat being controlled so as not to degrade the product.

For most pharmaceuticals and biologicals, the liquid product is sterilized by filtration before being filled into the dosage container aseptically. The containers must remain open during the drying process to allow water vapor to escape; therefore, they must be protected from contamination during transfer from the filling area to the freeze-drying chamber, while in the freeze-drying chamber and at the end of the drying process until sealed. Automated loading and unloading of product to and from the freeze-dryer shelves is now state-of-the-art, where partially open vials are always under the auspices of Class 100 air and human intervention is eliminated.

Freeze-dryers are equipped with hydraulic or pneumatic internal-stoppering devices designed to push slotted rubber closures into the vials to be sealed while the chamber is still evacuated, the closures having been partially inserted immediately after filling, so the slots were open to the outside. If internal stoppering is not available or containers, such as ampoules, are used, filtered dry air or nitrogen should be introduced into the chamber at the end of the process to establish atmospheric pressure.

Table 26-8 provides some guidance on a typical formulation approach and initial cycle chosen to freeze-dry a typical product.

Table 26-8. Practical Aspects of Freeze-Drying

- Have appropriate analytical tools and methods in place for formulation characterization and stability studies
- Depend on literature, previous experience (if none, use consultants), and what is known about the active ingredient, design and develop initial formulations and conduct preliminary stability and compatibility studies
- Initial formulations should use commonly-known excipients used in freeze-drying
 - that produce acceptable cakes with rapid reconstitution times
 - that have known minimal collapse temperatures
 - that provide the desired finished product with respect to nature of the final solid (crystalline or amorphous)
- Solids content should be between 5% and 30% with a target of 10% to 15%
- Should have several initial formulations to evaluate and compare. Usually know the qualitative, but not quantitative composition of additives until after initial comparative stability studies have been conducted
- Determine the maximum allowable temperature permitted during freezing and primary drying
 - Know eutectic, glass transition, and/or collapse temperatures, as appropriate
- Select the appropriate size of vial and product fill volume
- Select the appropriate rubber closure
 - Low water vapor transmission
 - No absorption of oil vapor
 - Top design minimizes sticking to shelf during/after stoppering
- Determine appropriate processing parameters

(continued)

• Rate of freezing • Set point temperatures during all three phases • Need for annealing • Pressure during primary drying • Pressure during secondary drying • Stopper seating conditions (e.g., vacuum or gas)	
• Optimize formulation and process based on stability information during and after freeze-drying and after storage in dry state	
• Use a sample thief attachment for laboratory dryers to remove samples during the freeze-dry cycle in order to measure moisture, potency, or other parameters. Provides information for final selection of type and amount of stabilizer(s), if needed, and the cycle parameters necessary to provide an acceptable final moisture level in product	
• Typical freeze-dry formulation components	
Buffers:	Phosphate, citrate, acetate
Stabilizers:	Sucrose, trehalose, glycine
Bulking agents:	Mannitol, lactose
Collapse temperature modifiers:	Polymers, sugars
• Typical freeze-dry cycle (without knowing where to start)	
• Freezing phase • After loading, cool to 5°C • Decrease shelf temperature to -40°C • Hold for 2 hours • Primary drying phase • Must know collapse temperature (T_c) • Set shelf temperature approximately 20°C above T_c but making sure product temperature is 5°C below T_c • Maintain chamber pressure at 10% to 30% of vapor pressure of ice at the primary drying temperature (usually 100 to 200 microns) • Use temperature probes, pressure rise test, or dewpoint measurement to determine end of primary drying • Secondary drying • Use moderate to high vacuum (typically 100 microns) • Adjust shelf temperature to 25°C to 30°C for proteins; 35°C to 40°C for non proteins and hold for at least 4 hours • Adjust shelf temperature to 25°C or 5°C prior to stoppering, neutralizing and unloading	

FACTORS AFFECTING THE PROCESS RATE

From the diagram in Figure 26-29, it can be seen that the direction of heat and mass transfer causes the top of the product to dry first with drying proceeding downward to the bottom of the vial. Therefore, as drying proceeds, there exists a three-component or layer system in each vial—the upper dry product, the middle sublimation front, and the lower frozen liquid product. As the dried layer increases, it becomes a greater barrier or the source of greatest resistance to the transfer of mass out of the vials. This points out the importance of vial dimensions and volume of product per vial on the efficiency of the freeze-drying process. If large volumes of solution must be processed, the surface area relative to the depth may be increased, utilizing larger vials or by using such devices as freezing the container in a slanted position to increase the surface area.

The actual driving force for the process is the vapor pressure differential between the vapor at the surface where drying of the product is occurring (the drying boundary) and at the surface of the ice on the condenser. The latter is determined by the temperature of the condenser, as modified by the insulating effect of the accumulated ice. The former is determined by a number of factors, including:

1. The rate of heat conduction through the container and the frozen material, both relatively poor thermal conductors, to the drying boundary, while maintaining all of the product below its eutectic temperature.
2. The impeding effect of the increasing depth of dried, porous product above the drying boundary.
3. The temperature and heat capacity of the shelf itself.

The passageways between the product surface and the condenser surface must be wide open and direct for effective operation. The condensing surfaces in large freeze-dryers may be in the same chamber as the product or located in a separate chamber connected by a duct to the drying chamber. Evacuation of the system is necessary to reduce the impeding effect that collisions with air molecules would have on the passage of water molecules. However, the residual pressure in the system must be greater than the vapor pressure of the ice on the condenser, or the ice will be vaporized and pulled into the pump, an event detrimental to most pumps.

The amount of solids in the product, the ice crystal size, and their thermal conductance affect the rate of drying. The more solids present, the more impediment will be provided to the escape of the water vapor. The degree of supercooling (i.e., how much lower the product temperature goes below its equilibrium

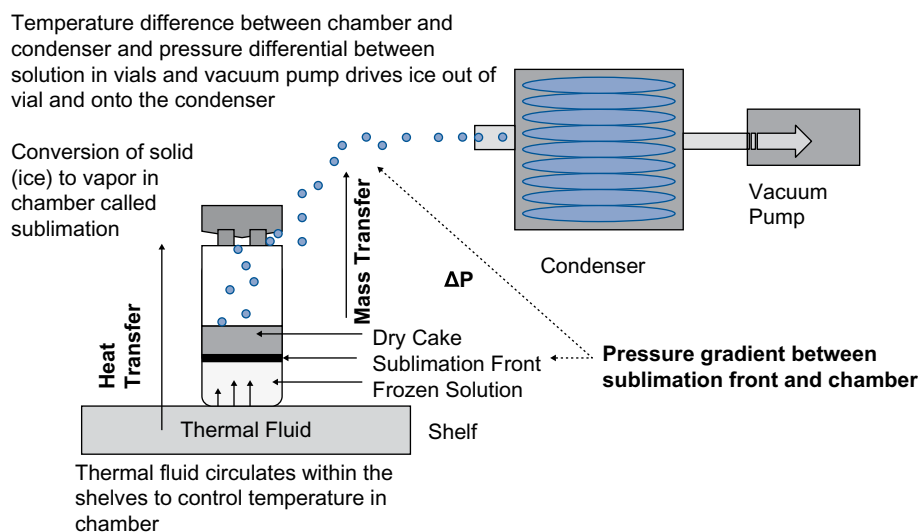


Figure 26-29. Schematic of heat and mass transfer in the freeze dryer.

freezing point before ice crystals first form) and the rate of ice crystallization define the freezing process and efficiency of primary drying. The larger the size of ice crystals formed, usually as a result of slow freezing, the larger the pore sizes are when the ice sublimates and, consequently, the faster the rate of drying. A high degree of supercooling produces a large number of small ice crystals, a small pore size when the ice sublimates in the dried layer, and a greater resistance to water vapor transport during primary drying. The poorer the thermal conducting properties of the solids in the product, the slower the rate of heat transfer through the frozen material to the drying boundary.

The rate of drying is slow, most often requiring 24 hours or longer for completion. The actual time required, the rate of heat input, and the product temperatures used must be determined for each product and then reproduced carefully with successive processes.

FACTORS AFFECTING FORMULATION

The active constituent of many pharmaceutical products is present in such a small quantity that, if freeze-dried alone, its presence would be hard to detect visually. In fact, the solids content of the original product, ideally, should be between 5% and 30%. Therefore, excipients are often added to increase the amount of solids. Such excipients are called 'bulking agents,' the most commonly used bulking agent in freeze-dried formulations is mannitol. However, most freeze-dried formulations must contain other excipients, due to the need to buffer the product and/or to protect the active ingredient from the adverse effects of freezing and/or drying. Thus, buffering agents, such as sodium or potassium phosphate, sodium acetate, and sodium citrate, are commonly used in freeze-dried formulations. Sucrose, trehalose, dextran, and amino acids, such as glycine, are commonly used lyoprotectants.

Each of these substances contributes to the appearance characteristics of the plug, such as whether dull and spongy or sparkling and crystalline, firm or friable, expanded or shrunken, and uniform or striated. Therefore, the formulation of a product to be freeze-dried must include consideration not only of the nature and stability characteristics required during the liquid state, both freshly prepared and when reconstituted before use, but also the characteristics desired in the dried plug.

MODIFICATIONS IN THE PROCESS AND EQUIPMENT

In some instances, a product may be frozen in a bulk container or in trays, rather than in the final container, and then handled as a bulk solid. Such a state requires a continuation of aseptic processing conditions as long as the product is exposed to the environment.

When large quantities of material are processed, it may be desirable to use ejection pumps in the equipment system. These draw the vapor into the pump and eject it to the outside, thereby eliminating the need for a condensing surface. Such pumps are expensive and usually practical only in large installations.

Available freeze-dryers range in size from small laboratory units (Fig. 26-28), with shelf surface areas of approximately 5 square feet to large industrial models with shelf surface areas of several hundred square feet. Their selection requires consideration of such factors as:

- Tray area required,
- Volume of water to be removed,
- How the chamber will be sterilized,
- Whether internal stoppering is required,
- Whether separate freezers will be used for initial freezing and condensation of the product, and
- Degree of automatic operation desired.

Other factors involved in the selection and use of equipment are considered in the literature.³⁵

Freeze-drying is now used for research in the preservation of human tissue and is finding increasing application in the food industry. Most biopharmaceuticals require lyophilization

to stabilize their protein content effectively. Therefore, many newer developments in the lyophilization process focus on the requirements of this new class of drug products.

QUALITY ASSURANCE AND CONTROL

The importance of undertaking every possible means to ensure the quality of the finished product cannot be overemphasized. Every component and step of the manufacturing process must be subjected to intense scrutiny to be confident quality is attained in the finished product. The responsibility for achieving this quality is divided appropriately in concept and practice into Quality Assurance (QA) and Quality Control (QC). QA relates to the studies made and the plans developed for ensuring quality of a product prospectively, with a final confirmation of achievement. QC embodies the carrying out of these plans during production and includes all of the tests and evaluations performed to ensure quality exists in a specific lot of product.

The principles for achieving quality are basically the same for the manufacture of any pharmaceutical. These are discussed in Chapter 3 (Quality Assurance and Control) During the discussion of preparation of injections in this chapter, mention was made of numerous quality requirements for components and manufacturing processes. Here, only selected tests characteristically required before a finished parenteral product is released are discussed briefly, including sterility, pyrogen, and particulate tests.

STERILITY TEST

All lots of injectables in their final containers must be tested for sterility, except products that are allowed to apply parametric release (i.e., terminally sterilized by a well-defined, fully validated sterilization process, has a sterility assurance level sufficient to omit the sterility test for release). The USP prescribes the requirements for this test for official injections. The FDA uses these requirements as a guide for testing official sterile products. The primary official test is performed by means of filtration, but direct transfer is used if membrane filtration is unsuitable. To give greater assurance that viable micro-organisms will grow, if present, the USP requires that all lots of culture media be tested for their growth-promotion capabilities. However, it must be recognized that the reliability of both test methods has the inherent limitations typical of microbial recovery tests. Therefore, it should be noted that this test is not intended as a thoroughly evaluative test for a product subjected to a sterilization method of unknown effectiveness. It is intended as a check test on the probability that a previously validated sterilization procedure has been repeated or to give assurance of its continued effectiveness. A discussion of sterility testing is given in Chapter 25 (Sterilization Processes and Sterility Assurance).

In the event of a sterility-test failure, the immediate issue concerns whether the growth observed came from viable micro-organisms in the product (true contamination) or from adventitious contamination during the testing (a false positive). The USP does not permit a retest, unless specific evidence is discovered to suggest contamination occurred during the test. Therefore, a thorough investigation must be launched to support the justification for performing the retest and assessing the validity of the retest results relative to release of the lot of product.

It should be noted that a 'lot,' with respect to sterility testing, is that group of product containers that has been subjected to the same sterilization procedure. For containers of a product that have been sterilized by autoclaving, for example, a lot would constitute those processed in a particular sterilizer cycle. For an aseptic filling operation, a lot would constitute all of those product containers filled during a period in which there was no change in the filling assembly or equipment and which is no longer than one working day or shift.

As stated previously, isolator technology has been applied to significantly reduce the incidence of false positives in the conductance of the sterility test. Figure 26-10 shows an example of a sterility testing isolator. Validation of isolator systems for sterility testing is described in USP <1208>.

PYROGEN TEST

The USP evaluates the presence of pyrogens in parenteral preparations by a qualitative fever response test in rabbits, the Pyrogen Test (Section <151>), and by the Bacterial Endotoxins Test (Section <85>). These two USP tests are described in Chapter 25 (Sterilization Processes and Sterility Assurance). Rabbits are used as test animals in Section <151>, because they show a physiological response to pyrogenic substances similar to that of man. Although a minimum pyrogenic dose (MPD), the amount just sufficient to cause a positive USP Pyrogen Test response, may produce uncertain test results, a content equal to a few times the MPD will leave no uncertainty. Therefore, the test is valid and has continued in use, since introduced by Seibert in 1923. It should be understood that not all injections may be subjected to the rabbit test, since the medicinal agent may have a physiological effect on the test animal such that any fever response would be masked.

The Bacterial Endotoxins Test (BET) is an *in vitro* test based on the formation of a gel or the development of color in the presence of bacterial endotoxins and the lysate of the amoebocytes of the horseshoe crab (*Limulus polyphemus*). The *Limulus* Amoebocyte Lysate (LAL) test, as it also is called, is a biochemical test performed in a test tube and is simpler, more rapid, and of greater sensitivity than the rabbit test. Figure 26-30 shows an example of a positive endotoxin test result in a test tube. Although it detects only the endotoxic pyrogens of gram-negative bacteria, these are the most prominent environmental microbial contaminants likely to invade sterile products. The test has been automated and can determine the quantitative amount of endotoxin in a sample. This test has enabled endotoxin limits to be established on finished products and bulk drug substances and excipients.

To provide standardization for the test, the USP has established a reference standard endotoxin (RSE) against which lots of the lysate is standardized. Thus, the sensitivity of the lysate is given in terms of endotoxin units (EU). Most USP injections have now been given limits in terms of EUs (e.g., Bacteriostatic Sodium Chloride Injection, 1.0 EU/mL), thus, indicating an increasing priority for the BET in testing for the presence of endotoxin in parenteral products and in medical devices.



Figure 26-30. Example of positive (left tube) endotoxin test.

PARTICULATE MATTER EVALUATION

Particulate matter in parenteral solutions has been recognized as unacceptable, since the user could be expected to conclude that the presence of visible dirt would suggest that the product is of inferior quality. Today, it is recognized that the presence of particles in solution, particularly if injected intravenously, can be harmful. Although data defining the extent of risk and the effects produced are still limited, it has been shown that particles of lint, rubber, insoluble chemicals, and other foreign matter can produce emboli in the vital organs of animals and man. Further, it has been shown that the development of infusion phlebitis may be related to the presence of particulate matter in intravenous fluids.

The particle size of particular concern has not been clearly delineated, but it has been suggested that, since erythrocytes have a diameter of approximately 4.5 μm , particles of more than 5 μm should be the basis for evaluation. This is a considerably smaller particle than can be seen with the unaided eye; approximately 50 μm is the lower limit, unless the Tyndall effect is used, whereby particles as small as 10 μm can be seen by the light scattered from them.

The USP specifies that good manufacturing practice requires each final container of an injection be subjected individually to a visual inspection and containers in which visible particles can be seen should be discarded. This 100% inspection of a lot of product is designed to prevent the distribution and use of parenterals that contain particulate matter. Therefore, all of the product units from a production line are currently being inspected individually, by human inspectors, under a good light, baffled against reflection into the eye, and against a black-and-white background. This inspection is subject to the limitation of the size of particles that can be seen, the variation of visual acuity from inspector to inspector, their emotional state, eye strain, fatigue, and other personal factors that will affect what is seen. However, it does provide a means for eliminating the few units that normally contain visible particles. Automated inspection machines are increasingly being used today.

The assessment of the level of particulate matter below the visible size of about 50 μm has become an increasingly used QC indicator of process cleanliness in the manufacture of injections. The tests used, however, are destructive of container units. Therefore, they are performed on appropriately selected samples of products. Further, all of these methods require very stringent, ultraclean preparation techniques to ensure accuracy in the counting and sizing of particles only in the product, rather than those that may have been introduced inadvertently during the sample preparation or the testing procedure.

The USP has identified two test methods in <788>, *Particulate Matter in Injections*. All LVIs for single-dose infusion and those SVIs for which the monograph specifies a limit (primarily those commonly added to infusion solutions) are subject to the specified limits given in Table 26-9. The first test used is the light obscuration test, which uses an electronic instrument designed to count and measure the size of particles by means of a shadow cast by the particle as it passes through a high-intensity light beam. If the injection formulation is not a clear, colorless solution (e.g., an emulsion) or it exceeds the limits specified for the light obscuration test, it is to be subjected to the microscopic count test. The latter method consists of filtering a measured sample of solution through a membrane filter under ultraclean conditions and then counting the particles on the surface of the filter, using a microscope and oblique light at 100 $\times$ magnification. The time requirements for performing the latter test are very long. These standards are readily met in the United States today by the manufacturers of LVIs and the specified SVIs.

Whether or not these standards are realistic, toxicologically has not been established; rather, the objective of the compendium is to establish specification limits that encourage the preparation of clean parenteral solutions, particularly those to be given intravenously.

Table 26-9. Subvisible Particulate Matter Limits in Injectable Products

Compendia	LVI/SVI	Method	≥10µm	≥25µm
USP	LVI	Light Blockage Microscope	25 part/mL 12 part/mL	3 part/mL 2 part/mL
USP	SVI	Light Blockage Microscope	6000 part/contain. 3000 part/contain.	600 part/contain. 300 part/contain.
EP	LVI SVI Soln SVI Powder	light Blockage light Blockage light Blockage	25 part/mL 6000 part/contain. 10000 part/contain.	3 part/mL 600 part/contain. 1000 part/contain.
BP	LVP	Coutler Counter Light Blockage	1000 part/mL≥2µm 500 part/mL≥2µm	100 part/mL≥5µm 80 part/mL≥5µm
JP	LVP	Microscope	20 part/mL	2 part/mL

It also should be realized that administration sets and the techniques used for preparing and administering intravenous infusion fluids may introduce substantial amounts of particulate matter into an otherwise clean solution. Therefore, the pharmaceutical manufacturer, the administration set manufacturer, the pharmacist, the nurse, and the physician must share responsibility for making sure the patient receives a clean intravenous injection.

CONTAINER/CLOSURE INTEGRITY TEST

Ampoules that have been sealed by fusion must be subjected to a test to determine whether or not a passageway remains to the outside; if so, all or a part of the contents may leak to the outside and spoil the package, or micro-organisms or other contaminants may enter. Changes in temperature during storage cause expansion and contraction of the ampoule and contents, and will accentuate interchange, if a passageway exists, even if microscopic in size.

This test is usually performed by producing a negative pressure within an incompletely sealed ampoule, while the ampoule is submerged entirely in a deeply colored dye solution. Most often, approximately 1% methylene blue solution is employed. After carefully rinsing the dye solution from the outside, color from the dye will be visible within a leak. Leakers, of course, are discarded.

Vials and bottles are not subjected to such a leaker test, because the sealing material (rubber stopper) is not rigid. Therefore, results from such a test would be meaningless. However, assurance of container-closure sealing integrity should be an integral part of product development, by developing specifications for the fit of the closure in the neck of the container, the physical characteristics of the closure, the need for lubrication of the closure, and the capping pressure.

SAFETY TEST

The National Institutes of Health requires, of most biological products, routine safety testing in animals. Under the Kefauver-Harris Amendments to the Federal Food, Drug, and Cosmetic Act, most pharmaceutical preparations are now required to be tested for safety. Because it is entirely possible for a parenteral product to pass the routine sterility test, pyrogen test, and chemical analyses, and still cause unfavorable reactions when injected, a safety test in animals is essential, particularly for biological products, to provide additional assurance that the product does not have unexpected toxic properties.

PACKAGING AND LABELING

A full discussion of the packaging of parenteral preparations is beyond the scope of this text. It is essential, of course, that the packaging provide ample protection for the product against physical damage from shipping, handling, and storage, as well as protecting light-sensitive materials from ultraviolet radiation.

PACKAGING

The USP includes certain requirements for the packaging and storage of injections:

1. The volume of injection in single-dose containers is defined as that which is specified for parenteral administration at one time and is limited to a volume of 1 L.
2. Parenterals intended for intraspinal, intracisternal, or peridural administration are packaged only in single-dose containers.
3. Unless an individual monograph specifies otherwise, no multiple-dose container shall contain a volume of injection more than sufficient to permit the withdrawal and administration of 30 mL.
4. Injections packaged for use as irrigation solutions or for hemofiltration or dialysis or for parenteral nutrition are exempt from the foregoing requirements relating to packaging. Containers for injections packaged for use as hemofiltration or irrigation solutions may be designed to empty rapidly and may contain a volume in excess of 1 L.
5. Injections intended for veterinary use are exempt from the packaging and storage requirements concerning the limitation to single-dose containers and to volume of multiple-dose containers.

LABELING

The labeling of an injection must provide the physician or other user with all information needed to ensure the safe and proper use of the product. Since all of this information cannot be placed on the immediate container and be legible, it may be provided on accompanying printed matter.

A restatement of the labeling definitions and requirements of the USP for injections is as follows:

The term 'labeling' designates all labels and other written, printed, or graphic matter upon an immediate container or upon, or in, any package or wrapper in which it is enclosed, with the exception of the outer shipping container. The term 'label' designates that part of the labeling upon the immediate container.

The label states the name of the preparation, the percentage content of drug of a liquid preparation, the amount of active ingredient of a dry preparation, the volume of liquid to be added to prepare an injection or suspension from a dry preparation, the route of administration, a statement of storage conditions, and an expiration date. The label must state the name of the vehicle and the proportions of each constituent, if it is a mixture, and the names and proportions of all substances added to increase stability or usefulness.

Also, the label must indicate the name of the manufacturer or distributor and carry an identifying lot number. The lot number is capable of providing access to the complete manufacturing history of the specific package, including each

single manufacturing step. The container label is so arranged that a sufficient area of the container remains uncovered for its full length or circumference to permit inspection of the contents.

Preparations labeled for use as dialysis, hemofiltration, or irrigation solutions must meet the requirements for injections, other than those relating to volume, and must also bear on the label statements that they are not intended for intravenous injection. Injections intended for veterinary use are so labeled.

REFERENCES

- Yalkowsky SH *et al.* Formulation-related problems associated with intravenous drug delivery. *J Pharm Sci.* (1998). 87: 787–796.
- Mottu F *et al.* Organic solvents for pharmaceutical parenterals and embolic liquids: a review of toxicity data. *PDA J Pharm Sci Tech.* (2000). 54: 456–469.
- Nema S *et al.* Excipients and their use in injectable products. *PDA J Pharm Sci Tech.* (1997). 51: 166–171.
- Powell MF *et al.* Compendium of excipients for parenteral formulations. *PDA J Parenteral Sci Tech.* (1998). 52: 238–311.
- Strickley RG. Parenteral formulations of small molecules therapeutics marketed in the United States, Part I. *PDA J Parenteral Sci Tech.* (1999). 53: 324–349.
- Strickley RG. Parenteral formulations of small molecules therapeutics marketed in the United States, Part II. *PDA J Parenteral Sci Tech.* (2000). 54: 69–96.
- Strickley RG. Parenteral formulations of small molecules therapeutics marketed in the United States, Part III. *PDA J Parenteral Sci Tech.* (2000). 54: 152–169.
- Rowe RC *et al.* (Ed.). *Handbook of Pharmaceutical Excipients*, 6th ed, Washington DC: The Pharmaceutical Press, American Pharmaceutical Association, 2009.
- Akers MJ. Excipient-drug interactions in parenteral formulations. *J Pharm Sci.* (2002). 91: 2283–2297.
- Sutton SVW, Porter D. Development of the Antimicrobial Effectiveness Test as USP Chapter <51>. *PDA J Parenteral Sci Tech.* (2002). 56: 300–311.
- Carpenter JF, Crowe JH. The mechanism of cryoprotection of proteins by solutes. *Cryobiology.* (1988). 25: 244–250.
- Carpenter JF *et al.* Rational design of stable lyophilized protein formulations: theory and practice. In: Carpenter JF, Manning MC, eds. *Rational Design of Stable Protein Formulations*. New York: Kluwer Academic, 2002.
- Joint Commission on Accreditation of Healthcare Organizations. *The Complete Guide*. Chicago: JCAHO, 1997.
- Trissel LA. *Handbook on Injectable Drugs*, 16th ed. Bethesda, MD: ASHP, 2010.
- Collentro WV. *Pharmaceutical Water System Design, Operation, and Validation*, Englewood, CO: Interpharm Press (now CRC Press), 1998.
- Jiang Y *et al.* Tungsten-induced protein aggregation: solution behavior. *J Pharm Sci.* (2009). 98: (12) 4695–4710.
- Iacocca RG *et al.* Factors affecting the chemical durability of glass used in the pharmaceutical industry. *AAPS PharmSciTech.* (2010). 11: 1340–1349.
- Hoehne M *et al.* Adsorption of monoclonal antibodies to glass microparticles. *J Pharm Sci.* (2011). 100: 123.
- Walther M *et al.* Pharmaceutical vials with extremely high chemical inertness. *PDA J Pharm Sci Tech.* (2002). 56: 124–129.
- Swift R. Glass containers for parenteral products. In: Nema S, Ludwig JD, eds. *Pharmaceutical Dosage Forms: Parenteral Medications*, 3rd ed, Vol 1. London: Informa, 2010, 287–304.
- Pang J *et al.* Recognition and identification of UV-absorbing leachables in EPREX pre-filled syringes: an unexpected occurrence at a formulation-component interface. *PDA J Pharm Sci Technol.* (2007). 61: 423–432.
- Vilivalam VD, DeGrazio FL. Plastic packaging for parenteral drug delivery. In: Nema S, Ludwig JD, eds. *Pharmaceutical Dosage Forms: Parenteral Medications*, 3rd ed, Vol 1. London: Informa, 2010, 305–323.
- Das REG *et al.* Monocyte activation test for pro-inflammatory and pyrogenic contaminants of parenteral drugs: test design and data analysis. *J Immun Methods.* (2004). 288: 165–177.
- Lysfjord J (Ed.). *Practical Aseptic Processing: Fill and Finish*, Vols 1 and 2. Bethesda, MD: Parenteral Drug Association/Davis Healthcare, 2009.
- Howorth H. Movement of airflow, peripheral entrainment, and dispersion of contaminants. *J Parenteral Sci Tech.* (1988). 42: 14–19.
- Akers MJ. Good Aseptic Practices: Education and Training of Personnel Involved in Aseptic Processing, In: MJ Groves, Murty R, eds. *Aseptic Pharmaceutical Manufacturing II*. Englewood, CO: Interpharm Press (now CRC Press), 1995: 181–221.
- US Department of Health and Human Services (2004). *Guidance for Industry: Sterile Drug Products Produced by Aseptic Processing—Current Good Manufacturing Practice*. <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm070342.pdf>. (accessed September 2, 2011).
- Whyte W, Niven L. Airborne bacteria sampling: the effect of dehydration and sampling time. *J Parenter Sci Technol.* (1986). 40: 182.
- Akers JE. Science based aseptic processing. *PDA J Pharm Sci Tech.* (2002). 56: 283–290.
- Tyagi AK *et al.* IgG particle formation during filling pump operation: a case study of heterogeneous nucleation on stainless steel nanoparticles. *J Pharm Sci.* (2009). 98: (1) 94–104.
- Sharma B. Immunogenicity of therapeutic proteins. Part 2. Impact of container closures. *Biotechnol Adv.* (2007). 25: 318–324.
- Thirumangalathu R *et al.* Silicone oil- and agitation-induced aggregation of a monoclonal antibody in aqueous solution. *J Pharm Sci.* (2009). 98: 3167–3181.
- Esfandiary R *et al.* (2010). *Characterization of protein aggregation and adsorption on prefillable syringe surfaces*. <http://www.westpharma.com/na/en/support/Scientific%20Posters/Characterization%20of%20Protein%20Aggregation%20and%20Adsorption%20on%20Prefillable%20Syringe%20Systems.pdf>. (accessed August 10, 2010).
- Guazzo D. Container-closure integrity testing. In: Akers MJ ed. *Sterile Drug Products—Formulation, Packaging, Manufacturing, and Quality*. London: Informa Healthcare, 2010: 455–472.
- Nail SL *et al.* Fundamentals of freeze-drying. In: Nail SL, Akers MJ, eds. *Development and Manufacture of Protein Pharmaceuticals*. New York: Marcel Dekker, 2002: 281–360.
- PDA technical report no. 30: parametric release of pharmaceuticals terminally sterilized by moist heat. *PDA J Pharm Sci Tech* 1999; 53: 217–222.

BIBLIOGRAPHY

- Akers MJ. *Sterile Drug Products—Formulation, Packaging, Manufacturing, and Quality*. London: Informa Healthcare, 2010.
- Akers MJ *et al.* *Parenteral Quality Control*, 3rd ed. New York: Dekker, 2002.
- Avis KE *et al.* (Eds.). *Pharmaceutical Dosage Forms: Parenteral Medications*, 2nd ed, Vols 1-3. New York: Dekker, 1992–1993.
- Avis KE *et al.* (Eds.). *Biotechnology: Quality Assurance and Validation Drug Manufacturing Technology Series*, Vol 4. Boca Raton, FL: CRC Press, 1998.
- Barber T. *Control of Particulate Matter Contamination in Healthcare Manufacturing*. Englewood, CO: Interpharm Press (now CRC Press), 2000.
- Block SS (Ed.). *Disinfection, Sterilization and Preservation*, 4th ed. Philadelphia, PA: Lea & Febiger, 1991.
- Carpenter JF, Manning MC (Eds.). *Rational Design of Stable Protein Formulations*. New York: Kluwer Academic/Plenum, 2002.
- Carleton FJ, Agalloco JP (Eds.). *Validation of Aseptic Pharmaceutical Processes*, 2nd ed. New York: Dekker, 1998.
- Coles T. *Isolation Technology: A Practical Guide*. Englewood, CO: Interpharm Press (now CRC Press), 1999.
- Katdare A, Chaubal MV (Eds.). *Excipient Development for Pharmaceutical, Biotechnology, and Drug Delivery Systems*. New York, London: Informa Healthcare, 2006.

- Kuhlman H, Coleman D. Engineering aspects of WFI systems design. In: Avis KE, ed. *Sterile Pharmaceutical Products: Process Engineering Applications*. Buffalo Grove, IL: Interpharm Press (now CRC Press), 1995: 221–268.
- Jameel F, Hershenson S. *Formulation and Process Development Strategies for Manufacturing Biopharmaceuticals*. Hoboken, NJ: John Wiley and Sons, Inc., 2010.
- Jennings T. *Lyophilization: Introduction and Basic Principles*. Boca Raton, FL: CRC Press, 1999.
- Luungqvist B, Reinmuller B. *Clean Room Design: Mimimizing Contamination Through Proper Design*. Boca Raton, FL: CRC Press, 1996.
- Meltzer TH. *High Purity Water Preparation for the Semiconductor, Pharmaceutical and Power Industries*. Littleton, CO: Tall Oaks, 1993.
- Meltzer TH, Jornitz MW (Eds.). *Filtration in the Biopharmaceutical Industry*. New York: Dekker, 1998.
- Nail SL, Akers MJ (Eds.). *Development and Manufacture of Protein Pharmaceuticals*. New York: Kluwers Academic/Plenum, 2002.
- Nema S, Ludwig J (Eds.). *Pharmaceutical Dosage Forms: Parenteral Medications*, 3rd ed, 3 vols. London: Informa Healthcare, 2010.
- Phillips GB, Miller WS (Eds.). *Industrial Sterilization*. Durham, NC: Duke University Press, 1973.
- Swarbrick J, Boylan JC (Eds.). *Encyclopedia of Pharmaceutical Technology*, 2nd ed. New York: Dekker, 2002.
- Turco S, King RE. *Sterile Dosage Forms*, 3rd ed. Philadelphia, PA: Lea & Febiger, 1987.

Pharmaceutical Compounding – USP <797> Sterile Preparations

Catherine Cone, PharmD, BCPS; Linda A. Felton, PhD and Amy Bachyrycz, PharmD

INTRODUCTION 533

COMPOUNDING PERSONNEL: RESPONSIBILITIES IN IMPLEMENTING USP <797> 533

FACILITIES 535

COMPOUNDING STERILE PREPARATIONS 535

QUALITY ASSURANCE 537

CONCLUSION 538

APPENDIX A 540

INTRODUCTION

Medications that are compounded and injected into humans or other animals must be sterile and free from contaminants and pyrogen. Mortality and morbidity from contaminants being injected into the blood stream require that personnel involved in intravenous compounding follow more stringent guidelines when compounding sterile preparations. In 2004, the *United States Pharmacopeia – National Formulary (USP – NF)* published standards of practice for compounding sterile preparations. The chapter, referred to as USP <797>, was developed in part as a result of patient injuries and deaths that had occurred secondary to problems with medication delivery and sterile compounding.¹ Publication of USP <797> is meant to increase practitioners understanding of how sterile preparations must be prepared in order to achieve safe, accurate, quality sterile compounds for patient use.² The chapter was updated in 2008 to clarify risk levels, primary engineering controls (PECs), personnel training and evaluation, verification of compounding accuracy, finished preparation checks and tests, storage and beyond use dating, maintenance of sterility and purity of dispensed and distributed sterile preparations, and suggested standard operating procedures.² The 2008 USP <797> also introduced new topics and made some significant changes in the areas of immediate use sterile preparations, single use and multiple dose vials, hazardous compounding including radiopharmaceuticals and allergen extracts, environmental quality control, personnel training and competency evaluation.² Since its publication in 2008, the *International Journal of Pharmaceutical Compounding* and the American Society of Health Systems Pharmacy (ASHP) have published numerous articles to help practitioners implement chapter USP <797> into all aspects of sterile compounding.¹⁻²³ The objective of the USP <797> chapter is to “describe conditions and practices to prevent harm, including death, to patients that could result from (1) microbial contamination (nonsterility), (2) excessive bacterial endotoxins, (3) variability in the intended strength of correct ingredients that exceeds either monograph limits for official articles (sic) or 10% for nonofficial articles, (4) unintended chemical and physical contaminants, and (5) ingredients of inappropriate quality in compounded sterile preparations (CSPs).”² While many pharmacies have become compliant, others cite costs and lack of scientific justification to support these changes.^{3,4} Most hospitals and facilities have gradually implemented the standards from the original recommendations in 2004, and are now taking steps towards becoming compliant with the 2008 revisions.^{3,4}

In order to facilitate implementation of the standards, Baxter Healthcare Corporation and ASHP developed a discussion guide¹ that suggests among other things, that a gap analysis be completed. A gap analysis in this context is used to identify areas in need of improvement in sterile compounding facilities.¹ USP <797> requirements should be compared to the current practices of compounding personnel and environmental standards of

the compounding facility to determine where “gaps” exist. Successful completion of a gap analysis and subsequent revisions to address the identified gaps may result in a direct positive impact on the delivery of quality patient care and may be used to demonstrate compliance with the standards to regulatory authorities.⁵

COMPOUNDING PERSONNEL: RESPONSIBILITIES IN IMPLEMENTING USP <797>

Compounding personnel have the responsibility to ensure that sterile preparations do not harm patients.^{1,2,5-7} Prevention of morbidity and mortality from contaminated CSPs is of utmost importance and can help be assured by strictly adhering to aseptic technique standards. Pharmacists, technicians and others who compound sterile preparations must have appropriate education to ensure that USP <797> is understood and must have training to develop skills required for proper aseptic technique.² ASHP's *Compounding Sterile Preparations Video Guide to Chapter <797>* sets forth best practices for compounding personnel, centering on personnel education and skill development.⁵ All components of drug use from sourcing and delivering of the medication to the institution, compounding, and then delivery of the medication for therapeutic patient use are addressed. The importance of sterility of the preparation in preventing harm and fatalities cannot be understated. The following sections discuss the various components needed to ensure sterile compounding including equipment, personnel training, facilities, and appropriate product/device use.

LAMINAR AIR FLOW HOODS AND ASEPSIS

Clean air that is free of particulates is necessary when compounding in order to achieve sterile preparations. In order to achieve cleaner air, methodical use of a laminar air flow hood (LAFH) is a requirement. Air is double filtered prior to coming into contact with products placed on the counter of the LAFH. The first filter, or pre-filter, is usually located below the counter top of a horizontal LAFH (Figure 27-1). It removes large or gross contaminants from the air. The air is then pressurized to insure constant flow through the second filter prior to washing over the products. The second filter is a high efficiency particulate arrestor (HEPA) that removes particulates from air as small as a 0.3 micron size. The HEPA filter also removes most microorganisms from the air. However, bacteria and viruses <0.3 microns can still persist in the air. Thus, personnel training to decrease transmission of these organisms to the product and ultimately to the patient becomes of paramount importance.^{5,7}

There are several types of LAFHs including horizontal (Figure 27-1), vertical (Figure 27-2), and compounding aseptic isolators (CAI) or compounding aseptic containment isolators (CACI). CAI and CACI are often referred to as glove box airflow hoods, as compounding personnel must put their hands into gloves that then can be manipulated inside the airflow hood (or inside the box). Generally, horizontal LAFHs are the easiest for

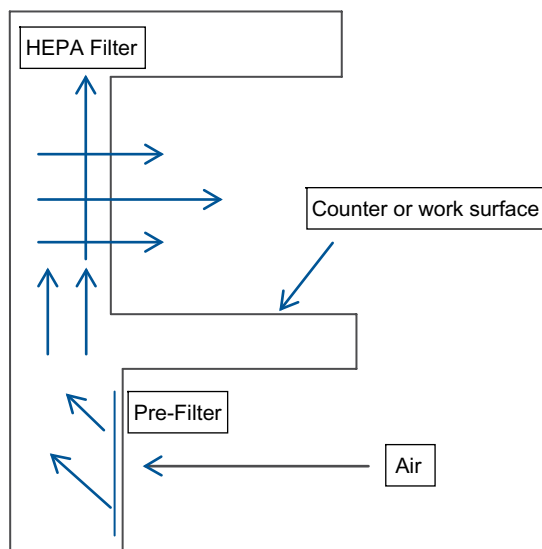


Figure 27-1. Horizontal Laminar Airflow Hood.

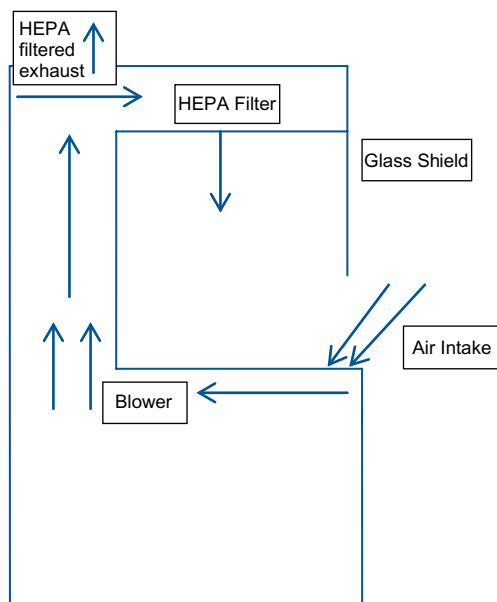


Figure 27-2. Vertical Laminar Airflow Hood.

personnel to use, followed by vertical, and then CAI or CACI.² LAFHs, CAIs, and CACIs should be serviced every 6 months and whenever they are moved.² The filters should be checked for particulates or tears during these services and changed as needed. If damage is done at any time to the filters, they must be immediately changed to ensure air quality. As recommendations may vary, refer to the LAFHs manufacturer instructions for more details. Appropriate placement and use of the LAFHs is discussed in the basics of aseptic compounding and facilities sections.

BASICS OF ASEPTIC COMPOUNDING

Personnel who compound sterile preparations must have training from experts to develop the skills necessary for proper aseptic technique.² Touch contamination poses the greatest microbial risk to patient safety when compounding sterile preparations. Pharmacy and technician schools should have both didactic and hands-on experiences including aseptic skill development, prior to students rotating through settings where sterile compounding is conducted. Skill development should include proper aseptic technique, an understanding of how LAFHs op-

erate in conjunction with other environmental controls to promote an aseptic environment, filling and checking medication orders for appropriateness of dose and therapy, accuracy of fill, any potential contaminants or leaks, and proper labeling.^{2,3,5} In all facilities, it is recommended that a second check of the medication order and preparation should be completed by another individual, preferably one who did not process the order, prior to delivery of the CSP for patient use. The competence of individuals should be tested prior to making CSPs for patient use.^{2,5} A discussion on competency testing can be found in the quality assurance portion of this chapter.

PERSONNEL HYGIENE AND GARBING

As touch contamination poses a great risk for microbial contamination of sterile preparations, standard operating procedures should be developed to include proper washing and garbing with personal protective equipment (PPE).^{2,5,8} Prior to entering the compounding area, personnel must cleanse and garb. Hand hygiene should consist of the use of warm water and anti-septic (examples include antimicrobial washes and povidine-iodine) gently rubbed from hands to elbows for 30 seconds. Nails should be gently scrubbed and then a nail pick used to remove debris from under the nails. This should be followed by rinsing, and then drying with a lint-free disposable towel. As humans shed up to 10^6 squamous cells per hour,² care must be taken to ensure that skin is not damaged during this process. Damaged skin can result in increased shedding and thus heightens the risk of contamination of CSPs. Additionally, skin damage could promote the development of infections which could also increase the risk of CSP contamination. Individuals with rash, sunburn, weeping sores, conjunctivitis, and active respiratory infections must not compound sterile products.² Additionally, jewelry including any visible piercings must be removed. Artificial nails may harbor microbes and skin cells that cleaning cannot remove so artificial nails are excluded from use by compounding personnel. Natural nails should be kept clean and trimmed. As cosmetics can shed particles, no make-up is allowed in compounding areas. Finally, outer clothes (jackets, shawls, etc.) should be removed prior to garbing. Institutions may require personnel to change into scrubs prior to washing and garbing. Scrubs should be cleaned professionally and separately to reduce lint and particles that could contaminate the compounding area.^{2,5,8}

In USP <797>, it is recommended that garbing be completed following a head-to-toe pattern with a couple of exceptions. This allows any particles or contaminants to fall to the ground or be covered up by PPE. The first element of this process is covering the hair with a head cover. If facial hair exists a special beard cover must be worn. Second, facial covers or surgical masks are donned such that the nose and mouth are completely covered. Eye shields can then be added, especially if irritants are to be compounded. Third, shoe covers must be placed over shoes carefully to prevent dirt from touching the outside of the shoe cover or from tracking dirt into the sterile environment. Fourth, hands and arms must be washed from hands to elbow. Fifth, a non-shedding gown with snug fitting wrist and neck closures is donned. This is followed by use of an anti-septic hand cleaner, prior to the placement of sterile gloves. The sterile gloves should be placed such that the outside of the gloves is never touched and the end of the gloves covers the wrist of the gown. The anti-septic hand cleaner is used again over the sterile gloves to decrease potential for touch contamination. Additional garb is recommended for hazardous compounding by the Department of Health and Human Services National Institute for Occupational Safety and Health (NIOSH). As sterile chemotherapy gloves are not currently manufactured, double gloving is recommended. After donning the chemotherapy protective glove (nonsterile), hand anti-septic should be used, followed by sterile glove placement over the chemotherapy glove. A final anti-septic hand cleanser should be used over the

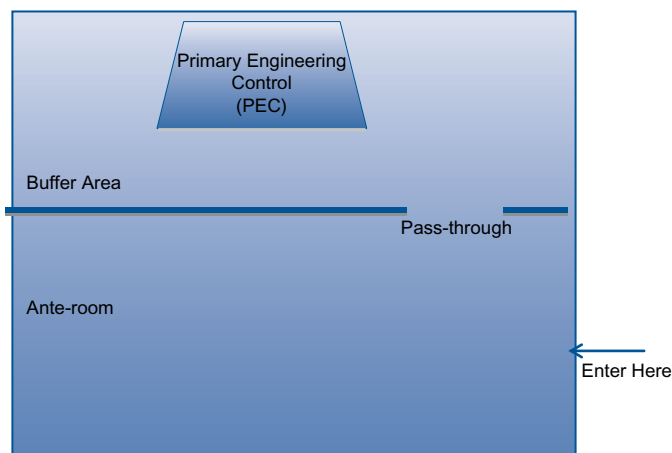


Figure 27-3. Example of Sterile Compounding Facilities.

sterile glove prior to entering the buffer area (Figure 27-3) for compounding.^{2,5,8,9}

FACILITIES

RISK LEVELS

The air quality in a sterile compounding area is classified according to the number of particles per meter cubed of air, as shown in Table 27-1. Figure 27-3 shows the recommended set-up of the sterile compounding facilities. The first area personnel enter prior to washing and garbing is the ante-room. The ante-room must be ISO class 8 or better. After washing and garbing are completed and medications and supplies are gathered, personnel then enter the buffer area. The buffer area must be ISO class 7 or better and the PEC must be ISO class 5 or better.^{2,5}

Risk level is determined by the potential risk for contamination of a CSP. USP <797> defines low-risk CSPs as ones that (1) are compounded within a LAFH of ISO class 5 or better using only sterile ingredients, (2) involve only transfer, measuring or mixing of not more than 3 commercially manufactured sterile products (e.g., bags or vials), and (3) involve manipulations limited to aseptically opened ampules, or penetration of sterile products and disinfected stoppers with sterile needles and syringes.² In the absence of passing a sterility test, the storage period of a low-risk CSP cannot exceed 48 hours if kept at controlled room temperature, 14 days if refrigerated, or 45 days if frozen.^{2,10} Medium-risk CSPs are compounded with sterile ingredients as well. However, the difference lies in the complexity of the process including time needed to compound the CSP and number of manipulations needed to complete the order.¹⁰ USP <797> defines medium-risk CSPs as (1) multiple individual or small doses of sterile products that are combined or pooled to prepare a CSP that will be administered either to multiple patients or to one patient on multiple occasions, (2) compounding processes that include complex aseptic manipulations, and (3) compounding processes that require an unusually long duration to complete.² In the absence of passing a sterility test, the storage period of medium-risk CSPs cannot exceed 30 hours if kept at controlled room temperature, 9 days if refrigerated, or 45 days if frozen.^{2,10} Examples of medium-risk CSPs include batches of sterile preparations that are compounded for multiple patients or a parenteral nutrition that is compounded with multiple ingredients.¹⁰

Table 27-1. ISO Classification Levels

ISO 8	3,520,000/m ³
ISO 7	352,000/m ³
ISO 6	35,200/m ³
ISO 5	3,520/m ³

High-risk level conditions occur when CSPs are compounded from non-sterile bulk ingredients that then must be sterilized for use, CSPs that are exposed to less than an ISO class 5 level air quality for more than 1 hour, CSPs where measuring and mixing of ingredients occurs in non-sterile devices before sterilization is completed, and CSPs where an assumption has been made that packages of bulk ingredients contain at least 95% by weight of their active chemical moiety and have not been contaminated or adulterated between uses.^{2,10} High-risk may also be characterized by personnel who do not meet appropriate garbing and gloving requirements or where non-sterile water-containing preparations are stored for greater than 6 hours before sterilization occurs.^{2,5} A majority of institutions that compound sterile preparations are classified as a low- to medium-risk level.

SET-UP OF FACILITIES AND ENVIRONMENTAL CONTROLS

Ideally, the sterile compounding facility layout should appear as in Figure 27-3, where the ante-room contains a hand washing station, PPE/garbing, and high use materials such as sterile syringes, needles, admixture solutions, and medications.⁵ In Figure 27-3, the buffer area is shown separating the ante-room from the PEC. The PEC can be a LAFH, CAI or CACI. The buffer areas should be kept at an ISO class 7 level or better.^{2,5,11-12} USP <797> does allow for some exceptions to this rule, although it is discouraged.² The PECs are designed to reduce the exposure of CSPs to the buffer area (ISO class 7 level). The PEC should be kept at an ISO class 5 level or better.² Placement of the PEC is paramount to maintaining the air quality; it should not be in an area where a draft could push air from the surrounding environment into the hood.^{2,11-12} When compounding in a PEC, the “critical sites” are all surfaces or openings on a CSP (e.g., vial septa, injection port, beaker, opened ampules, needle hubs, etc.) exposed and at risk of direct contact with air (whether from the PEC or buffer area), moisture, or touch contamination. The critical sites are the primary areas to safeguard for CSPs as they represent potential ports of entry for contaminants such as micro-bacterial, viral, or foreign matter.^{2,5,10-12}

CLEANING AND DISINFECTING

In addition to air quality, the buffer area and PEC must be kept in immaculate condition.⁵ In order to accomplish this, USP <797> recommends that the floors, walls, ceilings and PECs be cleaned regularly.² As some substances or ingredients do not dissolve in the disinfectants typically used, cleaning the PEC can be accomplished with sterile water and lint free disposable towels. For disinfecting the PEC, 70% isopropyl alcohol or 3% hydrogen peroxide can be used to maintain the air quality in the PEC and to kill any microbes that may be present. The PEC should be cleaned and disinfected between shifts, after each batch or every 30 minutes whichever comes first, when spills occur, and when contamination is suspected.^{2,5,11-12} Floors must be mopped daily, walls and ceilings cleaned weekly, and shelves cleaned monthly at a minimum. Mops must be dedicated to the clean room only. Mops that have disposable pads are recommended to decrease the transfer of dirt or other contaminants. Trash should be emptied daily and as needed. Care should be taken not to disturb the air while bags are removed and replaced from the trash cans. Personnel should also be trained to decontaminate and wipe down all products with an appropriate disinfectant prior to bringing them in to the ante-room. Additionally, all products should be wiped again prior to transferring them into the buffer area and PEC for use in compounding.¹¹⁻¹²

COMPOUNDING STERILE PREPARATIONS

BEYOND USE DATING

Compounding sterile preparations requires many environmental controls as previously described. Personnel working with CSPs must have the knowledge, skills, and motivation to fol-

Table 27-2. Beyond Use Dates

USP Temperature Ranges	Low-risk CSP	Medium-risk CSP	High-risk CSP
Controlled Room Temperature	48 hours	30 hours	24 hours
Refrigerated	14 days	9 days	3 days
Frozen	45 days	45 days	45 days

low these processes in order to make them effective.^{5,12} Understanding the risk levels in relation to the preparations viability (ability to remain sterile) after leaving the PEC is important for patient safety. Equally as important is the potency and purity of the compounded preparation. Beyond use dating (BUD) is a measure of the time interval from when the CSP is compounded until the compound starts to lose its purity and potency.^{2,13-14} BUDs are often confused with manufacturer's expiration dates. However, once a manufacturer's product leaves its original packaging and is mixed with other components, the expiration date ceases to exist and a BUD must be established.^{5,13-14} A BUD can be established via several routes. First, an extra CSP can be made during routine compounding and then tested to determine how long it retains its potency and purity. Many pharmacies contract out these services. Second, the BUD can be obtained from a reliable source (e.g., *Trissel's Handbook of Injectable Drugs*¹⁵ or other reference) that shows stability testing. Third, some manufacturers have done stability testing and reported this in their package inserts. In the absence of testing or evidence of the product's stability, USP <797> sets BUDs as shown in Table 27-2.^{2,5,13-14}

EXCEPTIONS TO USP <797> STANDARDS

In instances of life and death, CSPs may be compounded and injected in to a patient (e.g., during a patient emergency or code) without the use of LAFHs.² This exception to USP <797> standards applies only to such immediate use CSPs.^{2,16,24} Sterile solutions, needles, and syringes must still be utilized and sterile technique must be strictly maintained.

In rural areas, where facilities are not up to recommended standards, USP <797> allows for another exception. This exception is not to excuse institutions from the standards. However, when there is no clean room area available (i.e., no buffer area or ante-room) for the PEC, USP <797> states that the medications must be used within 12 hours of compounding.^{2,16} It should be noted that USP <797> has 4 requirements that must be met in order for this to be acceptable. First, the PEC must be of an ISO class 5 or better and must be in a segregated and restricted compounding area. Second, the segregated area must not have unsealed windows, doors to high traffic areas or be adjacent to construction sites, warehouses, food preparation areas, or similar sites. Third, policy and procedures must be written and followed. Sinks for washing cannot be located next to the PEC (as they may introduce contaminants including bacteria into the PEC). Finally, procedures for cleaning, disinfecting, personnel training, garbing, aseptic technique, and environmental air sampling must be written and followed.²

SINGLE-USE CONTAINERS AND MULTI-DOSE VIALS

In the 2008 revision of USP <797>, focus was placed on how single use containers and multi-dose vials (MDVs) should be used.^{2,5} Single use containers cannot be stored and re-used. These products contain no preservatives, and thus when punctured or opened, are subject to potential contamination. In contrast, MDVs contain preservatives and have been tested by the manufacturer for the length of time they retain their purity and potency once punctured. Unless otherwise specified, MDV have a BUD of 28 days.^{2,13}

COMPOUNDING DEVICES

Many pharmacies that compound parenteral nutrition (PN) use automated compounding devices (ACDs). PN requires a complex process of mixing multiple ingredients into a large volume bag meant to provide nutrition to patients via an intravenous route. ACDs are designed to help decrease the complexity of the compounding and reduce the time required to prepare PN preparations. However, care must be taken to ensure that these devices are kept in working order. ACDs should be checked regularly to ensure that the correct amounts of ingredients are delivered as programmed. For example, if the ACD is programmed to deliver 10 mL of water, personnel should weigh the water to check that it is 10 grams.² A daily record should be kept by compounding personnel of this check. It is recommended by USP <797> that these records should be reviewed at least weekly by the pharmacist-in-charge to avoid any potentially clinically significant cumulative errors.^{2,5,17}

Hazardous drugs

Another area where extra care must be taken when compounding sterile preparations is with hazardous drugs. A hazardous drug is defined as one that has been shown to potentially cause carcinogenicity, developmental or reproductive toxicity, or harm to organs in animal or human studies. The Department of Health and Human Services National Institute for Occupational Safety and Health (NIOSH) maintains a list of hazardous drugs in healthcare settings.^{18,25} Compounding personnel should review this list and develop and follow standard operating procedures regarding the handling and compounding with these agents. Routes, time, and extent of potential exposure must be considered when working with hazardous drugs. Although most pharmacists and technicians are careful while compounding of these agents, other routes of exposure must be considered. For example, when an anti-neoplastic agent is removed from its original carton, personnel should be wearing PPE to prevent potential inadvertent exposure from the bottles or packaging materials.²⁵ For specific information on this topic, USP <797> and the *International Journal of Pharmaceutical Compounding on Hazardous Drugs* should be read.²⁵

Sterilization methods

For pharmacies that compound high-risk preparations, non-sterile components and medications must be sterilized prior to human use.² Methods for sterilizing compounded preparations and components include: filtration, autoclaving, and dry heating.¹⁹ The appropriate method must be employed to ensure that compounds do not lose their stability, strength, or purity during sterilization.^{2,13,20} Quality assurance testing of these methods must be undertaken to ensure that sterilization and depyrogenation occur.

General considerations for filters include the size of the filter and the type of fluid to be filtered. In order to filter out bacteria a pore size must be 0.22 micrometers or less. The fluid to be filtered should be a solution that is chemically and physically compatible with membrane filtration. It is preferred for medications that are thermo-labile. The fluid should be filtered rapidly and completely without filter replacement. When deciding if filtration is the best method for sterilization, a few items must be taken into consideration including: microbes and viruses that are <0.22 microns cannot be filtered out; medications in the form of suspensions are not appropriate for filtration as medication will be filtered out; solutions that are not fully dissolved will get filtered out; and filtration does not remove pyrogens, and thus heat must be used to render pyrogens inactive. A simple test to determine the integrity of a filter is the bubble point test.^{2,19} It is a nondestructive (will not affect the sterility, purity, or strength of the fluid) test that can be performed before or after filtration. The bubble point is the pressure required to force air or gas through the pores of the filter. When the bubble point

is reached, bubbles will appear. The pressure is documented and compared to the manufacturer's standard for that filter.

More effective methods of sterilization include autoclaving and dry heating.¹⁹ Autoclaving uses steam heat (water vapor under elevated pressure) to kill spores, bacteria, and viruses. Sealed vials and ampules of an aqueous solution will generate their own steam that will sterilize the inner contents. It is the preferred method. However, some CSPs may not be stable when exposed to steam heat. Thus dry heat may be used in these circumstances. Dry heat sterilization will also kill spores, bacteria and viruses via dehydration. This requires higher temperatures for longer periods of time, so medication stability must be considered. For components used in compounding, depyrogenation by dry heat is the preferred method. Dry heat works well for oil-based suspensions and solutions as well as bulk powders.

LABELING REQUIREMENTS

The compounded CSP must be labeled with the following information: the drug and ingredient(s) names, quantities, strength, concentration, and volumes; the time that the CSP was compounded with a BUD; the personnel who compounded and checked the CSP; how the CSP should be stored; any auxiliary labels; and any other instructions necessary for proper use of the preparation.² In addition, if the compounded preparation is for a specific patient, the patient's name and an identifier along with the instructions for use must be included on the label.⁵

VERIFICATION

All CSPs should undergo a final visual inspection of the labeling as well as product. This final check should be done by a pharmacist. It is prudent for all CSPs to be double checked by someone other than the personnel who compounded the CSP. The order should be placed alongside the CSP with quantities or amounts written and any calculations shown. Additionally, the CSP should be examined for any contaminants and color changes against a light and dark background, and for any leaks. If any doubt exists, the CSP should be discarded and a new CSP made. In cases where high-risk level compounding is involved, one CSP out of the batch must be sent out for sterility (required) and endotoxin (recommended) testing prior to the CSP being dispensed.¹²

STORAGE AND DELIVERY

Nurses, patients, physicians, or any other person to whom CSPs are dispensed should be educated on how to properly store CSPs. When CSPs are delivered to medication rooms (e.g., in institutional settings), pharmacy personnel should educate those who administer the CSPs to monitor room and refrigerator temperatures to ensure product integrity. Areas where CSPs are kept should be inspected monthly by pharmacy personnel and standard operating procedures (SOPs) should be in place to ensure that CSPs maintain their quality in the patient care setting until administration. When CSPs are returned to the pharmacy, any outdated CSPs (those past their BUD) should be properly disposed of. If CSPs are to be re-used, strict SOPs should be in place delineating the circumstances when this is appropriate.⁵

One area that is often overlooked is the delivery of CSPs via personnel who are not under direct administrative control by the pharmacy (e.g., truck delivery service). Basic training of these personnel should occur so that the quality of the CSP is maintained.⁵

QUALITY ASSURANCE

USP <797> requires that a quality assurance (QA) plan and formal QA programs are developed to ensure the continuous quality of compounded preparations.² Quality assurance programs should be formalized in writing and encompass all aspects of CSPs.^{5,21-22} Detailed procedures on monitoring and evaluation activities should be described. QA programs should

specify how results are reported and evaluated, and identify mechanisms for follow-up when results are out of range. In addition, descriptions of responsibilities for personnel maintaining the QA program(s) should be clearly identified. Documentation is key to providing evidence to accrediting agencies that a QA plan and programs are in place and being routinely followed. Quality assurance programs should be detailed in your SOPs.

STANDARD OPERATING PROCEDURES

Standard operating procedures are important for maintaining continuous quality care to patients, especially when there is high turnover or when resources are scarce. Developing and implementing SOPs can be daunting to personnel. USP <797> details the competencies, conditions, practices, and quality assurances that are required and recommended in an appendix. These should be used to develop SOPs for sterile compounding facilities. Appendix 1 summarizes these by main topic. For greater details, USP <797> and an article by Okeke and Allen can be reviewed.^{2,23}

PERSONNEL TRAINING AND QUALIFICATIONS

Personnel who compound sterile products must be adequately trained. USP <797> states that theoretical principles and practical skills of aseptic technique in achieving and maintaining ISO Class 5 level conditions should come from expert personnel through hands-on, audio-video instruction, and professional publications.² It further states that personnel shall perform didactic reviews, pass written tests as well as media-fill challenge tests of aseptic technique initially and annually thereafter. Media-fill challenge testing uses a microbiological growth medium (broth) to simulate admixture compounding. This testing shows objective evidence that personnel and processes are in place to produce sterile products. Personnel should also pass finger/touch testing using agar plates that contain compounds to de-activate antimicrobial substances on the glove. Personnel who fail either of these tests or written exams will need to be re-instructed and re-evaluated prior to compounding sterile preparations. For high-risk level compounding, these processes should occur semi-annually.^{2,5,7,8}

PATIENT AND CAREGIVER TRAINING

When intravenous therapy is required after discharge from an institutional setting, patient and caregiver training becomes essential for the safe use of CSPs. Elements of training programs for patients and caregivers should include written and visual materials as well as hands on practice (skill development) with the CSPs and devices.^{2,22} Training should include how to check the CSP to ensure its integrity, how to handle and store the CSP, how to administer the CSP using proper techniques, how to care for catheters, dressings and maintain patency of injection site, how to use any devices associated with the delivery of CSPs, how to monitor for complications or adverse effects, how to dispose of waste, and how to seek emergency services if needed. SOPs must be in place so that all personnel, patients and caregivers understand the process surrounding safe and effective use of CSPs, and patient and caregiver understanding of these processes should be assessed prior to dispensing the CSP and any devices.^{2,22,23}

ADVERSE EVENT REPORTING AND PATIENT MONITORING

Pharmacists must participate actively in adverse event reporting and patient monitoring. In many cases, the pharmacist is the primary point of contact for the patient and caregiver with the healthcare team. When CSPs and devices are dispensed to patients, pharmacists should inquire about adverse events and how the treatment is working for the patient. Additionally, pharmacists should be available to answer patients' questions on CSPs, devices and to report problems to physicians and others.^{5,22}

Table 27-3. Temperature Ranges for Compounded Sterile Products

Environment	Temperature Range
Controlled Room Temperature	20°C to 25°C (68°F to 77°F)
Controlled Cold Temperature	2°C to 8°C (36°F to 46°F)
Freezing Temperature	–25°C to –10°C (–13°F to 14°F)

ENVIRONMENTAL MONITORING

Testing of the air quality should be performed every 6 months to ensure that the PEC, buffer area, and ante-room area remain at an ISO class 5, 7, and 8 or less, respectively.^{2,11-12,21} Volumetric air sampling by impaction should be utilized to test for viable and nonviable particles. Commercial products and services are available for this purpose. LAFH, CAI and CACI should also be tested using agar plates to determine sterility in the ISO Class 5 environment. Each facility should have a sampling plan that should describe the sample location(s), method of collection, frequency of sampling, volume of air sampled (in the case of air sampling), and the time of the day that the activity will be performed. If compounding is occurring at the time of sampling, this should be noted in the records. Records must be maintained and reviewed by supervising personnel to ensure that the environmental controls comply with USP <797> standards. These records may also be reviewed by accrediting agencies. In addition to air quality control, pressure differential or airflow monitoring and recording must be completed at a minimum of once daily. For low- to medium-risk compounding, pressure differential or airflow displacement can be measured. Facility design determines which measurement is utilized. In the case of airflow displacement, the PEC, buffer area and ante-room are located together without any separation. Airflow displacement should measure at least 40 feet per minute for this facility design.² Pressure differential monitoring is utilized for facilities where the PEC and buffer area are separated from the ante-room. A positive pressure differential of at least 0.02-inch water column is required.² In compounding facilities with positive pressure environments, the air flows from the buffer room to the ante-room to the general pharmacy. The ante-room typically has an intake panel that facilitates this pressure differential by pulling air from the buffer into the ante-room. Digital electronic devices to measure the pressure differential can be purchased. These devices must be placed appropriately to ensure that the desired pressure differentials are measured. Temperature in buffer rooms should be maintained to prevent microorganisms from growing and for personnel comfort. Additionally, personnel often wear multiple layers of clothing, head covers, and facemasks when they compound CSPs that could cause sweating or general discomfort.⁵ Temperature in the ante-room and buffer area should be maintained cooler than the general pharmacy areas for these reasons. Although not required by USP <797>, other environmental parameters may be monitored such as light, humidity, and sound. Humidity should be maintained between 35% to 65%. Light and sound levels should meet state and federal requirements and be comfortable for personnel working in the environment.⁵

MEDICATION AND CSP STORAGE

Medications should be stored according to manufacturer specifications. Table 27-3 shows the temperature range according to USP-NF requirements. These must be monitored daily at a minimum.²

CONCLUSION

In an effort to control the purity and sterility of intravenous medications, the USP-NF developed rules and recommendations termed USP Chapter <797>. These rules and recommendations

apply to all injected medications whether made by pharmacy, nursing, medicine, dental, or any other health care provider. The costs associated with implementation have created some controversy. However, a majority of institutions have come into compliance with the 2004 standards, and are now in the process of implementing the 2008 standards. The risks associated with injectable medications are high when standards are not followed. Quality patient care must be the number one priority. Developing QA programs and following standard operating procedures are necessary to maintain continuous quality care to patients.

REFERENCES

1. American Society of Health-Systems Pharmacy (2010). *The ASHP discussion guide on USP Chapter <797>: Compounding sterile preparations*. http://www.ashp.org/s_ashp/docs/files/HACC_797guide.pdf (accessed 11 February 2011).
2. *United States Pharmacopeia 31 – National Formulary 26* (2008). Chapter USP <797> Pharmaceutical compounding sterile preparations.
3. Allen LV, Okeke CC. Basics of compounding: considerations for implementing United States Pharmacopeia Chapter <797> Pharmaceutical Compounding – Sterile Preparations, Part 6: 2008 Revisions to Chapter <797>. *IJPC* 2008; 12(2): 136–144.
4. McElhiney L. Compliance by hospitals and healthcare facilities with the revised United States Pharmacopeia Chapter <797>. *IJPC* 2009; 13(2): 94–98.
5. Kienle PC. *Compounding Sterile Preparations: ASHP's Video Guide to Chapter <797> Companion Guide and Assessment Record*, 1st edn. American Society of Health-System Pharmacists, 2011.
6. Allen LV, Okeke CC. Basics of compounding for the implementation of United States Pharmacopeia Chapter <797> Pharmaceutical Compounding – Sterile Preparations, Part 1. *IJPC* 2007; 11(3): 230–236.
7. Allen LV, Okeke CC. Basics of compounding: the responsibilities of compounding personnel in implementing United States Pharmacopeia Chapter <797> Pharmaceutical Compounding – Sterile Preparations, Part 2. *IJPC* 2007; 11(4): 314–323.
8. Allen LV, Okeke CC. Basics of compounding: considerations for implementing United States Pharmacopeia Chapter <797> Pharmaceutical Compounding – Sterile Preparations, Part 15: Personnel training, cleansing, and garbing. *IJPC* 2009; 13(5): 420–426.
9. Allen LV, Okeke CC. Basics of compounding: considerations for implementing United States Pharmacopeia Chapter <797> Pharmaceutical Compounding – Sterile Preparations, Part 4: Considerations in selection and use of disinfectants and antiseptics. *IJPC* 2007; 11(6): 492–499.
10. Allen LV, Okeke CC. Considerations for implementing United States Pharmacopeia Chapter <797> Pharmaceutical Compounding – Sterile Preparations, Part 3: Risk levels. *IJPC* 2007; 11(5): 404–411.
11. Allen LV, Okeke CC. Basics of compounding: considerations for implementing United States Pharmacopeia Chapter <797> Pharmaceutical Compounding – Sterile Preparations, Part 13: Environmental quality and control. *IJPC* 2009; 13(3): 234–238.
12. Allen LV, Okeke CC. Basics of compounding: considerations for implementing United States Pharmacopeia Chapter <797> Pharmaceutical Compounding – Sterile Preparations, Part 14: Environmental quality and control (continued). *IJPC* 2009; 13(4): 322–329.
13. Allen LV, Okeke C. Basics of compounding: considerations for implementing USP Chapter <797> – Sterile Preparations, Part 19: Storage and beyond-use dating, maintaining sterility, purity, and stability of dispensed and distributed compounded sterile preparations. *IJPC* 2010; 14(3): 232–239.
14. Allen LV, Okeke CC. Basics of compounding: considerations for implementing United States Pharmacopeia Chapter <797> Pharmaceutical Compounding – Sterile Preparations, Part 8: Stability and beyond-use dating. *IJPC* 2008; 12(4): 344–53.

15. Trissel LA. *Handbook on Injectable Drugs*, 16th edn. American Society of Health-System Pharmacists, 2011.
16. Allen LV, Okeke CC. Basics of compounding: considerations for implementing United States Pharmacopeia Chapter <797> Pharmaceutical Compounding – Sterile Preparations, Part 7: Microbial contamination risk levels, 12-hour use, and immediate use. *IJPC* 2008; 12(3): 252–257.
17. Allen LV, Okeke C. Basics of compounding: considerations for implementing United States Pharmacopeia Chapter <797> Pharmaceutical Compounding – Sterile Preparations, Part 18: Verification of automated compounding devices for parenteral nutrition compounding and finished preparation release checks and tests. *IJPC* 2010; 14(2): 142–149.
18. Department of Health and Human Services; Centers for Disease Control and Prevention; National Institute for Occupational Safety and Health (2010). NIOSH List of Antineoplastic and Other Hazardous drugs in Healthcare Settings. <http://www.cdc.gov/niosh/docs/2010-167/>. (accessed 11 November 2011).
19. Allen LV, Okeke CC. Basics of compounding: considerations for implementing United States Pharmacopeia Chapter <797> Pharmaceutical Compounding – Sterile Preparations, Part 12: Sterilization methods. *IJPC* 2009; 13(2): 146–148.
20. Allen LV, Okeke CC. Basics of compounding: considerations for implementing United States Pharmacopeia Chapter <797> Pharmaceutical Compounding – Sterile Preparations, Part 5: Verification of accuracy, sterility, and sterilization of compounded preparations. *IJPC* 2008; 12(1): 54–60.
21. Allen LV, Okeke C. Considerations for implementing United States Pharmacopeia Chapter <797> Pharmacy Compounding Sterile Preparations, Part 17: Elements of quality control. *IJPC* 2010; 14(1): 60–64.
22. Allen LV, Okeke C. Basics of compounding: considerations for implementing USP Chapter <797> Pharmacy Compounding – Sterile Preparations, Part 20: Patient or caregiver training, patient monitoring and adverse events reporting, and quality-assurance program. *IJPC* 2010; 14(4): 328–331.
23. Okeke CC, Allen LV. Basics of compounding: considerations for implementing United States Pharmacopeia Chapter <797> Pharmaceutical Compounding – Sterile Preparations, Part 16: Suggested standard operating procedures. *IJPC* 2009; 13(6): 546–553.
24. Allen LV, Okeke CC. Basics of compounding: considerations for implementing United States Pharmacopeia Chapter <797> Pharmaceutical Compounding – Sterile Preparations, Part 9: Immediate-use compounded sterile preparations. *IJPC* 2008; 12(5): 440–443.
25. Allen LV, Okeke CC. Basics of compounding: considerations for implementing United States Pharmacopeia Chapter <797> Pharmaceutical Compounding – Sterile Preparations, Part 12: Sterilization methods. *IJPC* 2009; 13(1): 56–62.

APPENDIX A

Suggested Standard Operating Procedures²

1. CSP Microbial Contamination Risk Levels
 - a. Low-Risk, Medium-Risk, and High-Risk
2. Personnel Training and Evaluation in Aseptic Manipulation Skills
3. Immediate Use CSPs or 12 Hours CSPs (if compounded at your institution)
4. Single-dose and Multiple-dose Containers
5. Hazardous Drug Compounding (if compounded at your institution)
6. Radiopharmaceutical Use (if compounded at institution)
7. Allergen Extracts (if compounded at institution)
8. Verification of Compounding Accuracy and Sterility Processes
 - a. Sterilization Methods by Risk Levels
 - b. Depyrogenation Processes
9. Environmental Quality and Controls
 - a. Critical Site Management
 - b. PEC, Buffer Area, Ante-Room, and ISO Class Levels
- c. Viable and Nonviable Air Testing
- d. Particle Counts
- e. Pressure Differential Monitoring
- f. Growth Media used in Testing
- g. Facility Design and Environmental Controls
- h. Placement of PEC
- i. Cleaning and Disinfecting the Sterile Compounding Areas
- j. Personnel Cleansing and Garbing
- k. Personnel Training and Competency Evaluations and Procedures
 - l. Action Levels, Documentation, and Data Evaluation
10. Finished Preparation Release Checks and Tests
11. Storage and Beyond Use Dating
12. Maintaining Sterility, Purity, and Stability of Dispensed and Distributed CSPs
13. Patient and Caregiver Training
14. Patient Monitoring and Adverse Event Reporting

From United States Pharmacopeia 31 – National Formulary 26 (2008). Chapter USP <797> Pharmaceutical compounding sterile preparations.

Chapter 28

Ophthalmic Preparations

Masood Chowhan, PhD; John C. Lang, PhD and Paul Missel, PhD

INTRODUCTION	541	SUBCONJUNCTIVAL INJECTIONS	551
ANATOMY AND PHYSIOLOGY OF THE EYE	542	RETROBULBAR INJECTIONS	551
BIOAVAILABILITY	545	PREPARATION	551
TYPES OF OPHTHALMIC DOSAGE FORMS	547	PHARMACY COMPOUNDING	552
DRUG ADMINISTRATION	549	STERILIZATION	552
METHODS OF USE FOR TOPICAL ADMINISTRATION	550	OPHTHALMIC PREPARATION CHARACTERISTICS	553
IRRIGATING SOLUTIONS	550	PACKAGING	555
INTRAOCULAR INJECTIONS	550	ANTIMICROBIAL PRESERVATIVES	556
INTRAVITREAL INJECTIONS/IMPLANTS	551	OTC PRODUCTS FOR DRY EYE	558
JXTASCLERAL INJECTIONS	551	CONTACT LENSES AND THEIR CARE	559
INTRACAMERAL INJECTIONS	551	GUIDELINES FOR SAFETY AND EFFICACY TESTING	562
IONTOPHORESIS	551	SUMMARY	562

INTRODUCTION¹

Ophthalmic preparations are specialized dosage forms designed to be instilled onto the external surface of the eye (topical), administered inside the eye (intraocular) or adjacent to it (periocular, e.g., juxtasceral or subtenon), or used in conjunction with an ophthalmic device. The latter include preparations used in conjunction with surgical implantation (such as an intraocular lens) and dry eye formulations compatible with a punctal appliance (e.g., a punctal plug), and extends to a variety of solutions used in the maintenance of contact lenses. The preparations may have any of several purposes (e.g., therapeutic, prophylactic, or palliative for topically administered agents) but include mechanical, chemical, and biochemical actions of agents used in the care of ocular appliances and tissue prophylaxis during or following surgery. Because of the dangers associated with the administration or repetitive administration of intraocular and periocular preparations, their suitability is restricted to therapeutic applications or surgical adjuncts.

The versatility of dosage forms of ophthalmic preparations allows the clinician to choose the form most suitable for the function desired. Therapeutically active formulations can be designed to provide extended action for convenience or for reduction in risk of repetitive administration, improved bioavailability of the agent, or improved delivery to a targeted tissue. The residence of an ocular preparation can range from the few seconds needed for tears to clear an irritating substance; to hours for a gel, a gel-forming solution, or an ointment; to months or years for an intraocular or periocular dosage form. A preparation may be strictly therapeutic or may serve in prophylaxis. The latter includes surgical adjuncts to maintain the health of fragile cells, and postsurgical or post-trauma preparations designed to prevent or reduce the likelihood of infection. Another form of prophylaxis, one for a device, is the antisoiling function provided by some contact lens solutions.

Ophthalmic preparations are similar to parenteral dosage forms in their requirement for sterility as well as considerations for osmotic pressure (tonicity), preservation, tissue compatibility, the avoidance of pyrogens in intraocular dosage forms, particulate matter, and suitable packaging.

Topical therapeutic dosage forms have customarily been restricted to solutions, suspensions, and ointments. But with advances in materials science, the range of ophthalmic dosage forms has expanded significantly to include gels, either preformed or spontaneous gels responsive to the ocular environment, and ocular inserts, both forms reducing dosage frequency. These are most often multidose products, containing suitable preservative(s) to meet compendial preservative effectiveness test [e.g., *US Pharmacopeia* (USP),² *European Pharmacopoeia*,³ or *Japanese Pharmacopoeia*⁴] requirements. Now, however, single-dose units (also referred to as unit-dose products) that are preservative-free preparations have become available, generally packaged in form-fill-seal plastic containers with 0.25 mL to up to 0.8 mL, have become available. These unit-dose containers are designed to be discarded after a single use or after a single day's use if the container has a reclosable feature and the product is so labeled.

Injections and implants have been developed for intraocular drug delivery. Irrigating solutions and viscoelastic gels are available specifically for adjunctive use in ophthalmic surgery. Specialized formulations are now available for use in the care of contact lenses. The designs of these preparations, meeting all of the requirements for safety, efficacy, component compatibility, tissue acceptability, storage, shipping, and shelf-life, are beyond the scope of this review. Nonetheless, a description of the requirements and the designs for some of these formulations should be illustrative and didactic.

From a historical perspective, preparations intended for treatment of eye disorders can be traced to the writings of the Egyptians, Greeks, and Romans. In the Middle Ages, *collyria* were referred to as materials that were dissolved in water, milk, or egg white and used as eyedrops. One such collyrium contained the mydriatic substance belladonna to dilate the pupils of milady's eyes for cosmetic purpose.

From the time of belladonna collyria, ophthalmic technology progressed at a pharmaceutical snail's pace until after World War II. Before World War II and into the 1950s, ophthalmic preparations were mostly compounded by the pharmacist for immediate use. Not until 1953 was there a legal requirement by the US Food and Drug Administration (FDA) that all manufactured ophthalmic solutions be sterile. The range of medicinal

agents to treat eye disorders was limited, as was the state of eye surgery and vision correction, which was limited to eyeglasses. In the past 50 years, a modern pharmaceutical industry specializing in ophthalmic preparations has developed to support the advances in diagnosis and treatment of eye diseases, in eye surgery, and in contact lens design and use. Because of the variety of ophthalmic products readily available commercially, the pharmacist now is rarely required to compound a patient's ophthalmic prescription. More important, however, is that the pharmacist appreciate even subtle differences in formulations that may impact efficacy, comfort, compatibility, or suitability of a preparation for particular patients.

Currently and in the future, in addition to the advances in dosage-form technology, drug molecules will be designed and optimized specifically for ophthalmic application. New therapies may become available for preventing blindness caused by degenerative disease – including age-related macular degeneration (AMD), macular edema, and diabetic retinopathy. Biotechnological products may also become available to treat causes of multifactorial eye disorders like glaucoma. Such specialized therapeutic agents also will require carefully designed and compatible dosage forms.

Because dosage forms are fashioned to complement the requirements of the therapeutic agent, and the latter are selected for their action upon particular tissues so as to modify their function, we will now turn to a description of ocular tissues and their physiology.

ANATOMY AND PHYSIOLOGY OF THE EYE

In many ways the human eye is an ideal organ for studying drug administration and disposition, organ physiology, and function. Unlike many bodily organs, most of its structure can be inspected without surgical intervention. Its macroscopic responses can be investigated by direct observation. Its miraculous function, so intricate and complex, converting a physical electromagnetic stimulus into a chemical signal that is coupled to distant neurons for signal processing by an electrochemical wave, is detectable by sensitive instruments attached to external tissues. The basis for the function and protection of this important organ that links humans to

their external environment is the tissues composing it. The structures to be described are illustrated in Figures 28-1 to 28-3. Figure 28-1 provides a horizontal section of the eyeball identifying the major structures and their interrelationships.⁵ Figures 28-2 shows in greater detail the anterior portion of the eye and eyelids, in vertical section, emphasizing some of the structures associated with tear apparatus.⁶ Figure 28-3 emphasizes the flow of tears into the nasal structures.⁶ This brief introduction will focus on the anatomical structures composing the eye and their function.

EYELIDS

Eyelids serve two purposes: mechanical protection of the globe and creation of an optimum milieu for the cornea. The inner surfaces of the eyelids and the outermost surfaces of the eye are lubricated by the tears, a composite of secretions from both lacrimal glands and specialized cells residing in both the bulbar (covering the sclera) and palpebral (covering the inner surface of the lids) conjunctiva. The antechamber has the shape of a narrow cleft directly over the front of the eyeball, with pocket-like extensions upward and downward. The pockets are called the superior and inferior fornices (vaults), and the entire space, the conjunctival cul-de-sac. The elliptical opening between the eyelids is called the palpebral fissure, and the corner of the eyes where the eyelids meet are the canthi.

OVERVIEW OF STRUCTURE AND FUNCTION OF THE EYEBALL

Structure

The eyeball is housed in the bones of the skull, joined to form an approximately pyramid-shaped housing for the eyeball, called the orbit. The wall of the human eyeball (bulbus, globe) is composed of three concentric layers that envelop the fluid and lenticular core.⁷⁻⁹

OUTER FIBROUS LAYER

The outer scleral layer is tough and pliable but only slightly elastic. The anterior third is covered by the conjunctiva, a clear, transparent, mucous surface. The most anterior portion of the outer layer forms the cornea, a structure so regular and the

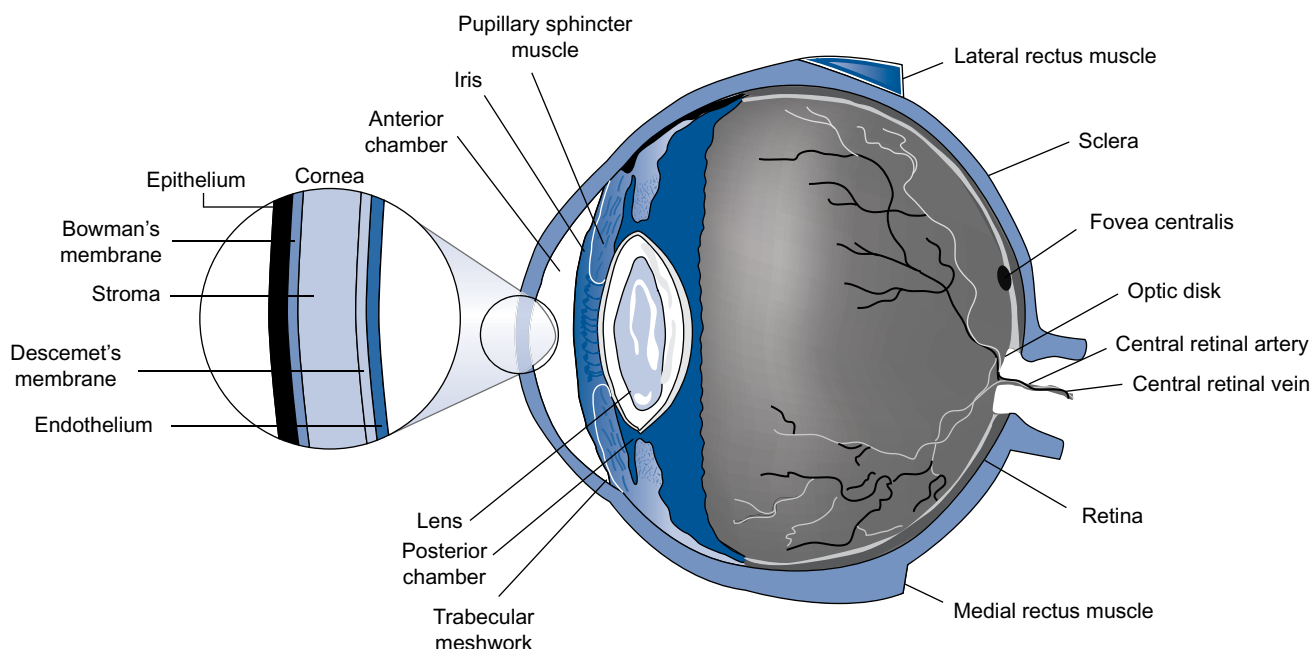


Figure 28-1. A cutaway horizontal section of the eyeball, illustrating the important anatomic structures and their interrelationships diagrammatically. The different layers of the cornea are illustrated in the magnified view. Relative sizes are suggestive and not proportional. The diameter of a mature eyeball is generally slightly greater than one inch. (Courtesy, Alcon Laboratories, Inc., Fort Worth, TX).

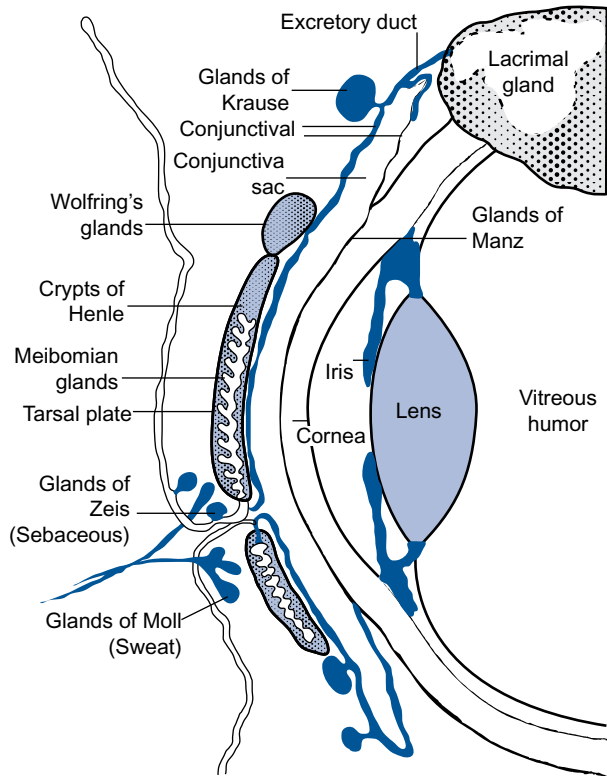


Figure 28-2. The front half of the eye, in vertical section, identifying the important structures associated with cornea and the front of the eye, including the eyelids and the glands associated with tears.

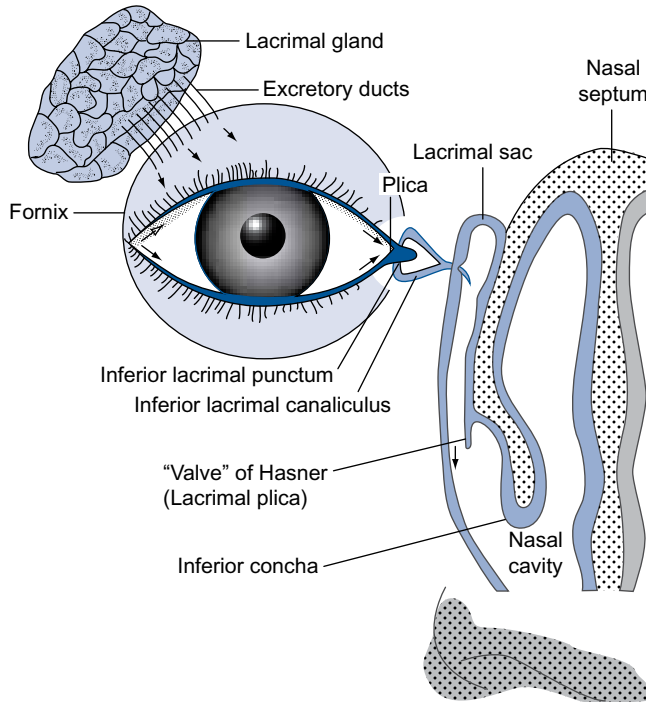


Figure 28-3. The structures associated with the tears and lacrimal flow and access to the nasolacrimal system.

water content so carefully adjusted that it acts as a transparent window. It is devoid of blood vessels. Over the remaining two thirds of the globe, the fibrous collagen-rich coat is opaque (the *white* of the eye) and is called the sclera. It contains the microcirculation, which nourishes the tissues of this anterior

segment, and is usually white except when irritated vessels become dilated.

The cornea, slightly thicker than the sclera and ranging in thickness from 500 micrometers to 1 mm, consists of five identifiable layers. Proceeding from the most anterior layer, these are the hydrophobic stratified squamous epithelium, which is underlaid by the Bowman membrane, then the stroma and the Descemet membrane, and then the innermost layer, the endothelium. The stroma is a hydrophilic elastic network of highly organized connective tissue and is the thickest layer of the cornea. The fibrous collagen-rich Descemet membrane separates the stroma from the single squamous-cell layer of endothelium, the location of the pump that keeps the cornea in its relatively dehydrated, transparent state. The cornea functions as a bilayer barrier, with the hydrophobic epithelium as the primary barrier to hydrophilic molecules and the hydrophilic stroma as the primary barrier to hydrophobic molecules. A schematic drawing of the cornea is provided in Figure 28-1 and its relation to other anterior tissues in Figure 28-2.

MIDDLE VASCULAR LAYER

The middle vascular layer, or uvea, provides nourishment to the eye and consists (from the back of the eye to the front) of the choroid, the ciliary body, and the iris. The choroid consists of a pigmented vascular layer, colored by melanocytes and traversed by medium-sized arteries and veins, with the choriocapillaris containing a network of small vessels that nourish the neural retina. The ciliary body contains muscles that control the extension of the lens, allowing visual accommodation, as well as the ciliary processes that secrete aqueous humor into the posterior chamber to maintain the intraocular pressure (IOP) that, in turn, keeps the eyeball fully expanded. The pigmented iris is a ring of muscular tissue around the pupil, a round centric hole that acts as a variable aperture to control pupil diameter and, thereby, the level of light entering the eye. The canal of Schlemm, one of the important paths for outflow of the aqueous humor, resides in the angle of the iris. Bruch's membrane separates the choroid from the retina.

NEURAL RETINA

This innermost layer of the eyeball is a complex tissue that supports the harvesting of light through the collective action of photoreceptors and nerve cells specialized for distinguishing white from black (rods) or for discerning color (cones). In addition, the retina consists of cells that support metabolism (like the heavily pigmented retinal pigmented epithelium, which purges photoreceptors of spent molecules and metabolites, and regenerates the *cis*-retinal), provide structure (astrocytes and Müller cells), or contribute to the primary function of photo-detection/signal processing (the ganglion cells that begin to process the electrochemical information transmitted from the photoreceptors).

OCULAR CORE

Within the globe, the crystalline lens spans the interior fluid-filled center close to the iris and is anchored by zonule fibers to the ciliary body. The lens is composed of a single layer of replicating epithelial cells that, with age, flatten into layers of long, thin crystalline-filled lamellar fibers. The lens is the only tissue in the body that retains all cells ever produced, a fact that contributes to age-related alterations in size, clarity, and extensibility. A tough, thin transparent membrane called the capsule covers the outermost layer of the lens.

The aqueous and vitreous humors are interposed between the solid structures of the eye. The clear, fluid aqueous humor fills the globe anterior to the lens and is primarily responsible for maintaining correct IOP. The gel-like vitreous humor accounts for most of the weight of the eye and resides posterior to the lens, in direct contact with the retina.

FUNCTION

The eyeball houses the optical apparatus that causes the formation on the neural retina of inverted, reduced images of the outside world.

Dimensional Stability

DIMENSIONAL STABILITY

The optical function of the eye calls for stability of its dimensions, which is provided partly by the fibrous outer coat but more effectively by the intra ocular pressure (IOP), which exceeds the pressure prevailing in the surrounding tissues. This IOP is the result of a steady production of specific fluid, the aqueous humor, which originates from the ciliary processes anterior to the lens and exits the eye by an intricate system of outflow channels. The resistance encountered during this passage and the rate of aqueous production are the principal factors determining the level of the IOP. In addition to this hydromechanical function, the aqueous humor acts as a carrier of nutrients, substrates, and metabolites for the avascular tissues of the eye.¹⁰

OPTICAL PATHWAY

The optical pathway consists, in sequence, of the precorneal tear film, the cornea, the aqueous humor, the pupil, the crystalline lens, the vitreous humor, and the retina. The chief refraction of light for the eye occurs at the outer surface of the cornea, where the index of refraction changes from that of air (1.00) to that of precorneal substance (1.38). After traversing the cornea, light passes through the clear aqueous humor to the pupil, where the amount of light entering the eye is regulated by the pupillary diameter, and to the second refractive element of the eye, the lens, whose variable focal length allows objects both near and far to be brought into focus (accommodation). The shape of the lens, controlled by the muscles of the ciliary body, refracts and focuses the reduced inverted image on the retina. That image is sharp and clear, in part because of the high transparency of the vitreous humor, which, because of its gel-like state, keeps debris and cells from entering the pathway for the light. The image formed by the electromagnetic light signal on the neural retina is converted to a chemical signal, changing *cis*-retinal to *trans*-retinal, which in turn is processed by neural cells into an electrochemical signal trans-ported through axons to central nerve cell bodies. Continuous maintenance of the photoreceptors and neural tissues in a state capable of supporting the conversion of the excitation associated with a retinal image to processable information is integral to the functioning of the retina and is the reason retinal tissue is among the most rapidly metabolizing tissues in the body. The dependence of the neural retina on the metabolic support provided by the underlying cell layer, the retinal pigmented epithelium, explains why damage to the tissue—such as detachment of the retina or diminished blood supply—can result in nearly immediate and permanent loss of vision.¹¹

TISSUES RESPONSIBLE FOR REFRACTION

Any alteration in the shape or transparency of the *cornea* interferes with the formation of a clear image; therefore, any pathological process, however slight, may interfere seriously with the resolving power or visual acuity of the eye. Transparency of the cornea is largely attributable to its organized laminar arrangement of cells and fibers, and the absence of blood vessels. The normal cornea possesses no blood vessels except at the sclerocorneal junction, the limbus. The cornea must therefore derive its nutrition by diffusion and must have certain permeability characteristics; it also receives nourishment from the fluid circulating through the chambers of the eye and from the air. That the normal cornea is devoid of blood vessels is an important feature in surgical grafting.

Cloudiness of the cornea may occur as a result of disease (e.g., excess pressure in the eyeball, a symptom associated with

glaucoma); the presence of scar tissue due to injury or infection; or deficiency of oxygen or excess hydration, such as may occur during the wearing of improperly fitted contact lenses. A wound of the cornea may heal as an opaque patch that can be a permanent impairment of vision unless it is located in the periphery of the cornea.

The corneal nerves do not supply all forms of sensation to the cornea, but pain and cold are well supplied. The pain fibers have a very low threshold, which makes the cornea one of the most sensitive areas on the surface of the body. It now is agreed that the cornea possesses a true sense of touch; nerve endings supplying the sensation of heat appear to be lacking.

The corneal epithelium provides an efficient barrier against bacterial invasion. Unless its continuity has been broken by an abrasion (a traumatic opening or defect in the epithelium), pathogenic bacteria, as a rule, cannot gain a foothold. Trauma, therefore, plays an important part in most of the infectious diseases of the cornea that occur exogenously. A means of detecting abrasions on the corneal surface is afforded by staining the cornea with sodium fluorescein. Any corneal abrasion is subject to infection.

As with the cornea, any change in the transparency of the *lens* as a result of age or disease can significantly affect visual clarity. Loss of flexibility of the lens can reduce visual accommodation and cause difficulty in focusing on near objects. Also, upon aging or as a result of trauma, the lens may generate opacities caused by the oxidation and crosslinking of lens proteins. When cataract surgery is required to restore clarity to cloudy vision, the natural lens is removed and replaced by an artificial one. The capsule, though, is preserved so that it can provide scaffolding for the implanted synthetic intraocular lens.

Lacrimal System

The conjunctival and corneal surfaces are covered and lubricated by a precorneal tear film, a fluid secreted by the conjunctival and lacrimal glands.¹¹ The clear, watery secretion of the lacrimal gland, delivered through a number of fine ducts into the conjunctival fornix, contains numerous salts, glucose, other organic compounds, and approximately 0.7 percent protein, including the enzyme lysozyme. Small accessory lacrimal glands are situated in the conjunctival fornices. The tear film, compatible with both aqueous and lipid ophthalmic preparations, is composed of a thin outer lipid layer, a thicker middle aqueous layer, and a thin inner mucoid layer. It is renewed during blinking and, when blinking is suppressed, may dry in patches. It seems to be unaffected by the addition of concentrations of up to 2 percent sodium chloride to conjunctival fluid. A pH below 4.0 or above 9.0 causes derangement of the film. The film affects the movement of contact lenses and forms more easily on hydrophilic than on hydrophobic prostheses.

The innermost mucin protein layer of the film is especially important in maintaining the stability of the film and is postulated to be held in place by the microvilli of the corneal epithelial cells. Sebaceous meibomian glands of the eyelids secrete an oily fluid that forms the outer layer of the tears, helps to prevent overflow at the lid margin, and reduces evaporation from the exposed surfaces of the eye.

Spontaneous blinking replenishes the fluid film by pushing a thin layer of fluid ahead of the lid margins as they come together. The excess fluid is directed into the lacrimal lake—a small, triangular area lying in the angle bound by the innermost portions of the lids. The skin of the eyelids is the thinnest in the body and folds easily, thus permitting rapid opening and closing of the palpebral fissures, at velocities of tens of centimeters per second. The movement of the eyelids includes a narrowing of the palpebral fissures in a zipper-like action from lateral to medial canthus. This aids transport or movement of fluid toward the lacrimal lake and elimination of unwanted contaminants.

Tears are drained from the lacrimal lake through two small openings, the superior and inferior puncta, which drain into connecting small tubes, the lacrimal canaliculi, which

themselves join at the common canaliculus that leads into the upper part of the nasolacrimal duct, the beginning of which is the lacrimal sac, as shown in Figure 28-3. The drainage of tears into the nose does not depend merely on gravity. Fluid enters and passes along the lacrimal canaliculi by capillary action, aided by aspiration resulting from contraction of muscles embedded in the eyelids, and by peristalsis in the muscles near the canaliculi. When the lids close, as in blinking, contraction of the muscle causes dilatation of the upper part of the lacrimal sac and compression of its lower portion. Tears aspirated into the sac are forced down the nasolacrimal duct toward its opening into the nose. As the lids open, the muscle relaxes. The upper part of the sac then collapses and forces fluid into the lower part, which at the same time is released from compression. Thus, the act of blinking exerts a suction force-pump action in removing tears from the lacrimal lake and emptying them into the nasal cavity. Lacrimation is induced reflexively by stimulation of nerve endings of the cornea or conjunctiva. This reflex is abolished by anesthetization of the surface of the eye and by disorders affecting its nerve function.

The normal cul-de-sac is maintained free of pathogenic organisms in part by the chemical action of enzymes, such as lysozyme, which normally destroy saprophytic organisms with limited action against pathogens, and in part by the continuous physical flow of normally sterile secretions, which constantly wash the bacteria, dust, etc., away from the eye down into the nose. In certain diseases or upon aging, the lacrimal gland, like other glandular structures in the body, may undergo involution, with the result that the lacrimal fluid becomes scanty. Changes in the conjunctival glands may lead to alteration in the character of the secretion, so that quality as well as quantity of tears may be abnormal. This can lead to symptoms of dryness, burning, and general discomfort, and ultimately may interfere with visual acuity.

Summary

This brief overview of the structure and function of the eye should provide a basis for understanding the highly integrated tissues comprising this miraculous organ and suggest the importance of providing medications that in no way impair the balance of functions required for maintaining normal functioning of the eye.

BIOAVAILABILITY

THERAPEUTIC TARGETS

Bioavailability of pharmacological agents is dictated by the ocular structure and physiology just discussed. But bioavailability also is controlled by physical constraints and tissue biochemistry (to be discussed in the next subsections), by physical and chemical characteristics of the therapeutic agents, and by the preparations by which they are presented. The bioavailability and potential efficacy of an agent also are determined by the therapeutic targets, which are governed by disease etiology.

For example, treatment of a superficial infection of the cornea, while possibly requiring sustained delivery to enable less frequent administration of antimicrobial therapy, does not involve special considerations of corneal permeation or access to the target tissue. In this circumstance, product design might be directed toward the reduction of any corneal or scleral transport, in that these would be regarded as drug lost from the target site. On the other hand, a recalcitrant case of uveitis may require both topical and systemic administration of anti-inflammatory agents so as to eliminate the condition. In this case, however, transport to the uvea, an internal tissue, is desirable.

Chronic diseases associated with aging, like glaucoma, may be amenable to routine administration of a topical medication. The target, however, may be determined by the particular mechanism for treating the disease. For instance, a drug influencing the generation of aqueous humor may target the iris

ciliary body, whereas a drug influencing the outflow of aqueous humor might target the trabecular meshwork that opens into the canal of Schlemm. The inherent characteristics of the drug also will determine, in part, any need for sustained delivery. In general, a lipophilic agent will be absorbed readily into the lipophilic corneal epithelium, whereas an ionic or hydrophilic agent will be absorbed more slowly. But corneal permeation may still be a significant barrier to delivery at the target tissue even for a lipophilic drug, because transport through the largely aqueous stroma is still required. The capacity of the epithelium for the drug, in part controlled by its partitioning characteristics, may govern any need for sustained administration as well as the dosing regimen. The solubility of the drug, especially for often relatively insoluble hydrophobic lipophilic agents, can limit the size of the reservoir of drug administered, or the exposure to the drug, where the latter is defined as the area under the curve in a concentration-vs.-time plot of drug residence.

Recently, treatments for ocular diseases of the deep tissues have invoked techniques more commonly utilized in targeted systemic delivery. For example, the treatment of retinal cytomegalovirus infection, accompanying end-stage HIV infections, has engendered treatments delivering drug directly to the vitreous humor, with either intracameral injection or an implant. Photodynamic therapy utilizes a radiation (generally light) to activate a drug delivered systemically. Angiostatic agents for the treatment of wet AMD are delivered from an implant. As Higuchi, years ago, and others more recently have described, the rate of delivery from such devices can be controlled by the design of the device, by drug characteristics (primarily solubility), or by a balance of both.¹²

Nonetheless, eyedrops remain the most common modality for administration of therapeutic agents, and so considerations of the quantitative relationships governing the balance of effects controlling access of these agents to their target tissues are of significance and will be summarized in the next few sections. The approach will be primarily hydrodynamic,^{13,14} though molecular theories are now available to give rationale for the macroscopic transport laws.¹⁵ More intricate transport characteristics, such as those occurring when membranes actively transport the therapeutic agent, are less common than simple passive diffusion and therefore will be neglected here. The approach highlights the interrelationships of phenomena whose subtle control in commercial products improves their efficacy. While transmembrane transport is most relevant to transcorneal delivery, some of the same considerations apply to active drugs delivered from implants, because these drug reservoirs serve only as a source for the agents that must traverse tissues or tissue boundaries to reach their target sites.¹⁶

CORNEAL ABSORPTION AND DRUG ACCESS

Physical and Chemical Considerations

Under normal conditions the human tear volume averages about 7 microliters.¹⁷ The estimated maximum volume of the cul-de-sac is about 30 microliters, with drainage capacity far exceeding lacrimation rate. The outflow capacity accommodates the sudden large volume resulting from the instillation of an eyedrop. Most commercial eyedrops range from 25 to 50 microliters in volume.

Within the rabbit cul-de-sac, the drainage rate has been shown to be proportional to the instilled drop volume. Multiple drops administered at intervals produced higher drug concentrations. Ideally, a high concentration of drug in a minimum drop volume is desirable. Patton¹⁸ has shown that approximately equal tear film concentrations result from the instillation of 5 microliters of 1.61×10^{-2} M pilocarpine nitrate or from 25 microliters of 1.0×10^{-2} M solution. The 5 microliters contains only 38 percent as much pilocarpine, yet its bioavailability is greater because of decreased drainage loss. Human responses can be expected to be similar, and this is supported by studies of gamma scintigraphy.¹⁹ However, human tear responses are

much more significant, and variable, than those in some *in vivo* models. When the therapeutic agents themselves are irritating, excess tearing may occur that may influence bioavailability in affected individuals. There is a practical limit to the concept of minimum dosage volume. There is a difficulty in designing and producing a dropper configuration that will deliver small volumes reproducibly.²⁰ Also, the patient often cannot detect the administration of such a small volume. This sensation, or lack of sensation, is particularly apparent at the 5.0- to 7.5-microliter dose-volume range.

The concept of dosage-volume drainage and cul-de-sac capacity directly affects the prescribing and administering of separate ophthalmic preparations. The first drug administered may be diluted significantly by the administration of the second. On this basis, combination drug products for use in ophthalmology have considerable merit.

Though this is generally not a major concern, an instilled drug can be subject to protein binding in the tear fluid and metabolic degradation by enzymes such as lysozyme. Stability, however, is always a primary consideration, both in the bottle and in the tissues. Rapid conversion of a drug to a metabolite is generally to be avoided unless by design (prodrugs).

Corneal Absorption

Penetration of drugs administered topically occurs primarily through the cornea. Drugs administered by instillation must penetrate the eye and do so primarily through the cornea. Corneal absorption is generally more effective than scleral or conjunctival absorption, in which drug is removed by circulatory flow into the general circulation. This is also true because most therapeutic agents tend to be reasonably lipophilic. In the case of more hydrophilic materials, which as mentioned earlier are more slowly absorbed into the corneal epithelium, there is evidence to suggest that a scleral route may be more common. A practical example comes from studies of carbonic anhydrase inhibitors,²¹ a class of drug used in the treatment of glaucoma.

Many ophthalmic drugs are weak bases and are conveniently applied to the eye as aqueous solutions of their salts. If the molecule is maintained in an ionized state, for example, by employing a weakly acidic buffered vehicle, stability of the drug often can be prolonged. If the buffer is weak, little discomfort is generally experienced at the time of instillation. Once neutralized by the tears, the fraction of free base may increase, and this may be the form of the drug most readily absorbed by the corneal epithelium. Nonetheless, during transport through the hydrophilic stroma, an important fraction of the agent may be ionized. Similarly, the form transported through the endothelium may be the free base, and that presented to the ciliary body (one of the sites of pharmacological action) in transport from the aqueous humor may be the ionized salt. This scenario is described simply to suggest the complexity that can be exploited in targeting an ophthalmic drug, as well as complementary considerations involved in providing a preparation capable of the requisite stability for a conventional two-year shelf-life.

The cornea can be penetrated by ions to a small but measurable degree. Under comparable conditions, the permeabilities are similar for all ions of low molecular weight, which suggests that the passage is through extracellular spaces. The diameter of the largest particles that can pass across the cellular layers seems to be in the range of 10 to 25 Å. Some effort has been directed toward increasing the upper molecular-weight limit for water-soluble therapeutic agents by utilizing a class of molecules referred to as penetration, or permeation, enhancers. These molecules are selected for their capacity to increase transiently permeation of larger molecules while producing minimal discomfort or toxicity.

Because highly water-soluble drugs generally penetrate the cornea less readily and the cornea is known to be a membrane including both hydrophilic and lipophilic barrier layers, most

effective penetration is obtained with drugs having both lipophilic and hydrophilic properties. As an example, highly water-soluble steroid phosphate esters penetrate the cornea poorly. Better penetration is achieved with the poorly soluble but more lipophilic steroid alcohol; still greater absorption is seen with the steroid acetate form.

Work on transport of drugs through the cornea, dating from the work of Edelhauser *et al.* in 1965,¹³ and Lee and Robinson in 1979,²² and Lee in 1990,²³ has indicated the interplay of physical properties that regulate access of drugs to ophthalmic tissues. The hydrodynamic principles are well known and will be outlined here. From the physical chemistry of passive diffusion across a simple membrane capable of sustaining a concentration difference (and discussed in other chapters of this volume), the main mechanism of entry into the eye,

$$J = -D dC_m/dx \quad (1)$$

according to Fick's first law of diffusion, where J is the flux of drug across the membrane (amount per area per time), D is the diffusion coefficient, and the derivative of C_m is the spatial concentration gradient in the membrane (at steady state, constant across the membrane).²⁴ Even from this simple relationship, the consequences of the physical properties of the drug can be inferred. As the solubility of the drug increases the gradient will increase, so the driving force for entry of the agent into the aqueous humor will be increased. Experiment similarly has shown specific characteristics of the diffusion coefficient; namely, the diffusion coefficient decreases with increasing molecular size (and hence molecular weight of the compound). So all else being equal, we need to balance the increased specificity and targeting of a larger active molecule against the loss in rate of membrane transport. The equation also makes clear that any means for maintaining the concentration on the donor side of the membrane, thereby sustaining the gradient term for a longer period of time, will increase the mean total flux.

To perceive the consequences of partitioning in a more complicated bilayer membrane (a good model for the cornea in that the major layers serving as barriers to drug entry are the epithelium and stroma), we need to look at the diffusion equation for a bilayer:

$$J \cong \frac{PC_w}{\frac{Pl_s}{D_s} + \frac{l_e}{D_e}} \quad (2)$$

where P is the distribution coefficient (e.g., the octanol:water partition coefficient, wherein it is a good approximation of the partitioning into the epithelium), the subscripts designate either stroma (s) or epithelium (e), l is the length representing the thickness of the layer, and C_w is the concentration of drug at the donor side of the membrane, i.e., in the concentration of drug in the tears.²⁵ This equation is an approximation to the differential equation. One source of approximation is that it presumes that the concentration on the donor side (in the eye, the concentration of active ingredient in the tears) is much higher than on the receiver side of the membrane (the concentration in the aqueous humor). From this equation we can observe that as the partitioning increases, the effectiveness of the epithelium as a barrier diminishes and is replaced nearly exclusively by the stroma. Similarly, as the partitioning decreases, the epithelium becomes the dominant barrier. Of course, this is just a mathematical representation of an anticipated phenomenon. Clearly there is an advantage to increasing the lipophilicity of the agent so as to improve absorption and transport. A precaution to be kept in mind, however, is that ordinarily when the lipophilicity is increased, the aqueous solubility (which C_w may approach) generally decreases. This is an example of the need to balance a number of characteristics, as mentioned earlier.

There are circumstances when sustaining the presence and concentration of the drug in the tear volume is important, especially when the drug is sufficiently hydrophilic to severely limit absorption directly into the corneal epithelium. From the equations for a stirred-tank chemical reactor, which is undergoing a

steady-state flow of fluid with a solvent entering and a solution of mixed chemicals flowing from it, we expect:

$$C_w(t) = C_i \cdot \exp\left(\frac{-\dot{V}_T \cdot t}{V_T}\right) \quad (3)$$

where the time dependence of C_w is explicitly indicated, C_i is the initial concentration, V_T is the volume of the reactor (here the tear volume), and its derivative is the rate of flow through the reactor, essentially the rate of tear generation. A number of characteristics of drug delivery are apparent from this equation. One is that if the drug causes discomfort and increases the rate of tearing ($\dot{V}_T$ in the equation), then the concentration in the tears is depleted more rapidly. Analogously, if the preparation includes a means for retaining a drug carrier in the cul-de-sac, then only a fraction of the drug present is released free into the tears, and $C_w(t)$ no longer changes with time and, until the reservoir is depleted, may be approximately constant. Because what is important is the total amount of drug reaching the target tissue over a time period,

$$N_{\text{eff}} = A \cdot \int dt J(t) \quad (4)$$

the effect of reducing the loss rate from tear flow may counterbalance the consequence of reducing the total level in the tears. Note, for situations in which there is a means for sustaining drug release, Equations 2 and 3 have to be modified in any event, because C_i is no longer the total initial concentration but the fraction in the tears that is transportable through the cornea. This introduction should provide some notion of the types of complexity involved in achieving improved delivery of a therapeutic agent. When more is known about the details of ocular function, characteristics of the therapeutic agents, and vehicles, it is possible to solve the differential equations numerically so as to improve the design of topically applied medications.^{26–28}

DELIVERY TO THE OCULAR POSTERIOR

Optimizing drug delivery to the ocular posterior requires understanding the mechanisms by which substances are cleared by diffusion, by physiological processes of hydraulic, vascular, and lymphatic flows.^{29,30} Computational fluid dynamics can be a useful tool for predicting the influences of processes on the clearance of intravitreally injected substances³⁰ and delivery from devices implanted into the vitreous³¹ or immediately outside the globe in the juxtasclear space.³² This entails construction of an anatomically accurate model for the eye, which enables simulation of diffusion of drug in physiological flow fields with clearance mechanisms engaged. An example of such a simulation exercise modeling clearance of an intravitreal depot of triamcinolone suspension particles is illustrated in Figures 28.4 and 28.5. The significance of these activities, a subject of ongoing research, is difficult to overstate since the diseases being addressed—AMD, macular edema, and diabetic retinopathy—are sight-threatening. The current standard of treatment for AMD requires monthly injections of an antibody against vascular endothelial growth factor. Significant efforts are under way to develop alternate therapies to reduce the frequency of administration or implantation, providing durations of delivery ranging from months to years. A practical consequence is that the testing period for an implanted device is extended appreciably.

Finally, though the approach described in this section has been primarily macroscopic and phenomenological, microscopic models and calculations of transport based on them have provided additional detail about the molecular features of ocular drug transport. For example, it is generally believed that lipophilic drugs permeate the corneal epithelium by transcellular transport, whereas more hydrophilic drugs permeate the corneal epithelium by paracellular transport. A model of these different mechanisms has indicated that the transcellular transport probably is along the lipophilic bilayers of the cell outer membranes, whereas the paracellular transport is probably along

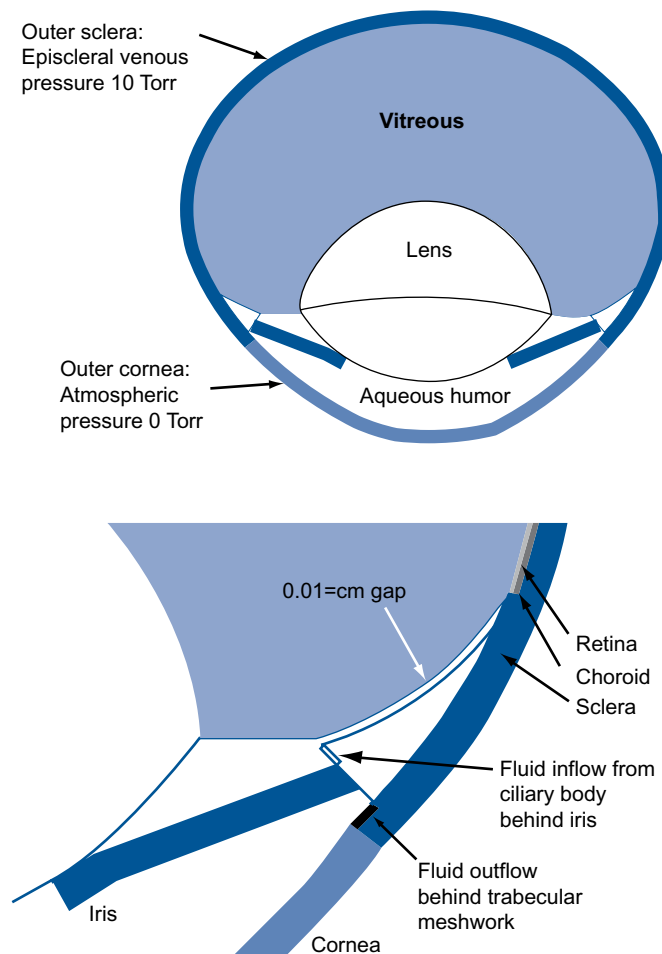


Figure 28-4. Ocular geometry used for computational fluid dynamic simulation of clearance of intravitreally injected substances. All tissue regions are treated as porous media except for the aqueous humor, which is treated as a simple liquid. The small gap between the ciliary body represents the Canal of Petit, a feature which becomes important for accurate simulation of the clearance of slowly diffusing materials. Aqueous fluid is produced at a rate of 3 μL / minute on the blue boundary of the iris as shown. Most of the fluid is eliminated on the outflow boundary behind the trabecular meshwork, simulating conventional outflow. A small amount of the fluid percolates throughout the remaining ocular structure. (From Missel PJ *et al.* Simulating dissolution of intravitreal triamcinolone acetonide suspensions in an anatomically accurate rabbit eye model. *Pharm Res* 2010; 27: 1530–1546.)

the extracellular space, the “pores” of a macroscopic model.²⁴ These models provide insight into mechanisms of transport and how these may be manipulated to improve drug delivery while maintaining safety.

TYPES OF OPHTHALMIC DOSAGE FORMS

Ophthalmic products include prescription and over-the-counter (OTC) drugs, products for the care of contact lenses, and products used in conjunction with ocular surgery. This section will focus on the pharmaceutical aspects of the various ophthalmic dosage forms encompassed by these types of products. The therapeutic uses of individual products can be found in several reference books along with the individual product’s labeling.^{33,34}

OPHTHALMIC SOLUTIONS

These are by far the most common dosage forms for delivering drugs to the eye. By definition, ingredients are completely soluble such that dose uniformity is not an issue, and there is little physical interference with vision. The principal disadvantage of solutions is their relatively brief contact time with the

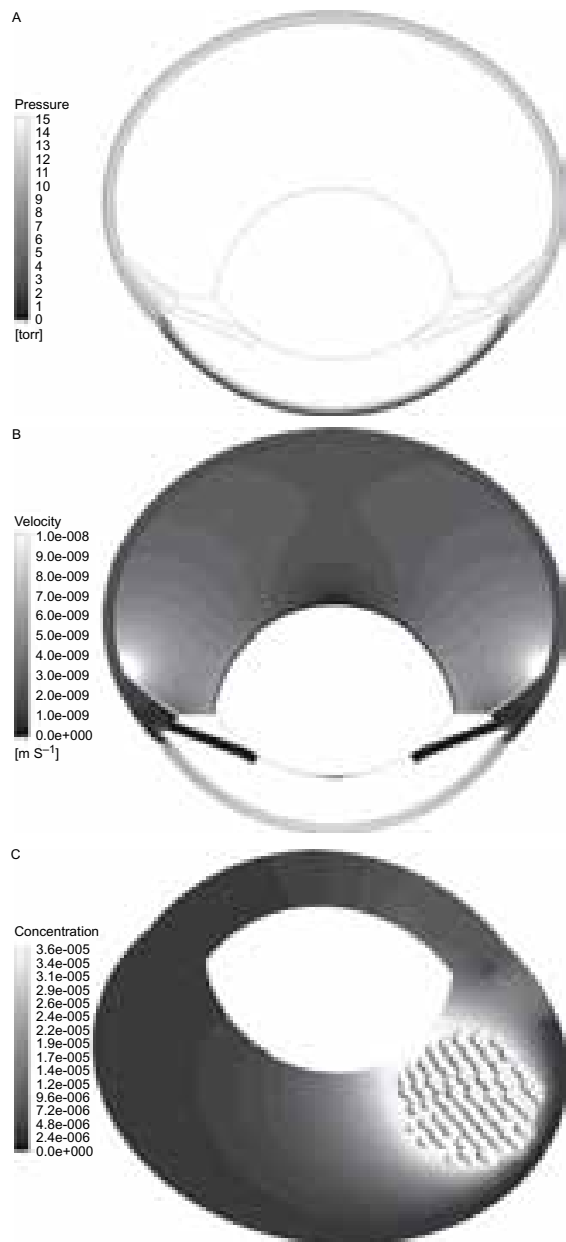


Figure 28-5. Computational fluid dynamic simulation results in the model for the rabbit eye. (a) Pressure distribution within a normotensive eye (Torr). (b) Velocity, clipped to a maximum value of 10^{-8} m/s to illustrate contours in the vitreous. (c) Dissolved triamcinolone acetonide concentration (grams/mL) after the dissolution process has achieved steady state for a suspension of 16 mg triamcinolone acetonide divided into 276 particles confined to a 100 microliter intravitreal depot. (From Missel PJ *et al.* Simulating dissolution of intravitreal triamcinolone acetonide suspensions in an anatomically accurate rabbit eye model. *Pharm Res* 2010; 27: 1530–1546.)

drug and the absorbing tissues of the external eye. Contact time can be increased by the inclusion of viscosity-imparting agents; however, their use is limited to relatively low viscosities so as to allow dispensing of the eyedrop from the container or eyedropper and to minimize excessive blurring of vision. A residue can be produced on the eyelashes and around the eye when any excess of a viscous solution spills out of the eye and dries. The residue usually can be removed easily by careful wiping with a moist towel to the closed eye.

GEL-FORMING SOLUTIONS

Ophthalmic solutions (usually water-based), which contain a polymer system that is a low-viscosity liquid in the container

but gels on contact with the tear fluid, have increased contact time and can provide increased drug absorption and prolonged duration of therapeutic effect. The liquid-to-gel phase transition can be triggered by a change in temperature, pH, ionic strength, or presence of tear proteins, depending on the particular polymer system employed. Timolol maleate gel-forming solutions formulated with specific patented gellan or xanthan gums have clinically demonstrated prolonged duration of IOP lowering, such that their dosing frequency can be reduced from twice to once a day.^{35,36} Review the labeling for commercial gel-forming solutions before dispensing for current instructions related to patient administration.

POWDERS FOR SOLUTIONS

Drugs that have very limited stability in aqueous solution can sometimes be prepared as sterile powders for reconstitution by the pharmacist before dispensing to the patient. The sterile powder should be aseptically reconstituted with the accompanying sterile diluent that has been optimized for dissolution, preservation, and stability. The pharmacist must convey to the patient any special storage instructions, including the expiration date.

OPHTHALMIC SUSPENSIONS

Suspensions are dispersions of finely divided, relatively insoluble drug substances in an aqueous vehicle containing suitable suspending and dispersing agents. The vehicle is, among other things, a saturated solution of the drug substance. Because of a tendency of particles to be retained in the cul-de-sac, the contact time and duration of action of a suspension could theoretically exceed that of a solution. The drug is absorbed from solution, and the solution concentration is replenished from retained particles. Each of these actions is a function of particle size, with solubility rate being favored by smaller size and retention favored by a larger size; thus, optimum activity should result from an optimum particle size.

For aqueous suspensions the parameters of intrinsic solubility and dissolution rate must be considered. The intrinsic solubility determines the amount of drug actually in solution and available for immediate absorption upon instillation of the dose. As the intrinsic solubility of the drug increases, the concentration of the drug in the saturated solution surrounding the suspended drug particle also increases. For this reason, any comparison of different drugs in suspension systems should include their relative intrinsic solubilities. The observed differences in their biological activities may be ascribed wholly or in part to the differences in this physical parameter. As the drug penetrates the cornea and the initial saturated solution becomes depleted, the particles must dissolve to provide a further supply of the drug. The requirement here is that the particles must undergo significant dissolution within the residence time of the dose in the eye if any benefit is to be gained from their presence in the dosing system.

For a drug whose dissolution rate is rapid, the dissolution requirement may present few problems, but for a slowly soluble substance the dissolution rate becomes critical. If the dissolution rate is not sufficiently rapid to supply significant additional dissolved drug, there is the possibility that the slowly soluble substance in suspension provides no more drug to the aqueous humor than does a more dilute suspension or a saturated solution of the substance in a similar vehicle. Obviously, the particle size of the suspended drug affects the surface area available for dissolution. Particle size also plays an important part in the irritation potential of the dosing system. This consideration is important, because irritation produces excessive tearing and rapid drainage of the instilled dose, as discussed earlier. It has been recommended that particles be smaller than 10 microns to minimize irritation to the eye. It should be kept in mind, however, that in any suspension system the effects of prolonged storage and changes in storage temperature might cause the smallest particles to dissolve and the largest particles to become larger.

The pharmacist should be aware of two potential difficulties inherent in suspension dosage forms. In the first instance, dosage uniformity nearly always requires brisk shaking to distribute the suspended drug. Adequate shaking is a function not only of the suitability of the suspension formulation but also—and most importantly—patient compliance. Studies have demonstrated that a significant number of patients may not shake the container at all; others may contribute a few trivial shakes. The pharmacist should use a “Shake Well” label and counsel the patient whenever an ophthalmic suspension is dispensed. An improved ophthalmic suspension has been developed for insoluble drugs such as steroids, which tend to cake upon settling.³⁷ The improved suspension controls the flocculation of the drug particles such that they remain substantially resuspended for months and provides for easy resuspension of any settled particles. Nonetheless, the pharmacist also should be aware of the possibility of crystal growth over time. This potential stability problem is especially problematic for drug substances whose solubility is significantly dependent on temperature. The majority of suspension products have a “Do Not Freeze” warning on the label, because they are likely to agglomerate on freezing and will not be resuspended by simple shaking.

A second and infrequently occurring characteristic of suspensions is the phenomenon of polymorphism, or the ability of a substance to exist in several different crystalline forms. A change in crystal structure may occur during storage, resulting in an increase (or decrease) in crystal size and alteration in the suspension characteristics, causing solubility changes reflected in increased or decreased bioavailability. Manufacturers of commercial suspensions take these possibilities into account in the development and testing of the final formulation and the labeled storage conditions.

In some cases a water-soluble drug has been converted to an insoluble form and formulated as a suspension to improve the drug's stability, compatibility, bioavailability, or patient tolerance. The insoluble forms of steroids such as prednisolone and dexamethasone have better ocular bioavailability and are considered more potent anti-inflammatories for topical ocular use. A resin-bound form of the β -blocker betaxolol has been formulated as a suspension and is prepared *in situ* using a carbomer polymer.³⁸ The novel suspension formulation improves both comfort and ocular bioavailability of betaxolol, the 0.25 percent suspension therapeutically equivalent to a 0.5 percent solution.

OPHTHALMIC OINTMENTS

Ophthalmic ointments are primarily anhydrous and contain mineral oil and white petrolatum as the base ingredients, the proportions of which can be varied to adjust consistency and the melting temperature. Dosage variability probably is greater than with solutions (though probably no greater than that with suspensions). Ointments will interfere with vision, and their use is usually limited to bedtime instillation. They remain popular as a pediatric dosage form and for postoperative use. The anhydrous nature of the base enables its use as a carrier for moisture-sensitive drugs. The petrolatum base can be made more miscible with aqueous components by the addition of liquid lanolins.

Ointments do offer the advantage of longer contact time and greater total drug bioavailability, albeit with slower onset and time to peak absorption. The relationship describing the availability of finely divided solids dispersed in an ointment base was given by Higuchi,³⁹ where the amount of solid (drug) released in unit time is a function of concentration, solubility in the ointment base, and diffusivity of the drug in the base.

OPHTHALMIC EMULSIONS

An emulsion dosage form offers the advantage of the ability to deliver a poorly water-soluble drug in a solubilized form as an eyedrop. The drug is dissolved in a nonaqueous vehicle, such as castor oil, and emulsified with water, using a nonionic surfactant and, if needed, an emulsion stabilizer. An emulsion with

water as the external phase can be less irritating and better tolerated by the patient than use of a purely nonaqueous vehicle. Such an emulsion is used to deliver cyclosporine topically for the treatment of chronic dry eye conditions.⁴⁰

OPHTHALMIC GELS

Gel-forming polymers, such as carbomer, have been used to develop aqueous, semisolid dosage forms, which are packaged and administered the same as ointments. The viscous gels have significantly increased topical residence time and can increase drug bioavailability and decrease dosage frequency, compared to solutions. Although they contain a large proportion of water, they can still cause blurring of vision. A carbomer gel of pilocarpine administered at bedtime has been shown to prolong the IOP-lowering effect in patients for up to 24 hours.⁴¹

OCULAR INSERTS

Ocular inserts have been developed in which the drug is delivered on the basis of diffusional mechanisms. Such a solid dosage form delivers an ophthalmic drug at a near-constant known rate, minimizing side effects by avoiding excessive absorption peaks. The delivery of pilocarpine by such an insert was commercialized in 1975 (Ocuser Pilo) by *Alsa Corporation*. The Ocuser is designed to be placed in the lower cul-de-sac to provide a weekly dose of pilocarpine, at which time the system is removed and replaced by a new one. The near zero-order rate delivery is based on the selection of a non-eroding co-polymer membrane enclosing the drug reservoir.⁴²

Ocular inserts are plagued with some of the same manipulative disadvantages as conventional eyedrops. The insert must be placed in the eye in a manner similar to the insertion of a contact lens. Additionally, the insert must be removed from the eye when exhausted of its drug content. Such manipulations can be difficult for the elderly patient. Nonetheless, such therapeutic inserts represent a notable commercialized scientific achievement in pharmaceutical sciences. The Ocuser Pilo product is no longer marketed, because the drug has largely been replaced in glaucoma therapy by topical β -blockers.

Ocular inserts that gradually erode in the tear fluid have been studied but not commercially developed as ocular drug delivery systems.⁴³ In theory, an erodible insert would be advantageous, because it would not require removal at the end of its therapeutic cycle, would provide precise unit dosing, and, if anhydrous, would not require a preservative. It may also increase ocular bioavailability and reduce the therapeutic dosage and possible systemic effects. The chief disadvantages may be related to patient use issues, control of erosion and drug release rates, and sterilization.

An erodible insert is available (LACRISERT) for treatment of dry eye. It is molded in the shape of a rod from a hydroxypropyl cellulose polymer, which is the active ingredient. When inserted into the lower cul-de-sac, the polymer imbibes tear fluid and forms a gel-like mass that gradually erodes while thickening the tear film over a period of several hours. The unit-dose insert is anhydrous, and no preservative is required, which is beneficial for some sensitive patients.

DRUG ADMINISTRATION

TOPICAL ADMINISTRATION

The instillation of eyedrops remains one of the less precise, yet more accepted means of topical drug delivery. The method of administration is cumbersome at best, particularly for the elderly, patients with poor vision who have difficulty seeing without eyeglasses, and patients with other physical handicaps. Perhaps surprisingly, most patients become quite adept at routine instillation.

The pharmacist should advise each patient to keep the following points in mind to aid in the instillation of eyedrops or ointments.

METHODS OF USE FOR TOPICAL ADMINISTRATION

HOW TO USE EYEDROPS

1. Wash hands.
2. With one hand, gently pull lower eyelid down.
3. If dropper is separate, squeeze rubber bulb once while dropper is in bottle to bring liquid into dropper.
4. Holding dropper above eye, drop medicine inside lower lid while looking up; do not touch dropper to eye or fingers.
5. Release lower lid. Try to keep eye open and not blink for at least 30 seconds.
6. If dropper is separate, replace on bottle and tighten cap.

PRECAUTIONS WHEN USING EYEDROPS

- If dropper is separate, always hold it with the tip down.
- Never touch dropper to any surface.
- Never rinse dropper.
- When dropper is at top of bottle, avoid contaminating cap when removed.
- When dropper is a permanent fixture on the bottle (i.e., when supplied by a pharmaceutical manufacturer to the pharmacist), the same rules apply to avoid contamination.
- Never use eyedrops that have changed color.
- If you are using more than one kind of drop at the same time, wait several minutes before using other drops.
- After instillation of drops, do not close eyes tightly and try not to blink more often than usual, as this removes the medicine from the place on the eye where it will be effective.

HOW TO USE OPHTHALMIC OINTMENTS/GELS

1. Wash hands.
2. Remove cap from tube.
3. With one hand, gently pull lower eyelid down.
4. While looking up, squeeze a small amount of ointment (about $\frac{1}{4}$ to $\frac{1}{2}$ inch) inside lower lid. Be careful not to touch tip of tube to eye, eyelid, fingers, etc.
5. Close eye gently and roll eyeball in all directions while eye is closed. Temporary blurring may occur.
6. The closed eyelid may be rubbed very gently by a finger to distribute the drug throughout the fornix.
7. Replace cap on tube.

PRECAUTIONS WHEN USING OPHTHALMIC OINTMENTS/GELS

- Take care to avoid contaminating cap when removed.
- When opening ointment tube for the first time, squeeze out the first $\frac{1}{4}$ inch of ointment and discard, as it may be too dry.
- It may be helpful to practice ointment use by positioning yourself in front of a mirror.
- Never touch tip of tube to any surface.
- If you have more than one tube of the same ointment, open only one at a time.
- If you are using more than one kind of ointment at the same time, wait about 10 minutes before using another ointment.
- To improve flow of ointment, hold tube in hand several minutes to warm before use.

NASALACRIMAL OCCLUSION

To limit absorption of topically applied drugs directly into the bloodstream from the highly vascular areas of the nasal cavity, the patient should be instructed to close the eye immediately after eyedrop instillation, place a finger between the eyeball and the nose, and apply pressure for several minutes. This will temporarily occlude both superior and inferior canaliculi, preventing nasolacrimal drainage.

INTRAOCULAR PREPARATIONS

Ophthalmic products introduced into the anterior chamber or vitreous chamber of the internal eye structure are specialized dosage forms, requiring additional pharmaceutical considerations in their formulation, packaging, and manufacture. They are essentially parenteral-type products requiring both sterility and nonpyrogenicity as well as strict control of particulate matter, to ensure that they are compatible with sensitive internal tissues critical to visual acuity. In some cases in which the product may be introduced into the sterile field during surgery, sterility of the exterior of the primary package is required. Because preservatives commonly used in topical ophthalmic products can be toxic to sensitive intraocular tissues, such as the corneal endothelium, intraocular products are designed and packaged as preservative-free single-use products.

Present technologies are summarized in the next subsections, but many novel technologies can be expected in the next several years with the advent of new therapeutics and approaches to delivery for such devastating diseases as AMD, diabetic retinopathy, and cystoid macular edema, diseases of the back of the eye.

IRRIGATING SOLUTIONS

During ocular surgery an irrigating solution is used to maintain hydration, ocular volume, and clarity of the cornea, providing the surgeon a clear view of the surgical field. The irrigating solution also provides a physiological medium for removing blood and cellular debris, and for replacing the natural aqueous intraocular fluid. A balanced salt solution (BSS) is the primary intraocular irrigating solution, which includes the key ionic components to maintain corneal endothelial integrity: sodium, potassium, calcium, magnesium, and chloride, as well as a neutral to slightly alkaline pH and osmolality of about 305 mOsm.⁴⁴ An enriched balanced salt solution (BSS Plus) is also available to provide enhanced physiological compatibility when required in an irrigating solution. The enriched solution contains oxidized glutathione, dextrose, and bicarbonate in addition to the critical ionic components.⁴⁵ The additional ingredients in the enriched solution require, for maximum shelf-life stability, that the product be packaged in a two-part system to be aseptically reconstituted just before use. Intraocular solutions do not contain a preservative and should be discarded after initial use. These irrigating solutions are also designed to be used without additional additives, because there is potential for intraocular toxicity. An acidic epinephrine injection containing sodium bisulfite antioxidant, when added to an intraocular irrigating solution and diluted as much as 500-fold, has been reported to produce intraocular toxicity.⁴⁶

INTRAOCULAR INJECTIONS

Products approved for direct injection into the eye include miotics and viscoelastics for ocular surgery, and two antiviral agents to treat cytomegalovirus retinitis. The miotics carbachol and acetylcholine are used at the end of cataract surgery to constrict the pupil, allowing the iris to cover the implanted intraocular lens. Both products are specially packaged such that the exterior of the primary vial is sterile. This is accomplished by use of a special outer container that is permeable to sterilant gas and prevents contamination of the exterior of the vial.

Solutions with viscoelastic properties are injected into the eye during surgery to provide a mechanical barrier between tissues and allow the eye surgeon more space for manipulation with less potential for trauma to very sensitive intraocular tissues. The primary viscoelastic substance is a highly purified fraction of sodium hyaluronate. Purification is required to remove foreign proteins to be non-antigenic and noninflammatory for intraocular use. The viscoelasticity can be varied by use of fractions and concentrations of different molecular weights. Chondroitin sulfate and purified hydroxypropyl

methylcellulose are also utilized as viscoelastic and viscoadherent surgical adjuncts. The viscoelastic products are packaged and sterilized such that the primary package can be placed in the sterile surgical field before use. Sodium hyaluronate viscoelastic products usually require refrigeration during storage to maintain their integrity.

INTRAVITREAL INJECTIONS/IMPLANTS

Two antivirals have been approved for treatment of the ocular sequelae of AIDS with direct emplacement in the vitreous cavity so as to provide high localized ocular therapeutic concentrations. A polymer-coated sterile tablet containing ganciclovir, called a Vitrasert,⁴⁷ is implanted by the surgeon in the vitreous cavity, where it releases drug over a period of several months and then is removed and replaced with a new tablet. The tablet is formulated with magnesium stearate as the drug carrier and coated with polymers that provide the prolonged drug release. It is important that, during handling, this polymer coating is not damaged and the special sterile packaging not be compromised. A sterile solution of fomivirsen is also available for intravitreal injection to treat cytomegalovirus retinitis. It is supplied in single-use vials and is injected intravitreally without requiring surgery but must be repeated every 2 to 4 weeks.

JUXTASCLERAL INJECTIONS

Certain disease conditions in the back of the eye are usually untreatable through topical administration to the eye. For example, cystoid macular edema (associated with trauma, chronic inflammation, and diabetes) results in swelling of the macula and, if left untreated, loss of vision. At present there is no effective treatment available for AMD, whose incidence and severity increase with age, resulting in progression from accumulation of proteins and glycosaminoglycans (drusen) to leakage of blood vessels (hemorrhage) and eventual blindness. For such conditions, localized intravitreal or juxtascleral injections may be preferable to systemic administration of drugs (either alone or accompanied by laser-initiated events, such as photodynamic therapy).

OTHER MODES OF ADMINISTRATION

Packs

These sometimes are used to give prolonged contact of a drug solution with the eye to maximize absorption. Sterile cotton pledgets, placed in the lower cul-de-sac, have been used in this manner, but more recently corneal shields and soft contact lenses, developed as bandages, have been utilized to protect the cornea during healing.⁴⁸ Corneal shields are made of collagen that is crosslinked to control the rate of erosion, usually 12 to 72 hours. Hydrophilic polymers used for soft contact lenses without vision correction have also been utilized as a drug reservoir. The soft lenses remain intact and therefore require removal.

INTRACAMERAL INJECTIONS

Injections may be directly into the anterior chamber (e.g., acetylcholine chloride, alpha-chymotrypsin, carbamylcholine chloride, certain antibiotics, and steroids) or directly into the vitreous chamber (e.g., amphotericin B, gentamicin sulfate, and certain steroids).

IONTOPHORESIS

This procedure keeps the solution in contact with the cornea by means of an eyecup bearing an electrode. An ionic drug thus is subjected to a current generated by a difference in electrical potential, which acts as an electrochemical driving force to transport the drug in the direction of the potential gradient.⁴⁹

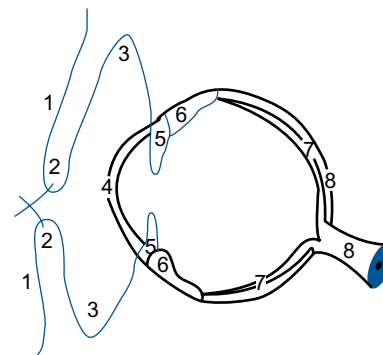


Figure 28-6. The locations for administration and target of therapy. Ointment: regions 1–5. Drops: regions 3–5. Parenteral injections—subconjunctival: regions 5–6, deep subtenons: regions 6–8, retrobulbar: region 8.

SUBCONJUNCTIVAL INJECTIONS

Subconjunctival injections (Figure 28-6)⁵⁰ are used frequently to introduce medications that, if applied topically, either do not penetrate into the anterior segment or penetrate too slowly to attain the concentration required. The drug is injected underneath the conjunctiva and probably passes through the sclera and into the eye by simple diffusion. The most common use of subconjunctival injection is for the administration of antibiotics in infections of the anterior segment of the eye. Subconjunctival injections of mydriatics and cycloplegics also are used to achieve maximal pupillary dilation or relaxation of the ciliary muscle. If the drug is injected underneath the conjunctiva and the underlying Tenon capsule in the more posterior portion of the eye, effects on the ciliary body, choroid, and retina can be obtained.

RETROBULBAR INJECTIONS

Drugs administered by retrobulbar injection (Figure 28-6) may enter the globe in essentially the same manner as the medications given subconjunctivally. The orbit is not well vascularized, and the possibility of significant removal via bloodstream effects from these injections is remote. In general, such injections are given for the purpose of targeting medications (e.g., antibiotics, local anesthetics, enzymes with local anesthetics, steroids, and vasodilators) to the posterior segment of the globe and affecting the nerves and other structures in that space for a sustained period of time.

PREPARATION

COMMERICAL MANUFACTURE

Pharmaceutical manufacturers provide finished ophthalmic products, manufactured and tested for quality according to stringent industrial and governmental standards. The products' sterility and, where required, their nonpyrogenicity, as well as other important quality standards, are assured by end-product testing of each batch and also by the use of validated manufacturing processes developed for each individual product. The products' manufacturing and quality control are governed largely by the current Good Manufacturing Practice (GMP) regulations⁵¹ promulgated and enforced by the FDA. If the product is a new or generic drug or medical device and has received FDA marketing approval, the manufacturer must also meet the requirements specified in the approved application for chemistry, microbiology, manufacturing, and quality control. Significant changes to the approved requirements must be submitted and approved by the FDA. For products not requiring prior FDA approval to market, such as monographed OTC drug products,⁵² the manufacturer must meet the quality standards of applicable

USP or NF compendial monographs in addition to the current GMP regulations.

The manufacturer is subject to pre-approval inspections and periodic GMP postapproval inspections by the FDA. The FDA also conducts inspections of the suppliers of active ingredients, both foreign and domestic, and certain inactive ingredients and packaging operations. The manufacture is subject to recalls of batches of product that do not meet quality requirements. In some cases a recall may result, not from actual failed test results, but from a lack of assurance that the product will continue to meet quality requirements through its shelf-life.

PHARMACY COMPOUNDING

Compounding of individual patient prescriptions by pharmacists has been and continues in some areas to be an integral part of pharmacy practice. The routine need to compound sterile ophthalmic products is no longer required with the broad range of commercially manufactured products available today. Most of the new and generic prescription ophthalmic products marketed today have been subjected to the FDA's rigorous requirements for proof of safety and efficacy or bioequivalence, as well as the manufacturing and quality requirements already described.

If the pharmacist is asked to compound a prescription for a noncommercial product, such as a preservative-free version or pediatric strength, he or she should be well versed in the preparation of sterile products, have the proper equipment and facilities, and be knowledgeable about the special requirements for ophthalmic formulations and packaging. Reference information on the standards and technology for pharmacy compounding, with special emphasis on preparation of sterile products, should be consulted.^{53,54} The pharmacist must also consult the rules and regulations of the applicable state board of pharmacy concerning sterile pharmacy compounding as well as any federal regulatory requirements promulgated by the FDA. Congress included in the 1997 FDA Modernization Act certain legal conditions for which compounding, as defined in the Act, would be exempt from FDA regulation. The pharmacy compounding section of the Act was subsequently litigated and overturned in its entirety, because it contained unconstitutional restrictions on commercial speech. The FDA may seek new legislation to address its concerns regarding manufacturing in the guise of compounding. In the meantime, the FDA has issued a guidance document on how the agency intends to regulate pharmacy compounding of human drugs considered to be outside the bounds of traditional pharmacy practice and in violation of the US Food, Drug, and Cosmetic Act of 1938 (FDCA; www.fda.gov/cder/pharmcomp), and has provided a Pharmacy Compounding Compliance Policy Guide (2002).

STERILIZATION

Common methods of sterilization include moist heat under pressure (autoclave), dry heat, filtration, gas sterilization, and ionizing radiation. Please refer to Chapter 25, where these sterilization procedures are described in detail.

DANGERS OF NONSTERILE MEDICATIONS

The possibility of serious ocular infection resulting from the use of contaminated ophthalmic solutions has been documented amply in the literature. Such solutions have repeatedly been the cause of corneal ulcers and even loss of eyesight. Contaminated solutions have been found to be in use in physicians' offices, eye clinics, and industrial infirmaries, and have been dispensed on prescription in community and hospital pharmacies. The microbes most frequently found as contaminants are staphylococci group. *Pseudomonas aeruginosa* is a less common contaminant, and the solution most often found contaminated is sodium fluorescein.

Ps. aeruginosa (*Bacillus pyocyaneus*; *Pseudomonas pyocyanea*; blue pus bacillus) is a very dangerous and opportunistic organism that grows well on most culture media and produces both toxins and antibacterial products. The latter tend to kill off other contaminants and allow the *Ps. aeruginosa* to grow in pure culture. This Gram-negative bacillus also grows readily in ophthalmic solutions, which may become the source of extremely serious infections of the cornea. It can cause complete loss of sight in 24 to 48 hours. At concentrations tolerated by tissues of the eye, most of the antimicrobial agents discussed in the following sections may be ineffective against some strains of *Ps. aeruginosa*.

A sterile ophthalmic solution in a multiple-dose container can be contaminated in a number of ways unless precautions are taken. For example, if a dropper bottle is used, the tip of the dropper while out of the bottle can touch the surface of a table or shelf if laid down, or it can touch the eyelid or eyelash of the patient during administration. If the Drop-Tainer (Alcon) type of bottle is used, the dropper tip can touch an eyelash or the cap while removed to permit administration, or its edge may touch a table or finger, and that edge can touch the dropper tip as the cap is replaced.

The solution may contain an effective antimicrobial, but the next use of the contaminated solution may occur before enough time has elapsed for all of the organisms to be killed, and living organisms can find their way through an abrasion into the corneal stroma. Once in the corneal stroma, any residual traces of antimicrobial agents are neutralized by tissue components, and the organisms find an excellent culture medium for rapid growth and dissemination through the cornea and the anterior segment of the eye.

OTHER ORGANISMS

Bacillus subtilis may produce a serious abscess when it infects the vitreous humor. The pathogenic fungus considered of particular importance in eye solutions is *Aspergillus fumigatus*. Other fungi or molds may cause harm by accelerating deterioration of the active drugs.

With regard to viruses, as many as 42 cases of epidemic keratoconjunctivitis were caused by one bottle of virus-contaminated tetracaine solution. Virus contamination is particularly difficult to control, because none of the preservatives now available is virucidal. Moreover, viruses are not removable by filtration. However, they are destroyed by autoclaving. The pharmacist and physician have not been made adequately aware of the dangers of transmitting viral infection via contaminated solutions. This is particularly pertinent to the adenoviruses (types III and VIII), now believed to be the causative agents of viral conjunctivitis, such as epidemic keratoconjunctivitis.

The danger of nonsterile preparations is exponentially increased for products intended for injection within the eyeball. Endophthalmitis and loss of vision can occur within a short time of onset of a bacterial infection.

METHODS

Steam Under Pressure

Terminal sterilization by autoclaving is an acceptable, effective method of sterilization; however, the solution or suspension components must be sufficiently heat-resistant to survive the procedure. If sterilization is carried out in the final container, the container also must be able to withstand heat and pressure. A recent addition to this technique is the so-called air-over-steam autoclave. This combination allows pressure adjustments to be made during the autoclave cycle. Pressure manipulations permit the autoclave sterilization of materials that, while heat-resistant, tend to deform (i.e., polypropylene containers). The sterilization cycle for a product should be carefully validated, and one must assure that sterilization temperature and time are monitored at the coldest spot of the autoclave load to assure sterility of the product.

Filtration

The USP states that sterile membrane filtration under aseptic conditions is the preferred method of sterilization. Membrane filtration offers the substantial advantage of room-temperature operation with none of the deleterious effects of exposure to heat or sterilizing gas.

Sterilization by filtration does involve the transfer of the finished sterile product into previously sterilized containers, using aseptic techniques. The membrane filtration equipment itself usually is sterilized as an assembly by autoclaving.

Several types of membrane filtration equipment are available for small-scale processing, as described in Chapter 25.⁵⁴ Particular interest has been shown in the Swinney adapter, fitted on a syringe, and in the *Millipore* Swinnex disposable filter units. Empty sterile plastic *squeeze* containers and sterile plastic filtration units can be purchased directly from the manufacturers, including Wheaton (polyethylene containers) and *Millipore* (Swinnex filter units). They permit extemporaneous preparation of ophthalmic solutions, which have a high probability of being sterile if the work is carried out under aseptic conditions. A supplementary device can permit automatic refilling of the syringe. The filter unit must be replaced after use. To avoid contamination, the pharmacist should fill sterile filtered solutions into presterilized containers closed with appropriate fitments and closed under laminar flow, using aseptic techniques. The Parenteral Drug Association has produced several audiovisual teaching materials for aseptic techniques and processing. The reader should refer to these guidelines for reference.

Gas

Gas sterilization of heat-sensitive materials may be carried out by exposure to ethylene oxide (EtO) gas in the presence of moisture. EtO gas for sterilization use is available commercially, diluted with either carbon dioxide or halogenated hydrocarbons. EtO sterilization requires careful consideration of conditions required to effect sterility. Temperature and pressure conditions are quite nominal in contrast to wet or dry heat; however, careful control of exposure time, EtO concentration, and moisture is essential.

Gas sterilization requires that specialized, but not necessarily elaborate, equipment be used. Gas autoclaves may range from very large walk-in units to small units on the scale of a laboratory bench that are suitable for small hospitals, laboratories, or pharmacies. Users of gas autoclaves must keep in mind the possibility of human toxicity and take care to restrict exposure to EtO during the loading, venting, and unloading of the sterilizer. EtO sterilization produces irritating by-products that remain as residues in or on the articles sterilized. Residues include ethylene glycol and ethylene chlorohydrin (when in contact with chloride ions), in addition to EtO itself. Aeration of the sterilized articles for at least 72 hours, preferably at 40°C to 50°C, will minimize such residues.

Ambient aeration time for sterilized polyethylene bottles should be about 48 hours. EtO is recommended for the sterilization of solid materials that will not withstand heat sterilization. The FDA has recommended maximum residues in the parts-per-million range for EtO, ethylene glycol, and ethylene chlorohydrin.

Extreme caution is necessary when using EtO gas sterilization. The previously accepted 12:88 mixture of EtO to Freon has been replaced with 100 percent EtO for environmental reasons. This gas is explosive, and all workers also should be protected from accidental exposure because of concern for carcinogenicity and other toxic reactions to EtO. This method at present is used as a last resort when no other methods can be used.

Radiation

Sterilization by exposure to ionizing radiation is an acceptable procedure for components of ophthalmic preparations or,

indeed, for the total product, such as certain ophthalmic ointments. Sources of radiation are twofold and include linear electron accelerators and radioisotopes. The linear accelerators produce high-energy electrons with very little penetrating power. Radioisotopes, particularly ⁶⁰Co (a source for gamma irradiation), are employed more widely than linear electron accelerators for sterilization. Sterilization by radiation may produce untoward effects, such as chemical changes in product components as well as changes in color or physical characteristics of package components. Gamma sterilization is currently the method used in place of EtO for sterilizing most of the packaging components. Gamma sterilization of containers may alter the surface characteristics, and this can increase degradation of drug molecules and/or some of the excipients used in the formulation. Some plastic manufacturers have developed special grades of polypropylene resin that are stabilized to withstand gamma irradiation.

OPHTHALMIC PREPARATION CHARACTERISTICS

CLARITY

Ophthalmic solutions, by definition, contain no undissolved ingredients and are essentially free from foreign particles. Filtration can enhance clarity in some cases. It is essential that the filtration equipment be clean and well rinsed to avoid introduction of particulate matter into the solution by equipment designed to remove it. Operations performed in clean surroundings, the use of laminar-flow hoods, and proper nonshedding garments will contribute collectively to the preparation of clear solutions essentially free from foreign particles. In many instances the same filtration step can produce both clarity and sterility. If viscosity-imparting polymers are used, a polish-filtering step may be necessary before the final filtration.

Both container and closure must be thoroughly clean, sterile, and non-shedding, so that neither one introduces particulate matter to the solution during prolonged contact for the duration of the shelf-life. Normally this is established by thorough stability testing, which also will indicate whether insoluble particles (by-products of drug degradation) have been generated. Solution formulations may also contain viscosity-imparting polymers that can diminish clarity. In these situations it may be important to both define the visual clarity of the product and monitor its stability. The *European Pharmacopoeia* describes visual clarity and recommends standards that can be used for clarity specifications.³

STABILITY

The stability of a drug in an ophthalmic product depends on a number of factors, including the chemical nature of the drug substance, whether it is in solution or suspension, product pH, method of preparation (particularly temperature exposure), solution additives, and type of packaging. A pharmaceutical manufacturer strives for a shelf-life measured in years at controlled room-temperature conditions, whereas the compounding pharmacist often is uncertain about the shelf-life of his or her preparation and thus provides relatively small quantities at one time, assigns a shelf-life in terms of days or weeks, and may specify refrigerated storage as a further precaution. The attainment of optimum stability often requires some compromises in the formulation, packaging, and preparation of the final product.

The product's pH is often the stability-controlling factor for many drugs. Drugs such as pilocarpine and physostigmine are both active and comfortable in the eye at a pH of 6.8; however, at this pH, chemical stability (or instability) can be measured in days or months. Either drug will lose a substantial amount of chemical stability in less than a year. On the other hand, at a pH of 5.0, both drugs are stable for a period of several years. (With regard to eye comfort at acidic pH, see the later discussion under Buffer and pH.)

In addition to optimal pH, if oxygen sensitivity is a factor, adequate stability may require inclusion of an antioxidant or special packaging. Plastic packaging, such as the low-density polyethylene containers (e.g., the DropTainer from Alcon) that are convenient for patient use, may prove detrimental to stability by permitting oxygen permeation, resulting in oxidative decomposition of the drug substance. Development of an epinephrine solution with 2 to 3 years' stability in a plastic package requires the use of a pH of about 3.0 for protection from oxidation, whereas an epinephrine borate solution formulated at a pH of about 7.0, which is more comfortable to the patient, requires an antioxidant system and the use of glass packaging. The prodrug of epinephrine, dipivefrin, significantly increases ocular bioavailability and is effective at one-tenth the concentration of epinephrine. The structure of the chemical derivative protects the active epinephrine portion from oxidation, so that it can be packaged in plastic. However, the labile ester linkage introduced in the prodrug requires that it be formulated at a pH of about 3.0 to minimize hydrolysis; even with this precaution, the shelf-life of dipivefrin stored at room temperature can be extended to no longer than 18 months.

Pharmaceutical manufacturers conduct comprehensive stability programs to assure the assigned expiration dating for each product. In addition to monitoring the standard chemical and physical stability of the pharmaceutical, they test the stability of the preservative by chemical means or by actual challenge of its efficacy with appropriate test organisms. Sterility is not a stability parameter *per se*, but each container-closure system can be tested by microbial challenge to assure integrity of the package against environmental contamination before opening.

Some of the newer classes of ophthalmic drugs, like prostaglandins, are very hydrophobic and have very low concentrations. For example, in the product Xalatan, latanoprost is present at 0.005 percent, and in the product Travatan, travoprost is present at 0.004 percent. Active agents at such low concentrations present a challenge for formulators, because the loss of even small amounts of drug (e.g., from adsorption losses to packaging) can become significant. Pharmacia's Xalatan requires refrigerated storage, and as indicated earlier, temperature cycling also can reduce the concentration of active drug. It is important that the pharmacist knows the properties of the drug substance, in order to maintain product quality throughout the shelf-life of the product.

BUFFER AND PH

Ideally, ophthalmic preparations should be formulated at a pH equivalent to the tear fluid value of 7.4. Practically, formulators seldom achieve this. Most active ingredients used in ophthalmology are salts of weak bases and are most stable at an acid pH. This property generally holds for suspensions of insoluble corticosteroids.

Optimum pH adjustment generally requires a compromise on the part of the formulator, who should select not only a pH that is optimal for stability but a buffer system that has adequate capacity to maintain pH within the stability range for the duration of the product shelf-life. Buffer capacity is the key in this situation.

It generally is accepted that a low (acid) pH *per se* necessarily will not cause stinging or discomfort on instillation. If the overall pH of the tears, after instillation, reverts rapidly to pH 7.4, discomfort is minimal. On the other hand, if the buffer capacity is sufficient to resist adjustment by tear fluid and the overall eye pH remains acid for an appreciable period of time, then stinging and discomfort may result. Consequently, buffer capacity should be adequate for stability but minimized, so far as possible, to allow only momentary disruption of the overall pH of the tear fluid. Special care in formulating intraocular products is required regarding their pH and buffer capacity. The corneal endothelium can tolerate much less deviation from physiological conditions, compared to the external corneal epithelium.⁵⁵

TONICITY

Tonicity refers to the osmotic pressure exerted by salts in aqueous solution. An ophthalmic solution is isotonic with another solution when the magnitudes of the colligative properties of the solutions are equal. An ophthalmic solution is considered isotonic when its tonicity is equal to that of a 0.9 percent sodium chloride solution (290 mOsm). However, the osmotic pressure of the aqueous intraocular fluid is slightly higher than that of normal tears, measuring about 305 mOsm.

In actuality the *external* eye is much more tolerant of tonicity variations than was at one time suggested and usually can tolerate solutions equivalent to a range of 0.5 to 1.8 percent sodium chloride. Given a choice, one will find that isotonicity is desirable and is particularly important in intraocular solutions.⁵⁶ However, in certain cases a nonisotonic topical product is desirable. Tear fluid in some cases of dry eye (keratoconjunctivitis sicca) is reported to be hypertonic, and a hypotonic artificial-tear product is used to counteract this condition. Hypertonic ophthalmic products are used to relieve corneal edema, and solutions and ointments containing 2 or 5 percent sodium chloride are available for this use.

The tonicity of ophthalmic (and parenteral) solutions has been the subject of intensive investigation over the years. These studies have resulted in the accumulation and publication of a large number of sodium chloride equivalents that are useful in calculating tonicity values. See Chapter 16 for a thorough discussion of the measurement and calculation of tonicity values.

VISCOSITY

Ophthalmic solution and suspension eyedrops may contain viscosity-imparting polymers to thicken the tear film and increase corneal contact time (i.e., reduce the rate of tear fluid drainage). For suspensions, the increased viscosity also retards the settling of particles between uses and at the same time maintains their suspension for uniform dosing. However, added viscosity may impede initial resuspension, particularly in a suspension that has a tendency to cake during storage. The hydrophilic polymers most often used for these purposes are methylcellulose, hydroxypropyl methylcellulose, hydroxyethyl cellulose, and polyvinyl alcohol—at concentrations that produce viscosities in the range of about 5 to 100 Centipoise. These polymers also themselves appear as the active ingredients in artificial-tear solutions for dry eye therapy because of their lubrication and moisturizing. Viscosity agents can have several disadvantages, in that they sometimes produce blurring of vision and can leave a residue on the eyelids. These effects are most often seen at the higher end of the viscosity range. The added viscosity can make filtration more difficult, particularly for the small pore-size filters used to sterilize solutions.

Newer ophthalmic dosage forms, such as gel-forming solutions and semisolid aqueous gels with their increased viscosity and gel elasticity, significantly improve drug bioavailability and duration of effect. With these advances, the frequency of dosing can be reduced and patient compliance improved. These newer dosage forms incorporate novel polymer systems with special rheological properties to enhance their effect. Their complex rheology and intricate dependence on environment, however, increase the complexity of the sterile manufacturing process.

ADDITIVES

Most ophthalmic dosage forms contain additives or pharmaceutical excipients as inactive ingredients. Because of the need for tissue compatibility in ophthalmics, particularly in intraocular products, however, the use of additives is perhaps less common.

The most common inactive ingredient is the product's vehicle. For topical dosage forms, Purified Water USP is used. Because of the requirement for nonpyrogenicity, Water for Injection USP is used for intraocular products. While a mineral oil and petrolatum combination is the vehicle used for ophthalmic

ointments, nonaqueous liquids are rarely used in topical eye-drops because of their potential for ocular irritation and poor patient tolerance. Some mineral and vegetable oils have served as vehicles for very moisture-sensitive or poorly water-soluble drugs. Use of the purest grade of oil, such as that used for parenteral products, is a requirement.

Multiple-dose topical ophthalmic products commonly contain microbiological preservatives, which will be discussed in a later section. Other commonly used additives in topical eye products are ingredients to adjust pH and tonicity, and to buffer pH, in addition to the viscosity agents previously discussed. Ingredients to adjust pH and tonicity and to buffer pH are essentially the same as those used in parenteral products. Less commonly used additives are antioxidants, such as sodium bisulfite, ascorbic acid, and acetylcysteine.

Topical eye products sometimes incorporate surfactants to disperse insoluble ingredients or to aid in solubilization. Formulators use them in the smallest concentration possible to achieve the desired function, because they can be irritating to sensitive ocular tissues. Nonionic surfactants are preferable, because they are generally less irritating than ionic surfactants. Polysorbate 80 is the surfactant used to prepare an ophthalmic emulsion. Polyoxyl 40 stearate and polyethylene glycol function to solubilize a drug in an anhydrous ointment so that it can be filter-sterilized. Formulators often add surfactants to stabilize more hydrophobic drugs, for example, preventing loss to adsorption on the container walls. For example, the nonionic surfactant like polyoxyl 40 hydrogenated castor oil (HCO-40) stabilizes travoprost, a prostaglandin derivative.⁵⁷ Similarly, Cremophore EL stabilizes diclofenac in the Voltaren formulation marketed by Novartis.

The FDA has published a list of all inactive ingredients used in approved drug products on its Internet website at www.fda.gov/cder. The list includes dosage forms and concentration ranges.

PACKAGING

Currently, almost all commercially available ophthalmic products are packaged in plastic containers. Obvious advantages—ease of use, little breakage, less spillage—have led to universal acceptance of these plastic packaging components, consisting of bottle, fitment, and closure. Alcon was the first company to introduce these packaging components, identified as a DropTainer for ophthalmic products, in the late 1940s and then saw them adopted by the industry as the standard for packaging topical ophthalmic products. These bottles are generally of low-density polyethylene, either without any colorants or with opacifying agents or other colorants for light protection. Polypropylene or high-density polyethylene resins are also used to meet specific product requirements. The fitments determine drop size of the product and may contain additional features to prevent streaming of product at the time of use. Caps or closures are generally made from polypropylene and basically seal the container to prevent contamination or leakage of the product.

The FDA and other health authorities are also concerned about leachable impurities extracted either from the packaging components themselves or even occasionally from the label adhesives or ink used in printing the labels. Normally, as a part of a stability program, actual package-formulation compatibility studies are required, including monitoring any extractables that come out in the product as a function of the duration of storage. Selection of compatible packaging material has become a critical issue for newer drug products, especially those in which the concentration of active is extremely low, like prostaglandins, described earlier. The prostaglandin travoprost, the active ingredient in Travatan, required reengineering of the total packaging, starting with the resin (syndiotactic polypropylene) to stabilize the drug, as well as altering size and shape to maintain squeezability of the bottle.⁵⁸ Travoprost was found not to be stable when stored in low-density polyethylene containers, the compound being lost to adsorption onto and into surfaces,

even at room temperature. Consequently, chemical assay of marketed product must be conducted in a manner that does not deplete the concentration.

Specialized containers, like those developed by Merck for their timolol/pilocarpine combination product, maintain two solutions in different chambers, adjusted separately to provide optimal chemical stability. At the time of dispensing the product, fractions of both solutions are combined and reconstituted by the pharmacist. For example, the final timolol/pilocarpine product, once reconstituted, would be at a nearly physiological pH, would provide pilocarpine as stable for one month at room temperature, and will be found comfortable to patients. Alcon also has developed a proprietary package⁵⁹ that can keep unstable or reactive components separate during storage, and that permits the pharmacist, at the time of dispensing to the patient, to rupture a diaphragm, allowing the disparate fractions to be mixed. After mixing, the product is expected to remain stable for a month at room temperature. This container is designed to handle either liquid/liquid two-part or solid/liquid two-part components.

Pharmacists also should be aware of unit-dose products currently available in form-fill-seal containers that are preservative free. These products are discarded after single use. These products are most suited for patients who are sensitive or allergic to common preservatives and require chronic administration of a product. Specialized manufacturing technology involves melting a plastic resin to form the container walls, filling with the drug solution (usually 0.2 to 0.8 mL), and then sealing to maintain sterility. At the time of use the tip is broken off, the solution is dosed, and then the package is discarded. For technical reasons relating to manufacturing and filling capabilities, as well as for the need to minimize evaporation rate, these containers contain an excess volume of the product. The disadvantage is that patients may desire to use the entire volume and risk using an accidentally contaminated product. Recently, some unit-dose containers have been modified to allow the product to be recapped after use. These containers are designed and labeled to be discarded after a single day's use (12 hours) in order to reduce the risk of significant contamination, yet make them more economical for the patient.

In only a very few instances are glass containers still in use, usually because of stability limitations. Large-volume intraocular solutions of 250 and 500 mL have been packaged in glass, but even these parenteral-type products are beginning to be packaged in specially fabricated polyethylene/polypropylene containers or flexible bags. Type 1 glass vials with appropriate stoppers are used for intraocular ophthalmic products administered by injection. These packaging components should meet the same requirements as parenteral products. Products injected intraocularly also are required to meet endotoxin limits. Readers should refer to Chapter 26, Parenteral Preparations, and Chapter 27, Pharmaceutical Compounding – USP <797> Sterile Preparations for more details.

Plastic packaging, usually low-density polyethylene, is by no means interchangeable with glass, however. Plastic packaging is permeable to a variety of substances, including light and air. The plastic package may contain a variety of extraneous substances such as mold-release agents, antioxidants, reaction quenchers, and the like, which may leach out of the plastic and into the contained solution. Label glues, inks, and dyes also may penetrate polyethylene. Conversely, confined volatile or lipophilic materials may permeate from solution into or through plastic walls. Sterilization, depending on the method, may influence the resin properties; for example, gamma irradiation may increase acid extractables or generate sites that may degrade certain drugs. For these products EtO gas sterilization may be the only remaining option. However, gas sterilization also may degrade active molecules or leave harmful residuals, and certainly requires aeration under forced air to remove the traces of the gas and volatile residues. Whatever process is selected will need to be validated and monitored, assuring the process neither degrades the active ingredient nor generates toxic residues.

Table 28-1. Ophthalmic Cap Color Coding

Color	Pharmaceutical Class
Yellow or blue	β -blockers
Gray	Nonsteroids
Pink	Steroids
Brown or tan	Anti-infectives
Orange	Carbonic anhydrase inhibitors
Turquoise	Prostaglandins
Red	Mydriatics
Green	Miotics

Patients may be prescribed more than one ophthalmic medication for the same or different conditions, and this can lead to confusion as to which medication is for which use. Historically, red caps have been used for mydriatic drops and green caps for miotic drops, such as pilocarpine. The FDA now requires the use of certain colored caps on several additional types of ophthalmic drugs as a result of a cooperative effort between the FDA, the ophthalmic industry, and the Academy of Ophthalmology. The intent is to help patients prevent medication errors and improve patient compliance. The pharmacist should counsel the patient or caregiver about the purpose of the cap color coding and the importance of opening only one container at a time so that the cap is replaced on the correct container. A listing of the current coding is provided in Table 28-1.

Glass containers (using type 1 glass) remain a convenient package material for extemporaneous preparation of ophthalmic solutions. Type 1 glass should be used. The container, well rinsed with sterile distilled water, can also be sterilized by autoclaving. Presterilized droppers normally are available, presterilized and packaged in a convenient blister pack.

Invariably, packaging of ophthalmic ointments consists of metal tubes with an ophthalmic tip. Autoclaving with either heat or EtO can conveniently sterilize such tubes. In rare cases of metal reactivity or incompatibility, it may be necessary to package ointments in tubes lined with epoxy or vinyl plastic.

Regardless of the form of packaging, consumer protection requires that some type of tamper-evident feature be included. The common tamper-evident feature used on most ophthalmic preparations is the moisture- or heat-sensitive shrink band. Some form of identification on the band should be present, so that its disruption or absence constitutes a warning that tampering, either accidental or purposeful, has occurred.

The eyecup, an ancillary packaging device, fortunately seems to have gone the way of the community drinking cup. Use of an eyecup will inevitably spread or aggravate eye infections. Pharmacists should discourage their use, just as they should take the time to instruct patients in the proper use and care of eye medications. While ophthalmic administration may seem simple enough, it can be a foreign and difficult task for many people. The suggestions and precautions given earlier (under Types of Ophthalmic Dosage Forms) may be useful in instructing patients.

ANTIMICROBIAL PRESERVATIVES⁶⁰

A preservative is a substance or mixture of substances added to a product formulation to prevent the growth of, or to destroy, microorganisms introduced accidentally once the container is opened for use.^{61,62} Addition of the preservative is not intended to be a means of preparing a sterile solution, for which other appropriate techniques exist (discussed elsewhere).

Preservatives are used for topical ophthalmic products packaged in multiple-dose containers unless the formulation itself is self-preserving, as is the case with some antimicrobial products, like the ophthalmic solution Vigamox (moxifloxacin). FDA regulations (21 CFR 200.50) allow unpreserved ophthalmic liquid

products to be packaged in multiple-dose containers only if they are packaged and labeled in a manner that affords adequate protection and minimizes the hazards resulting from accidental contamination during patient use. This can be accomplished by using a reclosable container with a minimum number of doses, which is to be discarded after 12 hours from initial opening and which should be labeled appropriately.

Solutions intended for intraocular use should not contain preservatives because of the risk of irritation and damage to these delicate tissues. Medicines intended for chronic topical applications in patients who cannot tolerate preservatives should also be free of preservatives. Unit-dose ophthalmic solutions are especially useful for patients with sensitivity to preservatives who nevertheless require daily medication, such as glaucoma patients or individuals requiring chronic application of palliatives for dry eye. Ophthalmic solutions prepared and packaged for a single application (i.e., a unit dose) need not contain a preservative, because they are not intended for reuse.

The need for proper control of ophthalmic solutions to prevent serious contamination was recognized in the 1930s. The first preservative recommended for use in ophthalmics was chlorobutanol, as an alternative to daily boiling.

The selection of an ophthalmic preservative can be difficult in part because of the relatively small number of suitable candidates. There is, of course, no such thing as an ideal preservative; however, the following criteria may be useful in preservative selection.

1. The agent should have a broad spectrum, being active against Gram-positive and Gram-negative organisms as well as fungi. The agent should exert a rapid bactericidal activity, particularly against known virulent organisms such as *Ps. aeruginosa* strains.
2. The agent should be stable over a wide range of conditions, including autoclaving temperatures and pH range.
3. Other components of the preparation as well package systems should be compatible with the agent.
4. The preservative should cause neither toxicity nor irritation, within a reasonable margin of safety.

Only through evaluating preservative substances can as a part of the total ophthalmic preparation in the proposed package can producers establish the adequacy of the preservative. Criteria for preservative effectiveness in ophthalmic products are official compendial requirements in the *US Pharmacopeia*, *European Pharmacopeia*, and the *Japanese Pharmacopoeia*. Health authorities expect that products will meet these preservative effectiveness criteria throughout their approved shelf-lives, in the final packaging, and at the recommended storage conditions.²⁻⁴

In addition to its effectiveness as an immediate measure, formulators must also ascertain a preservative's adequacy or stability as a function of time. They often accomplish this by measuring both chemical stability and preservative effectiveness over a given period of time and under varying conditions.

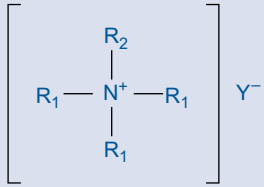
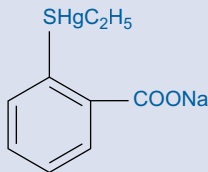
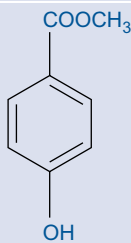
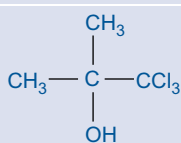
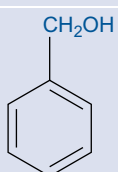
Many of these test procedures are not totally applicable to the preparation of an extemporaneous ophthalmic solution. In such a situation the pharmacist must make selections based upon known conditions and physical and chemical characteristics, guided by those used in commercial ophthalmic preparations. In these circumstances it would be prudent to prepare only minimum volumes for short-term patient use.

The choice of preservatives suitable for ophthalmic use is surprisingly narrow. The classes of compounds available for such use are described in Table 28-2.⁶³ In each case or category there are specific limitations and shortcomings.

QUATERNARY AMMONIUM COMPOUNDS

Benzalkonium chloride (BAC) is a typical quaternary ammonium compound and is the most common preservative used in ophthalmic preparations, being present in over 65 percent of commercial ophthalmic products. Note the structure of BAC is

Table 28-2. Ophthalmic Preservatives⁶³

Type	Typical Structure	Concentration Range	Incompatibilities
Quaternary ammonium compounds		0.004–0.02 %, 0.01% most common	Soaps Anionic materials Salicylates Nitrates
Organic mercurials (thimerosal example)		0.001–0.01%	Certain halides with phenylmercuric salts
Parahydroxy benzoates (methyl ester example)		Maximum 0.1%	Adsorption by macromolecules; marginal activity
Chlorobutanol		0.5 %	Stability is pH-dependent; activity concentration is near solubility
Aromatic alcohols		0.5–0.9 %	Low solubility in water; marginal activity

a slight departure from the general one provided in Table 28-2 in which there are three small alkyl chains on one longer alkyl chain (the R's). For BAC one of the R1's is a benzyl group. Despite its widespread use, the compound has definite limitations. As a cationic surface-active material of high molecular weight, it is not compatible with anionic compounds. It is incompatible with salicylates and nitrates, and may be inactivated by high-molecular-weight nonionic compounds. Conversely, BAC has excellent chemical stability and very good antimicrobial characteristics. Given the alternatives, it is often better to modify a formulation to eliminate the incompatibility than to include a compatible but less effective preservative. The presence of a surfactant, for example, may require higher levels of BAC to achieve an adequate level of preservation.

The literature on BAC is somewhat mixed; however, this is to be expected given the wide variation in test methods and, indeed, the chemical variability of BAC itself. The official substance is defined as a mixture of alkyl benzyldimethylammonium chlorides, including all or some of the group ranging from $n\text{-C}_8\text{H}_{17}$ through $n\text{-C}_{16}\text{H}_{33}$. The $n\text{-C}_{12}\text{H}_{25}$ homologue content is at least 40 percent on an anhydrous basis.

Reviews⁶⁴ of BAC indicate that it is well suited for use as an ophthalmic preservative. Certain early negative reports have been shown to be erroneous; in some cases adverse tissue reactions were attributed to BAC when, in fact, a totally different compound was used as the test material. Although BAC is the most common quaternary preservative, others occasionally mentioned include benzethonium chloride and cetyl pyridinium chloride. All are official compounds. More recently,

quaternary ammonium functionality has been attached to soluble, reasonably high-molecular-weight polymers. These agents possess good antimicrobial effectiveness with fewer compatibility problems than the official quaternary preservatives.

Search for milder, safer, gentler preservatives for ophthalmic products, specifically artificial-tear and chronic-use products, has been a challenge for many companies. As a result of this research, some commercial products now incorporate these improved agents as preservatives. Alcon has introduced many lens care and ophthalmic products preserved with polyquaternium-1 (polyquad), a newer preservative of the same class as BAC but with less cytotoxicity. More effective than BAC, these agents also are effective at a lower concentration. Their concentration in the product can range from 0.005 percent to 0.0005 percent and yet still meet compendial preservative efficacy requirements. Polyquad has been used widely in many lens care and artificial-tear products.

OXIDIZING AGENTS

Systems based on either sodium perborate or a stabilized oxychloro complex are being used as preservatives based on their ability to generate a mild oxidative and cytotoxic effect in aqueous media. Sodium perborate produces hydrogen peroxide as the oxidative species, and the stabilized oxychloro complex is a mixture of oxychloro species but primarily comprises chlorite and a trace of chlorine dioxide. Once in the eye, the active agents in either system are spontaneously reduced to harmless by-products and have been marketed as so-called disappearing preservatives in OTC products for dry eye treatment.

ORGANIC MERCURIALS

It is generally stated that phenylmercuric nitrate or phenylmercuric acetate, in 0.002 percent concentration, should replace BAC as a preservative for salicylates and nitrates, and in solutions of salts of physostigmine and epinephrine that contain 0.1 percent sodium sulfite. The preferred range of concentrations employed is 0.002 to 0.004 percent. Phenylmercuric borate sometimes is used in place of the nitrate or acetate.

Phenylmercuric nitrate has the advantage over some other organic mercurials of not being precipitated at a slightly acid pH. As with other mercurials, it is slow in its bactericidal action and it also produces sensitization reactions. Phenylmercuric ion is incompatible with halides, for it forms precipitates.

The effectiveness of phenylmercuric nitrate against *Ps. aeruginosa* is questionable; researchers have found that pseudomonads survive after exposure to a concentration of 0.004 percent for longer than a week.

Development of iatrogenic mercury deposits in the crystalline lens has occurred, resulting from use of miotic eyedrops containing 0.004 percent phenylmercuric nitrate, three times daily, for periods of 3 to 6 years. Although there was no impairment of vision, the yellowish brown discoloration of the lens capsule is reportedly permanent.

Thimerosal (Merthiolate; Lilly) is an organomercurial with bacteriostatic and antifungal activity that functions as an antimicrobial preservative in concentrations of 0.005 to 0.02 percent. Its action, as with other mercurials, has been reportedly slow.

PARAHYDROXYBENZOIC ACID ESTERS

Mixtures of methylparaben and propylparaben are sometimes used, as antimicrobial preservatives, mostly in ophthalmic ointments; the preferred maximum concentration of methylparaben is in the range of 0.1 percent, its NOEL, while that of propylparaben approaches its solubility in water (>0.04 percent). They are not considered efficient bacteriostatic agents and are slow in their antimicrobial action. Ocular irritation and stinging have been attributed to their use in ophthalmic preparations.

SUBSTITUTED ALCOHOLS AND PHENOLS

Chlorobutanol is reportedly effective against both Gram-positive and Gram-negative organisms, including *Ps. aeruginosa* and some fungi. It is broadly compatible with other ingredients, and manufacturers normally choose a concentration of 0.5 percent. One of the products of hydrolysis is hydrochloric acid, which causes a decrease in the pH of aqueous solutions. This decomposition occurs rapidly at high temperatures and slowly at room temperature, in unbuffered solutions that were originally neutral or alkaline. Therefore, ophthalmic solutions that contain chlorobutanol should be buffered between pH 5.0 and 5.5, and generally packaged in glass containers. At room temperature it dissolves slowly in water, and although heat dissolves it more rapidly, loss by vaporization and decomposition is accelerated.

A combination of chlorobutanol and phenylethyl alcohol (0.5 percent of each) has been reported to be more effective against *Ps. aeruginosa*, *Staphylococcus aureus*, and *Proteus vulgaris* than either antimicrobial singly. In addition, dissolving chlorobutanol in phenylethyl alcohol before dissolution in water eliminates the need to heat the solution.

OTC PRODUCTS FOR DRY EYE

Dry eye is a condition that can be either an annoying irritation in its mild form or a painful and destructive pathology in its severe form, damaging the corneal surface and interfering with vision. Possible causes of dry eye include decreased tear production, increased tear evaporation rate, or an abnormality of the tear film that decreases its natural capacity to protect and lubricate the epithelial tissues. It is one of the most common

complaints of patients seeking treatment from their eye-care doctor. This condition is somewhat unique in that it is largely treated with OTC monograph drug products, and the pharmacist is often asked to assist in its selection.⁶⁵ These products are variously known as artificial tears, ocular lubricants, demulcents, or emollients, and are available as solutions, gels, or ointments. Also unique is that these dry eye products are often the essential viscous-vehicle component of therapeutic ophthalmic products. Several reference books include the compositions of the various commercial dry eye products.^{33,34}

OTC dry eye products contain water-soluble polymers as active ingredients that protect and lubricate the mucous membrane surfaces of the conjunctiva and cornea, providing temporary relief of the symptoms of dryness and irritation. The polymers thicken the tear film and decrease the rate of tear loss in addition to lubricating and protecting the tissues. Mucin is a natural lubricating component of tears, and thus, dry eye products are expected to provide a mucomimetic effect. The active-ingredient demulcent (lubricant) polymers in these products, along with their permitted concentrations and labeling, are defined in an OTC monograph issued by the FDA.⁶⁶ The monograph does not specify viscosity grades, and therefore, a wide range of product viscosities is available. The more viscous products can provide longer duration of relief; however, they are also more likely to blur vision and leave a residue of polymer on the eyelids. The OTC products vary in their choice of polymer or combination of polymers, concentrations, and viscosity. The majority of OTC dry eye products contain one or more of the following permitted polymers: celluloses such as hydroxypropyl methylcellulose (hypromellose) or carboxymethylcellulose sodium; polyol liquids such as glycerin, polyethylene glycol, or polysorbate 80; polyvinyl alcohol; or povidone. Dextran 70 can be used only in combination with another permitted demulcent polymer.

The inactive components of OTC dry eye products are similar to those used in therapeutic ophthalmic products and require equally careful selection, keeping in mind that the monograph products are labeled for use as often as needed and thus must be nonirritating and physiologically compatible. In the moderate to severe forms of dry eye, the product may be dosed as often as every hour, and in the especially severe cases, as often as every 15 minutes. The preferred vehicle is aqueous for replenishing natural tears and dissolving the active and inactive ingredients. The vehicle is usually adjusted to be isotonic with sodium chloride, and in some cases other salts are used as well as nonionic ingredients, such as dextrose and mannitol. Some cases of dry eye have been reported to produce tears that are hypertonic, and several products feature a slightly hypotonic formula to restore the natural tear osmotic pressure. The products generally are adjusted or buffered to provide a pH of about 7.4, but rarely outside of the range of about 7.0 to 8.0.

Storage of solutions of some polymers, notably polyvinyl alcohol, can lead to a more acidic pH. The preservative chlorobutanol also produces an acidic pH as it degrades in solution. The bicarbonate included in some products is a more physiologically compatible buffer that requires a hermetically sealed secondary package for stability. Some products include both monovalent and divalent chloride salts in their vehicles in an attempt to mimic the electrolyte composition of natural tears.

The majority of dry eye products contain a preservative and are packaged in multiple-dose plastic ophthalmic containers. As in therapeutic products, BAC is the most widely used preservative. Newer preservatives having improved compatibility with ocular tissues, such as Polyquad, are increasingly in use, as well as the “disappearing” preservatives sodium perborate and Purite. Sodium perborate forms hydrogen peroxide, which can be irritating in the eye but quickly degrades to oxygen and water. Purite is a stabilized oxychloro complex that degrades to sodium chloride and water when exposed to long-wavelength ultraviolet light.

Preservative-free products are also available and are gaining popularity, particularly for patients with chronic dry eye

conditions and those who are sensitive to preservatives. These products are packaged in form-fill-seal unit-dose containers, but manufacturing considerations dictate that they contain enough product for several doses while bearing labels indicating that they are to be discarded after a single dose. This problem has led to the introduction of a modified package with a reclosable cap allowing its use as a small-volume multiple-dose container for preservative-free products when for safety reasons they are labeled to be discarded no longer than 12 hours after first opening.

The large majority of dry eye products are solution dosage forms administered as eyedrops, but gel-like and ointment dosage forms are also available. People who suffer with dry eye or who have a defect in lid closing that causes increased tear evaporation benefit from the these ointments, which consist primarily of a mixture of petrolatum and mineral oil that imparts emollient properties. Lanolin is sometimes added to the base to provide some water miscibility. The ratio of oils can be varied for consistency and melting temperature. Patients primarily use the ointments as a night-time medication, because they usually cause a blurring of vision. Several products bearing labels identifying them as gels or liquid gels are aqueous-based dosage forms containing one or more of the permitted lubricant polymers. The gel dosage form identifier is used to designate a more viscous product; however, the dosage directions read the same as for the eyedrop products (instill one or two drops as needed), as this dosage statement is the only one permitted for dry eye products marketed using the permitted lubricant polymers under the OTC monograph.

One disadvantage of most OTC dry eye products is their relatively short duration of effect. Several products have become available that incorporate novel delivery system technology into these products to prolong the duration of the lubricant and protectant effects of the monograph active ingredients. One product, AquaSite,⁶⁷ incorporates the polymer polycarbophil, and in a more recent product, Systane,⁶⁸ a modified hydroxypropyl guar polymer forms a gel-like matrix when the product components interact with the tear fluid.

CONTACT LENSES AND THEIR CARE

HISTORICAL BACKGROUND⁶⁹⁻⁷¹

Evidence suggests that the concept of altering corneal power was first envisioned by Leonardo da Vinci early in the sixteenth century. In the next century, more than a hundred years later, René Descartes described a device, a glass-filled tube in direct contact with the cornea, capable of implementing this concept, though it prohibited blinking and so was not a practical solution. In the early nineteenth century the British astronomer Sir John Herschel described the mathematics of these devices and proposed a means of treating very irregular corneas by using a glass capsule filled with a gelatin solution. Not until 1888 was the original concept executed by the artificial-eye maker, Albert Müller. He made a glass protective shell for the cornea of a lagophthalmic patient who had carcinoma of the upper lid. The patient wore the device for 20 years, and corneal clarity was maintained.

But perhaps the first contact lenses, scleral lenses resting on the bulbar conjunctiva beyond the limbal ring, were fabricated and fit by the German ophthalmologist A.E. Fick, working in Zurich late in the nineteenth century. Contact lenses without scleral portions (corneal lenses) were in existence at least as early as 1912, when they were manufactured by Carl Zeiss. Glass prostheses produced by Fick and others, while conceptually a step forward, suffer from reasonably rapid deterioration in the tear fluid. However, they had the advantage that the glass was readily wetted by tears.

Experimentation in the twentieth century led J. Dallos, working in Budapest, to perceive the importance of tear flow underneath the contact lens. Dallos also took impressions of

human eyes to improve the fidelity of the ground lens to the shape of the cornea. The first contact lenses, scleral lenses made of plastic in the late 1930s with polymethyl methacrylate (PMMA) from Rohm and Haas, are attributed to W. Mullen and T. Obrig. The advantages of durability and weight reduction far outweighed the slight differences in optical properties. And in 1948, K. Tuohy filed a patent for a plastic corneal contact lens. With the advent of PMMA, a flush-fitting shell became possible, a concept developed in England in the 1950s by Ridley. The first corneal lenses to have any measure of commercial success were developed in the early 1950s by F. Dickinson and W. Sohnges. Its thickness was about 0.2 mm, considered to be fairly thick. Thinner lenses, of about 0.1 mm, were introduced in the early 1960s.

The first soft contact lenses consisted of silicone, an elastomer nearly devoid of water but with good permeability to oxygen and carbon dioxide. However, the first soft contact lenses to be commercialized were of hydroxyethyl methacrylate hydrogels, developed by O. Wichterle and D. Lim in Czechoslovakia in the early 1960s. Continuous improvements have progressed to the present day, including milestones such as silicone acrylate rigid gas-permeable (RGP) lenses in the 1970s, disposable inexpensive lenses in the 1980s, and daily disposable and silicone hydrogel lenses in the 1990s.

With the advent of commercialized, relatively comfortable, inexpensive corneal contact lenses, the need for lens care products developed beginning in the 1950s and, as summarized below, a great variety of functionality now exists, depending on the requirements of the individual products.

From a regulatory perspective, during the period from their approval by the FDCA of 1938 until the US Medical Device Amendments to the FDCA of 1976, contact lenses and contact lens care products were regulated as drugs. Although during this period a device was defined in the FDCA, many medical devices were thought to be “used for the cure, mitigation, or prevention of disease,” leading to their classification as drugs. With increased activity in industry and significance of biomedical devices in maintaining health, congressional awareness arose regarding the advantages for greater regulation. The amendments made provision for three classes of device, with different requirements for each. Those with least risk were assigned to class I, and those associated with increasing risk, to a progressively higher class. If safety and efficacy of a device could be reasonably assured by “general controls” (including restrictions on design, production, storage, maintenance, and use), then the device was assigned to class I, which may be exempted from the requirement for a 510(k) (the section of the FDCA that allows for clearance of class II medical devices; see later) but still require company or product registrations. Under the 1976 amendments to the FDCA, contact lenses and lens care products were reclassified as class III devices, requiring a premarket approval application and approval by the FDA before marketing.

However, by 1990, following years of experience with certain devices in the marketplace and recognition that little purpose was being served duplicating regulatory investigation of substantially equivalent safe and effective devices, Congress passed new legislation. The US Safe Medical Devices Act of 1990 authorized reclassification of certain contact lens and lens care products as class II devices, requiring special controls in the form of a 510(k) submission and clearance. The products covered by this reclassification include solutions, dry products, tablets, and disinfecting units used in caring for and disinfecting soft (hydrophilic) and RGP contact lenses. In March 1994, the down-classification of two types of daily-wear contact lenses, from class III to class II, was implemented by a rule in the *Federal Register*. In July of 1997 a similar down-classification of certain contact lens care products, including multipurpose solutions and in-eye contact lens solutions, was implemented by a rule in the *Federal Register*. Extended-wear contact lenses remain governed by class III regulations.

Table 28-3. Contact Lens Classes, Characteristics, and Support Products

Lens Type	Chemical Classification	Major Characteristics	Typical Support Products
Hard, rigid, hydrophobic	PMMA (polymethyl methacrylate)	Negligible gas permeability, low water content, medium wettability	Wetting solutions Soaking solutions Cleaning solutions Combination Artificial tears
Soft, flexible, hydrophilic	HEMA (hydroxyethyl methacrylate)	High water content, low gas permeability, good wettability	Cleaning solutions Disinfection solutions
Flexible hydrophobic	Silicone rubber	Good gas permeability; poor wettability	Wetting solutions Cleaning solutions
	Silicone vinylpyrrolidone	Good gas permeability; good wettability	Soaking solutions
Rigid, hydrophilic	CAB (cellulose acetate butyrate)	Good gas permeability; good wettability	Wetting solutions Cleaning solutions Soaking solutions Rewetting solutions

CONTACT LENSES, CURRENT ART

As a consequence of these years of development, the consumer currently can choose from an enormous selection of contact lenses and lens materials. There are basically four classes of lenses, and the identification of these classes is both by polymer type and lens characteristics: hard contact lenses, RGP contact lenses, soft hydrophilic contact lenses, and silicone-based flexible hydrophobic lenses. These classes are summarized in Table 28-3. Hard contact lenses are generally fabricated from PMMA or polysilicone acrylate polymers or co-polymer blends containing one of these materials. RGP contact lenses are generally consist of more polarizable and oxygen-permeable materials such as cellulose acetate butyrate, *t*-butyl styrene, or silicone polymers or co-polymers. These lenses generally provide visual acuity superior to that of the more flexible soft contact lenses. The US Adopted Names (USAN) designates these hydrophobic lens types with less than 10 percent water by the suffix *focon*. However, RGP lenses have a low oxygen permeability and so should be removed and cleaned after a day's wear; moreover, because of greater lens awareness, especially a new wearer, should wear them routinely (i.e., not intermittently). They are durable, however, and so may be preferred because the replacement schedule is extended. RGP lenses with higher oxygen transmissibility tend to be preferred, especially in comparison with hard contact lenses, because the improved exchange of gases assists in maintaining the health of the cornea.

The third class of contact lenses, soft hydrophilic lenses, is fabricated from such a wide variety of materials and their blends that the FDA has grouped them according to two parameters, ionicity and water content (either low or high, less than or greater than 50 percent water):

- Group I includes the nonionic polymers of low water content, such as polymacon, a simple poly(hydroxyethyl methacrylate) (HEMA),
- Group II includes the nonionic polymers of high water content, such as alphafilcon A (a co-polymer of five monomeric units: HEMA, vinyl pyrrolidone, ethylene dimethacrylate, hydroxycyclohexyl methacrylate, and a substituted vinyl carbonate),
- Group III includes ionic polymers of low water content, such as bufilcon (a terpolymer of HEMA, dimethyl oxobutyl acrylamide, and the trimethacrylate ester of ethyl *t*-butane triol with some level of hydrolysis of the esters to the free acid) or ocufilcon A (a terpolymer with HEMA, methacrylic acid, and ethylene dimethacrylate), and
- Group IV includes ionic polymers of high water content, such as a higher water content ocufilcon B/C or etafilcon (a terpolymer of HEMA, the salt of methacrylic acid, and the trimethacrylate ester of ethyl *t*-butane triol).

The final class of contact lenses, considered either a group of harder soft lenses or a group of softer RGP lenses with high oxygen solubility and transmission, is that of the silicone hydrogel lenses. These silicone and fluorosilicone materials are candidates for continuous extended-wear applications.

Such diverse functionalities provide distinct properties in addition to water content, ranging from oxygen and metabolite transport, hardness, and flexibility, through processability, to durability and stability. More complete lists of the materials used in contact lens manufacture are available, but in general the USAN Council designates these hydrophilic lens types with greater than 10 percent water by the suffix *filcon*.⁷²⁻⁷⁴ While the nomenclature adequately distinguishes monomer components, and the alphabetical trailing designation indicates differences in ratios of monomer units, one must remember that differences in crosslinker (often ethylene glycol dimethacrylate), initiator, catalyst, filler, or color additive are not delineated.

While this is a reasonable overview of the categories of choices for the materials composing contact lenses, it is still an incomplete list with regard to the properties that concern both the consumer and his or her practitioner in selecting a contact lens. In addition to the chemical composition, the following choices are necessary: (1) duration of wear (daily vs. continuous), (2) lens replacement schedule (daily vs. biweekly vs. monthly vs. quarterly), (3) lens design (spherical, toric, orbifocal), (4) colored or clear, (5) means of manufacture (lathe-cut or cast-molded), (5) edge design (which can affect lens motion, overall fit, comfort, and acuity) and lens thickness, (6) fitting characteristics, like the selection available for base curves and lens diameter, and (7) any surface treatment or conditioning. With this multiplicity of choices and the subtlety of competing requirements, the recommendation of the practitioner is vital.

But ultimately, the consumer's personal metrics—comfort, convenience, and cost—also remain significant factors in lens selection rates. For example, patients with drier eyes are more likely to prefer group I lenses.

In addition to their use in vision correction, soft contact lenses, like collagen shields, find occasional use for therapeutic purposes, serving as bandages providing protection in treatment of bullous keratopathy (or other forms of corneal edema or other corneal irregularities or sources of cicatrization) or for lid abnormalities, and with ulcers, chemical burns, grafts, or dry eye. They are also efficacious in sustaining the delivery of therapeutic agents, as discussed previously. Patients will have to care for lenses that are used for these therapeutic purposes with the same preparations as those needed for lenses prescribed for refractive correction.

In summary, both doctors and consumers of contact lenses have a broad range of alternatives. Current lens materials and

the solutions to maintain them (next section) provide a wide selection of properties suitable for individual ocular health requirements and preferences.

CONTACT LENS CARE PRODUCTS

General Considerations

With the previous two sections for background, we are ready to discuss those preparations produced for the care of contact lenses. For all of the specialized solutions to be discussed, it is generally necessary (1) to adjust pH, tonicity (osmolality), surface and interfacial tensions, and viscosity; (2) preserve and maintain sterility; (3) assure stability and shelf-life; and (4) package and sterilize appropriately. In the United States these solutions are class II medical devices.⁷¹

Wetting Solutions

These are lubricating and cushioning preparations designed to furnish a hydrophilic coating over the characteristically hydrophobic surface of PMMA, silicon, acrylate, and other rigid lens surfaces. Typically, wetting solutions include an acceptable viscosity-imparting agent, a surfactant, and a preservative. The surface-activity and viscosity effects may be obtained from a single compound. Agents commonly used include cellulose derivatives, polyvinyl pyrrolidone, polyvinyl alcohol, and polyethylene glycol derivatives. The need for surface-active agents, which facilitate wetting of lenses and spreading of tears, is greater for hard contact lenses because of their hydrophobic surface characteristics. Preservatives include those acceptable for ophthalmic use, including polyquad and a polyhexylmethylene biguanide, called PHMB or DYMED.

Cleaning Solutions

Cleaning solutions commonly are used to remove surface contaminants—lipids, protein, organic salts, and the like. Cleaning is accomplished by the use of surfactants that preferably are nonionic or amphoteric. Viscosity-imparting agents generally are not included. Some cleaning agents also include a mild abrasive: silica for RGPs or plastic particles for soft lenses.

Adequate cleaning of hydrophilic lenses is a far more complex and challenging problem than for hard-lens cleaning. Because of their permeability characteristics, contaminants penetrate into the lens structure and may bind chemically or physically to the HEMA or ionic portions of the lens material. Contaminants may be surface films or crystals, amorphous aggregates of protein material, cellular debris, or insoluble inorganic salts.

Cleaning products generally are specific to the lens material and require FDA clearance, with proof of cleaning efficacy and safety. Cleaners are based on surface activity, enzyme action, or even abrasant action, in which case the abrasant material must be softer than the lens itself. Adequate daily cleaning of hydrophilic lens material can enhance disinfection but in some cases may not be necessary, given the choices of today's powerful disinfecting solutions and enzymatic cleaners, available in either a tableted or liquid form. Most recently, daily-wear planned-replacement lenses have found wide acceptance. Conventionally their successful use is reliant on enzymatic cleaning together with special disinfectants. However, extended-wear lenses, soiling at a lower rate, require only rewetting drops.

The significance of adequate cleaning cannot be overstated. Improperly or inadequately cleaned and maintained lenses are perhaps the most important contributors to complications associated with the wear of contact lenses such as discomfort, loss of visual acuity, conjunctivitis, and keratitis.

There are two types of cleaning solutions, daily and weekly cleaners. Weekly-cleaning solutions may include proteolytic enzymes like subtilisin, or a broad-spectrum less toxic enzyme like pancreatin (which contains protease, lipase, and amylase activity), or concentrated surfactants. Daily-cleaning solutions achieve cleaning using less aggressive surface-active agents and polymers, often nonionic but nonsolubilizing agents that will be

less toxic if carried over into the eye. Examples include Tween 20 or 80 (types of polysorbate detergents) and tyloxapols (surfactants). A daily cleaning solution even containing a nontoxic enzyme is available (SupraClens).

The goal is to provide thorough cleaning of the lenses for the benefits described previous without causing any degradation of the lenses.

Disinfecting Systems

Disinfection of the first hydrophilic lens approved by the FDA was accomplished using a heating device that generated steam from a saline solution. The latter was either prepared by the user or available from the manufacturer. Subsequent to the so-called thermal systems, requiring heating at 80°C for 10 minutes, disinfection solutions were developed that met the requirements for FDA approval while not damaging the more delicate soft lens materials. Because of the sorption characteristics of hydrophilic lens materials, many of the accepted ophthalmic preservatives are unsatisfactory for use in soft-lens disinfecting systems, including the ubiquitous BAC. Once again, however, the use of a quaternary disinfectant covalently bonded to a soluble, relatively high-molecular-weight polymer has met with some success. These have to a large extent replaced chemical antimicrobial agents like sorbic acid, thimerosal (containing mercury), and hydrogen peroxide, among many others. Perhaps the most common in current use are a cationic polymer, polyquad, and a polyhexylmethylene biguanide.

In addition to possessing satisfactory disinfecting activity, such a preparation must be isotonic, be in an acceptable pH range, be nonreactive (nonbinding) with lens materials, and over a normal use period induce or bring about no physical, chemical, or optical changes in the lens. It is, of course, sterile and safe for use in the eye, even though direct instillation into the eye is not intended.

Soaking Solutions

Soaking or storage solutions, as the name suggests, serve in storage and hydration of hard or RGP lenses but, most importantly, in disinfection of such lenses. Disinfection should be rapid and as complete as possible, making use, once again, of acceptable ophthalmic preservative substances. Soaking solutions typically contain chlorhexidine (gluconate), benzalkonium, or quaternary/polymer compounds, enhanced by sodium edetate. Soaking solutions are intended to be rinsed off lenses before insertion.

Rewetting Solutions

Solutions intended to rewet hard or RGP lenses *in situ* are referred to as rewetting solutions or lubricating drops. Such preparations are intended to reinforce the wetting capacity of the normal tear film. Early products of this type tended to be somewhat viscous wetting solutions, acceptable for direct instillation into the eye. More recent preparations mimic tears more accurately, and their viscosity may be rather low, to improve user acceptability, or adjusted to be responsive and to improve retention and duration of the wetting characteristics.

Multipurpose Solutions

By combining the actions of two or more of the solutions just described, these products simplify regimens, improve compliance, generally increase the comfort of contact lens products, and greatly increase the convenience of using contact lenses. For soft contact lenses, multipurpose solutions have to combine only the functions of cleaning and disinfection, whereas a complete multipurpose solution for a hard or RGP lens must also provide some level of wetting. Suffice it to say that these solutions, which contain an intricate balance of ingredients, have grown in popularity significantly because of the greatly increased level of convenience they provide. Their efficacy often precludes the need for separate cleaning solutions for many patients.

GUIDELINES FOR SAFETY AND EFFICACY TESTING

The FDA periodically issues or updates guidelines describing recommended test procedures for contact lens care products. The reader is advised to review the most recent guidelines for appropriate requirements.

Current guidelines cover all lens care products. Applicable guidelines are determined by the type of product (e.g., saline solution, cleaning solution, multipurpose solution, and rewetting drops) and the claim (e.g., disinfection in a regimen as opposed to a stand-alone process). Safety testing includes determinations of cytotoxicity, oral toxicity, potential for sensitization, and single to multiple exposure in an animal (rabbit) eye. The tests designated depend on the active ingredient (e.g., if it is a new entity or has been in previous products) and its concentration (e.g., if it is the same or different from that used in a currently approved product). Efficacy of preservation, disinfection, and cleaning, compatibility with lens or lens types, and clinical trials are also required, with their type and scope dependent on product type, characteristics, ingredients, and claims.

SUMMARY

There has been considerable progress in ophthalmic pharmaceuticals and in lens care products during the last decade. Very substantial advances have occurred in designing vehicles and packaging for highly potent active ingredients presented at very low concentrations; in increasing ophthalmic bioavailability and controlling factors influencing ophthalmic drug absorption; in the design of implants and means of delivering them for providing therapeutic agents to the retina and other deep ophthalmic tissues; and in devising robust yet delicately balanced multipurpose solutions for contact lens wearers. Continuing advances in the general field of ophthalmic pharmaceuticals and pharmacokinetics can be expected to assist in maintaining and improving ocular health.

ACKNOWLEDGEMENTS: The authors gratefully acknowledge insights provided by colleagues, noting especially Renee Garofalo, Terry Dagnon, and Denise Rodeheaver.

REFERENCES

- Lang JC, Roehrs RE, Jani R. Ophthalmic preparations. In: Beringer P, ed. *Remington: The Science and Practice of Pharmacy*, 21st edn. Baltimore: Lippincott Williams & Wilkins, 2006: Chap 28.
- US Pharmacopeia, 26th revision, and *National Formulary*, 21st edn. Rockville, MD: US Pharmacopeial Convention, 2003.
- European Pharmacopoeia*, 4th edn. Strasbourg: Directorate for the Quality of Medicines of the Council of Europe (EDQM), Council of Europe, 2001.
- Japanese Pharmacopoeia* 14th edn. Tokyo: Pharmaceuticals and Medical Devices Agency, 2001.
- Alcon Research Ltd, Global Graphics Dept., Fort Worth, TX.
- Botelho SY. Tears and the lacrimal gland. *Sci Am* 1964; 211: 78–86.
- Slatt BJ, Stein HA. *Ophthalmic Assistant: Fundamentals and Clinical Practice*. St Louis: Mosby, 1983.
- Vaughan D, Asbury T. *General Ophthalmology*, 11th edn. Los Altos, CA: Lange Medical Publications, 1986.
- Bron AS et al. *Wolff's Anatomy of the Eye and Orbit*, 8th edn. New York: Chapman and Hall Medical, 1997.
- Berman ER. *Biochemistry of the Eye*. New York: Plenum Press, 1991: Chap 4.
- Spaeth GL, ed. *Ophthalmic Surgery Principles and Practice*, 3rd edn. Philadelphia: WB Saunders, 2003.
- Lee TW-Y, Robinson JR. Controlled-release drug-delivery systems. In: Gennaro AR, ed. *Remington: The Science and Practice of Pharmacy*. Baltimore: Lippincott Williams & Wilkins, 2000: Chap 47.
- Edelhauser HF et al. *In vitro* ion and water movement in corneas of Rainbow Trout. *Invest Ophthalmol* 1965;4: 290–296.
- Schoenwald RD, Ward RL. Corneal penetration behavior of beta-blocking agents I: physiochemical factors. *J Pharm Sci* 1983;72: 1266–1272.
- Edwards A, Prausnitz MR. Predicted permeability of the cornea to topical drugs. *Pharm Res* 2001;18: 1497–1508.
- Lang JC et al. Design and evaluation of ophthalmic pharmaceutical products. In: Banker GS, Rhodes CT, eds. *Modern Pharmaceuticals*, 4th edn. New York: Marcel Dekker, 2002.
- Shell JW. Pharmacokinetics of topically applied ophthalmic drugs. *Surv Ophthalmol* 1982;26: 207–218.
- Patton TF. Pharmacokinetic evidence for improved ophthalmic drug delivery by reduction of instilled volume. *J Pharm Sci* 1977;66: 1058–1059.
- Vaidyanathan G et al. Scintigraphic evaluation of the ocular disposition of ¹⁸F-imirestat in rabbits. *Pharm Res* 1990;7: 1198–1200.
- Santvliet LV, Ludwig A. Influence of the dropper tip design on the size of eye-drops. *Pharm Ind* 2001;63: 402–409.
- Conroy CW, Maren TH. The ocular distribution of methazolamide after corneal and scleral administration: the effect of ionization state. *J Ocul Pharm Ther* 1998;14: 565–573.
- Lee VHL, Robinson JR. Mechanistic and quantitative evaluation of precorneal pilocarpine disposition in albino rabbits. *J Pharm Sci* 1979;68: 673–684.
- Lee VHL. New directions in the optimization of ocular drug delivery. *J Ocul Pharmacol* 1990;6(2): 157–164.
- Lang JC, Stiemke MM. Biological barriers to ocular delivery. In: Reddy IK, ed. *Ocular Therapeutics and Drug Delivery, A Multidisciplinary Approach*. Basel: Technomic Publishing AG, 1996: 51–132.
- Martin AN et al. *Physical Pharmacy: Physical Chemical Principles in the Pharmaceutical Sciences*. Philadelphia: Lea & Febiger, 1983.
- United States Patent No. 4,861,760 (Merck Manual).
- United States Patent No. 5,212,162 (Lang).
- United States Patent No. 6,174,524 (Bawa).
- Gaudana R et al. Ocular drug delivery. *AAPS J* 2010;12: 348–360.
- Missel PJ et al. Simulating dissolution of intravitreal triamcinolone acetonide suspensions in an anatomically accurate rabbit eye model. *Pharm Res* 2010; 27: 1530–1546.
- United States Patent No. 7,621,907 (Rodstrom).
- United States Patent No. 6,416,777 (Yaacobi).
- Ophthalmic Drug Facts, Facts and Comparisons*. St. Louis: Mosby, 2003.
- Physicians' Desk Reference for Ophthalmic Medicines*, 31st edn. 2003.
- United States Patent No. 4,861,760 (Manzuel).
- United States Patent No. 6,174,524 (Bawa).
- United States Patent No. 5,461,081 (Ali).
- United States Patent No. 4,911,920 (Jani).
- Higuchi T. Rate of release of medicaments from ointment bases containing drugs in suspension. *J Pharm Sci* 1961;50: 874–875.
- United States Patent No. 5,474,979 (Ding).
- March WF et al. Duration of effect of pilocarpine gel. *Arch Ophthalmol* 1982; 100: 1270–1271.
- Shell JW, Baker RW. Diffusional systems for controlled release of drugs to the eye. *Ann Ophthalmol* 1975; 7: 1037–1045.
- Mitra AK, ed. *Ophthalmic Drug Delivery Systems*. New York: Marcel Dekker, 1993: Chap 17.
- Edelhauser HF et al. Intraocular irrigating solutions. Their effect on the corneal endothelium. *Arch Ophthalmol* 1975;93: 648–657.
- Edelhauser HF et al. Intraocular irrigating solutions. A comparative study of BSS Plus and lactated Ringer's solution. *Arch Ophthalmol* 1978;96: 516–520.
- Slack JW et al. A bisulfate-free epinephrine solution. *Am J Ophthalmol* 1990;110: 77–82.
- United States Patent No. 5,378,475 (Smith).

48. Mitra AK, ed. *Ophthalmic Drug Delivery Systems*, 2nd edn. New York: Marcel Dekker, 2003: Chap 10.
49. Mitra AK, ed. *Ophthalmic Drug Delivery Systems*, 2nd edn. New York: Marcel Dekker, 2003: Chap 12.
50. Aronson SD, Elliott JH. *Ocular Inflammation*. St Louis: Mosby, 1972: 899.
51. Title 21, *Code of Federal Regulations*, Chap 1, Vol 4, Parts 210 & 211.
52. Title 21, *Code of Federal Regulations*, Chap1, Vol 5, Part 330.
53. Allen LV Jr. *The Art, Science and Technology of Pharmaceutical Compounding*, 3rd edn. Washington, DC: American Pharmacists Association, 2008.
54. *US Pharmacopeia*, revision 26, Chapter <1206> Sterile Drug Products for Home Use. Rockville, MD: US Pharmacopeial Convention, 2003.
55. Block SS. *Disinfection, Sterilization, and Preservation*. Malvern, PA: Lea & Febiger, 1991.
56. Gonnering R *et al.* The pH tolerance of rabbit and human corneal endothelium. *Invest Ophthalmol Vis Sci* 1979; 18: 373–390.
57. Edelhauser HF *et al.* Osmotic tolerance of rabbit and human corneal endothelium. *Arch Ophthalmol* 1981; 99: 1281–1287.
58. United States Patent Nos. 5,849,792 and 6,011,062 (Schneider).
59. United States Patent No. 6,235,781 (Weiner).
60. United States Patent Nos. 5,474,209 (Valet) and 5,782,345 (Guasch).
61. Title 21, *Code of Federal Regulations*, Chap 1, Part 200, Subpart C, Section 200.50.
62. *United States Pharmacopeia* 26, Chapter § 513 Antimicrobial Effectiveness Testing. United States Pharmacopeial Convention, 2003.
63. Hoover JE. *Dispensing of Medication*, 8th edn. Easton, PA: Mack Publishing Co, 1976: 237.
64. Mullen W *et al.* Ophthalmic preservatives and vehicles *Surv Ophthalmol* 1973;17: 469–483
65. Berardi R *et al.*, eds. *Handbook of Nonprescription Drugs*, 13th edn. Washington, DC: American Pharmacists Association, 2002.
66. Title 21, *Code of Federal Regulations*, Chap 1, Vol 5, Part 349.
67. United States Patent No. 5,188,826 (Chandrasekaran)
68. United States Patent No. 6,403,609 (Asgharian).
69. Mandel RB. *Contact Lens Practice*. Springfield, IL: Charles C Thomas, 1981.
70. Efron N, ed. *Contact Lens Practice*. Boston: Butterworth Heinemann, 2002.
71. Randeri KJ *et al.* Lens care products. In: Swarbrick J, Boylan JC, eds. *Encyclopedia of Pharmaceutical Technology*, Vol 8. New York: Marcel Dekker, 1993: 361–402.
72. *Contact Lens Quarterly*. Irvine, CA: Jobson Publishing, Frames Data, Inc.
73. *Tyler's Quarterly*, Little Rock, AR: Tyler's Quarterly, Inc.
74. Expert Committee on Nomenclature and Labeling of the USP Council of Experts. *USP and USAN Dictionary*. Rockville, MD: US Adopted Names Council, Rockville, MD.

Chapter 29

Medicated Topicals

Lawrence H. Block, PhD

INTRODUCTION 565

THE SKIN AND ITS APPENDAGES 565

DRUG DELIVERY TO THE SKIN AND ITS APPENDAGES—BASIC PRINCIPLES 566

THE FORMULATION OF DOSAGE FORMS AND DELIVERY SYSTEMS FOR TOPICAL DRUG APPLICATION 569

CONVENTIONAL TOPICAL DOSAGE FORMS—SEMISOLIDS 571

OTHER DOSAGE FORMS FOR APPLICATION TO THE SKIN AND ITS APPENDAGES 572

RECTAL, VAGINAL, AND URETHRAL ROUTES OF ADMINISTRATION 574

INTRODUCTION

The application of medicinal substances to the skin and its appendages (e.g., nails) or various body orifices is a concept as old as humanity. The papyrus records of ancient Egypt describe a variety of such medications for external use. Galen described the use in Roman times of a forerunner to today's vanishing creams.

Medications are applied in a variety of forms reflecting the ingenuity and scientific imagination of pharmacists through the centuries. New modes of drug delivery have been developed to remedy the shortcomings of earlier vehicles or, more recently, to optimize drug delivery. Conversely, some external medications have fallen into disuse because of changes in the practice of medicine.

Medications are applied to the skin or its appendages or inserted into body orifices in liquid, semisolid, or solid form. Accordingly, emphasis in this chapter has been placed on (1) the epidermal, dermal, and transdermal (percutaneous) routes, and (2) the rectal, vaginal, and urethral routes and their corresponding formulations. Nasal and intranasal products, ophthalmics, and topical aerosol products will not be discussed in this chapter.

THE SKIN AND ITS APPENDAGES

ANATOMIC AND PHYSIOLOGIC ASPECTS

The Skin

The skin often has been referred to as the largest of the body organs: an average adult's skin has a surface area of about 2 m². It is probably the heaviest organ of the body. Its accessibility and the opportunity it affords to maintain applied preparations intact for a prolonged time have resulted in its increasing use as a route of drug administration, whether for local, regional, or systemic effects.

Anatomically, human skin may be described as a stratified organ with three distinct tissue layers: the epidermis, the dermis, and the subcutaneous fat layer (Figure 29-1).

Epidermis, the outermost skin layer, consists of stratified squamous epithelial cells. Keratinized, flattened remnants of these actively dividing epidermal cells accumulate at the skin surface as a relatively thin region (about 10–20 μ m thick) termed the stratum corneum, or horny layer. The horny layer is itself lamellar with the keratinized cells overlapping one another, linked by intercellular bridges and compressed into about 15 layers. The lipid-rich intercellular space in the stratum corneum consists of lamellar matrices with alternating hydrophilic layers and lipophilic bilayers formed during the process of keratinization. The region behaves as a tough but flexible coherent membrane.

The stratum corneum also is markedly hygroscopic—far more so than other keratinous materials, such as hair or nails. Immersed in water, the isolated stratum corneum swells to about three times its original thickness, absorbing about four to five times its weight in water in the process. The stratum corneum functions as a protective physical and chemical barrier and is only slightly permeable to water. It retards water loss from underlying tissues, minimizes ultraviolet light penetration, and limits the entrance of microorganisms, medications, and toxic substances from without. The stratum corneum is abraded continuously. Thus, it tends to be thicker in regions more subject to abrasion or the bearing of weight. Its regeneration is provided by rapid cell division in the basal cell layer of the epidermis. Migration or displacement of dividing cells toward the skin surface is accompanied by differentiation of the epidermal cells into layers of flat, laminated plates, as noted above. An acidic film, (pH ranging between 4 and 6.5, depending on the area tested) made up of emulsified lipids, covers the surface of the stratum corneum.

The dermis can be described as a gel structure involving a fibrous protein matrix, embedded in an amorphous, colloidal, ground substance. Protein, including collagen and elastin fibers, is oriented approximately parallel to the epidermis. The dermis supports and interacts with the epidermis, facilitating its conformation to underlying muscles and bones. Blood vessels, lymphatics, and nerves are found within the dermis, although only nerve fibers reach beyond the dermal ridges or papillae into the germinative region of the epidermis. Sweat glands and hair follicles extending from the dermis through the epidermis provide discontinuities in an otherwise uniform integument.

The subcutaneous fat layer serves as a cushion for the dermis and epidermis. Collagenous fibers from the dermis thread between the accumulations of fat cells, providing a connection between the superficial skin layers and the subcutaneous layer.

THE SKIN'S APPENDAGES – HAIR, SEBACEOUS GLANDS, AND SWEAT GLANDS

Human skin has numerous surface openings (i.e., hair follicles and sweat glands) extending well into the dermis (see Figure. 29-1). Hair follicles, together with the sebaceous glands that empty into the follicles, make up the pilosebaceous unit. Human hair consists of compacted keratinized cells formed by the follicles. The hair follicles, surrounded by sensory nerves, are able to provide a sensory function. Human hair varies enormously within the same individual, even within the same specific body area. Follicular density varies considerably as well, from values of about 250 follicles per cm² for the scalp to 50 per cm², or less, for the thigh and other relatively nonhirsute areas. Follicular density is determined genetically (no new follicles are formed after birth). One characteristic human trait is that although most of the body hairs never develop beyond the rudimentary vellus state, the only hairless areas are confined,

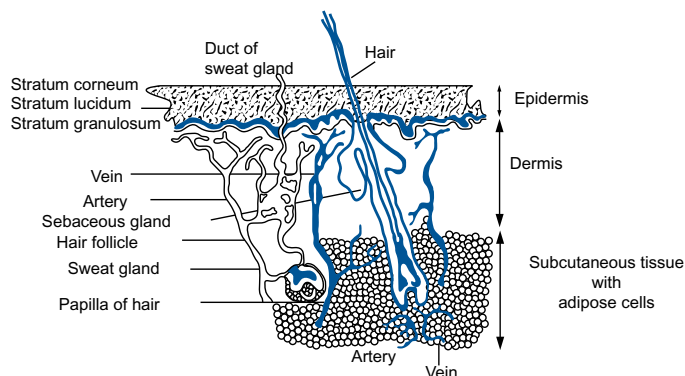


Figure 29-1. Vertical section of human skin.

primarily, to the palmar and plantar surfaces. Individual hairs can vary in microscopic appearance, diameter, cuticle appearance, and even presence or absence of medulla.

Sebaceous glands are similar anatomically and functionally but vary in size and activity according to location. Population in the scalp, face, and anogenital areas may vary from 400 to 900 per cm^2 . Fewer than 100/ cm^2 are found in other areas. Sebaceous glands are richly supplied with blood vessels.

Sebaceous cells synthesize and accumulate lipid droplets. This accumulation results in enlarged cells that fragment to form sebum. Sebum is made up of a mixture of lipids: triglycerides (57.5 percent), wax esters (26 percent), squalene (12 percent), cholesterol esters (3.0 percent), and cholesterol (1.5 percent).

The sebaceous gland, containing sebum, cell debris, and microorganisms such as *Propionibacterium acnes*, is connected to the pilosebaceous canal by a duct of squamous epithelium. When access to the surface is blocked and bacteria multiply, the result is the comedo of acne.

Sweat glands are classified as either apocrine or eccrine. Apocrine glands are secretory but are not necessarily responsive to thermal stimulation. Such glands do not produce sweat in the normal sense of the word. Apocrine glands, however, often are associated with eccrine sweat glands, particularly in the axilla. Eccrine sweat glands are coiled secretory glands, equipped with a blood supply, extending from the dermis to the epidermal surface. Eccrine sweat glands function to regulate heat exchange in humans. As such, they are indispensable to survival. About 3 million eccrine glands are distributed over the human body. Distribution varies from less than 100 to more than 300 per cm^2 .

THE SKIN'S APPENDAGES—NAILS

Nails facilitate our ability to use our fingers and toes to grasp and manipulate objects, protect our fingers and toes from physical trauma, and, from a cosmetic perspective, may be indicative of our socioeconomic status. Thus, human nail functionality extends beyond that of corresponding structures (i.e., nails, hooves, and claws) in other animals. The nail unit includes the nail folds, nail matrix, nail bed, and the skin under the free edge of the nail plate, i.e., the hyponychium. The nail plate (about 0.25–0.6 mm thick) consists of about 25 layers of highly interconnected keratinized cells and detaches from the nail bed at the hyponychium. Unlike the stratum corneum of the skin, the nail plate has been described as a substantially crosslinked, minimally lipoidal hydrogel with poor permeability to drug molecules.

DRUG DELIVERY TO THE SKIN AND ITS APPENDAGES—BASIC PRINCIPLES

THE SKIN

Drugs are applied to the skin to elicit one or more of four general effects: an effect on the skin surface, an effect within the

stratum corneum, a more deep-seated effect requiring penetration into the epidermis and dermis, or a systemic effect resulting from delivery of sufficient drug through the epidermis and the dermis to the vasculature to produce therapeutic systemic concentrations.

Drug activity on the skin surface may be in the form of a protective film (e.g., a zinc oxide cream or a sunscreen), an action against surface microorganisms, or a cleansing effect. Surface films may be somewhat occlusive and provide a moisturizing effect by diminishing loss of moisture from the skin surface. In such instances the film or film formation *per se* fulfills the objective of product design. The action of antimicrobials against surface flora requires more than simple delivery to the site. The vehicle must facilitate contact between the surface organisms and the active ingredient. Skin cleansers employ soaps or surfactants to facilitate the removal of superficial soil.

Drug effects within the stratum corneum are seen with certain sunscreens; *p*-aminobenzoic acid is an example of a sunscreensing agent that both penetrates and is substantive to stratum corneum cells. Skin moisturization takes place within the stratum corneum. Whether it involves the hydration of dry outer cells by surface films or the intercalation of water in the lipid-rich, intercellular laminae, the increased moisture results in the softening of the skin. Keratolytic agents, such as salicylic acid, act within the stratum corneum to cause a breakup or sloughing of stratum corneum cell aggregates. This is particularly important in conditions of abnormal stratum corneum such as psoriasis, a disease characterized by thickened, scaly plaques.

The stratum corneum also may serve as a *reservoir phase* or depot, wherein topically applied drug accumulates due to partitioning into or binding with skin components. This interaction can limit the subsequent migration of the penetrant unless the interaction capacity of the stratum corneum is surpassed by providing excess drug. The drug depot may also serve to sustain the flux of drug into the living epidermis and beyond. Examples of drugs that exhibit significant skin interaction include benzocaine, estrogens, scopolamine, and corticosteroids.

Given the mechanical resistance offered by the stratum corneum to topically applied drugs, the penetration of a drug into the viable epidermis and dermis may be difficult to achieve. But, once transepidermal permeation or percutaneous absorption has occurred, the continued diffusion of drug into the dermis is likely to result in drug transfer into the microcirculation of the dermis and then into general circulation. Nonetheless, it is possible to formulate drug delivery systems that provide substantial localized delivery without achieving correspondingly high systemic concentrations. Limited studies in man of topical triethanolamine salicylate, minoxidil, and retinoids—among others—demonstrate the potential of this approach.

Unwanted systemic effects stemming from the inadvertent transdermal penetration of drugs have been reported for a wide variety of compounds (e.g., benzocaine, hexachlorophene, lindane, corticosteroids, or *N,N*-diethyl-*m*-toluamide) over the years. With the commercial introduction of transdermal drug delivery systems for numerous drugs (e.g., clonidine, 17β -estradiol, fentanyl, lidocaine, methylphenidate, nitroglycerin, nicotine, oxybutynin, rivastigmine, rotigotine, scopolamine, selegiline, and testosterone), transdermal penetration is now generally regarded as an opportunity rather than a nuisance.

Nonetheless, for a drug to qualify as a candidate for systemic delivery after topical application, it must satisfy additional requirements beyond the exhibition of good skin permeation. Successful candidates for transdermal drug delivery should be nonirritating and nonsensitizing to the skin. Since relatively little drug may reach systemic circulation over a relatively long time, drug candidates should be relatively potent drugs. In addition, the limitation to relatively potent drugs can ease problems of formulation, since the amount of drug that can be incorporated in the formulation may be limited by physicochemical considerations, such as solubility.

Percutaneous absorption involves the transfer of drug from the skin surface into the stratum corneum, under the aegis of a concentration gradient, and its subsequent diffusion through the stratum corneum and underlying epidermis, through the dermis, and into the microcirculation. The skin behaves as a passive barrier to diffusing molecules. Evidence for this includes the fact that the impermeability of the skin persists long after the skin has been excised. Furthermore, Fick's law is obeyed in the vast majority of instances.

Molecular transport through the various strata of the skin is limited by the diffusional resistances encountered in each of the regions. The total diffusional resistance (R_{skin}) to drug permeation through the skin can be described as the sum of diffusional resistances offered by each of the successive regions of the skin:

$$R_{skin} = \sum (R_{sc} + R_e + R_{pd})$$

where R is the diffusional resistance, and the subscripts sc , e , and pd refer to the stratum corneum, epidermis, and papillary layer of the dermis, respectively. In addition, resistance to transfer into the microvasculature limits the systemic delivery of drug.

By and large, the greatest resistance to drug penetration is met in the stratum corneum (i.e., diffusion through the stratum corneum tends to be the rate-limiting step in percutaneous absorption).³

The role of hair follicles and sweat glands in percutaneous absorption must be considered, though their effect is minimized by the relatively small fractional areas occupied by these appendages (at least in humans). On the other hand, liposomal vehicles and microbead (3–10 μm diameter) suspensions appear to accumulate selectively in pilosebaceous and perifollicular areas. In the very early stages of absorption, transit through the appendages may be comparatively large, particularly for lipid-soluble molecules and those whose permeation through the stratum corneum is relatively low. Surfactants and volatile organic solvents, such as ethanol, have been found to enhance drug uptake via the transfollicular route.

Rather than characterizing drug transfer into and through the skin in terms of the diffusional resistances encountered, one could define permeation in terms of the *pathways* followed by the diffusing species. Drug permeation through the intact skin of humans involves either an intercellular or transcellular path in the stratum corneum, for the most part, rather than the so-called shunt pathways (i.e., the transglandular or transfollicular routes).

The conventional wisdom is that for the most part, lipophilic compounds transfer preferentially into the lipoidal intercellular phase of the stratum corneum, while relatively more hydrophilic compounds transfer into the intracellular domain of the stratum corneum. One should keep in mind that the often-postulated biphasic character of the horny layer—with hydrophilic cells in a lipophilic matrix—is overly simplistic: the hydrophilic cells themselves are enclosed within lipid bilayer membranes, while the lipophilic matrix comprises intercellular lipids that are, in fact, present in lamellar structures that “sandwich in” hydrophilic layers. The intercellular pathway is *bicontinuous*, consisting of both a nonpolar and a polar diffusion pathway between the corneocytes. Thus, attempts to predict permeability constants from oil:water or solvent:water partition coefficients have had limited success.

The stratum corneum can be regarded as a passive diffusion membrane but not an inert system; it often has an affinity for the applied substance. The adsorption isotherm often is linear in dilute concentration ranges. The correlation between external and surface concentrations is given in terms of the tissue:solvent distribution coefficient K_m :

$$K_m \text{ is the } \frac{\text{solute absorbed per cm}^3 \text{ of tissue}}{\text{solute in solution per cm}^3 \text{ of solvent}} = \frac{C_m}{C_s}$$

The integrated form of Fick's law is given as

$$J_s = \frac{K_m DC_s}{\delta}$$

and

$$K_p = \frac{K_m D}{\delta}$$

where K_p is the permeability coefficient, J_s is the steady state flux of solute, C_s is the concentration difference of solute across the membrane, δ is the membrane thickness, and D is the average membrane diffusion coefficient for the solute.

Percutaneous absorption of a drug can be enhanced by the use of occlusive techniques or by the use of so-called penetration enhancers.^{4–6}

Occluding the skin with wraps or impermeable plastic film such as Saran Wrap prevents the loss of surface water from the skin. Since water is absorbed readily by the protein components of the skin, the occlusive wrap causes greatly increased levels of hydration in the stratum corneum. The concomitant swelling of the horny layer ostensibly decreases protein network density and the diffusional path length. Occlusion of the skin surface also increases skin temperature ($\sim 2^\circ\text{--}3^\circ$), resulting in increased molecular motion and skin permeation.

Hydrocarbon bases that occlude the skin to a degree will bring about an increase in drug penetration. However, this effect is trivial compared with the effects seen with an occlusive skin wrap. Occlusive techniques are useful in some clinical situations requiring anti-inflammatory activity, and occlusive wrappings have been commonly used with steroids. Since steroid activity can be enhanced by two orders of magnitude by skin occlusion, it is possible to depress adrenal function unknowingly.

Transdermal drug delivery systems, with their occlusive backing, can effect increased percutaneous absorption as a result of increased skin temperature and hydration. Investigators have shown that exposure of the surrounding skin area to localized heating (e.g., via hot tubs and saunas) or cooling could cause extensive changes in transdermal drug bioavailability, presumably due to changes in regional cutaneous blood flow and subsequent systemic uptake.

One additional consequence of occlusion of the skin surface, whether by a transdermal delivery system or a hydrophobic film, is that an aqueous film may form at the formulation-skin interface. This aqueous film or interface could result in decreased transfer efficiency and, in the case of a transdermal delivery system, a loss of adhesion. Accordingly, the suppression of perspiration could enhance vehicle-skin partitioning efficiency and drug permeation.

The term “penetration enhancer” has been used to describe substances that facilitate drug absorption through the skin. Penetration enhancers with a direct effect on skin permeability include solvents, surfactants, and miscellaneous chemicals, such as urea and *N,N*-diethyl-*m*-toluamide (Table 29-1). While most such materials have a direct effect on the permeability of the skin, other so-called enhancers (e.g., polyols, such as glycerin and propylene glycol) appear to augment percutaneous absorption by increasing the thermodynamic activity of the penetrant, thereby increasing the effective escaping tendency and concentration gradient of the diffusing species. The mechanism of action of these enhancers is complex since these substances also may increase drug solubility. Nonetheless, the predominant effect of these enhancers is to either increase the stratum corneum's degree of hydration or disrupt its lipoprotein matrix. In either case, the net result is a decrease in resistance to drug (penetrant) diffusion. (The formulator should note that the inclusion of a penetration enhancer in a topical formulation mandates additional testing and evaluation to ensure the absence of enhancer-related adverse effects.)

Foremost among the solvents that affect skin permeability is water. Water is a factor even for *anhydrous* transdermal delivery systems due to their occlusive nature. Due to its safety and

Table 29-1. Penetration Enhancers

Solvents
Water
Alcohols
Methanol
Ethanol
2-Propanol
Alkyl methyl sulfoxides
Dimethyl sulfoxide
Decylmethyl sulfoxide
Tetradecylmethyl sulfoxide
Pyrrolidones
2-Pyrrolidone
N-Methyl-2-pyrrolidone
N-(2-Hydroxyethyl)pyrrolidone
Laurocapram
Miscellaneous solvents
Acetone
Dimethyl acetamide
Dimethyl formamides
Tetrahydrofurfuryl alcohol
Amphiphiles
I- α -Amino acids
Anionic surfactants
Cationic surfactants
Amphoteric surfactants
Nonionic surfactants
Fatty acids and alcohols
Miscellaneous
Clofibric acid amides
Hexamethylene lauramide
Proteolytic enzymes
Terpenes and sesquiterpenes
α -Bisabolol
d-Limonene
Urea
N,N-ethyl-m-toluamide

efficacy, water has been described as the ultimate penetration enhancer. Other solvents include the classic enhancer, dimethyl sulfoxide (DMSO), which is of limited utility because of its potential ocular and dermal toxicity, its objectionable taste and odor (a consequence of its absorption and subsequent biotransformation), and the need for concentrations in excess of 70% to promote absorption. Analogs of DMSO such as decylmethyl sulfoxide are used currently in some topical formulations. In contrast to many other solvents, laurocapram (Azzone) has been shown to function effectively at low concentrations ($\leq 5\%$). Furthermore, laurocapram's effect on skin permeability persists long after a single application, due apparently to its prolonged retention within the stratum corneum.

Surfactants, long recognized for their ability to alter membrane structure and function, can have a substantial effect on skin permeability. However, given the irritation potential of surfactants applied chronically, their utility as penetration enhancers is limited. Their effect on permeability may be complicated further by surfactant monomer aggregation to form micelles and the concomitant solubilization of the permeant. As the impact of surfactants on skin permeability of a penetrant is problematic, the effect of their inclusion in a formulation should be evaluated using appropriate *in vitro* and *in vivo* studies.

Even though *in vitro* studies of percutaneous transport may reflect the resistance of the skin to drug diffusion, such studies cannot characterize adequately the transfer of diffusing drug into the microvasculature of the dermis and its subsequent transfer into general circulation. There is evidence in the literature of delayed dermal clearance in the aged. This may reflect, in part, a decrease in older subjects in dermal capillary loop density, a

decrease in the rate and/or extent of dermal blood perfusion, or an increase in resistance to transfer into the capillaries.

The potential effect of blood-flow-limited percutaneous absorption underscores the possible occurrence of cutaneous interactions between vasoconstrictors or vasodilators and topically applied drugs intended for a systemic effect: bioavailability could be increased or diminished as a result. The assessment of the potency of corticosteroids by corticosteroid-induced skin blanching (vasoconstriction) lends credence to this issue.

Another factor to consider is that the stratum corneum of older subjects may offer less resistance to the penetration of topically applied drugs due to the age-related increase in photochemical and mechanical damage to the skin.

Given the substantial intersubject variations that occur in diffusional resistance and in dermal clearance, it is not surprising that *in vivo* studies of percutaneous absorption often demonstrate marked intersubject differences in systemic availability of drugs. Furthermore, the tendency to employ normal, healthy, young adults in bioavailability studies may not provide data that is indicative of drug permeation through the skin of older subjects or patients. Comprehensive studies of percutaneous absorption as a function of age continue to be warranted.

Cutaneous biotransformation, i.e., catabolic enzymatic activity in the viable epidermis, is substantial. In fact, the viable epidermis is metabolically more active than the dermis. If the topically applied drug is subject to biotransformation during skin permeation, local and systemic bioavailability can be affected markedly. Enzymatic activity in the skin, or for that matter in systemic fluids and tissues, can be taken advantage of to facilitate percutaneous absorption.

The effect of regional variations in skin permeability can be marked. It has been suggested that one ought to differentiate between two species of horny layer: the palms and soles (up to 600 μm thick), adapted for weight-bearing and friction; and the body horny layer (~ 10 – 20 μm thick), adapted for flexibility, impermeability, and sensory discrimination.

Overall, data suggest the following order for diffusion of simple molecules through the skin: plantar < palmar < arms, legs, trunk, dorsum of hand < scrotal and postauricular < axillary < scalp. Electrolytes in solution penetrate the skin poorly. Ionization of a weak electrolyte substantially reduces its permeability; e.g., sodium salicylate permeates poorly compared with salicylic acid. The development of iontophoretic devices in recent years may minimize this problem with ionic penetrants. For any specific molecule, the predictability of regional variations in skin permeability continues to elude investigators. Pharmacokinetic models that do not adequately reflect the anisotropicity of the skin's composition and structure, the skin's interactions with the drug and the vehicle, and the physiological parameters that affect transfer will not succeed in simulating or predicting the rate and extent of percutaneous drug absorption or transdermal delivery.

Classically, percutaneous absorption has been studied *in vivo*, using radioactively labeled compounds, or by *in vitro* techniques, using excised human or animal skin. *In vivo* studies have often made use of the skin-stripping method, which permits the estimation of the concentration or amount of the penetrating species as a function of depth of the stratum corneum. Layers of the stratum corneum can be removed or stripped successively away by the repeated application and removal of adhesive tape strips. Skin penetration of a drug and the effect of additives or formulation may be studied as a function of time and evaluated through analysis of successive individual skin strips that provide a temporal profile of skin penetration via appropriate dermatopharmacokinetic models.

Clearly, the evaluation of new chemical entities (NCEs) of indeterminate toxicity mandates *in vitro* testing. A diffusion cell often used for *in vitro* experiments is shown in Figure 29-2. In this system, the intact skin or the epidermis is treated as a semipermeable membrane separating two fluid media. The transport rate of a particular drug is evaluated by introducing

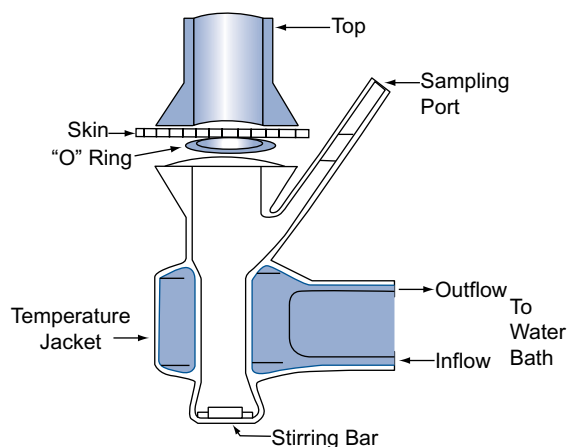


Figure 29-2. Schematic representation of a Franz diffusion cell. The top is open to ambient laboratory environment. (From Franz TJ. Percutaneous absorption. On the relevance of *in vitro* data. *J Invest Dermatol* 1975; 64: 190–195.)

the drug in solution on the stratum corneum side of the membrane, then measuring penetration by periodic sampling and analysis of the fluid in the receptor compartment.

Investigators have recognized that transport across an immersed, fully hydrated stratum corneum may not represent the absorption system or rate observed in *in vivo* studies. Percutaneous absorption across a fully hydrated stratum corneum may be an exaggeration. It may be more representative of enhanced absorption that is seen after *in vivo* skin is hydrated as a result of the application of an occlusive wrapping or material.

Numerous studies have demonstrated that *in vitro* studies of percutaneous absorption under controlled conditions are relevant to estimations of *in vivo* drug penetration. Although human skin is preferable for *in vitro* permeation studies, its availability is limited. Additional constraints apply if one is only willing to use freshly obtained viable human skin from surgical specimens or biopsies, as opposed to skin harvested from cadavers.

Concern has been voiced over the notorious variability in barrier properties of excised skin, whether animal or human. Factors responsible for the variability include the source and characteristics of the donor skin (e.g., elapsed time from death to harvesting of the skin, age and gender of the donor, and health of the skin prior to the donor's death), exposure of the skin to chemicals or mechanical treatment (e.g., shaving or clipping prior to harvesting of the skin), etc. The availability of a *living skin equivalent*—comprising a bilayered system of human dermal fibroblasts in a collagenous matrix, upon which human corneocytes have formed a stratified epidermis—offers an alternative, less variable, model for evaluating human skin permeation and biotransformation.

Skin-flap methods represent *in vivo* and *in vitro* techniques for evaluating percutaneous absorption in animals or animal models: the general approach entails the surgical isolation of a skin section of an animal such that the blood supply is singular; this ensures that drug can be collected and assayed in the vascular perfusate as it undergoes absorption from the skin surface. The perfused skin flap can be maintained in the intact animal or mounted in an *in vitro* perfusion system, all the while maintaining its viability.

Animals also have been used to detect contact sensitization, measure antimitotic drug activity, measure phototoxicity, and evaluate the comedogenic and comedolytic potential of substances. In each of these test procedures, be it a safety test or assay model, the animal is considered a substitute for humans. It is, therefore, important to realize that the animal is not man, even though man is the ultimate test animal. Animal testing

presents the investigator with unique advantages; lack of appreciation of the variables involved can destroy these advantages.

Any evaluation of a study of percutaneous absorption in animals must take species variation into consideration. Just as percutaneous absorption in man will vary considerably with skin site, so will absorption in various animal species. There is a consensus that rhesus monkeys and miniature pigs are good *in vivo* models for human percutaneous absorption, while smaller laboratory animals (e.g., mouse, rat, and rabbit) are not. Nonetheless, it should be stressed that percutaneous absorption studies in animals, either *in vivo* or *in vitro*, only can be approximations of activity in humans. The effect of species variation, site variability (about which little is known in animals), skin condition, experimental variables, and, of major importance, the vehicle, must be kept in mind.

In recent years, *in silico* or *in numero* modeling or computer simulation of percutaneous absorption has been advocated as a link between *in vitro* and *in vivo* studies. A number of relatively simplistic dermatopharmacokinetic models have been developed that do provide the formulator with some insight into transdermal drug delivery, in spite of the biological and physicochemical complexity of drug transport into and through the skin. By and large, these models are analogous to the classical pharmacokinetic models that have been employed to assess *in vivo* drug uptake and disposition. Some of the dermatopharmacokinetic models proposed differ from more classically oriented models in that drug transport in the vehicle and in the epidermis, particularly the stratum corneum, is modeled in accordance with Fickian diffusion. Thus, the formulator can anticipate the effect of variables such as the thickness of the applied (vehicle) phase, alterations in drug partitioning between the vehicle and the stratum corneum, and the frequency of re-application on the overall appearance of drug systemically as a function of time following topical application.

The study of the targeted delivery of drugs to follicles and/or sebaceous glands has become necessary in view of the selective uptake or deposition of antiacne drugs, such as tretinoin in pilosebaceous units. Fortunately, the anatomical and physiological correspondence of hamster ear pilosebaceous units to those in humans has facilitated studies of the cutaneous and pilosebaceous disposition of drugs following topical application.

UNGUAL AND PERUNGUAL DRUG DELIVERY

The properties of the nail unit (e.g., thickness and hydration) and the applied drug (e.g., molecular size, shape, electrical charge, and lipophilic-hydrophilic character) are critical factors affecting unguinal and perungual drug delivery. Nonetheless, the hardness and rigidity of the nail plate—coupled with its relatively slow growth rate—generally do not bode well for drug absorption into and through the nail. Given the prevalence of disorders of the nail unit (e.g., fungal infections and psoriasis) in the general population, their refractoriness to treatment, and issues of safety and toxicity of long-term dosing of systemically administered drugs for these disorders, mechanistic studies of topical drug delivery and delivery system formulation need to be continued.

THE FORMULATION OF DOSAGE FORMS AND DELIVERY SYSTEMS FOR TOPICAL DRUG APPLICATION⁷

The array of formulations and compositions employed for topical application confounds attempts at categorization. Considerable overlap is evident in the terminology used for topical products. Although the rheological properties of topical formulations have been suggested as a basis for discrimination among types of formulations, definitive differentiation and boundaries among formulation types are difficult to establish. Nonetheless, a distinction can be made between drug *dosage forms* and drug *delivery systems*. Dosage forms contain the active drug

ingredient in association with nondrug (usually inert) excipients that comprise the vehicle or formulation matrix. A conventional dosage form tends to be empirical in composition; its formulator's focus tends to emphasize stability and esthetics rather than efficacy. On the other hand, delivery systems involve a holistic approach to formulation that is optimized for the drug's relevant biopharmaceutic and pharmacokinetic characteristics in the patient population. Thus, a delivery system is formulated with functionality and efficacy in mind, not just stability and aesthetics.

CONVENTIONAL DOSAGE FORMS

The Skin

In many (if not most) clinical situations, the rate-limiting step is penetration of the drug across the skin barrier (i.e., percutaneous penetration through the skin alone). Diffusion of the drug from its vehicle should not be, unknowingly, the rate-limiting step in percutaneous absorption. Such a rate limitation or control may, of course, be an objective and the endpoint of specific drug optimization, but inappropriate formulation can reduce substantially the effectiveness of a topical drug substance.

In the formulation of a vehicle for topical drug application, many factors (drug stability, intended product use, site of application, and product type) must be considered in formulating a dosage form or delivery system that will release the drug readily when placed in contact with the skin. Further, the release characteristics of the vehicle depend on the physical-chemical properties of the specific drug substance to be delivered to the skin: drug release from a vehicle is a function of the drug's concentration and solubility in the vehicle, and the drug's partition coefficient between the vehicle and the skin. A vehicle optimized for delivery of hydrocortisone may be quite inappropriate for delivery of a different steroid.

Ostrenge *et al.*¹ discussed the significance of vehicle composition on the percutaneous absorption of fluocinolone acetonide and fluocinolone acetonide 21-acetate (fluocinonide) (see Figure 29-3). These investigators used propylene glycol/isopropyl myristate partition coefficients, *in vitro* (human) skin penetration, and finally *in vivo* vasoconstrictor studies to evaluate formulation variables. They concluded that

In general, an efficacious topical gel preparation is one in which (a) the concentration of diffusible drug in the vehicle for a given labeled strength is optimized by ensuring that all of the drug is in solution, (b) the minimum amount of solvent is used to dissolve the drug completely

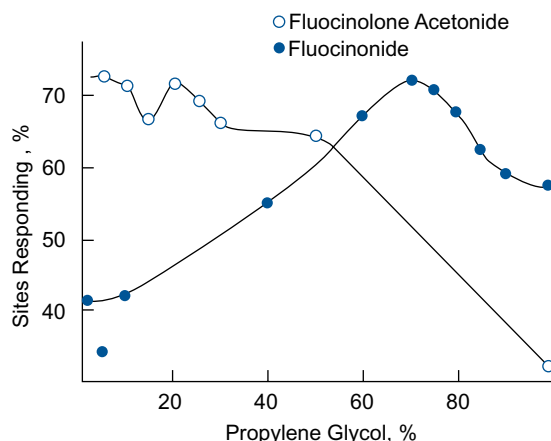


Figure 29-3. *In vivo* response as a function of vehicle composition (24-hour vasoconstriction). (From Ostrenge J *et al.* Significance of vehicle composition. I. Relationship between topical vehicle composition, skin permeability, and clinical efficacy. *J Pharm Sci* 1971; 60: 1175–1179.)

and yet maintain a favorable partition coefficient and (c) the vehicle components affect the permeability of the stratum corneum in a favorable manner.

The effect of propylene glycol concentration on *in vivo* vasoconstrictor activity is illustrated strikingly in Figure 29-3.

Experimental work of the kind described by Ostrenge, Steinmetz, and Poulsen provides a means of optimizing drug release from a vehicle and penetration of the drug into the skin. The formulator must then proceed to develop a total composition in which the drug is stable and causes no irritation to sensitive skin areas. Safety, stability, and effective preservative efficacy must be combined with optimum drug delivery in the total formulation.

The work of Chiang, Weiner, and others² on the physicochemical stability of topical drug-delivery systems *postapplication* has facilitated the exploration of additional formulation factors that are crucial to the success of topical formulations. Flynn notes that the functionality of topical drug delivery systems stands in stark contrast to that of transdermal drug-delivery systems; while both delivery systems are open systems kinetically due to the formulation-skin interface, they differ to a considerable extent *thermodynamically* because most topical formulations are left open to the air postapplication, while transdermal delivery systems are self-contained closed systems.

One study focused on a topical delivery system for minoxidil. The vehicle was a 60:20:20 ethanol:propylene glycol:water system, with just enough propylene glycol to maintain 2 percent minoxidil in solution, following evaporation of the more volatile ethanol and water. Minoxidil fluxes across human cadaver skin, measured as a function of minoxidil concentration, increased as the initial concentration of drug increased, but only to about 3 percent (Figure 29-4). At initial minoxidil concentrations greater than 3 percent, transport was disproportionately low, relative to initial concentration, due to early precipitation of the drug.

Evaporation and loss of volatile formulation components, such as water or ethanol postapplication, can be expected to affect topical drug-delivery system composition and performance. Chiang *et al.*² have shown that so-called nonvolatile excipients, e.g., propylene glycol, evaporate after topical application. Skin permeation by excipients also may occur after application leading to further compositional changes in the applied film on the skin surface. The impact of this evaporative and absorptive loss of adjuvants increases as the volume of the applied formulation is reduced. As Chiang *et al.*² note, “the momentary compositions, and thus delivery capabilities, of real vehicles are significantly influenced by the amounts applied.”

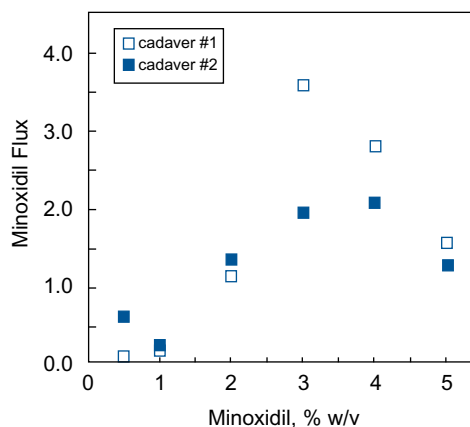


Figure 29-4. Minoxidil flux ($\times 10^4$ mg/cm²/h) through human cadaver skin as a function of minoxidil concentration in the topically applied formulation. (Data from Chiang C-M *et al.* Bioavailability assessment of topical delivery systems: effect of vehicle evaporation upon *in vitro* delivery of minoxidil from solution formulations. *Int J Pharm* 1989; 55: 229–236.)

CONVENTIONAL TOPICAL DOSAGE FORMS—SEMISOLIDS

THE USP CLASSIFICATION OF OINTMENT BASES

Ointments

The USP defines ointments as semisolid preparations intended for external application to the skin or mucous membranes. They usually, but not always, contain medicinal substances. The types of ointment bases used as vehicles for drugs are selected or designed to facilitate drug transfer into the skin. Ointments also contribute emolliency or other quasi-medicinal benefits. The USP recognizes four general classes of ointment bases: hydrocarbon bases (i.e., oleaginous bases), absorption bases, water-removable bases, and water-soluble bases. The selection of the optimum vehicle based on the USP classification *per se* may require compromises. For example, stability or drug activity might be superior in a hydrocarbon base; however, patient acceptability is diminished because of the greasy nature of the base. The water solubility of the polyethylene glycol bases may be attractive, but the glycol(s) may be irritating to traumatized tissue. For some drugs, activity and percutaneous absorption may be superior when using a hydrocarbon base; however, it may be prudent to minimize percutaneous absorption by the use of a less occlusive base. In other instances, activity and percutaneous absorption may be enhanced by using a hydrophilic base. These problematic aspects of bioavailability of drugs in topical formulations are discussed above.

HYDROCARBON (OLEAGINOUS) BASES

Hydrocarbon bases are usually petrolatum *per se* or petrolatum modified by waxes or liquid petrolatum to change viscosity characteristics. Liquid petrolatum gelled by the addition of polyethylene also is considered a hydrocarbon ointment base, albeit one with unusual viscosity characteristics.

Hydrocarbon ointment bases are classified as oleaginous bases along with bases prepared from vegetable fixed oils or animal fats. Bases of this type include lard, benzoinated lard, olive oil, cottonseed oil, and other oils. Such bases are emollient but generally require addition of antioxidants and other preservatives. They are now largely of historic interest.

Petrolatum USP is a tasteless, odorless, unctuous material with a melting range of 38°C to 60°C; its color ranges from amber to white (when decolorized). Petrolatum often is used externally, without modification or added medication, for its emollient qualities.

Petrolatum used as an ointment base has a high degree of compatibility with a variety of medicaments. Bases of this type are occlusive and nearly anhydrous and thus provide optimum stability for medicaments such as antibiotics. The wide melting range permits some latitude in vehicle selection, and the USP permits addition of waxy materials as an aid in minimizing temperature effects.

Hydrocarbon bases, being occlusive, increase skin hydration by reducing the rate of loss of surface water. Bases of this kind may be used solely for such a skin-moisturizing or emollient effect. Skin hydration, on the other hand, may increase drug activity, as discussed earlier.

A gelled mineral oil vehicle (Plastibase) represents a unique member of this class of bases that comprises refined natural products: When approximately 5 percent of low-density polyethylene is added to liquid petrolatum and the mixture is then heated and subsequently shock-cooled, a soft, unctuous, colorless material resembling white petrolatum is produced. The mass maintains unchanged consistency over a wide temperature range. It neither hardens at low temperatures nor melts at reasonably high temperatures. Its useful working range is between -15°C and 60°C. Excessive heat (above 90°C), will destroy the gel structure.

On the basis of *in vitro* studies, drugs may be released faster from a gelled mineral oil vehicle than from conventional petrolatum. This quicker release has been attributed to easier diffusion of drug through a vehicle with lower microscopic viscosity (i.e., a vehicle that is essentially a liquid) than through petrolatum.

Despite the advantages that hydrocarbon or oleaginous vehicles provide in terms of stability and emolliency, such bases have the considerable disadvantage of greasiness. The greasy or oily material may stain clothing and is difficult to remove. In terms of patient acceptance, hydrocarbon bases (i.e., ointments) rank well below emulsion bases such as creams and lotions.

ABSORPTION BASES

Absorption bases are hydrophilic, anhydrous materials or hydrous bases that have the ability to absorb additional water. The former are anhydrous bases, which absorb water to become water-in-oil (W/O) emulsions; the latter are W/O emulsions, which have the ability to absorb additional water. The word absorption in this context refers only to the ability of the base to absorb water.

Hydrophilic petrolatum USP is an anhydrous absorption base. Its W/O emulsifying property is conferred by the inclusion of cholesterol. This composition is a modification of the original formulation, which contained anhydrous lanolin. The lanolin was deleted because of reports of allergy; cholesterol was added. Inclusion of stearyl alcohol and wax adds to the physical characteristics, particularly firmness and heat stability.

Absorption bases, particularly the emulsion bases, impart excellent emolliency and a degree of occlusiveness on application. The anhydrous types can be used when the presence of water would cause stability problems with specific drug substances (e.g., antibiotics). Absorption bases also are greasy when applied and are difficult to remove. Both of these properties are, however, less pronounced than with hydrocarbon bases.

Commercially available absorption bases include Aquaphor (Beiersdorf) and Polysorb (Fougera). Nivea Cream (Beiersdorf) is a hydrated emollient base. Absorption bases, either hydrous or anhydrous, are seldom used as vehicles for commercial drug products. The W/O emulsion system is more difficult to deal with than the more conventional oil-in-water (O/W) systems, and there is, of course, reduced patient acceptance because of greasiness.

WATER-REMOVABLE (WATER-WASHABLE) BASES

Water-removable bases are O/W emulsion bases, commonly referred to as *creams*, and represent the most commonly used type of ointment base. By far the majority of commercial dermatologic drug products are formulated in an emulsion (or cream) base. Emulsion bases are washable and are removed easily from skin or clothing. Emulsion bases can be diluted with water, although such additions are uncommon.

As a result of advances in cosmetic chemistry the formulator of an emulsion base is faced with a bewildering array of potential ingredients. A glance at the cosmetic literature and such volumes as the *International Cosmetic Ingredient Dictionary and Handbook* impresses one with the enormous number and variety of emulsion-base components, particularly surfactants and oil-phase components. Many of these substances impart subtle but distinct characteristics to cosmetic emulsion systems. While desirable, many of these characteristics are not really necessary in drug dosage forms or delivery systems. Furthermore, the likelihood of drug-excipient interactions, either physical or chemical, increases substantially (as does the cost) as the number of formulation components is increased. Thus, the formulator of topical products should minimize the number of excipients in the formulation. Nonetheless, emulsion bases typically include antimicrobial preservatives, stabilizers (such as antioxidants, metal chelating agents, or buffers), and humectants (e.g., glycerin or propylene glycol), in addition to the emulsifiers, in order to ensure stability and efficacy.

Soaps and detergents (i.e., emulsifiers) have, overall, a damaging effect on the skin. Both anionic and cationic surfactants can cause damage to the stratum corneum in direct proportion to concentration and duration of contact. Nonionic surfactants appear to have much less effect on the stratum corneum.

WATER-SOLUBLE BASES

Soluble ointment bases, as the name implies, are made up of soluble components or may include gelled aqueous solutions. The latter often are referred to as gels, and in recent years have been formulated specifically to maximize drug availability.

Major components, and in some instances the only components, of water-soluble bases are the polyethylene glycols (PEGs). Patch tests have shown that these compounds are innocuous, and long-term use has confirmed their lack of irritation. PEGs are relatively inert, nonvolatile, water-soluble or water-miscible liquids or waxy solids identified by numbers that are an approximate indication of molecular weight. Polyethylene glycol 400 is a liquid superficially similar to propylene glycol, while polyethylene glycol 6000 is a waxy solid.

Polyethylene glycols of interest as vehicles include the 1500, 1600, 4000, and 6000 products, ranging from soft, waxy solids (polyethylene glycol 1500 is similar to petrolatum) to hard waxes. Polyethylene glycols, particularly 1500, can be used as a vehicle *per se*; however, better results often are obtained by using blends of high- and low-molecular-weight glycols, as in polyethylene glycol ointment NF. The water-solubility of polyethylene glycol vehicles does not ensure availability of drugs contained in the vehicle. Because hydrated stratum corneum is an important factor in drug penetration, the use of polyethylene glycol vehicles, which are anhydrous and nonocclusive, actually may hinder percutaneous absorption due to dehydration of the stratum corneum.

Aqueous gel vehicles containing water, propylene, and/or polyethylene glycol and gelled with a carbomer or a cellulose derivative also are classed as water-soluble bases. Bases of this kind, sometimes referred to as *gels*, may be formulated to optimize delivery of a drug, particularly steroids. In such a preparation propylene glycol is often used for its solvent properties as well as for its antimicrobial or preservative effects.

Gelling agents used in these preparations may be nonionic or anionic. Nonionics include cellulose derivatives, such as methylcellulose or hypromellose (hydroxypropyl methylcellulose). Sodium carboxymethylcellulose is an anionic cellulose derivative.

Carbomers are the NF designation for various polymeric acrylic acids, crosslinked with carbohydrates or polyalcohol derivatives that are dispersible but insoluble in water. When the acid dispersion is neutralized with a base, a clear, stable gel is formed. Carbomers for which monographs appear in the NF include carbomers 910, 934, 934P, 940, 941, and 1342, as well as the more complex carbomer copolymer, carbomer homopolymer, and carbomer interpolymers.

Other gelling agents employed in topical formulations include sodium alginate and the propylene glycol ester of alginic acid (Kelcoloid). Sodium alginate is a hydrophilic colloid that functions satisfactorily between pH 4.5 and 10; addition of calcium ions will gel fluid solutions of sodium alginate.

Gels can also be formed or stabilized by the incorporation of finely divided solids such as colloidal magnesium aluminum silicate (Veegum) or colloidal (fumed) silicon dioxide (Aerosil, Cab-O-Sil). These inorganic particulates can function as emulsifiers and suspending agents as well as gellants. Their compatibility with alcohols, acetone, and glycols makes them particularly useful in topical gel formulations.

PREPARATION

Ointment preparation or manufacture depends on the type of vehicle and the quantity to be prepared. The objective is the same (i.e., to disperse the drug uniformly throughout the

vehicle). Normally, the drug materials are either in finely powdered form or in solution before being dispersed in the vehicle.

INCORPORATION OF DRUG BY LEVIGATION

The incorporation of a drug powder in small quantities of an ointment (i.e., 30–90 g) can be accomplished by using a spatula and an ointment tile (either porcelain or glass). The drug material is levigated thoroughly with a small quantity of the vehicle or a miscible liquid component of the formulation (e.g., propylene glycol; light mineral oil) to form a concentrate. The concentrate then is diluted geometrically with the remainder of the base.

If the drug substance is water soluble, it can be dissolved in water and the resulting solution incorporated into the vehicle by use of a small quantity of lanolin, if the base is oleaginous. Generally speaking, an amount of anhydrous lanolin equal in volume to the amount of water used will suffice.

On a larger scale, mechanical mixers (e.g., Hobart mixers and pony mixers) are used. The drug substance in finely divided form usually is added slowly or sifted into the vehicle contained in the rotating mixer. When the ointment is uniform, the finished product may be processed through a roller mill to ensure complete dispersion and reduce any aggregates.

An alternative procedure involves preparing and milling a concentrate of the drug in a portion of the base. The concentrate then is dispersed in the balance of the vehicle, using a mixer of appropriate size. Occasionally, the base may be melted for easier handling and dispersing. In such cases the drug is dispersed and the base slowly cooled, using continuous agitation to maintain dispersion.

EMULSION FORMULATIONS

Emulsions are prepared generally by combining the “oil”-soluble ingredients (e.g., petrolatum, waxes, and fats) and heating the admixture to about 75°C (i.e., a temperature at which the oil-phase ingredients are molten). The “water”-soluble ingredients are combined separately and heated to slightly above 75°C. The aqueous phase then is added to the oil phase, slowly and with constant agitation. When the emulsion is formed, the mixture is allowed to cool, maintaining slow agitation.

At this stage in the process the medicinal ingredients usually are added as a concentrated slurry, which usually has been milled to reduce any particle aggregates. Volatile or aromatic materials generally are added when the finished emulsion has cooled to about 35°C. At this point, additional water may be added to compensate for any evaporative losses occurring during exposure and transfer at the higher temperatures of emulsion formation.

While the product remains in the tank in bulk, quality-control procedures are performed (e.g., for pH, active ingredients). If control results are satisfactory, the product is filled into the appropriate containers.

OTHER DOSAGE FORMS FOR APPLICATION TO THE SKIN AND ITS APPENDAGES

POULTICES (CATAPLASMS)

Poultices, or cataplasms, represent one of the most ancient classes of pharmaceutical preparations. A poultice is a soft, moist mass of meal, herbs, seed, etc., usually applied hot in a cloth. The consistency is gruel-like, which is probably the origin of the word poultice.

Cataplasms were intended to localize infectious material in the body or to act as counterirritants. The materials tended to be absorptive, which, together with heat, accounts for their popularity. None is now official in the USP. The last official product was Kaolin Poultice NF IX.

PASTES

The USP defines pastes as semisolid dosage forms that contain one or more drug substances intended for topical application.

Pastes are divided into fatty pastes (e.g., zinc oxide paste) and those made from a single-phase aqueous gel (e.g., carboxymethylcellulose sodium paste). Another official paste is triamcinolone acetonide dental paste.

The term *paste* is applied to ointments in which large amounts of solids have been incorporated (e.g., zinc oxide paste). In the past, pastes have been defined as concentrates of absorptive powders dispersed (usually) in petrolatum or hydrophilic petrolatum. These fatty pastes are stiff to the point of dryness and are reasonably absorptive considering they have a petrolatum base. Pastes often are used in the treatment of oozing lesions, where they act to absorb serous secretions. Pastes also are used to limit the area of treatment by acting both as an absorbent and a physical dam.

Pastes adhere reasonably well to the skin and are poorly occlusive. For this reason they are suited for application on and around moist lesions. The heavy consistency of pastes imparts a degree of protection and may, in some instances, make the use of bandages unnecessary. Pastes are less macerating than ointments.

Because of their physical properties pastes may be removed from the skin by the use of mineral oil or a vegetable oil. This is particularly necessary when the underlying or surrounding skin is traumatized easily.

POWDERS

Powders for external use usually are described as dusting powders. Such powders should have a particle size of not more than 150 μm (i.e., less than 100-mesh) to avoid any sensation of grittiness, which could irritate traumatized skin. Dusting powders usually contain starch, talc, and zinc stearate. Absorbable dusting powder USP is composed of starch treated with epichlorohydrin, with not more than 2.0% magnesium oxide added to maintain the modified starch in impalpable powder form; since it is intended for use as a lubricant for surgical gloves, it should be sterilized (by autoclaving) and packaged in sealed paper packets.

The fineness of powders often is expressed in terms of mesh size, with impalpable powders generally in the range of 100- to 200-mesh (149 to 75 μm). Determination of size by mesh analysis becomes increasingly difficult as particle size decreases below 200-mesh.

DRESSINGS

Dressings are external applications resembling ointments, usually used as a covering or protection. Petrolatum gauze is a sterile dressing prepared by adding sterile, molten, white petrolatum to precut sterile gauze in a ratio of 60 g of petrolatum to 20 g of gauze. Topical antibacterials are available in the form of dressings.

CREAMS

Creams are viscous liquid or semisolid emulsions of either the O/W or W/O type. Pharmaceutical creams are classified as water-removable bases in the USP and are described under *Ointments*. In addition to ointment bases, creams include a variety of cosmetic-type preparations. Creams of the O/W type include shaving creams, hand creams, and foundation creams; W/O creams include cold creams and emollient creams.

PLASTERS

Plasters are substances intended for external application, made of such materials and of such consistency as to adhere to the skin and attach to a dressing. Plasters are intended to afford protection and support and/or to furnish an occlusive and macerating action and to bring medication into close contact with the skin. Medicated plasters, long used for local or regional drug delivery, are the prototypical transdermal delivery system.

Plasters usually adhere to the skin by means of an adhesive material. The adhesive must bond to the plastic backing and to

the skin (or dressing) with proper balance of cohesive strengths. Such a proper balance provides for removal (adhesive breakage at the surface of application), thus leaving a clean (skin) surface when the plaster is removed.

THE EVOLUTION OF TRANSDERMAL DRUG DELIVERY SYSTEMS FROM CONVENTIONAL TOPICAL FORMULATIONS

Conventional medicated topicals (e.g., creams and ointments) seldom permit substantial systemic uptake of the drug or drugs incorporated therein. This is a consequence, in part, of the limited persistence or residence time of the topical formulation on the skin surface. In effect, a drug does not remain in contact with the absorbing surface long enough for sufficient drug to transfer into the skin and, ultimately, into systemic circulation. Furthermore, there is the concomitant problem of the gradual depletion of drug from the region of the topical formulation immediately adjacent to the skin surface and the corresponding reduction in the concentration gradient for drug transfer from the topical formulation to the skin.

The emergence of adhesive transdermal drug delivery systems (TDDSs) in the early 1980s permitted skin residence times to increase from hours to days. The novel matrix- or reservoir-formulations employed in these TDDSs also provided for the maintenance of relatively uniform concentrations of diffusible drug in the formulation, thereby preventing the formation of drug-depleted regions within the topical formulation and helping to ensure relatively constant drug-release rates over an extended period of time. As noted above, skin occlusion by the water-impermeable backing film of TDDSs further facilitates TDDS systemic efficacy by increasing skin hydration and temperature, with a corresponding increase in the rate and extent of skin permeation. The inclusion of skin-penetration enhancers in medicated topicals serves to decrease diffusional resistance and increase transport.

Nonetheless, these—by now—*conventional* TDDSs have their limitations: the increased residence time of occlusive TDDSs on the skin surface leads to an increased incidence of skin maceration and adverse cutaneous reactions. In addition, effective skin permeation is limited to relatively small (<1 kDa) lipophilic drug molecules. Thus, increasingly more attention is being placed on alternative TDDSs—e.g., electrically modulated systems and mechanical systems—which circumvent the need for partitioning and diffusion of the drug out of the formulation matrix and into and through the skin. *Electrically modulated* systems, or *electrotransport* systems, facilitate drug transport by an external electrical field. Electrotransport mechanisms include *iontophoresis*, *electroosmosis*, and *electroporation*. *Mechanically (physically) modulated* systems are exemplified by systems employing *phonophoresis* or those using *microneedle arrays* to achieve transdermal drug delivery.

ELECTRICALLY MODULATED TRANSDERMAL DELIVERY SYSTEMS

For some poorly absorbed (ionic) compounds, parenteral administration appears to be the only viable option for regional or systemic delivery, because chemical penetration enhancers (Table 29-1) often do not function well for these compounds. Given the increased risk of adverse reactions associated with the use of such enhancers, the increased evaluation of iontophoretic devices for the enhancement of topical drug delivery has been of great interest. Iontophoretic drug delivery implies the delivery of ionic drugs into the body by means of an electric current. While the stratum corneum forms the principal barrier to electrical conductivity—due, in part, to its lower water content—the skin also acts as a capacitor. Thus, biological tissues such as the skin provide for a reactive electrical circuit. Ionic transport through the skin in the presence of a uniform electric field can be described, in part, in accordance with the Nernst-Planck equation

$$J_i = -D \frac{dC}{dx} + \frac{DzeEC}{kT}$$

where J_i is the flux of ions across the membrane, C is the concentration of ions with valence z and electron charge e , dC/dx is the concentration gradient, E is the electric field, k is Boltzmann's constant, and T is the absolute temperature. Thus, the ionic flux is the sum of the fluxes that arise from the concentration gradient and the electric field. Given the complexity of the skin's composition, the thickness of the stratum corneum, and the occurrence of electroosmotic effects, the Nernst-Planck equation is only a first approximation of the overall transdermal flux of a solute. Faraday's law

$$\frac{Q}{t} = \frac{t_j i}{|z| F}$$

further characterizes the iontophoretic flux Q/t in terms of the current i (in amperes) and its duration t (in seconds), the transference number parameter t_j , and the Faraday constant F . Additional factors that influence the rate and extent of iontophoretic delivery through the skin include pH and ionic strength of the drug solution.

Although iontophoretic techniques have been shown to increase percutaneous absorption of ionizable or ionic drugs (including lidocaine, salicylates, and peptides and proteins, such as insulin) markedly, the clinical safety and efficacy of drug-delivery systems employing iontophoretic technology have yet to be evaluated fully.

Problematic aspects of electrotransport include cutaneous irritation or erythema and the effect of the electrical field on the integrity and stability of the formulation. Electrically induced alterations in the formulation generally arise as a result of iontophoresis (due to the increased flux of ions) or electroosmosis (due to the electrically induced convective transport of water molecules and associated electrically neutral solutes). The use of pulsed or intermittent current electrotransport systems has been suggested as an alternative to continuous current systems. *Electroporation*—the use of pulsed electrical current to provoke the transient formation of pores in biomembranes—also has been suggested as an alternative, or complement, to iontophoresis. In any event, the potential of electrically modulated drug-delivery systems for the effective transdermal delivery of large, polar, or ionic molecules (e.g., proteins, peptides, DNA) necessitates continued research in this field. One encouraging advance in this area is the development of flexible wafer-thin arrays of conductive layers or filaments for drug delivery systems that are less bulky and potentially more acceptable to patients.

MECHANICALLY MODULATED TRANSDERMAL DELIVERY SYSTEMS

Phonophoresis, or *sonophoresis*, is defined as the movement of drug molecules through the skin under the influence of ultrasound. In general, ultrasound frequencies 1 to 3 MHz and intensities of 0.01 to 2 W/cm² have been used with varying degrees of effectiveness,³⁶ although high-frequency, low-intensity ultrasound has been observed to increase transdermal drug flux and decrease percutaneous diffusional lag times. A more recent analysis of ultrasound-enhanced transdermal transport indicates that low-frequency sonophoresis is much more important than high-frequency sonophoresis in enhancing transport. Various thermal and nonthermal changes have been implicated to explain phonophoretically induced increases in drug transport through the skin. Although the effect of temperature increases on molecular diffusivity and flux is clear, nonthermal effects of ultrasound (e.g., cavitation) are less clear. Transient ultrasound-induced cavitation (i.e., the generation and oscillation of gas bubbles) in the stratum corneum apparently perturbs barrier permeability and solute transport in the aqueous regions of the stratum corneum. Evidence for this is the lack of correlation between phonophoretic permeability and permeant lipophilicity.

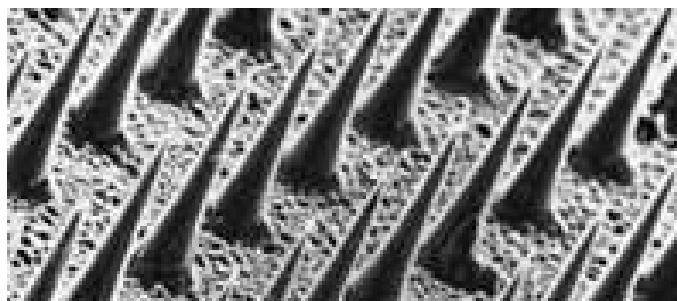


Figure 29-5. Electron micrograph of a microneedle array for transdermal drug delivery. (Courtesy of Georgia Institute of Technology; ©Georgia Tech Res Corp.)

Both hollow and solid drug-coated microneedle arrays (Figure 29-5) have been proposed as painless adjuncts to transdermal delivery systems. Microneedles (e.g., 150- μ m long) can penetrate the stratum corneum, thereby facilitating drug access to the living epidermis and dermis and ultimately to systemic circulation. In one embodiment of this approach, silicon microneedles—prepared by reactive ion etching microfabrication techniques originally developed for integrated circuits—leave holes about 1 μ m in diameter when removed from the skin. Other materials (e.g., stainless steel) have been proposed for transdermal needle array systems but are still undergoing evaluation for safety and efficacy. Possible problems with microneedle arrays include extent and duration of access of the perforated cutaneous tissue to pathogens and potentially hazardous chemicals on the skin surface.

NAILS

Conventional topical drug formulations for nail disorders have been limited, for the most part, to semisolids such as creams and ointments. The use of agents such as urea and salicylic acid to hydrate and soften the nail plate, thereby facilitating unguinal drug delivery, has not been uniformly successful. The most promising approaches to date involve the development of drug-containing nail lacquers or varnishes (transungual delivery systems) akin to cosmetic nail polishes. These formulations generally consist of a solution of the drug(s) in a quick-drying solution of a film-forming polymer. Following application of the drug-containing lacquer to the nail plate, the solvent evaporates, leaving the drug(s) in an adherent polymer film on the surface of the nail. Subsequent drug diffusion from the polymer film into the nail plate and the nail bed would be expected to yield effective unguinal concentrations of the drug, as long as drug partitioning from the polymer film into the nail unit was substantial enough. Thus, formulation factors to be considered include drug content and solubility in the nail lacquer, relative affinity of the drug for the polymer film and the nail unit, the diffusivity of the drug in the polymer film and the nail unit components, and the degree of hydration of the nail plate (this can be influenced by the polymer film-induced reduction of water loss to the environment from the nail surface).^{8,9}

RECTAL, VAGINAL, AND URETHRAL ROUTES OF ADMINISTRATION

ANATOMIC AND PHYSIOLOGIC ASPECTS OF THE RECTAL, VAGINAL, AND URETHRAL ROUTES OF ADMINISTRATION

The Rectal Route

The rectum is about 150 mm in length, terminating in the anal opening; its surface area is about 200 to 400 cm². In the absence of fecal matter, the rectum contains a small amount of fluid

(1 to 3 mL) of low buffering capacity. Fluid pH is purported to be about 7.2; because of the low buffer capacity, rectal pH will vary with the pH of the drug product or drug dissolved in it.

The bioavailability of rectally administered drugs is a relatively recent concern despite the antiquity of this dosage form; little was known about drug absorption or drug activity via suppository administration until recent years. Rectally instilled preparations, whether suppositories, foams, or solutions (enemas), tend to be confined to the rectum and sigmoid colon if the volume is less than about 50 mL. Foams tend to dissipate or spread to a lesser extent than solutions, particularly large-volume solutions (~100 to 200 mL). Although large-volume fluid formulations—solutions or enemas—may allow drug to reach the ascending colon, substantial intra- and intersubject variation is evident. Surveys of the literature indicate that rectal drug absorption from suppositories can be erratic and may be substantially different from absorption following oral administration. With only a few recent exceptions, suppository studies are based on either *in vivo* or *in vitro* data, with few attempts to correlate *in vitro* results with *in vivo* studies.

Major factors affecting the absorption of drugs from suppositories administered rectally are the following: anorectal physiology, suppository vehicle, and the physicochemical properties of the drug.

Most rectal suppositories today are torpedo-shaped, with the apex, or pointed end, tapering to the base, or blunt end, following the recommendation of H. S. Wellcome in 1893 that rectal suppositories should be inserted with the thicker end foremost so that when the anal sphincter contracts, expulsion is prevented. In the intervening 100 years or so, no study has correlated rectal suppository insertion with anorectal physiology until that of Abd-El-Maeboud *et al.*,¹⁰ who found that ease of insertion, retention, and lack of expulsion were enhanced when the suppository was inserted base or blunt end up. This was ascribed to reversed vermicular contractions of the external anal sphincter, which facilitate movement of the suppository upward into the rectum.

The rectal epithelium is lipoidal in character. The lower, middle, and upper hemorrhoidal veins surround the rectum. Only the upper vein conveys blood into the portal system; thus, drugs absorbed into the lower and middle hemorrhoidal veins will bypass the liver. Absorption and distribution of a drug therefore are modified by its position in the rectum, in the sense that at least a portion of the drug absorbed from the rectum may pass directly into the inferior vena cava, bypassing the liver.

Spreading characteristics of rectal formulations may be affected considerably by intraluminal rectal pressure—due, in part, to the weight of abdominal organs and to respiratory activity—and by periodic contractile activity of the rectal wall.

Based on available data, the bioavailability of a drug from a suppository dosage form depends on the physicochemical properties of the drug as well as the composition of the base. The drug-dissolution rate and, where appropriate, the partition coefficient and aqueous phase should be known.

For suppository formulation the relative solubility of the drug in the vehicle is a convenient comparison measure. Lipid-soluble drugs present in low concentration in a cocoa butter base will have little tendency to partition and diffuse into rectal fluids. Drugs that are only slightly soluble in the lipid base will partition readily into the rectal fluid. The partition coefficient between suppository base and rectal fluid thus becomes a useful measure. In water-soluble bases, assuming rapid dissolution, the rate-limiting step in absorption would be transport of the drug through the rectal mucosa.

In the absence of evidence of any substantial carrier-mediated uptake mechanisms, the predominant mechanism of colorectal mucosal permeation appears to involve transcellular passage across cell membranes in accordance with the pH-partition hypothesis. Ease of access to the rectal mucosa has encouraged the evaluation of absorption enhancers. A wide variety of substances have been investigated for their ability to enhance rectal

permeability to drugs. Agents such as EDTA have been used to chelate Ca^{2+} and Mg^{2+} in the vicinity of paracellular tight junctions and, thus, alter epithelial permeability. Other promoters of rectal absorption (e.g., bile salts and nonsteroidal anti-inflammatory agents, including aspirin, salicylic acid, and diclofenac) appear to exert their influence by affecting water influx and efflux rates across the rectal mucosa. Surfactants not only may modify membrane permeability but also may enhance wetting or spreading of the base and dissolution of the drug. In any event, it should be evident that whatever the mechanism, enhancing the rectal absorption of drugs—especially those that undergo presystemic elimination—could result in substantially reduced dosage requirements and decreased risk of adverse reactions.

Clearly, the bioavailability of a drug administered rectally depends on the nature of the drug and the composition of the vehicle or base. The physical properties of the drug can be modified to a degree, as can the characteristics of the base selected as the delivery system. Preformulation evaluations of physicochemical properties then must be confirmed by *in vivo* studies in animals and ultimately in the primary primate, the human. Dogs are probably the animal of choice in evaluating rectal drug availability. (The pig is a closer physiological match, but size and manageability argue in favor of the dog.) Blood and urine samples can be obtained from the dog, and rectal retention can be accomplished with facility. Smaller animals have been used; rabbits, rats, and even mice have been employed, but dosing and sampling become progressively more difficult.

Human subjects provide the ultimate measure of drug bioavailability. Subjects are selected on the basis of age, weight, and medical history. Subjects usually are required to fast overnight and evacuate the bowel prior to initiation of the study. Fluid volume and food intake usually are standardized in studies of this kind.

Given the difficulty of standardizing pharmacological endpoints, the usual measure of rectal drug bioavailability is the concentration of the drug in blood and/or urine as a function of time. A control group using oral drug administration provides a convenient means of comparing oral and rectal drug availability. While there is general agreement about drug absorption from the rectum, there is less agreement on dosage adequacy and the relationship between oral and rectal dosage. This state of affairs argues in favor of adequate studies to establish proper dosage and verify bioavailability.

The Vaginal Route

Passive drug absorption via the vaginal mucosa, as with other mucosal tissues, is influenced by absorption site physiology, absorption site pH, and the solubility and partitioning characteristics of the drug. The vaginal epithelial surface usually is covered with an aqueous film—emanating from cervical secretions—whose volume, pH, and composition vary with age, stage of the menstrual cycle, and location. Postmenarche, a vaginal pH gradient is evident, with the lowest values (pH ~4) near the anterior fornix and the highest (pH ~5) near the cervix. Vaginal pH increases with menstrual, cervical, and uterine secretions.

Following intravaginal administration, some drug absorption from the intact vaginal mucosa is likely, even when the drug is employed for a local effect. In fact, extensive drug absorption can occur from the vagina. Furthermore, systemic drug concentrations following vaginal dosing may be significantly higher than those after peroral administration of an equivalent dose; a reflection, in part, of decreased first-pass biotransformation following vaginal absorption. Nonetheless, the notion persists that the vaginal epithelium is relatively impermeable to drugs.

The widespread extemporaneous compounding of progesterone vaginal suppositories as well as the marketing of an intrauterine progesterone drug delivery system (Progestasert, Alza), has focused interest on systemic drug absorption following intravaginal administration. However, only limited reports of research on *in vitro* and *in vivo* aspects of vaginal absorption have appeared in the literature to date.

The Urethral Route

Minimal information has been published to date on the biopharmaceutical aspects of the urethral route of administration. Nonetheless, drug absorption from the urethral lumen via the urethral mucosa could be expected to be governed by the same principles that apply to drug absorption across other mucosal surfaces. The functional length of the urethra (from the bladder neck) in females is approximately 3 cm; in the male, approximately 4 to 5 cm.

RECTAL, VAGINAL, AND URETHRAL DOSAGE FORMS AND DELIVERY SYSTEMS

Conventional Dosage Forms—Suppositories

Suppositories are solid dosage forms of various weights and shapes, usually medicated, for insertion into the rectum, vagina, or urethra. After insertion, suppositories soften, melt, disperse, or dissolve in the cavity fluids.

The use of suppositories dates from the distant past, this dosage form being referred to in writings of the early Egyptians, Greeks, and Romans. Suppositories are suited particularly for administration of drugs to the very young and the very old, a notion first recorded by Hippocrates.

Rectal Suppositories. The USP describes rectal suppositories for adults as tapered at one or both ends and usually weighing about 2 g each. Infant rectal suppositories usually weigh about one-half that of adult suppositories. Drugs having systemic effects, such as sedatives, tranquilizers, and analgesics, are administered by rectal suppository; however, the largest single-use category is probably that of hemorrhoid remedies dispensed over the counter. The 2-g weight for adult rectal suppositories is based on use of cocoa butter as the base; when other bases are used, the weights may be greater or less than 2 g. An interesting variation on molded rectal suppositories is the so-called rectal “rocket,” for the treatment of hemorrhoidal conditions.¹¹

Vaginal Suppositories. The USP describes vaginal suppositories, or pessaries, as usually globular or oviform and weighing about 5 g each. Vaginal medications are available in a variety of physical forms, e.g., creams, gels, or liquids, which depart from the classical concept of suppositories.^{12,13} Vaginal tablets, or inserts prepared by encapsulation in soft gelatin, however, do meet the definition and represent convenience both of administration and manufacture.

Urethral Suppositories. Urethral suppositories, or bougies, are not described specifically in the USP, either by weight or dimension. Traditional values, based on the use of cocoa butter as a base, are as follows for these cylindrical dosage forms: diameter: 5 mm; length: 50 mm female, 125 mm male; weight: 2 g female, 4 g male. An intraurethral insert containing the prostaglandin alprostadil is available for the treatment of erectile dysfunction. The commercial formulation, described as a sterile micropellet (1.4 mm in diameter and 6 mm long), consisting of the drug and polyethylene glycol 1450, is inserted 3 cm deep into the urethra by use of a hollow applicator.

Suppository Bases (Vehicles). The ideal suppository base should meet the following general specifications:

- The base is nontoxic and nonirritating to mucous membranes.
- The base is compatible with a variety of drugs.
- The base melts or dissolves in rectal fluids.
- The base should be stable on storage; it should not bind or otherwise interfere with release or absorption of drug substances.

Rectal suppository bases can be classified broadly into two types: fatty and water soluble or water miscible. The traditional cocoa butter vehicle is immiscible with aqueous tissue fluids but melts at body temperature. Water-soluble or water-miscible vehicles also have been used. In general, formulators have been

reluctant to use glycerinated gelatin as a rectal suppository base because of its relatively slow dissolution. More typical of this class is the polyethylene glycol vehicle. Drug absorption from such dissimilar bases can differ substantially.

The USP lists the following as usual suppository bases: cocoa butter, cocoa butter substitutes (primarily vegetable oils modified by esterification, hydrogenation and/or fractionation), glycerinated gelatin, hydrogenated vegetable oils, mixtures of polyethylene glycols of various molecular weights, and fatty acid esters of polyethylene glycol.

Cocoa Butter and Other Fatty Bases. Theobroma oil, or cocoa butter, is a naturally occurring triglyceride. About 40 percent of the fatty acid content is unsaturated. As a natural material there is considerable batch-to-batch variability. A major characteristic of theobroma oil is its polymorphism (i.e., its ability to exist in more than one crystal form). While cocoa butter melts quickly at body temperature, it is immiscible with body fluids; this may inhibit the diffusion of fat-soluble drugs to the affected sites. Oleaginous vehicles, such as cocoa butter, seldom are used in vaginal preparations for esthetic reasons: many women consider them messy and prone to leakage.

If, in the preparation of suppositories, the theobroma oil is overheated (heated to about 60°), molded, and chilled, the suppositories formed will melt below 30°C. The fusion treatment of theobroma oil requires maximum temperatures of 40°C to 50°C to avoid a change in crystal form and melting point. Theobroma oil, heated to about 60°C and cooled rapidly, will crystallize in an alpha configuration characterized by a melting point below 30°C. The alpha form is metastable and will slowly revert to the beta form, with the characteristic melting point approaching 35°C. The transition from alpha to beta is slow, taking several days. The use of low heat and slow cooling allows direct crystallization of the more stable beta crystal form.

Certain drugs will depress the melting point of theobroma oil. This involves no polymorphic change, although the net effect is similar. Chloral hydrate is an example of a drug with a rectal dose (0.5 to 1.0 g, as a hypnotic agent) that causes a substantial melting-point depression. This effect can be countered by addition of a higher-melting wax, such as white wax or synthetic spermaceti. The amount to be added must be determined by temperature measurements. The effect of such additives on bioavailability also must be considered.

Various cocoa butter substitutes (e.g., hard fat and hydrogenated vegetable oil) are available commercially that offer a number of advantages over cocoa butter, such as decreased potential for rancidity and phase transition (melting and solidification) behavior tailored to specific formulation, processing, and storage requirements. However, as with cocoa butter, these semisynthetic glyceride mixtures are also subject to polymorphic transformations. Batch-to-batch variations of the physical properties of all of these bases, whether cocoa butter or cocoa butter substitutes, can play havoc with the final products' characteristics. The formulator should ensure that the melting and congealing behavior of these bases and the formulations prepared from them is evaluated thoroughly.

Water-Soluble or Dispersible Suppository Bases. These are of comparatively recent origin. The majority are composed of polyethylene glycols or glycol-surfactant combinations. Water-miscible suppository bases have the substantial advantage of lack of dependence on a melting point approximating body temperature. Problems of handling, storage, and shipping are simplified considerably.

Polymers of ethylene glycol are available as polyethylene glycol polymers (Carbowax, polyglycols) of assorted molecular weights. Suppositories of varying melting points and solubility characteristics can be prepared by blending polyethylene glycols of 1000, 4000, or 6000 molecular weight.

Polyethylene glycol suppositories, while prepared rather easily by molding, cannot be prepared satisfactorily by hand-rolling. The drug-glycol mixture is prepared by melting and then is cooled to just above the melting point before pouring into dry

unlubricated molds. Cooling to near the melting point prevents fissuring caused by crystallization and contraction. The USP advises that labels on polyethylene glycol suppositories should instruct patients to moisten the suppository before inserting it.

Water-miscible or water-dispersible suppositories also can be prepared using selected nonionic surfactant materials. Polyoxyl 40 stearate is a white, water-soluble solid, melting slightly above body temperature. A polyoxyethylene derivative of sorbitan monostearate is water-insoluble but dispersible. In using surfactant materials the possibility of drug-base interactions must be borne in mind. Interactions caused by macromolecular adsorption may have a significant effect on bioavailability.

PEG-based water-miscible suppository bases are exemplified by a low-melting formulation employing 96 percent PEG 1000 and 4 percent PEG 4000 and a more heat-stable formulation with 75 percent PEG 1000 and 25 percent PEG 4000. Both may be prepared conveniently by molding techniques.

Water-dispersible bases may include polyoxyethylene sorbitan fatty acid esters. These are either soluble (Tween, Myrj) or water-dispersible (Arlacel), used alone or in combination with other wax or fatty materials. Surfactants in suppositories should be used only with recognition of reports that such materials may either increase or decrease drug absorption.

Hydrogels, defined as macromolecular networks that swell but do not dissolve in water, have been advocated as bases for rectal and vaginal drug delivery. The swelling of hydrogels (i.e., the absorption of water), is a consequence of the presence of hydrophilic functional groups attached to the polymeric network. Cross-links between adjacent macromolecules result in the aqueous insolubility of these hydrogels.

The use of a hydrogel matrix for drug delivery involves the dispersal of the drug in the matrix, followed by drying of the system and concomitant immobilization of the drug. When the hydrogel delivery system is placed in an aqueous environment (e.g., the rectum or the vagina), the hydrogel swells, enabling the drug to diffuse out of the macromolecular network. The rate and extent of drug release from these hydrogel matrices depend on the rate of water migration *into* the matrix and the rate of drug diffusion *out of* the swollen matrix.

Hydrogels employed for rectal or vaginal drug administration have been prepared from polymers such as polyvinyl alcohol, hydroxyethyl methacrylate, polyacrylic acid, or polyoxyethylene. Although hydrogel-based drug-delivery systems have yet to appear in suppository or insert form commercially, research efforts in this direction are increasing, given their potential for controlled drug delivery, bioadhesion, retention at the site of administration, and biocompatibility.

Glycerinated Gelatin. Glycerinated gelatin is composed of water and the drug (10 percent w/w), glycerin (70 percent), and gelatin (20 percent). It is usually used as a vehicle for vaginal suppositories because rectal use may be associated with an osmotic laxative effect. Nonetheless, if rectal use is desired, a firmer suppository can be obtained by increasing the gelatin content to about 30 percent. PEG 400 or propylene glycol may be substituted for the glycerin if drug solubility is a problem. The type of gelatin employed is dependent on the drug's ionic compatibility: Type A gelatin is cationic, i.e., acidic in reaction, while Type B gelatin is anionic, i.e., alkaline in reaction. Glycerinated gelatin suppositories are prepared by dissolving the drug in the water, adding the glycerin or glycol, and then, with the aid of heat, dispersing the gelatin in the resultant solution. These suppositories must be formed by molding. If not for immediate use, they should contain a preservative such as methylparaben or propylparaben.

Suppository Preparation. Suppositories are prepared by rolling (hand-shaping), molding (fusion), and cold compression.

Rolled (Hand-Shaped) Suppositories. Hand-shaping suppositories is the oldest and the simplest method of preparing this dosage form. The manipulation requires considerable skill, yet avoids the complications of heat and mold preparation.

The general process can be described as follows:

Take the prescribed quantity of the medicinal substances and a sufficient quantity of grated theobroma oil. In a mortar, reduce the medicating ingredients to a fine powder or, if composed of extracts, soften with diluted alcohol and rub until a smooth paste is formed. The correct amount of grated theobroma oil then is added, and a mass resembling a pill is made by thoroughly incorporating the ingredients with a pestle, sometimes with the aid of a small amount of wool fat. When the mass has become plastic under the vigorous kneading of the pestle, it quickly is loosened from the mortar with a spatula, pressed into a roughly shaped mass in the center of the mortar, and then transferred with the spatula to a piece of filter paper that is kept between the mass and the hands during the kneading and rolling procedure. By quick, rotary movements of the hands, the mass is rolled to a ball, which immediately is placed on a pill tile. A suppository cylinder is formed by rolling the mass on the tile with a flat board, partially aided by the palm of the other hand, if weather conditions permit. The suppository *pipe* often will show a tendency to crack in the center, developing a hollow core. This occurs when the mass has not been kneaded and softened sufficiently, with the result that the pressure of the roller board is not carried uniformly throughout the mass but is exerted primarily on the surface. The length of the cylinder usually corresponds to about four spaces on the pill tile for each suppository, thus making the piece, when cut, practically a finished suppository except for the shaping of the point. When the cylinder has been cut into the proper number of pieces with a spatula, the conical shape is given it by rolling one end on the tile with a spatula, or in some cases even by shaping it with the fingers to produce a rounded point.

Compression-Molded (Fused) Suppositories. This method of suppository preparation also avoids heat. The suppository mass, such as a mixture of grated theobroma oil and drug, is forced into a mold under pressure, using a wheel-operated press. The mass is forced into mold openings, pressure is released, and the mold removed, opened, and replaced. On a large scale, cold-compression machines are hydraulically operated, water-jacketed for cooling, and screw-fed. Pressure is applied via a piston to compress the mass into mold openings.

Fusion or Melt Molding. In this method the drug is dispersed or dissolved in the melted suppository base. The mixture then is poured into a suppository mold, allowed to cool, and the finished suppositories removed by opening the mold. Using this procedure, one to hundreds of suppositories can be made at one time.

Suppository molds are available for the preparation of various types and sizes of suppositories. Molds are made of aluminum alloy, brass, or plastic and are available with anywhere from six to several hundred cavities.

The method of choice for commercial suppository production (Figure 29-6) involves the automated filling of molds or preformed shells by a volumetric dosing pump that meters the melt from a jacketed kettle or mixing tank directly into the molds or shells. Strips of preformed shells pass beneath the dosing pump and are filled successively, passed through cooling chambers (to promote solidification), sealed, and then packaged. Quality control procedures (e.g., weight, fill volume, and leakage) are conducted readily in-line.

Injection Molding. An alternative to the melt-and-pour processes described above is that of injection molding. This process is distinctive in that it makes use of the injection-molding technique developed for the fabrication of plastics. Polyethylene glycols are the excipients of choice in this process, with polyethylene oxide, povidone, or silicon dioxide added to adjust viscosity or plasticity. Long-chain saturated carboxylic acids also have been added to reduce the hygroscopicity inherent with the use of the polyethylene glycols. Typically, the molten excipient admixture is extruded or injected under pressure into precision-machined multicavity molds, followed by the ejection of the molded units from the mold cavities. Advantages claimed for this method include the wide range of shapes and

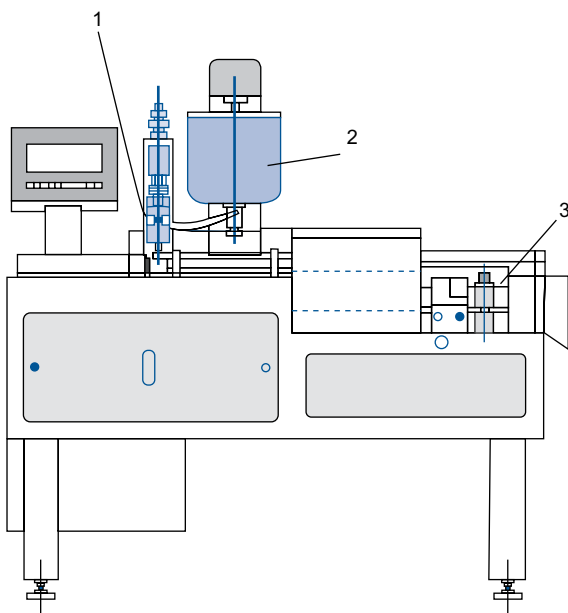


Figure 29-6. A cross-section of the Sarong SpA semiautomatic equipment for the production of suppositories in preformed plastic or foil shells. The fully jacketed piston-type *dosing pump* (1) meters the suppository melt in the jacketed tank (2) into preformed shells that pass directly beneath injection nozzles. The strips of filled preformed shells continue into a cooling chamber (3) prior to sealing and cartoning.

sizes that can be prepared at very high production rates with great precision.

Suppositories usually are formulated on a weight basis so that the medication replaces a portion of the vehicle as a function of specific gravity. If the medicinal substance has a density approximately the same as theobroma oil, it will replace an equal weight of oil. If the medication is heavier, it will replace a proportionally smaller amount of theobroma oil.

For instance, tannic acid has a density of 1.6 compared with cocoa butter (see Table 29-2). If a suppository is to contain 0.1 g tannic acid, then $0.1 \text{ g} \div 1.6$, or 0.062 g, cocoa butter should be replaced by 0.1 g of drug. If the blank weight of the suppository is 2.0 g, then $2.0 - 0.062 \text{ g}$, or 1.938 g, cocoa butter is required per suppository. The suppository will actually weigh $1.938 \text{ g} + 0.1 \text{ g}$, or 2.038 g. Table 29-2 indicates the density factor, or the density compared with cocoa butter, of many substances used in suppositories.

It is always possible to determine the density of a medicinal substance relative to cocoa butter, if the density factor is not available, by mixing the amount of drug for one or more suppositories with a small quantity of cocoa butter, pouring the mixture into a suppository mold and carefully filling the mold with additional melted cocoa butter. The cooled suppositories are weighed, providing data from which a working formula can be calculated as well as the density factor itself.

When using suppository bases other than cocoa butter, such as a polyethylene glycol base, it is necessary to know either the density of the drug relative to the new base or the densities of both the drug and the new base relative to cocoa butter. The density factor for a base other than cocoa butter is simply the ratio of the blank weights of the base and cocoa butter.

For instance, if a suppository is to contain 0.1 g tannic acid in a polyethylene glycol base, then $0.1 \text{ g} \div 1.6 \times 1.25$, or 0.078 g, polyethylene glycol base should be replaced by 0.1 g drug (the polyethylene glycol base is assumed to have a density factor of 1.25). If the blank weight is 1.75 g for the polyethylene glycol base, then $1.75 \text{ g} - 0.078 \text{ g}$, or 1.672 g, of base is required per suppository. The final weight will be $1.672 \text{ g base} + 0.1 \text{ g drug}$, or 1.772 g.

Table 29-2. Density Factors for Cocoa Butter Suppositories

Medication	Factor
Alum	1.7
Aminophylline	1.1
Aminopyrine	1.3
Aspirin	1.3
Barbital	1.2
Belladonna extract	1.3
Benzoic acid	1.5
Bismuth carbonate	4.5
Bismuth salicylate	4.5
Bismuth subgallate	2.7
Bismuth subnitrate	6.0
Boric acid	1.5
Castor oil	1.0
Chloral hydrate	1.3
Cocaine hydrochloride	1.3
Digitalis leaf	1.6
Gallic acid	2.0
Glycerin	1.6
Ichthammol	1.1
Iodoform	4.0
Menthol	0.7
Morphine hydrochloride	1.6
Opium	1.4
Paraffin	1.0
Peruvian balsama	1.1
Phenobarbital	1.2
Phenola	0.9
Potassium bromide	2.2
Potassium iodide	4.5
Procaine	1.2
Quinine hydrochloride	1.2
Resorcinol	1.4
Salicylic acid	1.3
Sodium bromide	2.3
Spermaceti	1.0
Sulfathiazole	1.6
Tannic acid	1.6
White wax	1.0
Witch hazel fluid extract	1.1
Zinc oxide	4.0
Zinc sulfate	2.8

^a Density adjusted taking into account white wax in mass.

When the dosage and mold calibration are complete, the drug-base mass should be prepared using minimum heat. A water bath or water jacket usually is used. The melted mass should be stirred constantly but slowly to avoid air entrapment. The mass should be poured into the mold openings slowly. Prelubrication of the mold will depend on the vehicle. Mineral oil is a good lubricant for cocoa butter suppositories. Molds should be dry for polyethylene glycol suppositories.

After being poured into tightly clamped molds, the suppositories and mold are allowed to cool thoroughly, using refrigeration on a small scale or refrigerated air on a larger scale. After thorough chilling, any excess suppository mass should be removed from the mold by scraping, the mold opened, and the suppositories removed. It is important to allow cooling time adequate for

suppository contraction. This aids in removal and minimizes splitting of the finished suppository.

Packaging and Storage. Suppositories often are packaged in partitioned boxes that hold the suppositories upright. Glycerin and glycerinated gelatin suppositories often are packaged in tightly closed screw-capped glass containers. Although many commercial suppositories are wrapped individually in aluminum foil or PVC-polyethylene, strip-packaging is commonplace.

Alternatively, suppositories may be molded directly into their primary packaging. In this operation the form into which the suppository mass flows consists of a series of individual molds formed from plastic or foil. After the suppository is poured and cooled, the excess is trimmed off, and the units are sealed and cut into 3s or 6s, as desired. Cooling and final cartoning then can be carried out.

Suppositories with low-melting ingredients are best stored in a cool place. Theobroma oil suppositories, in particular, should be refrigerated.

CONTRACEPTIVES

Contraceptives — in the context of this chapter — are employed in the form of creams, jellies, or aerosol foams, intended for vaginal use to protect against pregnancy. Contraceptive creams and jellies are designed to melt or spread, following insertion, over the vaginal surfaces. These agents act to immobilize spermatozoa.

Creams and jellies for contraceptive use may contain spermicidal agents such as nonoxynol 9, or they may function by a specific pH effect. A pH of 3.5 or less has an appreciable spermicidal effect. It is important to note that a final *in situ* pH of 3.5 or less is required; thus, the dilution effect and pH change brought about by vaginal fluids must be considered. To achieve the proper pH effect and control, buffer systems composed of acid and acid salts, such as lactates, acetates, and citrates, are commonly used.

PRESERVATIVES IN TOPICAL FORMULATIONS

Antimicrobial preservative substances are included in topical formulations to maintain the potency and integrity of product forms and to protect the health and safety of the consumer. The USP addresses this subject in its monograph *Microbiological Examination of Non-Sterile Products: Acceptance Criteria for Pharmaceutical Preparations and Substances for Pharmaceutical Use*. The USP designates acceptance criteria for total aerobic microbial count and total combined yeasts and mold

in topical formulations. The significance of microorganisms in nonsterile products should be evaluated in terms of the use of the product, the nature of the product, and the potential hazard to the user. Specific requirements are cited for cutaneous products (including transdermal formulations), which should be tested for the presence of *Pseudomonas aeruginosa* and *Staphylococcus aureus*. Products intended for vaginal administration should also be tested for *Candida albicans*.

REFERENCES

1. Ostrenga J *et al.* Significance of vehicle composition. I. Relationship between topical vehicle composition, skin permeability, and clinical efficacy. *J Pharm Sci* 1971; 60: 1175–1179.
2. Chiang C-M *et al.* Bioavailability assessment of topical delivery systems: effect of vehicle evaporation upon in vitro delivery of minoxidil from solution formulations. *Int J Pharm* 1989; 55: 229–236.
3. Machado M *et al.* Assessment of the variation of skin barrier function with anatomic site, age, gender and ethnicity. *Int J Cosmetic Sci* 2010; 32: 397–
4. Karande P *et al.* Design principles of chemical penetration enhancers for transdermal drug delivery. *Proc Nat Acad Sci* 2005; 102: 4688–4693.
5. Vergnaud J, Rosca I. Transdermal therapeutic systems. In: *Assessing Bioavailability of Drug Delivery Systems: Mathematical Modeling*. Boca Raton: CRC Press, Taylor and Francis Group, 2005.
6. Ceve G, Vierl U. Nanotechnology and the transdermal route: a state of the art review and critical appraisal. *J Contr Rel* 2010; 141: 277–299.
7. Buhse L *et al.* Topical drug classification. *Int J Pharm* 2005; 295: 101–112.
8. Elkeeb R *et al.* Transungual drug delivery: current status. *Int J Pharm* 2010; 384: 1–10.
9. Murdan S. Drug delivery to the nail following topical application. *Int J Pharm* 2002; 236: 1–26.
10. Abd-El-Maeboud KH, El-Naggar T. Rectal suppository: common-sense and mode of insertion. *Lancet* 1991; 338: 798–800.
11. Vail J. The rectal rocket: a two-day treatment for hemorrhoids. *Int J Pharm Compound* 2007; 11: 194–199.
12. Vermani K, Garg S. The scope and potential of vaginal drug delivery. *Pharm Sci Technol Today* 2000; 3: 359–364.
13. Mahalingam A *et al.* Design of a semisolid vaginal microbicide gel by relating composition to properties and performance. *Pharm Res* 2010; 27: 2478–2491.

Chapter 30

Oral Solid Dosage Forms

Ahmed Adel Sakr, PhD and Fars K. Alanazi, PhD

TABLETS 581

Medicinal substances are most frequently administered orally, by means of solid dosage forms, such as tablets and capsules. Large scale production methods used for their manufacture, as is described in this chapter, require the incorporation of other materials, in addition to the active ingredients. These additives are usually included in the formulation to facilitate handling, enhance the physical appearance, improve stability, and aid in the delivery of the medicament to the blood stream after administration. These materials, as well as the employed production methods, have been shown to potentially influence the absorption and/or bioavailability of the drugs.¹ In addition, the physicochemical characteristics of the drug substances may influence the physiological bioavailability from solid dosage forms.²

TABLETS

According to the USP, tablets are solid dosage forms, containing medicinal substances with or without suitable diluents. They may be classed, according to the method of manufacture, as compressed tablets or molded tablets. The vast majority of all tablets manufactured today are made by compression, and compressed tablets are the most widely used dosage form. Compressed tablets are prepared by the application of high pressures, utilizing steel punches and dies, to powders or granulations. Recently, punching of laminated sheets, electronic deposition methods, and three-dimensional printing methods have been used to make tablets. Tablets have been in widespread use, since the latter part of the 19th century, and their popularity continues. The term “compressed tablet” is believed to have been first used by John Wyeth and Brother, Inc. of Philadelphia. During this same period, molded tablets were introduced for use as hypodermic tablets, for the extemporaneous preparation of solutions for injection. Tablets remain popular as a dosage form, due to the advantages afforded both to the manufacturer (e.g., simplicity and economy of preparation, stability, and convenience in packaging, shipping, and dispensing) and to the patient (e.g., accuracy of dosage, compactness, portability, blandness of taste, and ease of administration). Although the basic mechanical approach for most tablet manufacture has remained the same, tablet technology has undergone great improvement and experimentation. Efforts are continually made to understand more clearly the physical characteristics of powder compaction and the factors affecting the availability of the drug substance from the dosage form after oral administration. Tableting equipment continues to improve in both production speed and the uniformity of tablets. Recent advances in tablet technology have been reviewed.^{3,4}

Although tablets are frequently discoid in shape, they also may be round, oval, oblong, cylindrical, or triangular. Other geometric shapes, such as diamonds and pentagons, and hexagons have also been used. They may differ greatly in size and weight, depending on the amount of drug substance present and the intended method of administration. Most commercial tablets can be divided into two general classes: whether they are made by compression or molding. Compressed tablets are prepared by large-scale production methods, whereas molded tablets involve small-scale operations. The various tablet types and abbreviations used in referring to them are listed here.

COMPRESSED TABLET (CT)

Compressed tablets are formed by compression and, in their simplest form, contain no special coating. They are made from powdered, crystalline, or granular materials, alone or in combination with binders, disintegrants, controlled-release polymers, lubricants, diluents, and, in many cases, colorants. The vast majority of tablets commercialized today are compressed tablets, either in an uncoated or coated state.

Sugar-Coated Tablets (SCT)

Sugar-coated tablets are compressed tablets surrounded by a sugar coating. Such coatings may be colored and are beneficial in covering up drug substances possessing objectionable tastes or odors and in protecting materials sensitive to oxidation. These coatings were once quite common, but lost commercial appeal due to the high cost of process validation. Recently, they have made a comeback, due to patient popularity and technical advances.

Film-Coated Tablets (FCT)

Film-coated tablets are compressed tablets covered with a thin layer or film of a water-soluble material. A number of polymeric substances with film-forming properties may be used. Film coating imparts the same general characteristics as sugar coating with the added advantage of the greatly reduced time required for the coating operation. Advances in material science and polymer chemistry have made these coatings the first-choice of formulators.

Enteric-Coated Tablets (ECT)

Enteric-coated tablets are compressed tablets coated with substances that resist solution in gastric fluid but disintegrate in the intestine. Enteric coatings can be used for tablets containing drug substances inactivated or destroyed in the stomach, for those that irritate the mucosa, or as a means of delayed release of the medication.

Multiple Compressed Tablets (MCT)

Multi-compressed tablets are compressed tablets made by more than one compression cycle. This process is best used when separation of active ingredients is needed for stability purposes or if the mixing process is inadequate to guarantee uniform distribution of two or more active ingredients.

Layered Tablets

Layered tablets are prepared by compressing additional tablet granulation on a previously compressed granulation. The operation may be repeated to produce multilayered tablets of two or more layers.

Press-Coated Tablets

Press coated tablets, also referred to as dry-coated tablets, are prepared by feeding previously compressed tablets into a special tableting machine and compressing another granulation layer around the preformed tablets. They have all the advantages of compressed tablets (i.e., slotting, monogramming, speed of disintegration), while retaining the attributes of sugar-coated tablets in masking the taste of the drug substance in the core tablets. An example of a press-coated tablet press is

the Manesty Drycota. Press-coated tablets can also be used to separate incompatible drug substances; in addition, they can provide a means of giving an enteric coating to the core tablets. Both types of multiple-compressed tablets have been widely used in the design of prolonged-action dosage forms.

Controlled-Release Tablets (CRT)

Compressed tablets can be formulated to release the drug slowly over a prolonged period of time. Hence, these dosage forms have been referred to as “prolonged-release” or “sustained-release” dosage forms as well. These tablets, as well as capsule versions, can be categorized into three types: 1) those that respond to some physiological condition to release the drug, such as enteric coatings; 2) those that release the drug in a relatively steady, controlled manner; and 3) those that combine combinations of mechanisms to release pulses of drug, such as repeat-action tablets. Chapter 32 describes the performance of these systems in more detail. Other names for these types of tablets are; Extended Release, Sustained Release, Prolonged Release, Delayed Release, and, in the case of pulsatile tablets, Repeat Action, Pulsatile Release, or Pulse Release.

Tablets for Solution (CTS)

Compressed tablets used for preparing solutions or imparting given characteristics to solutions must be labeled to indicate they are not to be swallowed. Examples of these tablets are Halazone Tablets for Solution and Potassium Permanganate Tablets for Solution.

Effervescent Tablets

In addition to the drug substance, effervescent tablets contain sodium bicarbonate and an organic acid, such as tartaric or citric. In the presence of water, these additives react, liberating carbon dioxide that acts as a disintegrator and produces effervescence. Except for small quantities of lubricants present, effervescent tablets are soluble.

Compressed Suppositories or Inserts

Occasionally, vaginal suppositories, such as Metronidazole tablets, are prepared by compression. Tablets for this use usually contain lactose as the diluent. In this case, as well as for any tablet intended for administration by means other than swallowing, the label must indicate the manner in which it is to be used.

Buccal and Sublingual Tablets

Buccal and sublingual tablets are small, flat, oval tablets. Tablets intended for buccal administration by inserting into the buccal pouch (the space between the lip and gum in the mouth) may dissolve or erode slowly; therefore, they are formulated and compressed with sufficient pressure to give a hard tablet. Progesterone tablets may be administered in this way. Some newer approaches have employed materials that act as bioadhesives to increase absorption of the drug. Other approaches use tablets that melt at body temperatures. The matrix of the tablet is solidified, while the drug is in solution. After melting, the drug is automatically in solution and available for absorption, thus, eliminating dissolution as a rate-limiting step in the absorption of poorly soluble compounds. Sublingual tablets, such as those containing nitroglycerin, isoproterenol hydrochloride, or erythrityl tetranitrate, are placed under the tongue. Sublingual tablets dissolve rapidly, and the drug substances are absorbed readily by this form of administration.

MOLDED TABLETS OR TABLET TRITURATES (TT)

Tablet triturates are usually made from moist material, using a triturate mold that gives them the shape of cut sections of a cylinder. Such tablets must be completely and rapidly soluble. The problem arising from compression of these tablets is the failure to find a lubricant that is completely water-soluble.

Dispensing Tablets (DT)

Dispensing tablets provide a convenient quantity of potent drug that can be incorporated readily into powders and liquids, thus, circumventing the necessity to weigh small quantities. These tablets are supplied primarily as a convenience for extemporaneous compounding and should never be dispensed as a dosage form.

Hypodermic Tablets (BT)

Hypodermic tablets are soft, readily soluble tablets and were originally used for the preparation of solutions to be injected. Since stable parenteral solutions are now available for most drug substances, there is no justification for the use of hypodermic tablets for injection. Their use in this manner should be discouraged, since the resulting solutions are not sterile. Large quantities of these tablets continue to be made, but for oral administration. No hypodermic tablets have ever been recognized by the official compendia.

COMPRESSED TABLETS (CT)

For medicinal substances, with or without diluents, to be made into solid dosage forms with pressure, using available equipment, it is necessary that the material, either in crystalline or powdered form, possess a number of physical characteristics. These characteristics include the ability to flow freely, cohesiveness, and lubrication. The ingredients, such as disintegrants designed to break the tablet up in gastrointestinal (GI) fluids and controlled-release polymers designed to slow drug release, ideally, should possess these characteristics or not interfere with the desirable performance traits of the other excipients. Since most materials have none or only some of these properties, methods of tablet formulation and preparation have been developed to impart these desirable characteristics to the material that is to be compressed into tablets. The basic mechanical unit in all tablet-compression equipment includes a lower punch that fits into a die from the bottom and an upper punch, with a head of the same shape and dimensions, which enters the die cavity from the top, after the tableting material fills the die cavity (Fig. 30-1). The tablet is formed by pressure applied on the punches and, subsequently, is ejected from the die. The weight of the tablet is determined by the volume of the material that fills the die cavity. Therefore, the ability of the granulation to flow freely into the die is important in ensuring a uniform fill, as well as the continuous movement of the granulation from the source of supply or feed hopper. If the tablet granulation does not possess cohesive properties, the tablet, after compression, will crumble and fall apart on handling. As the punches must move



Figure 30-1. Multi-tip punches and die. (Courtesy of Natoli.)

freely within the die and the tablet must be ejected readily from the punch faces, the material must have a degree of lubrication to minimize friction and allow the removal of the compressed tablets. There are three general methods typically used for commercial tablet preparation: the wet-granulation method, the dry granulation method, and direct compression. The method of preparation and the added ingredients are selected to give the tablet formulation the desirable physical characteristics, allowing the rapid compression of tablets. After compression, the tablets must have a number of additional attributes, such as appearance, hardness, disintegration ability, appropriate dissolution characteristics, and uniformity, which are also influenced by both the method of preparation and the added materials present in the formulation. In the preparation of compressed tablets, the formulator must also be cognizant of the effect that the ingredients and methods of preparation may have on the availability of the active ingredients and, hence, the therapeutic efficacy of the dosage form. In response to a request by physicians to change a dicumarol tablet, so it might be broken more easily, a Canadian company reformulated to make a large tablet with a score. Subsequent use of the tablet, containing the same amount of drug substance as the previous tablet, resulted in complaints that larger-than-usual doses were needed to produce the same therapeutic response. Conversely, literature reports indicate that the reformulation of a commercial digoxin tablet resulted in a tablet that, although containing the same quantity of drug substance, gave the desired clinical response at half its original dose. Methods and principles that can be used to assess the effects of excipients and additives on drug absorption have been reviewed.⁵⁻⁷

TABLET INGREDIENTS

In addition to the active or therapeutic ingredient, tablets contain a number of inert materials. The latter are known as additives or excipients. They may be classified according to the part they play in the finished tablet. The first group contains those that help to impart satisfactory processing and compression characteristics to the formulation. These include diluents, binders, glidants, and lubricants. The second group of added substances helps to give additional desirable physical characteristics to the finished tablet. Included in this group are disintegrants, surfactants, colors, and, in the case of chewable tablets, flavors and sweetening agents, and, in the case of controlled-release tablets, polymers or hydrophobic materials, such as waxes or other solubility-retarding materials. In some cases, antioxidants or other materials can be added to improve stability and shelf-life. Although the term "inert" has been applied to these added materials, it has become apparent that there is an important relationship between the properties of the excipients and the dosage forms containing them.⁸ Preformulation studies demonstrate their influence on stability, bioavailability, and the processes by which the dosage forms are prepared. The need for acquiring more information and use standards for excipients has been recognized in a joint venture of the Academy of Pharmaceutical Sciences and the Council of the Pharmaceutical Society of Great Britain. The result is called the *Handbook of Pharmaceutical Excipients*. This reference work is now distributed widely throughout the world.⁹

Diluents

Frequently, the single dose of the active ingredient is small, and an inert substance is added to increase the bulk to make the tablet a practical size for compression. Compressed tablets of dexamethasone contain 0.75 mg steroid per tablet; hence, it is obvious that another material must be added to make tableting possible. Diluents used for this purpose include dicalcium phosphate, calcium sulfate, lactose, cellulose, kaolin, mannitol, sodium chloride, dry starch, and powdered sugar. Certain diluents, such as mannitol, lactose, sorbitol, sucrose, and inositol, when present in sufficient quantity, can impart properties to some compressed tablets that permit disintegration in the mouth by chewing. Such tablets are commonly called chewable tablets.

Upon chewing, properly prepared tablets disintegrate smoothly at a satisfactory rate, have a pleasant taste and feel, and leave no unpleasant aftertaste in the mouth. Diluents used as excipients for direct compression formulas have been subjected to prior processing to give them flow ability and compressibility. These are discussed in the section Direct Compression. Most formulators of immediate-release tablets tend to consistently use only one or two diluents selected from the listed group in their tablet formulations. Usually, these are selected on the basis of experience and cost factors. However, in the formulation of new therapeutic agents, the compatibility of the diluents with the drug must be considered (e.g., calcium salts used as diluents for the broad-spectrum antibiotic tetracycline) have been shown to interfere with the drug's absorption from the GI tract. When drug substances have low water solubility, it is recommended that water-soluble diluents be used to avoid possible bioavailability problems. Highly adsorbent substances (e.g., bentonite and kaolin) are avoided in making tablets of drugs used clinically in small dosage, such as the cardiac glycosides, alkaloids, and synthetic estrogens. These drug substances may be adsorbed after administration. The combination of amine bases with lactose or alkaline salts with lactose in the presence of alkaline lubricant results in tablets that discolor on aging.

Microcrystalline cellulose (Avicel) is usually used as an excipient in direct-compression formulas. However, its presence in 5–15% concentrations in wet granulations has been shown beneficial in the granulation and drying processes in minimizing case-hardening of the tablets and in reducing tablet mottling. Many ingredients are used for several different purposes, even within the same formulation (e.g., cornstarch can be used in paste form as a binder). When added in drug or suspension form, it is a good disintegrant. Even though these uses are to achieve opposite goals, some tablet formulas use cornstarch in both ways. In some controlled-release formulas, the polymer hydroxypropyl methylcellulose (HPMC) is used both as an aid to prolong the release from the tablet, as well as a film-former in the tablet coating. Therefore, most excipients used in formulating tablets and capsules have many uses, and a thorough understanding of their properties and limitations is necessary to use them rationally.

Binders

Agents used to impart cohesive qualities to the powdered material are referred to as binders or granulators. They impart cohesiveness to the tablet formulation that ensures the tablet remain intact after compression, as well as improve the free-flowing qualities by the formulation of granules of desired hardness and size. Materials commonly used as binders include starch, gelatin, and sugars, such as sucrose, glucose, dextrose, molasses, and lactose. Natural and synthetic gums that have been used include acacia, sodium alginate, extract of Irish moss, panwar gum, ghatti gum, mucilage of isapol husks, carboxymethylcellulose, methylcellulose, polyvinylpyrrolidone, Veegum, and larch arabogalactan. Other agents considered binders under certain circumstances are polyethylene glycol, ethylcellulose, waxes, water, and alcohol. Synthetic binders, such as polyvinylpyrrolidone and HPMC are becoming more popular. The quantity of binder used has considerable influence on the characteristics of the compressed tablets. The use of too much binder or too strong a binder will make a hard tablet that will not disintegrate easily and will cause excessive wear of punches and dies. Differences in binders used for CT Tolbutamide resulted in differences in hypoglycemic effects observed clinically. Materials that have no cohesive qualities of their own require a stronger binder than those with these qualities. Alcohol and water are not binders in the true sense of the word, but, due to their solvent action on some ingredients, such as lactose, starch, and celluloses, they change the powdered material to granules, and the residual moisture retained enables the materials to adhere together when compressed. Binders are used both as a solution and in a dry form, depending on the other ingredients in the formulation and the method of preparation. However, several pregelatinized starches available are intended

to be added in the dry form, so water alone can be used as the granulating solution. The same amount of binder in solution will be more effective than if it were dispersed in a dry form and moistened with the solvent. By the latter procedure, the binding agent is not as effective in reaching and wetting each of the particles within the mass of powders. Each of the particles in a powder blend has a coating of adsorbed air on its surface, and it is this film that must be penetrated before the powders can be wetted by the binder solution. After wetting, a certain period of time is necessary to dissolve the binder completely and make it completely available for use. Since powders differ with respect to the ease with which they can be wetted and their rate of solubilization, it is preferable to incorporate the binding agent in solution. By this technique, it is often possible to gain effective binding with a lower concentration of binder. The direct-compression method for preparing tablets requires a material that is not only free-flowing but also sufficiently cohesive to act as a binder. This use has been described for a number of materials, including microcrystalline cellulose, microcrystalline dextrose, amylose, and polyvinylpyrrolidone. It has been postulated that microcrystalline cellulose is a special form of cellulose fibril in which the individual crystallites are held together largely by hydrogen bonding. The disintegration of tablets containing the cellulose occurs by breaking the intercrystallite bonds by the disintegrating medium.

Starch Paste—Cornstarch is used widely as a binder. The concentration may vary from 10% to 20%. It is usually prepared as it is to be used, by dispersing cornstarch in sufficient cold purified water to make a 5–10% w/w suspension and warming in a water bath with continuous stirring until a translucent paste forms. It has been observed that, during paste formation, not all of the starch is hydrolyzed. Starch paste then is not only useful as a binder, but also as a method to incorporate some disintegrant inside the granules.

Gelatin Solution—Gelatin is used as a 10–20% solution; gelatin solutions should be prepared freshly as needed and used while warm, or they will solidify. The gelatin is added to cold purified water and allowed to stand, until it is hydrated. It is then warmed in a water bath to dissolve the gelatin, and the solution is made up to the final volume on a weight basis to give the concentration desired.

Cellulosic Solutions—Various celluloses have been used as binders in solution form. Hydroxypropyl methylcellulose (HPMC) has been used widely in this regard. Typical of a number of celluloses, HPMC is more soluble in cold water than hot. It is also more dispersible in hot water than cold. Hence, to obtain a good, smooth gel that is free from lumps or “fisheyes,” it is necessary to add the HPMC in hot, almost boiling, water and, under agitation, cool the mixture down as quickly as possible, as low as possible. Other water-soluble celluloses, such as hydroxyethylcellulose (HEC) and hydroxypropylcellulose (HPC), have been used successfully in solution as binders.

Not all celluloses are soluble in water. Ethylcellulose can be used effectively, when dissolved in alcohol or as a dry binder that then is wetted with alcohol. It is used as a binder for materials that are moisture-sensitive.

Polyvinylpyrrolidone (PVP) can be used as an aqueous or alcoholic solution, and this versatility has increased its popularity. Concentrations range from 2% and vary considerably. It should be noted that binder solutions are usually made up to Weight, rather than volume. This is to enable the formulator to determine the weight of the solids that have been added to the tablet granulation in the binding solution. This becomes part of the total weight of the granulation and must be taken into consideration in determining the weight of the compressed tablet, which will contain the stated amount of the therapeutic agent. As can be seen by the list of binders in this chapter, most modern binders used in solution are polymeric. Because of this, the flow or spreadability of these solutions becomes important, when selecting the appropriate granulating equipment. The rheology of polymeric solutions is a fascinating subject in and of itself and should be considered for these materials.

Lubricants

Lubricants have a number of functions in tablet manufacture. They prevent adhesion of the tablet material to the surface of the dies and punches, reduce interparticle friction, facilitate the ejection of the tablets from the die cavity, and may improve the rate of flow of the tablet granulation. Commonly used lubricants include talc, magnesium stearate, calcium stearate, stearic acid, glyceryl behenate, hydrogenated vegetable oils, and polyethylene glycol (PEG). Most lubricants, with the exception of talc, are used in concentrations below 1%. When used alone, talc may require concentrations as high as 5%. Lubricants are, in most cases, hydrophobic materials. Poor selection or excessive amounts can result in waterproofing the tablets, resulting in poor tablet disintegration and/or delayed dissolution of the drug substance. The addition of the proper lubricant is highly desirable, if the material to be tableted tends to stick to the punches and dies. Immediately after compression, most tablets have the tendency to expand and will bind and stick to the side of the die. The choice of the proper lubricant will overcome this effectively. The method of adding a lubricant to a granulation is important, if the material is to perform its function satisfactorily. The lubricant should be divided finely by passing it through a 60- to 100-mesh nylon cloth onto the granulation. In production, this is called “bolting” the lubricant. After adding the lubricant, the granulation is tumbled or mixed gently to distribute the lubricant without coating the particles too well or breaking them down to finer particles. Some research has concluded that the order of mixing of lubricants and other excipients can have a profound effect on the performance of the final dosage form. Thus, attention to the mixing process itself is just as important as the selection of lubricant materials. These process variables can be seen in the prolonged blending of a lubricant in a granulation. Overblending materially can affect the hardness, disintegration time, and dissolution performance of the resultant tablets. The quantity of lubricant varies, being as low as 0.1% and, in some cases, as high as 5%. Lubricants have been added to the granulating agents in the form of suspensions or emulsions. This technique serves to reduce the number of operational procedures and, thus, reduce the processing time. In selecting a lubricant, proper attention must be given to its compatibility with the drug agent. Perhaps the most widely investigated drug is acetylsalicylic acid. Different tales varied the stability of aspirin significantly. Talc, with a high calcium content and a high loss on ignition, was associated with increased aspirin decomposition. From a stability standpoint, the relative acceptability of tablet lubricants for combination with aspirin was found to decrease in the following order: hydrogenated vegetable oil, stearic acid, talc, and aluminum stearate. The primary problem in the preparation of a water-soluble tablet is the selection of a satisfactory lubricant. Soluble lubricants reported to be effective include sodium benzoate, a mixture of sodium benzoate and sodium acetate, sodium chloride, leucine, and polyethylene glycol/Carbowax 4000. However, it has been suggested that formulations used to prepare water soluble tablets may represent a number of compromises between compression efficiency and water solubility. Although magnesium stearate is one of the most widely used lubricants, its hydrophobic properties can retard disintegration and dissolution. To overcome these waterproofing characteristics, sodium lauryl sulfate is sometimes included. One compound found to have the lubricating properties of magnesium stearate, without their disadvantage, is magnesium lauryl sulfate. Its safety for use in pharmaceuticals has not been established.

Glidants

A glidant is a substance that improves the flow characteristics of a powder mixture. These materials always are added in the dry state just prior to compression (i.e., during the lubrication step). Colloidal silicon dioxide Cab-o-sil (Cabot) is the most commonly used glidant and is used in low concentrations of

1% or less. Talc (asbestos-free) is also used and may serve the dual purpose of lubricant and glidant. It is especially important to optimize the order of addition and the mixing process for these materials, to maximize their effect and to make sure their influence on the lubricant(s) is minimized.

Disintegrants

A disintegrant is a substance or a mixture of substances added to a tablet to facilitate its breakup or disintegration after administration. The active ingredient must be released from the tablet matrix as efficiently as possible to allow rapid dissolution. Materials serving as disintegrants have been classified chemically as starches, clays, celluloses, alginates, gums, and cross-linked polymers. The oldest and still the most popular disintegrants are corn and potato starches that have been well dried and powdered.

Starch has a great affinity for water and swells when moistened, thus, facilitating the rupture of the tablet matrix. However, others have suggested that its disintegrating action in tablets is due to capillary action, rather than swelling; the spherical shape of the starch grains increases the porosity of the tablet, thus, promoting capillary action. Starch, 5%, is suggested, but, if more rapid disintegration is desired, this amount may be increased to 10% or 15%. Although it might be expected that disintegration time would decrease as the percentage of starch in the tablet increased, this does not appear to be the case for tolbutamide tablets. In this instance, there appears to be a critical starch concentration for different granulations of the chemical. When their disintegration effect is desired, starches are added to the powder blends in the dry state. A group of materials known as "super disintegrants" have gained in popularity as disintegrating agents. The name comes from the low levels (1–4%) at which they are completely effective. Croscarmellose, crospovidone, and sodium starch glycolate represent examples of a cross-linked cellulose, a cross-linked polymer, and a cross-linked starch, respectively. The development of these disintegrants fostered new theories about the various mechanisms by which disintegrants work. Sodium starch glycolate swells 7- to 12-fold in less than 30 seconds. Croscarmellose swells 4- to 8-fold in less than 10 seconds. The starch swells equally in all three dimensions, whereas the cellulose swells only in two dimensions, leaving fiber length essentially the same. It is postulated that the rate, force, and extent of swelling play an important role in those disintegrants that work by swelling. Cross-linked PVP swells little but returns to its original boundaries quickly after compression. Wicking, or capillary action, is also postulated to be a major factor in the ability of cross-linked PVP to function.^{10–13} In addition to the starches, large varieties of materials have been used and are reported to be effective as disintegrants. This group includes Veegum HV, methylcellulose, agar, bentonite, cellulose and wood products, natural sponge, cation-exchange resins, alginic acid, guar gum, citrus pulp, and carboxymethylcellulose. Sodium lauryl sulfate, in combination with starch, has also been demonstrated to be an effective disintegrant. In some cases, the apparent effectiveness of surfactants in improving tablet disintegration is postulated as due to an increase in the rate of wetting. The disintegrating agent is usually mixed with the active ingredients and diluents after granulation. In some cases, it may be advantageous to divide the disintegrant into two portions: one part is added to the powdered formula prior to granulation, and the remainder is mixed with the lubricant and added prior to compression. Incorporated in this manner, the disintegrant serves a double purpose: the portion added to the lubricant rapidly breaks down the tablet to granules, and the disintegrant mixed with the active ingredients disintegrates the granules into smaller particles.¹² Veegum has been shown more effective as a disintegrator in sulfathiazole tablets, when most of the quantity is added after granulation and only a small amount before granulation. Likewise, the montmorillonite clays were found to be good tablet disintegrants, when added to prepared granulations as powder.

They are much less effective as disintegrants, when incorporated within the granules.

Factors other than the presence of disintegrants can affect the disintegration time of compressed tablets significantly. The binder, tablet hardness, and the lubricant have been shown to influence the disintegration time. Thus, when the formulator is faced with a problem concerning the disintegration of a compressed tablet, the answer may not lie in the selection and quantity of the disintegrating agent alone. The evolution of carbon dioxide is also an effective way to cause the disintegration of compressed tablets. Tablets containing a mixture of sodium bicarbonate and an acidulant, such as tartaric or citric acid, will effervesce when added to water. Sufficient acid is added to produce a neutral or slightly acidic reaction when disintegration in water is rapid and complete.

One drawback to the use of the effervescent type of disintegrator is that such tablets must be kept in a dry atmosphere at all times during manufacture, storage, and packaging. Soluble, effervescent tablets provide a popular form for dispensing aspirin and noncaloric sweetening agents.

Coloring Agents

Colors in compressed tablets serve functions other than making the dosage form more esthetic in appearance. Color helps the manufacturer control the product during its preparation, as well as serves as a means of identification to the user. The wide diversity in the use of colors in solid dosage forms makes it possible to use color as an important category in the identification code developed by the AMA to establish the identity of an unknown, compressed tablet in situations arising from poisoning. All colorants used in pharmaceuticals must be approved and certified by the FDA. For several decades, colorants have been subjected to rigid toxicity standards, and, as a result, a number of colorants have been removed from an approved list of Food, Drug and Cosmetic Act (FD&C) colors, or "delisted." Several have been listed as well. Table 30-1 lists the colorants currently approved in the United States. Each country has its own list of approved colorants, and formulators must consider this in designing products for the international market. Any of the approved, certified, water-soluble FD&C dyes, mixtures of the same, or their corresponding lakes may be used to color tablets. A color lake is the combination by adsorption of a water-soluble dye to a hydrous oxide of a heavy metal, resulting in an insoluble form of the dye. In some instances, multiple dyes are used to give a purposefully heterogeneous coloring in the form of speckling to compressed tablets. The dyes available do not meet all the criteria required for the ideal pharmaceutical colorants. The photosensitivity of several of the commonly used colorants and their lakes has been investigated, as well as the protection afforded by a number of glasses used in packaging tablets. Another approach for improving the photostability of dyes has been in the use of ultraviolet-absorbing chemicals in the tablet formulations with the dyes. The Di-Pac line (Amstar) is a series of commercially available colored, direct-compression sugars. The most common method of adding color to a tablet formulation is to dissolve the dye in the binding solution prior to the granulating process. Another approach is to adsorb the dye on starch or calcium sulfate from its aqueous solution; the resultant powder is dried and blended with the other ingredients. If the insoluble lakes are used, they may be blended with the other dry ingredients. Frequently, during drying, colors in wet granulations migrate, resulting in an uneven distribution of the color in the granulation. After compression, the tablets have a mottled appearance, due to the uneven distribution of the color. Migration of colors may be reduced by drying the granulation slowly at low temperatures and stirring the granulation, while it is drying. The affinity of several water-soluble, anionic, certified dyes for natural starches has been demonstrated; in these cases, this affinity should aid in preventing color migration. Other additives have been shown to act as dye-migration inhibitors. Tragacanth (1%), acacia (3%),

Table 30-1. Color Additives Approved for Use in Drugs**Part 74, Subpart B: Color additives subject to batch certification**

21 CFR Section	Straight Color	Year ⁽²⁾ Approved	Uses and Restrictions
§74.1101	FD&C Blue No. 1	1969	Ingested drugs generally ⁽⁵⁾
§74.1102	FD&C Blue No. 2	1987	Ingested drugs
§74.1203	FD&C Green No. 3	1982	Drugs generally ⁽⁵⁾
§74.1205	D&C Green No. 5	1982	Drugs generally ⁽⁵⁾
§74.1303	FD&C Red No. 3	1969	Ingested drugs
§74.1306	D&C Red No. 6	1983	Drugs such that total with D&C Red No. 7 NTE ⁽⁷⁾ 5 mg/daily dose of drug
§74.1307	D&C Red No. 7	1983	Drugs such that total with D&C Red No. 6 NTE ⁽⁷⁾ 5 mg/daily dose of drug
§74.1321	D&C Red No. 21	1983	Drugs generally ⁽⁵⁾
§74.1322	D&C Red No. 22	1983	Drugs generally ⁽⁵⁾
§74.1327	D&C Red No. 27	1982	Drugs generally ⁽⁵⁾
§74.1328	D&C Red No. 28	1982	Drugs generally ⁽⁵⁾
§74.1330	D&C Red No. 30	1982	Drugs generally ⁽⁵⁾
§74.1333	D&C Red No. 33	1988	Ingested drugs, other than mouthwashes and dentifrices (NTE ⁽⁷⁾ 0.75 mg/daily dose of drug)
§74.1336	D&C Red No. 36	1989	Ingested drugs, other than mouthwashes and dentifrices, NTE ⁽⁷⁾ 1.7 mg/daily dose for drugs taken less than 1 yr; NTE ⁽⁷⁾ 1.0 mg/daily dose for drugs taken more than 1 yr
§74.1340	FD&C Red No. 40(3)	1971	Drugs generally ⁽⁵⁾
§74.1705	FD&C Yellow No. 5	1969	Ingested drugs generally ⁽⁵⁾
§74.1706	FD&C Yellow No. 6	1986	Drugs generally ⁽⁵⁾
§74.1710	D&C Yellow No. 10	1983	Drugs generally ⁽⁵⁾
		1984.	Modification of uses and restrictions

² The year approved is based on the date listed in the "Confirmation of Effective Date" notice for the action as published in the Federal Register.

⁵ Color additives permitted for general use may not be used in the area of the eye, in injections or in surgical sutures, unless such use is specified in the color additive listing regulation. Currently, no color additives are listed for use in injected products, such as tattoos or permanent makeup.

⁷ NTE—not to exceed

attapulgit (5%), and talc (7%) were effective in inhibiting the migration of FD&C Blue Nolin lactose. In using dye lakes, the problem of color migration is avoided, since the lakes are insoluble. Prevention of mottling can also be helped, by the use of lubricants and other additives that have been colored similarly to the granulation prior to their use. The problem of mottling becomes more pronounced as the concentration of colorants increases. Color mottling is an undesirable characteristic common to many commercial tablets.

Flavoring Agents

In addition to the sweetness that may be afforded by the diluent of the chewable tablet (e.g., mannitol or lactose) artificial sweetening agents may be included. Formerly, the cyclamates, either alone or in combination with saccharin, were used widely. With the banning of the cyclamates and the indefinite status of saccharin, new natural sweeteners are being sought. Aspartame (Pfizer), has found applications in pharmaceutical formulations. Sweeteners, other than the sugars, have the advantage of reducing the bulk volume, considering the quantity of sucrose required to produce the same degree of sweetness. Present in small quantities, they do not markedly affect the physical characteristics of the tablet granulation.

POWDER COMPACTION

Compressed tablets became a commercially viable and efficient dosage form with the invention of tablet machines. In 1843, William Brockendon, a British inventor, author, artist, and watchmaker, received British Patent #9977 for *Shaping Pills, Lozenges, and Black Lead by Pressure in Dies*. In over 150 years of tablet manufacture, the basic process has not changed. Surpris-

ingly, improvements have been made only with regards to speed of manufacture and quality control. The process of compaction has several identifiable phases. As can be seen in Figure 30-2, when powders undergo compression (a reduction in volume), the first process to occur is a consolidation of the powders. During this consolidation phase, the powder particles adopt a more efficient packing order. The second phase of the compaction process is elastic, or reversible deformation. If the force were removed during this phase, the powder would completely recover to the efficiently packed state. For most pharmaceutical powders, this phase is very short in duration and very difficult to identify on most instrumented tablet presses. The third phase of compaction is plastic, or irreversible, deformation of the powder bed. It is this phase that is the most critical in tablet

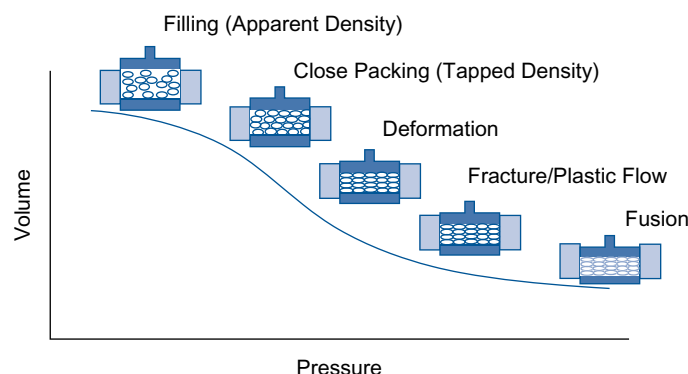


Figure 30-2. The stages of powder compaction.

formation. If too much force is applied to the powder, brittle fracture occurs. If the force is applied too quickly, fracture and de-bonding during stress relaxation can occur. In 1950, Stewart reported on the importance of plastic flow and suggested that, if a material has significant plastic flow under compression, it will be more likely to form a compact. Stress-relaxation data, using the Maxwell model of viscoelastic behavior, were evaluated, in an attempt to quantify the rate of plastic deformation of some direct compression excipients. The term “contact time” was used to describe the total time for which a moving punch applies a detectable force to the die contents during the compression and decompression event, excluding ejection.

Tablet Strength-Compression Pressure Profile

Most formulators use tablet hardness, or tensile strength, as a measure of the cohesiveness and/or the crushing strength/breaking strength of the tablet. With even the simplest of instrumented tablet presses, it is possible to plot tensile strength versus the force applied to the tablet. These plots can be useful in identifying forces that can cause fracture and can lead to a quick, tangible assessment of the compatibility of the formulation. However, there are many limitations to this method, as these plots cannot predict lamination or capping. In addition, the cohesiveness of a tablet can change upon storage, in either a positive or negative direction.

Tablet Friability

There have been many suggestions about how tablet friability tests should be performed. Many formulators believe this is an important indicator of cohesiveness but is of limited value in predicting failure in the field.

Changes in Bed Density During Compression

As applied stress (force) increases, elastic and plastic deformation of the particles occurs, which results in plastic flow and a reduction in inter- and intraparticulate void spaces. This lowers the overall compact density. For highly cohesive systems, the reduction in void space may yield a compact of sufficient strength for insertion into a capsules shell. However, the inherent cohesiveness for most drugs and excipients is not suitable alone for tablet manufacture. The Heckel equation is subsequently given; K can be considered equal to the reciprocal of the mean yield pressure, and A is a function of the original compact volume and is related to the densification and particle rearrangement prior to bonding:

$$\text{Log } [1/(1 - D)] = KP + A,$$

where D is the relative density at pressure P , and K and A are constants.

Changes in Surface Area During Compression

Bulk powders change their state of packing during compaction, and individual particles fracture and/or plastically deform. During this process, the surface area of the powders and the compact, in whole change. Conventional nitrogen absorption techniques can estimate these changes. Although this can be tedious, these measurements can give a means of examining lamination tendency.

Stress Relaxation

Stress relaxation, an experimental technique, consists of holding the compression process at a point of maximum compression and observing the compression force over various periods of time. By increasing the duration of this period (or dwell time), plastic flow is maximized, and tablet strength increases.

Stress Transmissions during Compression

If the stresses in the upper punch, lower punch, and die wall are monitored, a general plot can be constructed, showing the relationship between these forces.

GRANULATION METHODS

Wet Granulation

The most widely used and most general method of tablet preparation is the wet-granulation method. Its popularity is due to the greater probability that the granulation will meet all the physical requirements for the compression of good tablets. Its chief disadvantages are the number of separate steps involved, as well as the time and labor necessary to carry out the procedure, especially on a large scale. The steps in the wet method are weighing, mixing, wet massing, screening the damp mass, drying, dry screening, lubrication, and compression. The equipment involved depends on the quantity or size of the batch and the percent active ingredient per total weight of the tablets. Wet massing can be performed by:

1. Low Shear mixers/granulators,
2. High Shear mixers/granulators,
3. Fluid-Bed granulators/dryers,
4. Spray Dryers, or
5. Extruders and Spheronizers.

Low shear mixers include the barrel, cube, twin shell, double cone, slated double cone, ribbon, sigma blade, planetary mixers, etc. (Fig. 30-3). Although these mixers continue to be used in some older pharmaceutical factories, they are replaced, now, by the high shear mixers/granulators. High shear mixers are stationary shell mixers with a large mixer scraper blade, which mixes the ingredients, eliminates dead spots, and presents the ingredients with a high speed chopper blade, which intimately



Figure 30-3. Twin Shell Blender with Intensifier Bars. (Courtesy of Patterson Kelly Buflavak LLC.)

mixes the ingredients and breaks the lumps. The advantages of the high shear mixers/granulators are: simple, robust technology, production of high density granules, rapid and efficient mixing, low liquid requirements, and equipped for drying.

There are many models/trade names of high shear mixers in the pharmaceutical market, such as the Loedige, Littleford MGT, Diosna, Fiedler, Vector, Glatt, GEA, and Hüttlin Gentlewing. Figure 30-4 shows a typical high shear mixer granulator from Vector, and Figure 30-5 shows a typical Glatt high shear mixer. A relatively new high shear mixer from Hüttlin is the Gentlewing High Shear Mixer, which is available in laboratory and large scale sizes, either top driven or bottom driven. The Gentlewing technology applies the principle of a positive displacement impeller for blending a non-Newtonian medium. The impeller shape matches the contour of the mixing container, and its angled impeller plate ensures a forced mixing of the product. Rather than using the impulse of an impeller for mixing, the Gentlewing's uniform presence throughout the product container distributes the mixing energy, at lower speeds but higher torque, evenly throughout the product, reducing segregation issues typically caused by high dynamic forces (Fig. 30-6 and Fig. 30-7).

The active ingredient, diluent, and part of the disintegrant are mixed or blended well. Solutions of the binding agent are added or sprayed to the mixed powders with stirring. The powder mass is wetted with the binding solution, until the mass has the consistency of damp snow or brown sugar. If the granulation is over-wetted, the granules will be hard, requiring considerable pressure to form the tablets, and the resultant tablets may have a mottled appearance. If the powder mixture is not wetted sufficiently, the resulting granules will be too soft, breaking down during lubrication and causing difficulty during compression. In modern high shear mixers, the end point of wet granulation is no more dependent on the experience of the operator and the "hand" feeling of the wet mass, but is measured accurately by a built in torque meter. This consistency and reproducibility



Figure 30-4. Vector High Shear Mixer.



Figure 30-5. High Shear Mixer. (Courtesy of GLATT.)

from batch to batch are assured. The wet granulation mass is then forced through a 6- or 8-mesh screen. Small batches can be forced through by hand, using a manual screen. For larger quantities, one of several comminuting mills suitable for wet screening can be used. These include the Stokes oscillator, Colton rotary granulator, Fitzpatrick comminuting mill, Cone Mill, or Stokes tornado mill. In comminuting mills, the granulation is forced through the sieving device by rotating hammers, knives, or oscillating bars (Fig. 30-8–30-10). Most high-speed mixers are equipped with a chopper blade that operates independently of the main mixing blades and can replace the wet milling step (i.e., can obviate the need for a separate operation). For tablet

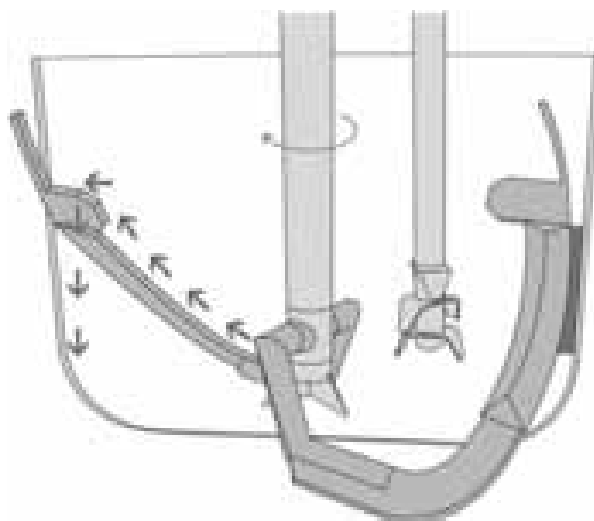


Figure 30-6. Gentlewing High Shear Mixer, diagram. (Courtesy of Hüttlin GmbH, A Bosch Packaging Technology Company.)



Figure 30-7. Gentlewing High Shear Mixer. (Courtesy of Hüttlin GmbH, A Bosch Packaging Technology Company.)

formulations in which continuous production is justified, extruders, such as the Reitz extruder, have been adapted for the wet-granulation process. Moist material from the wet milling step, traditionally, was placed on large sheets of paper on shallow wire trays and placed in drying cabinets with a circulating air current and thermostatic heat control. Although tray drying was the most widely used method of drying tablet granulations in the past, fluid-bed drying is now considered the standard. In drying tablet granulation by fluidization, the material is suspended and agitated in a warm air stream, while the granulation is maintained in motion. Drying tests comparing the fluidized bed and a tray dryer for a number of tablet granulations indicated that the former was 15 times faster than the conventional method of tray drying. In addition to the decreased drying time, the fluidization method is claimed to have other advantages, such as better control of drying temperatures, decreased handling costs, and the opportunity to blend lubricants and other materials into the dry granulation directly in the fluidized bed (Fig. 30-11). The application of microwave drying and infrared drying to tablet granulations has been reported as successful for most granulations tried. These methods readily lend themselves to continuous granulation operations. The study of drying methods for tablet granulations led to the development of the Rovac dryer system by Ciba Novartis pharmacists and engineers. The dryer is similar in appearance to the cone blender, except for the heating jacket and vacuum connections. By excluding oxygen and using the lower drying temperatures made possible by drying in a vacuum, opportunities for degradation of the ingredients during the drying cycle are minimized. A greater uniformity of residual moisture content is achieved, due to the moving bed, controlled temperature, and controlled time period of the drying cycle. Particle-size distribution can be controlled by varying the speed of rotation and drying tempera-



Figure 30-8. Fitzmill Comminutor. (Courtesy of Fitzpatrick Co.)

ture, as well as by comminuting the granulation to the desired granule size after drying. Recently, many machine manufacturers, such as Hüttlin, GEA, and Glatt, introduced the single pot technology in which the drying process can also take place in a vertical granulator specially equipped for this purpose. The drying is assisted by suitable measures, according to the product and the process, such as heated wall surfaces, gas stripping, and vacuum (Fig. 30-12).

In drying granulations, it is desirable to maintain a residual amount of moisture in the granulation. This is necessary to maintain the various granulation ingredients, such as gums, in a hydrated state. Also, the residual moisture contributes to the reduction of the static electric charges on the particles. In the selection of any drying process, an effort is made to obtain uniform moisture content. In addition to the importance of moisture content of the granulation in its handling, during the manufacturing steps, the stability of the products containing moisture-sensitive active ingredients may be related to the moisture content of the products. Previously, it was indicated that water-soluble colorants can migrate toward the surface of the granulation during the drying process, resulting in mottled tablets after compression. This is also true for water-soluble drug substances, resulting in tablets unsatisfactory as to content uniformity. Migration can be reduced by drying the granulation slowly at low temperatures or using a granulation in which the major diluent is present as granules of large particle size. The presence of microcrystalline cellulose in wet granulations also reduces migration tendencies. After drying, the granulation is reduced in particle size, by passing it through a smaller-mesh screen. Following dry screening, the granule size tends to be more uniform. For dry granulations, the screen size selected depends on the diameter of the punch. The following sizes are suggested:



Figure 30-9. Hammers. (Courtesy of Fitzpatrick Co.)

1. Tablets up to 3/16 inch diameter, use 20-mesh
2. Tablets 7/32 to 5/16 inch, use 16-mesh
3. Tablets 11/32 to 13/32 inch, use 14-mesh
4. Tablets 7/16 inch and larger, use 12-mesh

For small amounts of granulation, hand screens may be used and the material passed through with the aid of a stainless steel spatula. With larger quantities, any of the comminuting mills with screens corresponding to those just mentioned may be used. Note that the smaller the tablet, the finer the dry granulation to enable more uniform filling of the die cavity; large granules give an irregular fill to a comparatively small die cavity. With compressed tablets of sodium bicarbonate, lactose, and magnesium trisilicate, a relationship has been demonstrated between the particle size of the granulated material and the disintegration time and capping of the resultant tablets. For a sulfathiazole granulation, however, the particle-size distribution did not appear to influence hardness or disintegration. After dry granulation, the lubricant is added as a fine powder.

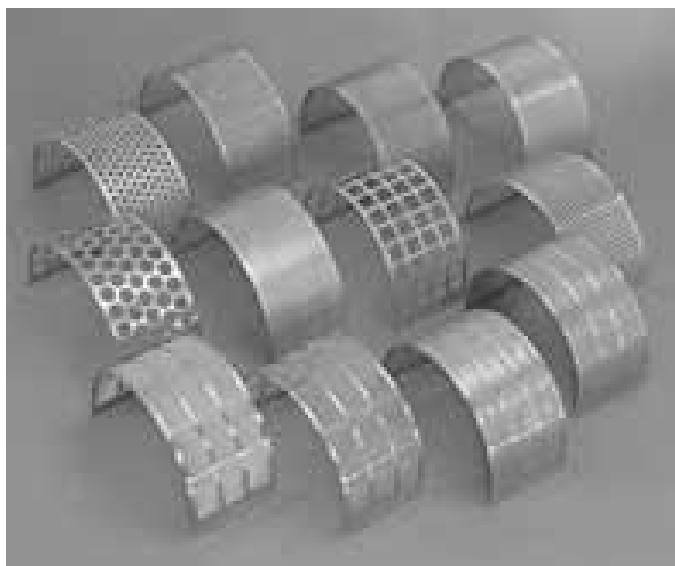


Figure 30-10. Screens. (Courtesy of Fitzpatrick Co.)

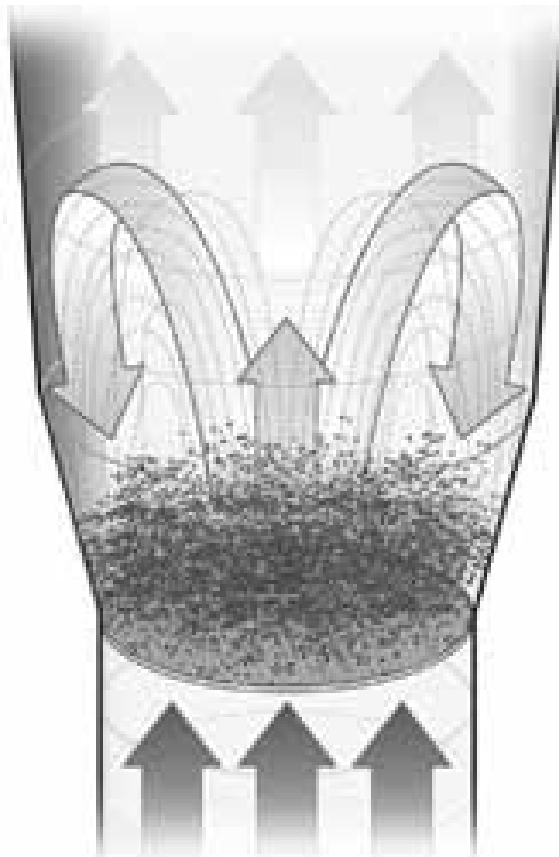


Figure 30-11. Schematic Fluid bed Dryer. (Courtesy of Glatt.)

It is screened onto the granulation through 60- or 100-mesh nylon cloth to eliminate small lumps, as well as to increase the covering power of the lubricant. As it is desirable for each granule to be covered with the lubricant, the lubricant is blended with the granulation very gently, preferably in a blender using a tumbling action, such as the Patterson Kelly Twin Shell Blender (Fig. 30-3). Gentle action is desired to maintain the uniform granule size, resulting from the granulation step. It has been claimed that too much fine powder is not desirable, because fine powder may not feed into the die evenly; consequently, variations in weight and density result. Fine powders, commonly designated as "fines," also blowout around the upper punch and down past the lower punch, making it necessary to clean the machine frequently. Fines, however, at a level of 10–20%, are, traditionally, sought by the tablet formulator. The presence of some fines is necessary for the proper filling of the die cavity. Now, even higher concentrations of fines are used successfully in tablet manufacture. Most investigators agree that no general limits exist for the amount of fines that can be present in a granulation; it must be determined for each specific formula. Many formulators once believed (and some still believe) that over blending resulted in an increased amount of fines and, hence, caused air entrapment in the formula. The capping and laminating of tablets associated with over blending lubricants was thought to be caused by these air pockets. Most scientists now recognize that a more plausible explanation has to do with the function of the lubricants themselves. Since the very nature of a lubricant tends to make surfaces less susceptible to adhesion, over blending prevents the intergranular bonding that takes place during compaction.

Fluid-Bed Granulation

A relatively new method for granulating evolved from the fluid-bed drying technology previously described. The concept was to spray a granulating solution onto the suspended particles,

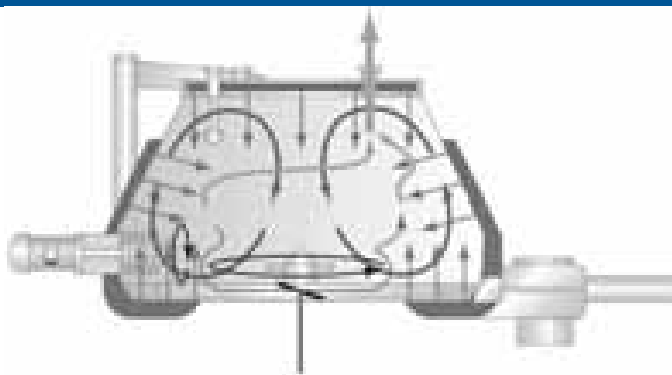


Figure 30-12. Single Pot Vertical Granulator. (Courtesy of Glatt.)

which then would be dried rapidly in the suspending air. The main benefit from this system is the rapid granulation and drying of a batch. The main firms that developed this technology are Glatt, Aeromatic (now GEA), Vector, and Hüttlin. The general design of these systems is the same with most companies (Fig. 30-13–30-16). In this method, particles of an inert material or the active drug are suspended in a vertical column with a rising air stream; while the particles are suspended, the common granulating materials in solution are sprayed into the column. There is a gradual particle buildup under a controlled set of conditions, resulting in a tablet granulation ready for compression after the addition of the lubricant. An obvious advantage exists, since granulating and drying can take place in a single piece of equipment. It should be noted, however, that many of the mixers discussed previously can be supplied with a steam jacket and vacuum and can provide the same advantage. In these systems, a granulating solution or solvent is sprayed into or onto the bed of suspended particles. The rate of addition of the binder, temperature in the bed of particles, temperature of the air, volume, and moisture of the air all play an important role in the quality and performance of the final product. Many scientists feel that this method is an extension of the wet-granulation method, as it incorporates many of its concepts. However, anyone who has developed a formulation in

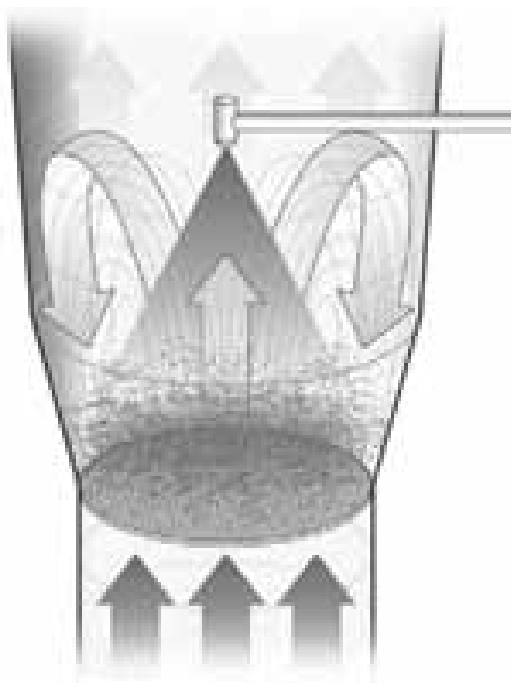


Figure 30-13. Fluid Bed Top Spray Granulator. (Courtesy of Glatt.)

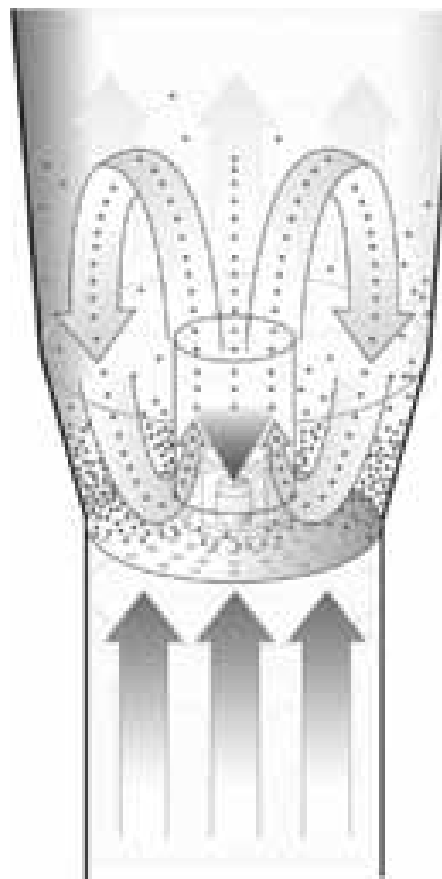


Figure 30-14. Bottom Spray Fluid Bed. (Courtesy of Glatt.)

a fluid-bed system knows that the many operating parameters involved make it somewhat more complex. In addition to its use for the preparation of tablet granulations, this technique has also been proposed for the coating of solid particles, as a means of improving the flow properties of small particles. Researchers have observed that, in general, fluid-bed granulation yields a less dense particle than conventional methods, and this can affect subsequent compression behavior. The Merck facility at Elkton, VA, was the first completely automated tablet production facility in the world. The entire tablet-manufacturing process, based on a wet-granulation method, was computer-controlled. The system weighed the ingredients, blended, granulated, dried, and lubricated to prepare a uniform granulation of specified particle size and particle-size distribution. The computer directed the compression of the material into tablets with exacting specifications for thickness, weight, and hardness. After compression, the tablets were coated with a water-based film coating. The computer controlled and monitored all flow of material. The plant represented the first totally automated pharmaceutical manufacturing facility. However, due to shifting market trends, the burdens of process validation, and changes to processes, totally automated processes are, generally, not used today. Instead, many production operations focus on computer-controlled and monitored unit operations, such as seen in various tableting machines and granulators. Equipment suppliers work closely with individual pharmaceutical companies in designing specialized and unique systems. Newer developments from the machine manufacturing industries improved the process efficiency by modifying the air distribution plate and introducing the new Diskjet with a tangential air exit, which ensures optimum exchange of materials and energy, as well as reduced process time for drying, granulating, and coating (Fig. 30-17). Total containment systems are available from Glatt, GEA, and Hüttlin. Thus, a high

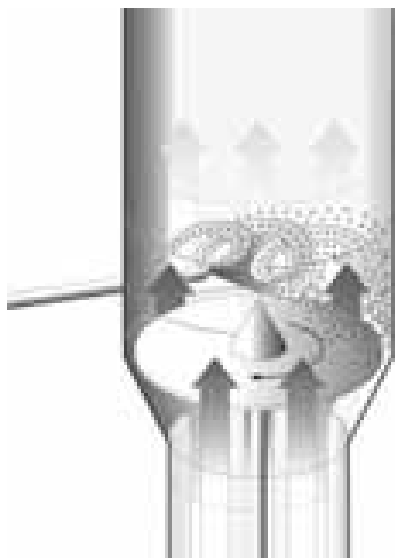


Figure 30-15. Tangential Spray Fluid Bed. (Courtesy of Glatt.)

shear mixer is connected to a cone mill and then a fluid bed dryer. Charging and discharging take place under totally dust and contamination free conditions (Fig. 30-18).

Dry Granulation

When tablet ingredients are sensitive to moisture or are unable to withstand elevated temperatures during drying and when the tablet ingredients have sufficient inherent binding or cohesive properties, slugging may be used to form granules. This method is referred to as dry granulation, precompression, or double-compression. It eliminates a number of steps but still includes weighing, mixing, slugging, dry screening, lubrication, and compression. The active ingredient, diluent (if required), and part of the lubricant are blended. One of the constituents, either the active ingredient or the diluent, must have cohesive properties. Powdered material contains a considerable amount of air; under pressure, this air is expelled, and a fairly dense piece is formed. The more time allowed for this air to escape, the better the tablet or slug. When slugging is used, large tablets are made as slugs, because fine powders flow better into large cavities. Also, producing large slugs decreases production time: 7/8-1 in are the most practical sizes for slugs. Sometimes, to obtain the pressure desired, the slug sizes are reduced to 3/4 in. The punches should be flat-faced. The compressed slugs are com-



Figure 30-16. Fluid-Bed. (Courtesy of Hüttlin GmbH, A Bosch Packaging Technology Company.)



Figure 30-17. Disk Jet. (Courtesy of Hüttlin GmbH, A Bosch Packaging Technology Company.)

minuted through the desirable mesh screen either by hand or, for larger quantities, through the Fitzpatrick or similar comminuting mill. The lubricant remaining is added to the granulation and blended gently, and the material is compressed into tablets. Aspirin is a good example of where slugging is satisfactory. Other materials, such as aspirin combinations, acetaminophen, thiamine hydrochloride, ascorbic acid, magnesium hydroxide, and other antacid compounds, may be treated similarly. Results comparable to those accomplished by the slugging process are also obtained with compacting mills. In the compaction method, the powder densified passes between high-pressure rollers that compress the powder and remove the air. The densified material is reduced to a uniform granule size and compressed into tablets after the addition of a lubricant. Excessive pressures that may be required to obtain cohesion of certain materials may result in a prolonged dissolution rate. Compaction mills available include the Chilsonator (Fitzpatrick), Roller Compactor (Vector), and the Compactor Mill (Allis-Chalmers; Fig. 30-19).

Direct Compression

As its name implies, direct compression consists of compressing tablets directly from component materials, without modifying the physical nature of the materials themselves. Formerly, direct compression, as a method of tablet manufacture, was reserved for a small group of crystalline chemicals having all the physical characteristics required for the formation of a good tablet. This group includes chemicals, such as potassium salts (chlorate, chloride, bromide, iodide, nitrate, and permanganate), ammonium chloride, and methenamine. These materials possess cohesive and flow properties that make direct compression possible. Since the pharmaceutical industry is constantly making efforts to increase the efficiency of tableting operations and reducing costs by using the smallest amount of floor space and labor as possible for a given operation, increasing attention is given to this method of tablet preparation. Approaches used to make this method more universally applicable include the introduction of formulation additives capable of imparting the characteristics required for compression and the use of force-feeding devices to improve the flow of powder blends.

For tablets in which the drug itself constitutes a major portion of the total tablet weight, it is necessary that the drug possess those physical characteristics required for the formulation to be compressed directly. Direct compression for tablets containing 25% or less of drug substances can frequently be used by formulating with a suitable diluent that acts as a carrier or

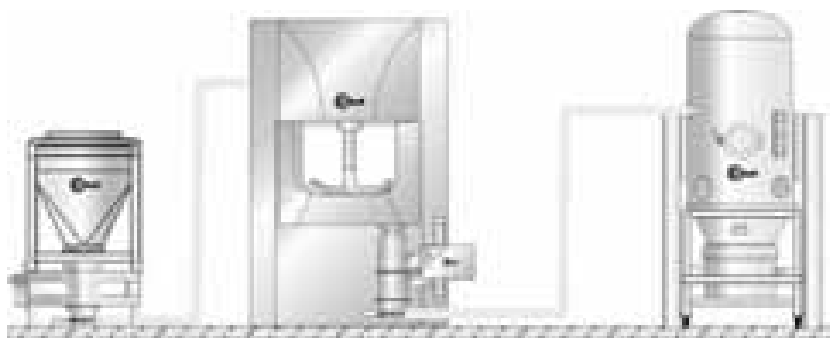


Figure 30-18. Horizontal granulation line consisting of a Top Driven Granulator and a Fluid bed Dryer. (Courtesy of Glatt.)

vehicle for the drug. Direct-compression vehicles or carriers must have good flow and compressible characteristics. These properties are imparted to them by a preprocessing step, such as wet granulation, slugging, spray-drying, spheronization, or crystallization. These vehicles include processed forms of most of the common diluents, including dicalcium phosphate dihydrate, tricalcium phosphate, calcium sulfate, anhydrous lactose, spray-dried lactose, pregelatinized starch, compressible sugar, mannitol, and microcrystalline cellulose. These commercially available, direct compression vehicles may contain small quantities of other ingredients (e.g., starch) as processing aids. Dicalcium phosphate dihydrate (Di-Tab, JRS), in its unmilled form, has good flow properties and compressibility. It is a white, crystalline agglomerate insoluble in water and alcohol. The chemical is odorless, tasteless, and nonhygroscopic. Since it has no inherent lubricating or disintegrating properties, other additives must be present to prepare a satisfactory formulation. Compressible sugar consists mainly of sucrose that is processed

to have properties suitable for direct compression. It also may contain small quantities of dextrin, starch, or invert sugar. It is a white crystalline powder with a sweet taste and complete water solubility. It requires the incorporation of a suitable lubricant at normal levels for lubricity. The sugar is used widely for chewable vitamin tablets because of its natural sweetness. One commercial source is Di-Pac (Amstar), prepared by the co crystallization of 97% sucrose and 3% dextrins. Some forms of lactose meet the requirements for a direct-compression vehicle. Hydrated lactose does not flow, and its use is limited to tablet formulations prepared by the wet-granulation method. Both anhydrous lactose and spray-dried lactose have good flowability and compressibility and can be used in direct compression, provided a suitable disintegrant and lubricant are present. Mannitol is a popular diluent for chewable tablets, due to its pleasant taste and mouth feel, resulting from its negative heat of solution. In its granular form (ICI Americas), it has good flow and compressible qualities. It has low moisture content and is not hygroscopic. The excipient that has been studied extensively as a direct compression vehicle is microcrystalline cellulose (Avicel, FMC). This non-fibrous form of cellulose is obtained by spray-drying washed, acid-treated cellulose and is available in several grades, ranging in average particle size from 20 to 250 μm . It is water-insoluble, but the material has the ability to draw fluid into a tablet by capillary action; it swells on contact and, thus, acts as a disintegrating agent. The material flows well and has a degree of self-lubricating qualities, thus, requiring a lower level of lubricant than other excipients. Recently, FMC introduced Avicel DG, which occurs as a white, odorless powder, containing 75% of microcrystalline cellulose and 25% anhydrous dibasic calcium phosphate. The wet dispersion and spray-drying of microcrystalline cellulose and anhydrous dibasic calcium phosphate results in an intimate physical combination, which cannot be achieved by traditional dry blending. JRS PHARMA introduced PROSOLV, which is coprocessed silicified microcrystalline cellulose. JRS PHARMA also introduced to the pharmaceutical excipient market PROSOLV EASY TAB, which is composed of coprocessed microcrystalline cellulose, colloidal silicon dioxide, sodium starch Glycolate, and sodium stearyl fumarate, which is claimed to be the ideal direct compression complete excipient. Forced-flow feeders are mechanical devices, available from pharmaceutical equipment manufacturers, designed to deaerate light and bulky material. Mechanically, they maintain a steady flow of powder moving into the die cavities under moderate pressure. By increasing the density of the powder, higher uniformity in tablet weights is obtained (Fig. 30-20, 30-21). Recently, many companies have reversed their optimism for some direct-compression systems. Some formulations made by direct compression were not as forgiving as the older, wet-granulated products were. As raw material variations occurred, especially with the drug, many companies found themselves with poorly compactable formulations. Interest in direct compression is also stimulating basic research on the flowability of powders with and without additives.

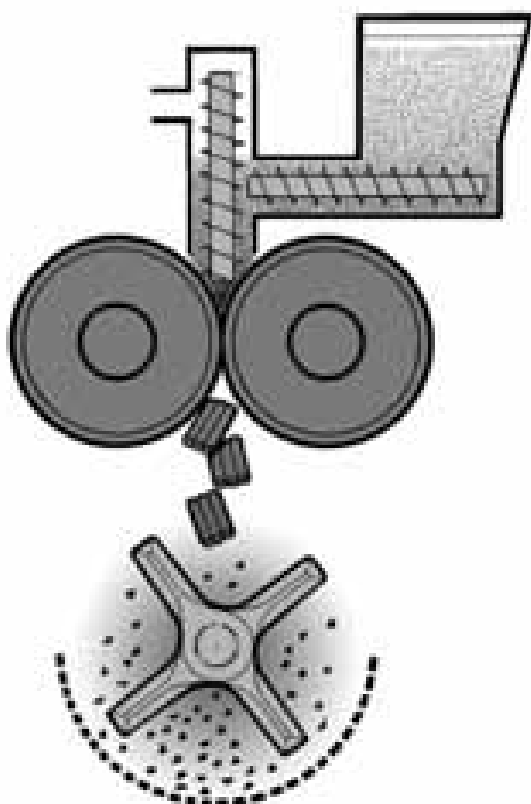


Figure 30-19. Roller Compactor Granulator. (Courtesy of Fitzpatrick Co.)



Figure 30-20. Forced Flow Feeder. (Courtesy of Fette.)

Related Granulation Processes

Spheronization—Spheronization, a form of pelletization, refers to the formation of spherical particles from wet granulations. Since the particles are round, they have good flow properties when dried. They can be formulated to contain sufficient binder to impart cohesiveness for tableting. Spheronization equipment, such as the Marumerizer (Luwa) and the CFGranulator (Vector), are commercially available for small-scale manufacture, on up to commercial sized equipment. A wet granulation containing the drug substance, diluent (if required), and binder is passed, first, through an extruding machine to form rod-shaped cylindrical segments, ranging in diameter from 0.5 to 12 mm. The segment diameter and the size of the final spherical particle depend on the extruder screen size. After extrusion, the segments are placed into the Marumerizer, where they are shaped into spheres by centrifugal and frictional forces on a rotating plate. The pellets are then dried by conventional methods, mixed with suitable lubricants, and compressed into tablets or used as capsule-fill material. Microcrystalline cellulose has been shown to be an effective diluent and binder in granulations to be spheronized. The advantages of the process include the production of granules, regular in shape, size, and surface characteristics; low friability, resulting in fewer fines and less dust; and the ability to regulate the size of the spheres within a narrow particle-size distribution. Spheres can also be produced by fluid-bed granulation techniques and other specialized equipment, such as the CFGranulator (Vector). These processes, however, must begin with crystals or nonpareil seeds followed by buildup. Exact results, such as sphere density, are different for the various methods and could be important in product performance. These processes can be run as batches or continuously.

Spray-Drying—A number of tableting additives suitable for direct compression have been prepared by the drying process known as spray-drying. The method consists of bringing together a highly dispersed liquid and a sufficient volume of hot air to produce evaporation and drying of the liquid droplets. The feed liquid may be a solution, slurry, emulsion, gel, or paste, provided it is pumpable and capable of being atomized.



Figure 30-21. Forced Flow Feeder. (Courtesy of Bosch Packaging Technology Ltd., Manesty.)

The feed is sprayed into a current of warm filtered air. The air supplies the heat for evaporation and conveys the dried product to the collector; the air is then exhausted with the moisture. As the liquid droplets present a large surface area to the warm air, local heat and transfer coefficients are high. The spray-dried powder particles are homogeneous, approximately spherical in shape, nearly uniform in size, and frequently hollow. The latter characteristic results in low bulk density, with a rapid rate of solution. Being uniform in size and spherical, the particles possess good flowability. The design and operation of the spray-dryer can vary many characteristics of the final product, such as particle size and size distribution, bulk and particle densities, porosity, moisture content, flowability, and friability. Among the spray-dried materials available for direct compression formulas are lactose, mannitol, and flour. Another application of the process in tableting is spray-drying the combination of tablet additives as the diluent, disintegrant, and binder. The spray-dried material is then blended with the active ingredient or drug, lubricated, and compressed directly into tablets. Since atomization of the feed results in a high surface area, the moisture evaporates rapidly. The evaporation keeps the product cool and, as a result, the method is applicable for drying heat-sensitive materials. Among heat-sensitive pharmaceuticals successfully spray-dried are the amino acids; antibiotics, such as aureomycin, bacitracin, penicillin, and streptomycin; ascorbic acid, cascara extracts, and liver extracts; pepsin and similar enzymes; protein hydrolysates; and thiamin. Frequently, spray-drying is more economical than other processes, since it produces a dry powder directly from a liquid and eliminates other processing steps as crystallization, precipitation, filtering or drying, particle-size reduction, and particle classifying. By the elimination of these steps, labor, equipment costs, space requirements and possible contamination of the product are reduced. Intrinsic factor concentrate obtained from hog mucosa was previously prepared by Lederle / American Home Products, using a salt-precipitation process, followed by freeze-drying. By using spray-drying, it was possible to manufacture a high-grade material by a continuous process. The spherical particles of the product facilitated its subsequent blending with vitamin B12. Similar efficiencies have been found in processes producing magnesium trisilicate and dihydroxyaluminum sodium carbonate; both chemicals are used widely in antacid preparations. Encapsulation of chemicals can also be achieved using spray-drying equipment. The process is useful in coating one material on another to protect the interior substance or to control the rate of its release. The substance to be coated can be either liquid or solid, but must be insoluble in a solution of the coating material. The oil-soluble vitamins A and D can be coated with a variety of materials, such as acacia gum to prevent their deterioration. Flavoring oils and synthetic flavors are coated to give the so-called dry flavors.

Spray-Congeaing—Also called spray-chilling, spray-congealing is a technique similar to spray-drying. It consists of melting solids and reducing them to beads or powder, by spraying the molten feed into a stream of air or other gas. The same basic equipment is used as with spray-drying, although no source of heat is required. Either ambient or cooled air is used, depending on the freezing point of the product. For example, mono-glycerides and similar materials are spray-congealed with air at 50°F. A closed-loop system with refrigeration cools and recycles the air. Using this process, drugs can be dissolved or suspended in a molten wax and spray-congealed; the resultant material can then be adapted for a prolonged-release form of the drug. Among the carbohydrates used in compressed tablets, mannitol is the only one that possesses high heat stability. Mannitol melts at 167 and, either alone or in combination with other carbohydrates, can be fused and spray-congealed. Selected drugs have been shown to be soluble in these fused mixtures, and the resultant spray-congealed material possesses excellent flow and compression characteristics.

TABLET MACHINES

As mentioned previously, the basic mechanical unit in tablet compression involves the operation of two steel punches within a steel die cavity. The tablet is formed by the pressure exerted on the granulation by the punches within the die cavity or cell. The tablet assumes the size and shape of the punches and dies used. Although round tablets are used more generally, oval, square, triangular or other irregular shapes and capsule-form may be used (Fig. 30-22). Likewise, the curvature of the faces of the punches determines the curvature of the tablets. Punch faces with ridges are used for compressed tablets scored for breaking into halves or fourths, although it has been indicated that variation among tablet halves is significantly greater than among intact tablets. However, a patented formulation for a tablet scored to form a groove that is one-third to two-thirds the depth of the total tablet thickness is claimed to give equal parts containing substantially equal amounts of the drug substance. Tablets, engraved or embossed with symbols or initials, require punches with faces embossed or engraved with the corresponding designs. The use of the tablet, sometimes, determines its shape; effervescent tablets are usually large, round, and flat, whereas vitamin tablets are frequently prepared in capsule-shaped forms. Tablets prepared using deep-cup punches are round and, when coated, take on the appearance of pills. Veterinary tablets often have a bolus shape and are much larger than those used in medical practice. The quality-control program for punches and dies, frequently referred to as tooling, instituted by large pharmaceutical companies, emphasizes the importance of their care in modern pharmaceutical production. To produce physically perfect, compressed tablets, an efficient punch-and-die program must be set up. Provisions for inspection of tooling, parameters for cost-per-product determination, product identification, and tooling specifications must all be considered. A committee of the Industrial and Pharmaceutical Technology Section of the American Pharmacists Association established a set of dimensional specifications and tolerances for standard punches and dies. Regardless of the size of the tableting operation, the attention that must be given to the proper care of punches and dies should be noted. They must be highly polished and kept free from rust and imperfections. In cases in which the material pits or abrades the dies, chromium-plated dies have been used. Dropping the punches on hard surfaces chips their fine edges. When the punches are in the machine, the upper and lower punches should not be allowed to contact each other; otherwise, a curling or flattening of the edges that is one of the causes of capping results. This is especially necessary to observe in the case of deep-cup punches. When the punches are removed from the machine, they should be washed thoroughly in warm, soapy water and dried well with a clean cloth. A coating of grease or oil should be rubbed over all parts of the dies and punches to protect them from the atmosphere. They should be stored carefully in boxes or paper tubes. Natoli wrote an excellent review



Figure 30-22. Punches and Dies with various shapes. (Courtesy of NATOLI.)

on tooling for pharmaceutical processing in which he addressed tooling options, standards, configurations, recent innovations, and production problems due to tooling. A significantly important part of the review is the listing of possible causes and corrective actions for tablet problems related to tooling.¹⁴

Single-Punch Machines

The simplest tableting machines available are those having the single-punch design. A number of models are available. Although most of these are power-driven, several hand-operated models are available. Compression is accomplished on a single-punch machine. The feed shoe filled with the granulation is positioned over the die cavity, which then fills. The feed shoe retracts and scrapes all excess granulation away from the die cavity. The upper punch lowers to compress the granulation within the die cavity. The upper punch retracts, and the lower punch rises to eject the tablet. As the feed shoe returns to fill the die cavity, it pushes the compressed tablet from the die platform. The weight of the tablet is determined by the volume of the die cavity; the lower punch is adjustable to increase or decrease the volume of granulation, thus, increasing or decreasing the weight of the tablet (Figs. 30-23, 30-24). For tablets having diameters larger than 1/2 inch, sturdier models are required. This is also true for tablets requiring a high degree of hardness, as in the case of compressed lozenges. The heavier models are capable of much higher pressures and are suitable for slugging.

Operation of Single-Punch Machines—In installing punches and dies in a single-punch machine, insert the lower punch first, by lining up the notched groove on the punch with the lower punch set screw and slipping it into the smaller bore in the die table; the setscrew is not tightened yet. The lower punch is differentiated from the upper punch in that it has a collar



Figure 30-23. Manufacture of 1, 2 or 3 layer tablets (Courtesy of Bosch Packaging Technology Ltd., Manesty.)

around the punch head. Slip the die over the punch head so the notched groove (with the widest area at the top) lines up with the die setscrew. Tighten the lower punch setscrew after seating the lower punch, by pressing on the punch with the thumb. Tighten the die setscrew, making certain the surface of the die is flush with the die table. Insert the upper punch, again lining up the grooved notch with the upper punch setscrew. To be certain that the upper punch is seated securely, turn the machine over by hand, with a block of soft wood or wad of cloth between the upper and lower punches. When the punch is seated, tighten the upper punch setscrew. Adjust the pressure so the upper and lower punches will not come in contact with each other when the machine is turned over. Adjust the lower punch so it is flush with the die table at the ejection point. Install the feed shoe and hopper. After adding a small amount of granulation to the hopper, turn the machine over by hand and adjust the pressure until a tablet is formed. Adjust the tablet weight until the desired weight is obtained. The pressure will have to be altered concurrently with the weight adjustments. Remembered, as the fill is increased, the lower punch moves farther away from the upper punch, and more pressure has to be applied to obtain comparable hardness. Conversely, when the fill is decreased, the pressure has to be decreased. When all the adjustments have been made, fill the hopper with granulation and turn on the motor. Hardness and weight should be checked immediately, and suitable adjustments made if necessary. Periodic checks should be made on the tablet hardness and weight during the running of the batch, at 15- to 30-minute intervals. When the batch has been run off, turn off the power and remove loose dust and granulation with the vacuum cleaner. Release the pressure from the punches. Remove the feed hopper and the feed shoe. Remove the upper punch, the lower punch, and the die. Clean all surfaces of the tablet machine, and dry well with clean cloth. Cover surfaces with thin coating of grease or oil prior to storage. As tablets are ejected from the machine, after compression, they are usually accompanied by powder and uncompressed granulation. To remove this loose dust, the tablets are passed over a screen, which may be vibrating, and cleaned with a vacuum line.

Rotary Tablet Machines

For increased production, rotary machines offer great advantages. A head carrying a number of sets of punches and dies revolves continuously, while the tablet granulation runs from the hopper, through a feed frame, and into the dies placed in a large, steel plate revolving under it. This method promotes a uniform fill of the die and, therefore, an accurate weight for the tablet. Compression takes place as the upper and lower punches pass between a pair of rollers, as can be seen in Figure 30-25. This action produces a slow squeezing effect on the material in the



Figure 30-24. Modern Single Punch Tablet Press. (Courtesy of Bosch Packaging Technology Ltd., Manesty.)

die cavity from the top and bottom, and so gives a chance for the entrapped air to escape. The lower punch lifts up and ejects the tablet. Adjustments for tablet weight and hardness can be made, without the use of tools, while the machine is in operation. One of the factors that contributes to the variation in tablet weight and hardness during compression is the internal flow of the granulation within the feed hopper. On most rotary machine models, there is an excess pressure release that cushions each compression and relieves the machine of all shocks and undue strain. The punches and dies can be readily removed for inspection, cleaning, and to insert different sets to produce a great variety of sizes and shapes. Most rotary tablet presses measure a tablet's weight either by measuring its variation in tablet height at precompression or by the force at main compression. They compress a volume of granules, captured in a die, between two rollers, using an upper and lower punch. By changing the distance between the rollers, one adjusts the force used to compress tablets. Once the distance between the rollers is set, the compression force will stay the same. Several factors affect the precise amount of granules captured within each die. For example, granule size, size distribution, and variation in punch length can have an effect. In addition, an excessively high rotational speed allows the granules insufficient time to fall into the die. Variations in the amount of granules in each die result in tablets of different weights and densities. An inconsistency in the maximum compression force can result in inconsistent tablet properties, thus, affecting the efficacy of the dosage form.

Compression to equal force (EF) is a new concept that allows tablets to be compressed at the same peak compression force, independent of tablet weight. This method relies on the use of an air piston (Fig. 30-26). The air piston is installed at the precompression and main compression station and allows the pre- and main compression rollers to be floating rather than fixed. The piston moves vertically in a cylinder filled with compressed air. The air pressure in the cylinder is preset and

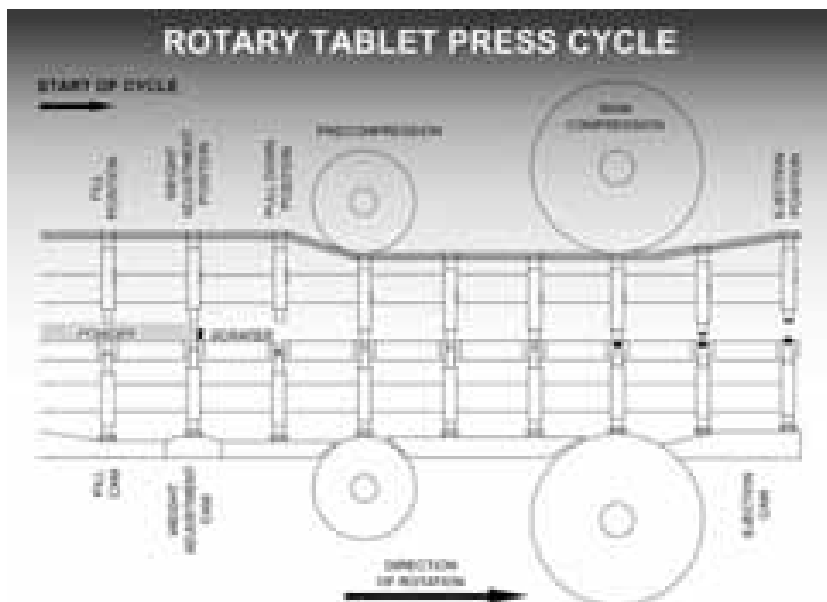


Figure 30-25. Rotary Tablet Press Cycle. (Courtesy of Natoli.)

kept constant by a pressure-regulating valve and an expansion chamber. Because the surface of the cylinder and the air pressure are constant, the force is also constant, irrespectively of the position of the floating roller. If the air pressure in the cylinder is such that this maximum compression force is higher than the actual compression force, the system compresses tablets to equal thickness, leading to high variation in tablet hardness. Conversely, the press can be set to make the compression roller move up at each compression. In this arrangement, all tablets are compressed at the same peak force, providing a much reduced variation in tablet hardness. The tablet press is equipped with a linear variable displacement transducer sensor that measures compression-roller movement accurately. A heavy tablet causes the compression roller to move more, and a light tablet causes less movement. This provides a tablet weight control method, called "tablet weight control by displacement," with a linear relationship between tablet weight and displacement. Accordingly, more accurate tablet weight for smaller tablets than the conventional tablet weight control by force is assured. Floating compression rollers also provide an effective means to increase dwell at high turret revolutions. Increasing the dwell time at pre- and main compression and at high turret speed is necessary to improve deaeration and uniform distribution of compaction stress in the die prior to final compaction at high tablet output. This provides reduced risk of capping, better bonding of layers, in the case of bilayer tablets, and improved tablet tensile strength. When compressed to equal force, variations in tablet weight do not affect tensile strength. Therefore, tablets maintain a consistent clinical efficacy. Equal force tableting is also of interest when compacting coated pellets or when producing of Orally Disintegrating Tablets.

Operation of Rotary Machines—Before inserting punches and dies, make certain the pressure has been released from the pressure wheel. The die holes should be cleaned thoroughly, making certain the die seat is completely free of any foreign materials. Back off all die locks, and loosely insert dies into the die holes, then tap each die securely into place, with a fiber of soft metal rod through the upper punch holes. After all the dies have been tapped into place, tighten each die locks crew progressively and securely. As each screw is tightened, the die is checked to see it does not project above the die table. Insert the lower punches through the hole made available, by removing the punch head. Turn the machine by hand until the punch bore coincides with the plug hole. Insert each lower

punch in its place progressively. Insert the upper punches by dropping them into place in the head. Each punch (upper and lower) should be coated with a thin film of mineral oil before insertion into the machine. Adjust the ejection cam, so the lower punch is flush with the die table at the ejection point. After insertion of the punches and dies, adjust the machine for the tablet weight and hardness. The feed frame should be attached to the machine along with the feed hopper. Add a small amount of the granulation through the hopper and turn the machine over by hand. Increase the pressure by rotating the pressure wheel until a tablet is formed. Check the weight of the tablet, and adjust the fill to provide the desired tablet weight. Most likely, more than one adjustment of the fill will be necessary before obtaining the acceptable weight. When the fill is decreased, the pressure must be decreased to provide the same hardness in the tablet. Conversely, when the fill is increased, the pressure must be increased to obtain comparable hardness. Fill the hopper with the granulation and turn on the power. Check tablet weight and hardness immediately after the mechanical operation begins, and make suitable adjustments, if necessary. Check these properties routinely and regularly at 15- to 30-minute intervals, while the machine is in operation. When the batch has been run, turn off the power. Remove the hopper and feed frame from the machine. Remove loose granulation and dust with a vacuum line. Remove all pressure from the wheel. Remove the punches and dies in the reverse order of that used in setting up the machine. First, remove the upper punches individually, then the lower punches, and finally the dies. Wash each punch and die in alcohol, and brush with a soft brush to remove adhering material. Dry them with a clean cloth, and cover them with a thin coating of grease or oil before storing.

High-Speed Rotary Tablet Machines

The rotary tablet machine has evolved, gradually, into models capable of compressing tablets at high production rates (see Figs. 30-27, 30-28). This has been accomplished by increasing the number of stations (i.e., sets of punches and dies) in each revolution of the machine head, improving feeding devices, and on some models installing dual compression points. Rotary machines with dual compression points are referred to as double rotary machines, and those with one compression point, single rotary. One-half of the tablets are produced 180° from the tablet chute. They travel outside the perimeter and

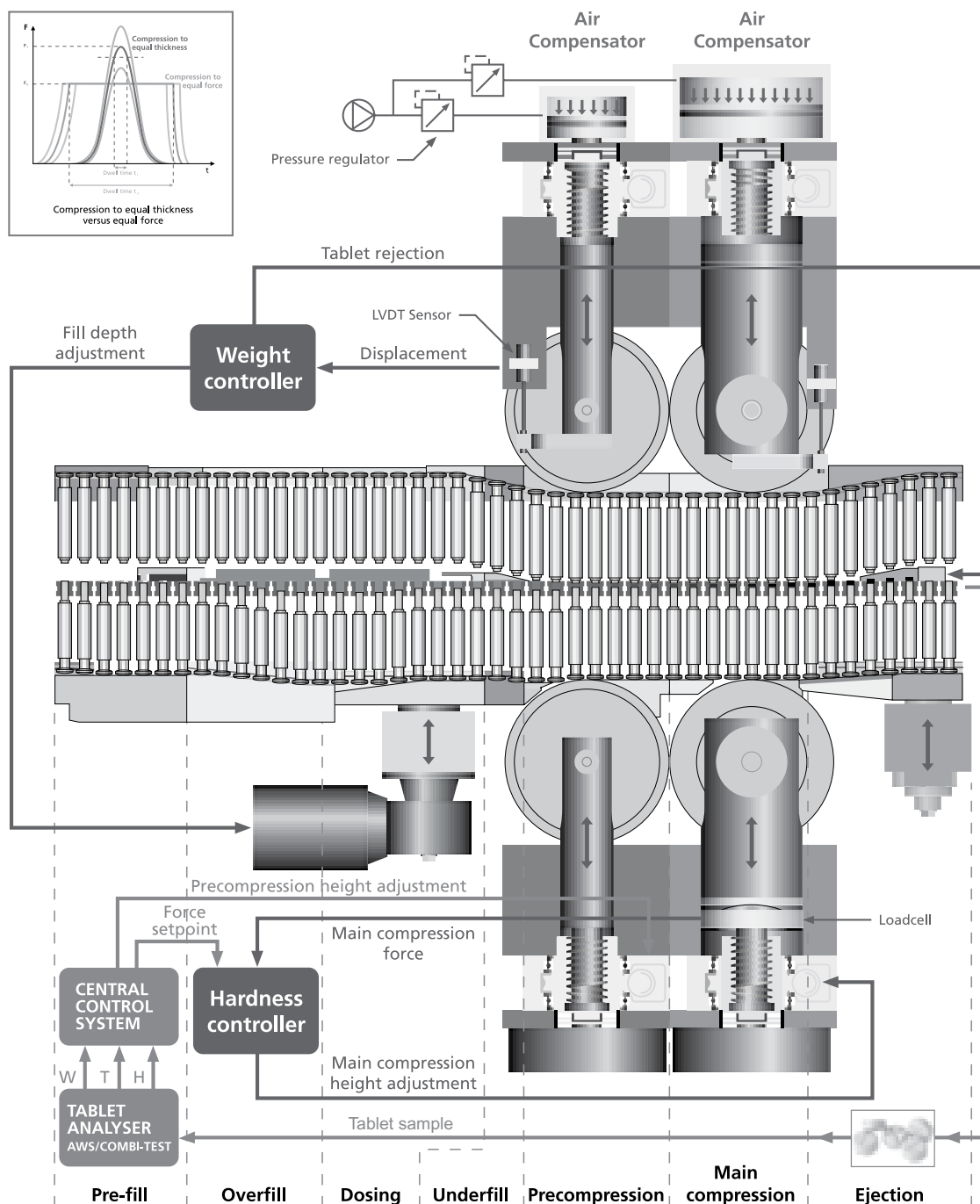


Figure 30-26. Rotary Tablet Compression Cycle. (Courtesy of Courtoy.)

discharge with the second tablet production. Although these models are mechanically capable of operating at production rates of more than 5000 tablets per minute, the actual speed still depends on the physical characteristics of the tablet granulation and the rate that is consistent with compressed tablets having satisfactory physical characteristics. The main difficulty in rapid machine operation is ensuring adequate filling of the dies. With rapid filling, dwell time of the die cavity beneath the feed frame is insufficient to ensure the requirements of uniform flow and packing of the dies. Various methods of force-feeding the granulation into the dies have been devised to refill the dies in the very short dwell time permitted on the high-speed machine. Presses with pre-compression rollers permit the partial compaction of material before final compaction. This provides for partial deaeration and particle

orientation of material before final compression. This helps in the direct compacting of materials and reduces laminating and capping due to entrapped air.

Multilayer Rotary Tablet Machines

Rotary tablet machines have also been developed into models capable of producing multiple-layer tablets; the machines are able to make 1-, 2-, or 3-layer tablets (Oystar Manesty). Stratified tablets offer a number of advantages. Incompatible drugs can be formed into a single tablet by separating the layers containing them with a layer of inert material. It has permitted the formulation of time-delay medication and offers a wide variety of possibilities in developing color combinations that give the products identity. Originally, the tablets were prepared by a single-compression method. The dies were filled



Figure 30-27. Modern Rotary Tablet Press. (Courtesy of Courtoy.)

with the different granulations in successive layers, and the tablet was formed by a single compression stroke. The separation lines of the tablets prepared by this method tended to be irregular. In the machines now available for multilayer produc-



Figure 30-28. Modern Tablet Press. (Courtesy of FETTE.)

tion, the granulation receives a precompression stroke after the first and second fill, which lightly compacts the granulation and maintains a well-defined surface of separation between each layer. The operator is able to eject either precompressed layer, with the machine running, at any desired speed, for periodic weight and analysis checks. Other multiple-compression presses can receive previously compressed tablets and compress another granulation around the preformed tablet. An example of a press with this capability is the Manesty Drycota (Oystar Manesty). Pressure coated tablets can be used to separate incompatible drug substances and to give an enteric coating to the core tablets.

Capping and Splitting of Tablets

The splitting or capping of tablets is one of great concern and annoyance in tablet making. It is quite difficult to detect, while the tablets are processed, but can be detected easily by vigorously shaking a few in the cupped hands. A slightly chipped tablet does not necessarily mean that the tablet will cap or split. There are many factors that may cause a tablet to cap or split: excess fines or powder, which traps air in the tablet mixture; deep markings on tablet punches. Many designs or "scores" on punches are too broad and deep. Hairline markings are just as appropriate as deep, heavy markings: worn and imperfect punches. Punches should be smooth and buffed. Nicked punches often cause capping. The development of fine feather edges on tablets indicates wear on punches: worn dies. Dies should be replaced or reversed. Dies that are chrome plated or have tungsten carbide inserts wear longer and give better results than ordinary steel dies; too much pressure. By reducing the pressure on the machines, the condition may be corrected; unsuitable formula. It may be necessary to change the formula; moist and soft granulation. This type of granulation will not flow freely into the dies, thus, giving uneven weights and soft or capped tablets; poorly machined punches. Uneven punches are detrimental to the tablet machine itself and will not produce tablets of accurate weight. One punch out of alignment may cause one tablet to split or cap on every revolution.

Instrumented Tablet Presses

Compressional and ejectional forces involved in tablet compression can be studied by attaching strain gauges to the punches and other press components involved in compression. The electrical output of the gauges has been monitored by telemetry or use of a dual-beam oscilloscope equipped with camera. Instrumentation permits a study of the compaction characteristics of granulations, their flowabilities, and the effect of formulation additives, such as lubricants, as well as differences in tablet press design. Physical characteristics of tablets, such as hardness, friability, disintegration time, and dissolution rate, are influenced not only by the nature of the formulation, but by the compressional force as well. The rate and duration of compaction forces can be quantified. The rate of force application has a profound effect on powder consolidation within the die and, hence, efficiency of packing and powder compaction. The rate of release of force, or decompression, has a direct effect on the ability of the tablet to withstand relaxation. A prominent hypothesis, fostered by suggestion that capping and laminating of tablets is caused by too-rapid stress relaxation or decompression. This explains why slowing a tablet press and using tapered dies is useful in such situations. Most prominent pharmaceutical scientists have embraced this theory and have largely discounted air entrapment as a cause of capping and laminating. As compaction force rises, the steel tooling actually compresses in accommodation to the forces applied. The forces used to produce a tablet are considerable and should be monitored and understood. Therefore, definition of the compressional force and duration of force (dwell time), giving a satisfactory tablet for a formulation, provides an in-process control for obtaining both tablet-to-tablet and lot-to-lot uniformity. Instrumentation

has led to the development of on-line, automatic, electromechanical tablet weight-control systems capable of continuously monitoring the weights of tablets as they are produced. Most commercial presses today can be delivered with some sort of instrumentation attached. When tablet weights vary from preset limits, the monitor will automatically adjust the weight control mechanism to re-establish weights within acceptable limits. If the difficulty continues, the unit will activate an audible warning signal or an optional shut-down relay on the press. Most production model tablet presses come equipped with complete instrumentation (optional) and with options for statistical analysis and print out of compression/ejection signals. The techniques and applications of press instrumentation have been reviewed.¹⁵

Contamination

Although good manufacturing practices used by the pharmaceutical industry for many years have stressed the importance of cleanliness of equipment and facilities for the manufacture of drug products, the penicillin contamination problem resulted in renewed emphasis on this aspect of manufacturing. Penicillin, as either an airborne dust or residual quantities remaining in equipment, is believed to have contaminated unrelated products in sufficient concentrations to cause allergic reactions in individuals hypersensitive to penicillin. This resulted in the industry spending millions of dollars to change or modify buildings, manufacturing processes, equipment, and standard operating procedures to eliminate penicillin contamination.

With this problem has come renewed emphasis on the dust problem, material handling, and equipment cleaning in dealing with drugs, especially potent chemicals. Any process using chemicals in powder form can be a dusty operation; the preparation of compressed tablets and encapsulation fall in this category. In the design of tablet presses attention is being given to the control and elimination of dust generated in the tableting process. In modern presses, shown in Figures 30-27 and 30-28, the pressing compartment is completely sealed off from the outside environment, making cross-contamination nearly impossible. The pressing compartment can be kept dust-free by the air supply and vacuum equipment developed for the machine. It removes airborne dust and granular particles that have not been compressed, thus, keeping the circular pressing compartment and the upper and lower punch guides free of dust. Drug manufacturers have the responsibility to make certain that micro-organisms present in finished products are unlikely to cause harm to the patient and will not be deleterious to the product. An outbreak of *Salmonella* infections in Scandinavian countries was traced to thyroid tablets prepared from contaminated thyroid powder. This concern, eventually, led to the establishment of microbial limits for raw materials of animal or botanical origin, especially those that readily support microbial growth and are not rendered sterile during subsequent processing. Harmful micro-organisms, when present in oral products, include *Salmonella* spp, *Escherichia coli*, and certain *Pseudomonas* spp, such as *P. aeruginosa*, and *Staphylococcus aureus*. The compendia have microbial limits on raw materials, such as aluminum hydroxide gel, cornstarch, thyroid, acacia, and gelatin. These represent examples of the industry's efforts to conform to the intent of current good manufacturing practice, as defined by the FDA.

Molded Tablets or Tablet Triturates (TT)

Tablet triturates are small, discoid masses of molded powders weighing 30–250 mg each. The base consists of lactose, B lactose, mannitol, dextrose, or other rapidly soluble materials. It is desirable in making tablet triturates to prepare a solid dosage form that is rapidly soluble; as a result, they are generally softer than compressed tablets. This type of dosage form

is selected for a number of drugs, due to its rapidly dissolving characteristic. Nitroglycerin in many concentrations is prepared in tablet triturate form, since the molded tablet rapidly dissolves, when administered by placing under the tongue. Potent alkaloids and highly toxic drugs, used in small doses, are prepared as tablet triturates that can serve as dispensing tablets used as the source of the drug in compounding other formulations or solutions. Narcotics, in the form of hypodermic tablets, were originally made as tablet triturates, because they rapidly dissolve in sterile water for injection prior to administration. Today, with stable injections of narcotics available, there is no longer any justification for their use in this manner. Although many hypodermic tablets are currently made, they are used primarily for oral administration. Tablet triturates are made by forcing a moistened blend of the drug and diluent into a mold, extruding the formed mass, which is allowed to dry. This method is essentially the same as it was when introduced by Fuller in 1878. Hand molds may vary in size, but the method of operation is essentially the same. Molds consist of two plates made from polystyrene plastic, hard rubber, nickel-plated brass, or stainless steel. The mold plate contains 50–500 carefully polished perforations. The other plate is fitted with a corresponding number of projecting pegs or punches that fit the perforations in the mold plate. The mold plate is placed on a flat surface, the moistened mass is forced into the perforations, and the excess is scraped from the top surface. The mold plate is placed over the plate with the corresponding pegs and lowered. As the plates come together, the pegs force the tablet triturates from the molds. They remain on the tops of the pegs until dry, and they can be handled. In some hand molds, the pegs are forced down onto the plate holding the moist trituration.

TABLET CHARACTERISTICS

Compressed tablets may be characterized or described by a number of specifications. These include the diameter size, shape, thickness, weight, hardness, disintegration time, and dissolution characteristics. The diameter and shape depend on the die and the punches selected for the compression of the tablet. Generally, tablets are discoid in shape, although they may be oval, oblong, round, cylindrical, or triangular. Their upper and lower surfaces may be flat, round, concave, or convex to various degrees. The concave punches, used to prepare convex tablets, are referred to as shallow, standard, and deep cup, depending on the degree of concavity. The tablets may be scored in halves or quadrants to facilitate breaking, if a smaller dose are desired. The top or lower surface may be embossed or engraved with a symbol or letters that serve as an additional means of identifying the source of the tablets. These characteristics, along with the color of the tablets, tend to make them distinctive and identifiable with the



Figure 30-29. Electric Tablet Hardness Tester. (Courtesy of Erweka.)



Figure 30-30. Electric Hardness Tester. (Courtesy of Key.)

active ingredient that they contain. The remaining specifications assure the manufacturer that the tablets do not vary from one production lot to another. In the case of new tablet formulations their therapeutic efficacy is demonstrated through clinical trials, and it is the manufacturer's aim to reproduce the same tablet with the exact characteristics of the tablets that were used in the clinical evaluation of the dosage form. Therefore, from the control viewpoint these specifications are important for reasons other than physical appearance.

Tablet Hardness or Crushing Strength

The resistance of the tablet to chipping, abrasion, or breakage under conditions of storage, transportation, and handling, before usage, depends on its hardness. In the past, a rule of thumb described a tablet to be of proper hardness if it was firm enough to break with a sharp snap, when it was held between the second and third fingers and using the thumb as the fulcrum, yet didn't break when it fell on the floor. For obvious reasons and control purposes, a number of attempts have been made to quantitate the degree of hardness. A small and portable hardness tester was manufactured and introduced in the mid-1930s by Monsanto. It now is distributed by the Stokes Div (Pennwalt) and may be designated as either the Monsanto or Stokes hardness tester. The instrument measures the force required to break the tablet, when the force generated by a coil spring is applied diametrically to the tablet. The force is measured in kilograms. The Strong-Cobb hardness tester, introduced in 1950, also measures the



Figure 30-31. Tablet Friability Testers. (Courtesy of Erweka.)



Figure 30-32. Tablet Thickness Tester. (Courtesy of Starrett.)

diametrically applied force required to break the tablet. In this instrument, the force is produced by a manually operated air pump. As the pressure is increased, a plunger is forced against the tablet placed on anvil. The final breaking point is indicated on a dial, calibrated into 30 arbitrary units. The hardness values of the Stokes and Strong-Cobb instruments are not equivalent. Another instrument is the Pfizer hardness tester, which operates on the same mechanical principle as ordinary pliers. The force required to break the tablet is recorded on a dial and expressed in either kilograms or pounds of force. Currently, the most widely used apparatuses to measure tablet hardness or crushing strength are the electrically operated equipment, which eliminates the operator variability. Newer equipment are also available with printers. Manufacturers, such as SOTAX, Key, Van Kel, Erweka, Dr. Schleuniger Pharmatron, and others, make electrically driven hardness testers (Figs. 30-29, 30-30). Hardness (or more appropriately, crushing strength) determinations are made throughout the tablet runs, to determine the need for pressure adjustments on the tableting machine. If the tablet is too hard, it may not disintegrate in the required period of time or meet the dissolution specification; if it is too soft, it will not withstand handling during subsequent processing, such as coating or packaging and shipping operations.

Friability

A tablet property related to hardness is friability, and the measurement is made by use of the Roche friabilator. Rather than a measure of the force required to crush a tablet, the instrument is designed to evaluate the ability of the tablet to withstand abrasion in packaging, handling, and shipping. A number of tablets are weighed and placed in the tumbling apparatus, where they are exposed to rolling and repeated shocks, resulting from free-falls within the apparatus. After a given number of rotations, the tablets are weighed, and the loss in weight indicates the ability of the tablets to withstand this type of wear (Fig. 30-31). The friability test is now included in the USP.

A similar approach is taken by many manufacturers, when they evaluate a new product in the new market package, by sending the package to distant points and back, using various methods of transportation. This is called a "shipping test." The condition of the product on its return indicates its ability to withstand transportation handling.

Tablet Thickness

The thickness of the tablet from production-run to production-run is controlled carefully. Thickness can vary with no change in weight, due to difference in the density of the granulation and the pressure applied to the tablets, as well as the speed

of tablet compression. Not only is the tablet thickness important in reproducing tablets identical in appearance, but also to ensure that every production lot will be usable with selected packaging components. If the tablets are thicker than specified, a given number may no longer be contained in the volume of a given size bottle. Tablet thickness also becomes an important characteristic in counting tablets using filling equipment. Some filling equipment uses the uniform thickness of the tablets as a counting mechanism. A column containing a known number of tablets is measured for height; filling is accomplished by continually dropping columns of tablets of the same height into bottles. If thickness varies throughout the lot, the result will be variation in count. Other pieces of filling equipment can malfunction, due to variation in tablet thickness, since tablets above specified thickness may cause wedging of tablets in previously adjusted depths of the counting slots. Tablet thickness is determined with a caliper or thickness gauge that measures the thickness in millimeters. Plus or minus 5% may be allowed, depending on the Standard Operating Procedures (Fig. 30-32).

UNIFORMITY OF DOSAGE UNITS

Over the past 40 years, the limits for uniformity of dosage units have been upgraded and made stricter. The reader is advised to refer to the most recent Pharmacopeia for complying with the regulations where the test is done.

The term “uniformity of dosage unit” is defined as the degree of uniformity in the amount of the drug substance in a tablet. The uniformity of dosage units can be demonstrated by either one of two methods, Content Uniformity (CU) or Weight Variation. The most recent USP 34 is harmonized with the Japanese and European Pharmacopoeias.¹⁶

CU testing was developed to ensure the content consistency of active pharmaceutical ingredients (API) within a narrow range around the label claim in dosage units. CU is related to efficacy, especially for low drug content where manufacturing loss is the dominant problem; whereas, for potent and narrow therapeutic indexed drugs, CU is also associated with safety.

By default, all tablets and capsules, as dosage units, are requested to pass the CU acceptance criteria, by direct individual analysis of each sampled unit, otherwise weight variation (WV) procedure is to be applied, if dosage units contain 25.0 mg or 25.0% W/W of API or more, or when the process development and validation data support its application. CU by WV is determined by multiplying the composite assay results by the ratio of each individual unit weight to the average weight of the sample.

Current CU testing has a confirmatory nature and is based on calculation of acceptance value (AV) of two stage sampling plan by applying the following equation:

$$AV = (|M - \bar{x}| + ks)$$

The multiplier k is either 2.4 or 2.0, for sample sizes of 10 or 30, respectively. s is the standard deviation. The reference value M takes different values, according to the sample mean ($\bar{x}$) and the target content per dosage unit at the time of manufacturing (T). The calculated AV should be less than 15.0 (L_1), and no single unit deviates by more than L_2 % (generally, 25%) from the reference value M . Theoretically, this new concept aims to assure dosage units to have drug content not critically (within $\pm 15\%$) deviating from the nominal dose value.

Tablet Disintegration

It is recognized generally that the *in vitro* tablet disintegration test does not necessarily bear a relationship to the *in vivo* action of a solid dosage form. To be absorbed, a drug substance must be in solution, and the disintegration test is a measure only of the time required under a given set of conditions for a group of tablets to disintegrate into particles. Generally, this test is useful as a quality-assurance tool for conventional dosage forms. In the present disintegration test, the particles are those that pass



Figure 30-33. USP Disintegration Tester. (Courtesy of SOTAX.)

through a 10-mesh screen. Regardless of the lack of significance as to *in vivo* action of the tablets, the test provides a means of control in ensuring that a given tablet formula is the same as regards disintegration from one production batch to another. The disintegration test is used as a control for tablets intended to be administered by mouth, except for tablets intended to be chewed before being swallowed or tablets designed to release the drug substance over a period of time. Exact specifications are given for the test apparatus, in as much as a change in the apparatus can cause a change in the results of the test. The USP apparatus consists of a basket-rack assembly, a 1000-mL, low-form beaker, 138–160 mm in height and having an inside diameter of 97–115 mm for the immersion fluid, a thermostatic arrangement for heating the fluid between 35°C and 39°C, and a device for raising and lowering the basket in the immersion fluid at a constant frequency rate between 29 and 32 cycles per minute through a distance of not less than 53 mm and not more than 57 mm. The volume of the fluid in the vessel is such that, at the highest point of the upward stroke, the wire mesh remains at least 15 mm below the surface of the fluid and descends to not less than 25 mm from the bottom of the vessel on the downward stroke. At no time should the top of the basket-rack assembly become submerged. The time required for the upward stroke is equal to the time required for the downward stroke, and the change in stroke direction is a smooth transition, rather than an abrupt reversal of motion. The basket-rack assembly moves vertically along its axis. There is no appreciable horizontal motion or movement of the axis from the vertical.

Basket-Rack Assembly—The basket-rack assembly consists of six open-ended transparent tubes, each 77.5 ± 2.5 mm long and having an inside diameter of 20.7–23 mm and a wall 1.0–2.8 mm thick; the tubes are held in a vertical position by two plates, each 88–92 mm in diameter and 5–8.5 mm in thickness, with six holes, each 22–26 mm in diameter, equidistant from

the center of the plate and equally spaced from one another. Attached to the under surface of the lower plate is a woven stainless steel wire cloth, which has a plain square weave with 1.8- to 2.2-mm apertures and with a wire diameter of 0.57-0.66 mm. The parts of the apparatus are assembled and rigidly held by means of three bolts passing through the two plates. A suitable means is provided to suspend the basket-rack assembly from the raising and lowering device, using a point on its axis (Fig. 30-33).

The design of the basket-rack assembly may be varied somewhat, provided the specifications for the glass tubes and the screen mesh size are maintained. The use of disks is permitted only where specified or allowed in the monograph. If specified in the individual monograph, each tube is provided with a cylindrical disk 9.5 ± 0.15 mm thick and 20.7 ± 0.15 mm in diameter. More details about disk design are in the USP. The endpoint of the test is indicated when any residue remaining is a soft mass with no palpably soft core. The plastic discs help force any soft mass that forms through the screen. For compressed, uncoated tablets the testing fluid is usually purified water USP at 37°C, but in some cases the monographs direct that Simulated Gastric Fluid TS be used. If one or two tablets fail to disintegrate, the test is repeated using 12 tablets. Of the 18 tablets then tested, 16 must have disintegrated within the given period of time. The conditions of the test are varied somewhat for coated tablets, buccal tablets, and sublingual tablets. Disintegration times are included in the individual tablet monograph. For most uncoated tablets, the period is 30 minutes, although the time for some uncoated tablets varies greatly from this. For coated tablets up to 2 hours may be required, whereas, for sublingual tablets, such as CT Isoproterenol Hydrochloride, the disintegration time is 3 minutes. This is harmonized with the corresponding texts of the European Pharmacopoeia and/or the Japanese Pharmacopoeia. (For the exact conditions of the test, consult the USP.)

Dissolution Test

Currently, for most tablets, the monographs direct compliance with limits on dissolution, rather than disintegration. Since drug absorption and physiological availability depend on having the drug substance in the dissolved state, suitable dissolution characteristics are important properties of a satisfactory tablet. Like the disintegration test, the dissolution test for measuring the amount of time required for a given percentage of the drug substance in a tablet to go into solution under a specified set of conditions is an *in vitro* test. It is intended to provide a step toward the evaluation of the physiological availability of the drug substance, but, as described currently, it is not designed



Figure 30-34. USP Dissolution Tester, Paddle Method. (Courtesy of Erweka.)



Figure 30-35. USP Dissolution Tester, Flow Through. (Courtesy of SOTAX.)

to measure the safety or efficacy of the tablet being tested. Both the safety and effectiveness of a specific dosage form must be demonstrated, initially, by means of appropriate *in vivo* studies and clinical evaluation. Like the disintegration test, the dissolution test does provide a means of control in ensuring a given tablet formulation is the same as regards dissolution as the batch of tablets shown, initially, to be clinically effective. It also provides an *in vitro* control procedure to eliminate variations among production batches. Figures 30-34 and 30-35 show some popular commercial dissolution testers. (Refer to Chapter 6 for a complete discussion of dissolution testing.)

Validation

In this era of increasing regulatory control of the pharmaceutical industry, manufacturing procedures cannot be discussed without the mention of some process-validation activity. By way of documentation, product testing, and, perhaps, in-process testing as well, manufacturers can demonstrate that their formulas and processes perform in the manner expected and that they do so reproducibly. Although the justification for requiring validation is found in the regulations relating to *Current Good Manufacturing Practices for Finished Pharmaceuticals*, as well as other sources, there is still much room for interpretation, and the process varies from one company to another. General areas of agreement are that the validation activity must begin in Research and Development and continue through product introduction. Documentation is the key. In general, three batches represent an adequate sample for validation. The FDA has rejected historical data or retrospective validation. They require that new products be validated from beginning to end, a process called prospective validation.¹⁷

PROCESS ANALYTICAL TECHNOLOGY (PAT)

One definition for PAT is the “systems for continuous analysis and control of manufacturing processes based on real-time measurements, or rapid measurements during processing, of quality and performance attributes of raw and in-process materials and processes to assure acceptable end product quality at the completion of the process.”¹⁸ This requires a synergy of multiple systems, including process analytical chemistry tools, information management tools, feedback process control strategies, and strategies for product/process design and optimization. The PAT initiative in pharmaceuticals began in the 1990s. In 1993, an AOAC International Symposium was held entitled “Pharmaceutical Process Control and Quality

Assessment by Non-Traditional Means.” Then, in 2001, an Advisory Committee for Pharmaceutical Science (ACPS) meeting was held and discussed PAT. This meeting prompted 20/FDA Science Board endorsement of the formation of a PAT subcommittee within ACPS.¹⁸ The PAT subcommittee is comprised of four working groups: 1) benefits, technology, definitions/terminology, 2) process and analytical validation, 3) chemometrics, and 4) product/process development. These working groups are composed of people from the FDA, industry, and academia. In addition, a new journal was published to address the expanding role of PAT in the pharmaceutical industry, *The Journal of Process Analytical Technology* with its first issue in September/October 2004. It is apparent that PAT will become an important part of the pharmaceutical industry, as it has in other industries. The reason is that PAT can improve efficiency and capability of pharmaceutical processes, while verifying and improving product quality. Results the pharmaceutical industry anticipates from optimal application of PAT include reduced risk of scrap and recalls, improved process understanding with quality “by design,” improved capacity utilization and reduced cycle times, and eventual reduction in product development time. The current methods to verify product quality, final product testing at the end of manufacturing and some process operation ranges, cause low manufacturing efficiency and require a very high level of regulatory scrutiny.¹⁸ For example, if a batch fails final product release testing for product quality control, these results are not known until weeks after the batch was manufactured. In that time period, multiple batches could have been manufactured with the same problems and would have to be rejected. In contrast, PAT allows for rapid identification of out of the ordinary situations, when the critical product properties are defined up front. Therefore, the time to respond to problems is greatly reduced, and much time and money can be saved. PAT also allows for statistical process control analysis (trending) of critical in-process product attributes.¹⁹

CAPSULES

Capsules are solid dosage forms in which the drug substance is enclosed in either a hard or soft, soluble container or shell of gelatin or other suitable material. The soft gelatin capsule was invented by Mothes, a French pharmacist, in 1833. During the following year, DuBlanc obtained a patent for his soft gelatin capsules. In 1848, Murdock patented the two-piece hard gelatin capsule. More recently, capsule shells made from hydroxypropyl methylcellulose have become available commercially. These capsules contain less residual moisture than gelatin capsules and thus may have applications in filling hydrolytic drugs. In addition, these capsules offer an excellent alternative for strict vegetarians. Gelatin, however, due to its unique properties, remains the primary composition material for the manufacture of capsules. The gelatin used in the manufacture of capsules is obtained from collagenous material by hydrolysis. There are two types of gelatin, Type A, derived mainly from pork skins by acid processing, and Type B, obtained from bones and animal skins by alkaline processing. Blends are used to obtain gelatin solutions with the viscosity and bloom strength characteristics desirable for capsule manufacture.

The encapsulation of medicinal agents remains a popular method for administering drugs. Capsules are tasteless, easily administered, and easily filled either extemporaneously or in large quantities commercially. In prescription practice, the use of hard gelatin capsules permits a choice in prescribing a single drug or a combination of drugs at the exact dosage level considered best for the individual patient. This flexibility is an advantage over tablets. Some patients find it easier to swallow capsules than tablets, therefore, preferring to take this form when possible. This preference has prompted pharmaceutical manufacturers to market products in capsule form, even though the products have already been produced in tablet form.

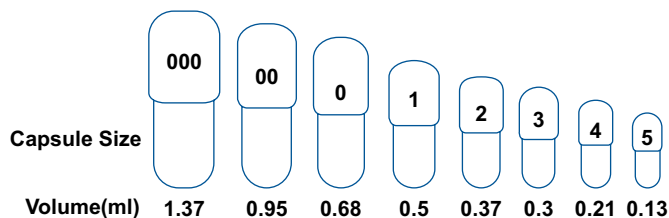


Figure 30-36. Hard gelatin capsules showing relative size.

Although the industry prepares approximately 75% of its solid dosage forms as compressed tablets, 23% as hard gelatin capsules, and 2% as soft elastic capsules, market surveys indicated a consumer preference of 44.2% for soft elastic capsules, 39.6% for tablets, and 19.4% for hard gelatin capsules.²⁰

HARD GELATIN CAPSULES

The hard gelatin capsule consists of two sections, one slipping over the other, thus, completely surrounding the drug formulation. Figure 30-36 illustrates the classic capsule shape and size. These capsules are filled by introducing the powdered material into the longer end or body of the capsule and then slipping on the cap. Hard gelatin capsules are made from gelatin and other materials used either to enable capsule formation or to improve their performance (see Table 30-2).²¹ Hard gelatin capsules contain 12–16% water, but the water content can vary, depending on the storage conditions. When the humidity is low, the capsules can become brittle; if stored at high humidities, the capsules can become flaccid, sticky, and lose their shape. Storage in high-temperature areas can also affect the quality of hard gelatin capsules. Gelatin capsules do not protect the fill materials from atmospheric water vapor, as moisture can readily diffuse through the gelatin wall.

Companies having equipment for preparing empty hard gelatin capsules include Lilly, Parke-Davis, Scherer, and Smith-Kline. The latter's production is mainly for its own use; the others are suppliers to the industry. With this equipment, stainless steel pins, set in plates, are dipped into the gelatin solution, which must be maintained at a uniform temperature and an exact degree of fluidity. If the gelatin solution varies in viscosity, it will correspondingly decrease or increase the thickness of the capsule wall. This is important, since a slight variation is sufficient to make either a loose or a tight joint. When the pins have been withdrawn from the gelatin solution, they are rotated, while being dried in kilns through which a strong blast of filtered air with controlled humidity is forced. Each capsule is removed from the pins, trimmed to uniform length, and joined, the entire process being mechanical.

Table 30-2. Gelatin Capsule Additives

Additives	Function and Examples
Coloring Agent	Aesthetic, Identification, Psychological effect and Light protection (e.g., E122, E127, E171)
Process Aids	Surfactants: help in gelatin to take the shape of the moulds Silicon Fluid: assistant in enteric coating
Performance Aids	Flavoring agent: improve patient acceptability
Preservative	Gelatin is good medium for bacterial and fungal growth (e.g., Sulfur dioxide)
Gelatin Extenders	To reduce the quantity of gelatin needed per capsule (e.g., hydroxylalkylstarch)
Protective Coating	To improve the capsule stability (i.e., moisture resistance and prevent capsule sticking)

Capsules are supplied in a variety of sizes. The hard, empty capsules (Fig. 30-36) are numbered from 000, the largest size that can be swallowed, to 5, which is the smallest. Larger sizes are available for use in veterinary medicine. The approximate capacity for capsules from 000 to 5 ranges from 600 to 30 mg, depending on the densities of the powdered drug materials. FDA issued a draft guidance for the industry in January 2011 about the maximum capsule bead size, which is 2.0 mm. This is based on approved product information and human studies that demonstrate food is chewed to approximately 2 mm in median particle size before swallowing.²²

Commercially filled capsules have the conventional oblong shape illustrated above, with the exception of capsule products from Lilly and SmithKline. For Lilly products, capsules are tapered at the base of the body to give the capsule a bullet-like shape; products encapsulated in this shape of capsule are called Pulvules. The SmithKline capsules differ in ends of both the cap and body are angular.

After hard gelatin capsules are filled and the cap applied, there are a number of methods used to ensure the capsules will not come apart, if subjected to vibration or rough handling, as in high-speed counting and packaging equipment. The capsules can be spot-welded by means of a heated metal pin pressed against the cap, fusing it to the body, or they may be banded with molten gelatin applied around the joint. Colored gelatin bands around capsules have been used for many years as a trademark by Parke-Davis for their line of capsule products, Kapseals. Another approach is the Snap-Fit and Coni-Snap capsules that use a pair of matched locking rings in the cap and body portions of the capsule. Prior to filling, these capsules are slightly longer than regular capsules of the same size. When the locking rings are engaged after filling, their length is equivalent to that of the conventional capsule. Following several tampering incidents, pharmaceutical companies now use a number of locking and sealing technologies in addition to tamper-resistant packaging.

For extemporaneous compounding, the pharmacist must determine the size of the capsule needed for a given prescription. The correct capsule size can be determined from the volume of the material to be held by a single capsule. For powdered materials, the body of the capsule is filled with the desired weight of powder and the cap replaced. Weight checks of the compounded capsules should be done to verify the correct amount of powder has been filled in each capsule.

In addition to the transparent, colorless, hard gelatin capsule, capsules are also available in various transparent colors, such as pink, green, reddish brown, blue, yellow, and black. If they are used, it is important to note the color, as well as the capsule size, on the prescription, so that a refilled prescription will duplicate the original. Colored capsules have been used chiefly by manufacturers to give a specialty product a distinctive appearance. Titanium dioxide is added to the gelatin to form white capsules or to make an opaque, colored capsule. Commercial prescription products must be given further identification by markings with special numbers to permit exact identification by the pharmacist or physician.

Extemporaneous Filling Methods

When filling capsules on prescription, the usual procedure is to mix the ingredients by trituration, reducing them to a fine and uniform powder. The principles and methods for the uniform distribution of an active medicinal agent in a powder mixture are discussed in Chapter 23. Granular powders do not pack readily in capsules, and crystalline materials, especially those that consist of a mass of filament-like crystals, such as the quinine salts, are not fitted easily into capsules, unless powdered. Eutectic mixtures that tend to liquefy may be dispensed in capsules, if a suitable absorbent, such as magnesium carbonate, is used. Potent drugs given in small doses are mixed with an inert diluent, such as lactose, before filling into capsules. When incompatible materials are prescribed together, it is sometimes

possible to place one in a smaller capsule and then enclose it with the second drug in a larger capsule.

The powder is placed on paper and flattened with a spatula, so the layer of powder is not greater than about 1/3 the length of the capsule being filled. This helps keep both the hands and capsules clean. The cap is removed from the selected capsule and held in the left hand; the body is pressed repeatedly into the powder until filled. The cap is replaced, and the capsule is weighed. In filling the capsule, the spatula is helpful in pushing the last quantity of the material into the capsule. If each capsule has not been weighed, there is likely an excess or a shortage of material, when the specified number of capsules have been packed.

Today, many manual and automatic capsule filling machines are available for increasing the speed of the capsule filling operation. Figure 30-37 illustrates a capsule-filling machine that is supplied by Torpac Inc. (ProFiller® 100). Many community pharmacists find this a useful apparatus, and some pharmaceutical manufacturers use it for small-scale production of specialty items. It is a 100-hole, hand-held capsule filling system available in all capsule sizes from 000 to 5, including el sizes. Production rates of 2000–3200 capsules per hour, within 1–2% of target capsule weight, are possible. Entire capsules (cap and body) are placed in the machine; the lower plate holds the capsule bodies in place and makes it possible to remove the caps. The powder is filled in the bodies; the accuracy of the fill depends on the selection of capsule size and the amount of pressure applied during filling. The capsule caps are then rejoined with the bodies. For large batches, the hand-operated machine (ProFiller® 3800) can be used, as illustrated in Figure 30-38. It is a bench-top filler with a bench-top Orienter (loader) that



Figure 30-37. Hand operated capsule machine. (Courtesy of Torpac Inc.)

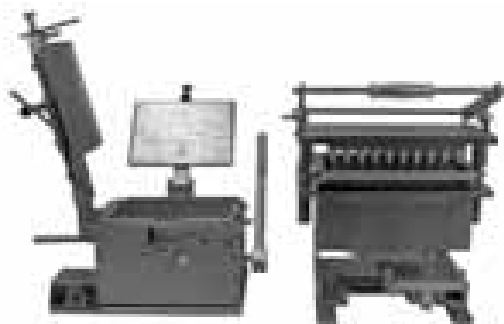


Figure 30-38 Hand operated capsule machine, Model 3008-bs. (Courtesy of Torpac Inc.)

fills 300 capsules per try with an output of up to 7200 to 9000 capsules/hr.

Machine Filling Methods

Large-scale capsule filling equipment operates on the same principle as the manual machines described above, namely the filling of the body of the capsule. Compared with tablets, powders for filling into hard gelatin capsules require a minimum of formulation efforts. The active ingredient is mixed with a sufficient volume of a diluent to fill the body with the desired amount of drug. Since powder flow is of great importance in the rapid and accurate filling of the capsule bodies, glidants, such as silica derivatives or the stearates, are also used frequently. As is true with tablets, additives present in the capsule formulation can influence drug release. Tablets and capsules of a combination product containing triamterene and hydrochlorothiazide in a 2:1 ratio were compared clinically. The tablet caused approximately twice as much excretion of hydrochlorothiazide and three times as much triamterene as the capsule.²³ The manner of operation of the machine can influence the volume of powder filled into the base of the capsule; therefore, the weights of the capsules must be checked routinely as they are filled (see Table 30-3).

Hard capsules are usually filled with active ingredient in dry powder form, yet other preparation may also be filled in capsules, including tablets, pellets, and even certain types of liquids. If liquids are filled in hard gelatin capsule, a banding step is generally necessary to prevent leakage of the fill. The filling of capsules is carried out in four steps: 1) separating the cap from the capsule body, 2) filling the body, 3) rejoining the capsule cap and body, and 4) discharge of the filled capsules. Filling the hard capsules with powders varies greatly from one manufacturer to another. The following are the four methods for filling the powder into hard capsules:

The Plate Method—The plate method is consider the basic and traditional filling method and is probably the oldest system used to fill hard gelatin capsules. It depends on overlaying the powder or granules over the capsule body in the dosing disc and then a using tamping device for dose punching and filling. The excess material is removed before replacing the caps. H Höfliger has developed a method, dependent on using free flowing material combined with a tamping punch (Fig. 30-39). In this method, the powder is compressed and tamped against the

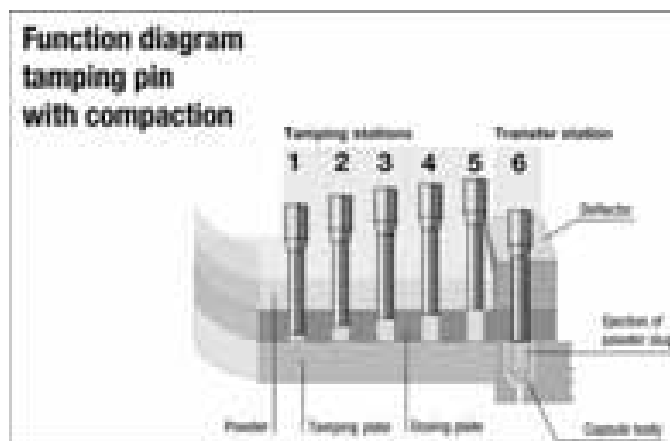


Figure 30-39. Filling capsule by tamping method. (Courtesy of Harro Höfliger Verpackungsmaschinen GmbH.)

base plate of the dosage disc, which is designed to give a certain fill-weight in the capsule.

Although the filling weight can be adjusted slightly, by changing the tamping force, the dosing amount is more or less defined by the thickness of the dosing discs. Height adjustable dosing discs are on the market, but tests have shown that, due to the nature of the design, only easy to fill powders can be processed on this type of device. Furthermore, the constancy of the powder bed height influences the fill weight. Without extra measures, the tamp filling principle is not suitable for dosages below 30 mg. Powders that tend to segregate are problematic when dosing discs are above a certain diameter.

In comparison to the vacuum drum filler, there is more residual powder in the powder bowl at the end of a production run. Thus, this filling method, as well as the dosator nozzle principle, is usually not first choice, if the powder is expensive and batch sizes are small. Irrespective of these limitations, the tamp filling method is a very robust and easy to maintain system.

Intermittent Compression Filling—This method is used by Zanasi and other companies. In this method the powder is fed into the dosage hopper, and then the dosage tube enters the powder bed. This leads to form a unit mass inside the tube, which is lifted by the dosator and ejected into the capsule body

Table 30-3. Capsule Fill Chart

CAPSULE FILL WEIGHTS (MG) BASED ON SIZE AND DENSITY										
POWDER DENSITY (g/ml)	CAPSULE VOLUME (ML)									
	0.95	0.78	0.68	0.54	0.5	0.37	0.3	0.25	0.21	0.13
	CAPSULE SIZE									
	00	0e1	0	1e1	1	2	3	4e1	4	5
0.3	285	234	204	162	150	111	90	75	63	39
0.4	380	312	272	216	200	148	120	100	84	52
0.5	475	390	340	270	250	185	150	125	105	65
0.6	570	468	408	324	300	222	180	150	126	78
0.7	665	546	476	378	350	259	210	175	147	91
0.8	760	624	544	432	400	296	240	200	168	104
0.9	855	702	612	486	450	333	270	225	189	117
1	950	780	680	540	500	370	300	250	210	130
1.1	1045	858	748	594	550	407	330	275	231	143
1.2	1140	936	816	648	600	444	360	300	252	156
1.3	1235	1014	884	702	650	481	390	325	273	168
1.4	1330	1092	952	756	700	518	420	350	294	182
1.5	1425	1170	1020	810	750	555	450	375	315	159

(Fig. 30-40). One of the important factors that should be considered in this method is producing sufficient compression force to form an united mass and perfectly transferring it to the capsule body.

The dosator procedure is suitable for powder dosing quantities of ≥ 20 mg; however, it has only limited usage with very cohesive powders, with a high level of fine particles. Dosator pins can be manufactured with coated front surfaces or from materials to which the respective powders are less likely to adhere (e.g., certain polymers, which are certified by the FDA as product-contact materials.)

Continuous Compression Filling—This method was first developed by mG2 company. It is similar to the intermittent method in principle, yet it differs in application of high filling speeds. In this method, the dosator dips into the powder, and then fills the capsule body. The movement is rapid and continuous.

Vacuum Filling—This method was developed first by Perry Industries Inc. In this filling process the dosator draws the powder by suction, not like the previous process (i.e., punch). Some compression takes place by suction, allowing the material to stay in the dosator by vacuum pressure. The dosator then delivers the powder into the capsule body, after releasing the vacuum and applying positive pressure. One important technique of vacuum filling is called vacuum-drum dosing.

The principle construction of such a dosing system is represented in Figure 30-41. In the center is the rotating dosing drum, which consists of a core with a connection to the vacuum/pressure system and an outer casing with one or more rows of holes, which form the dosing chambers. Each dosing cham-

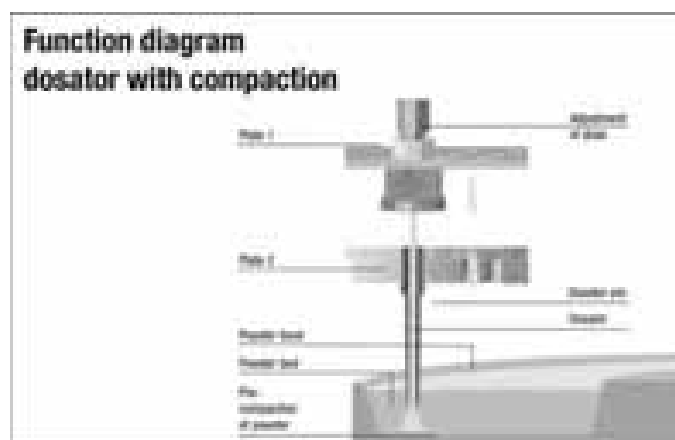


Figure 30-40. Intermittent compression filling of capsules. (Courtesy of Harro Hofliger Verpackungsmaschinen GmbH.)

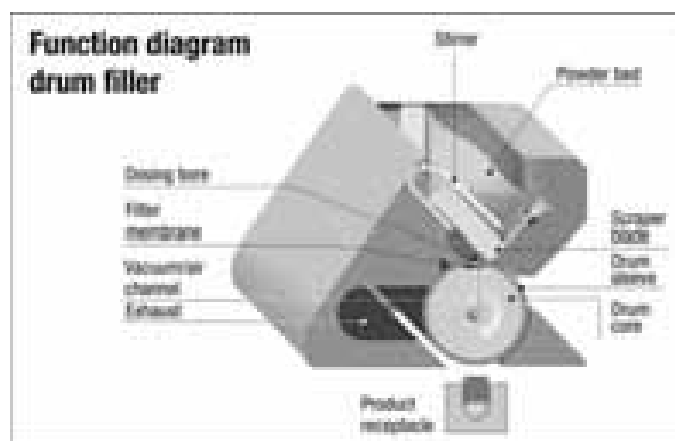


Figure 30-41. Filling capsule by Vacuum-Drum Method. (Courtesy of Harro Hofliger Verpackungsmaschinen GmbH.)

ber is positioned over a cross drilled hole in the core that is connected with the air system. The openings turned toward the core of the chamber are closed by a suitable filter membrane. The filter material forms the separation of the chamber to the vacuum/pressure system and prevents the powder particles from being drawn off during dosing.

The filling cycle starts when the dosing cavities are positioned beneath the powder bed and vacuum is applied. The powder is forced into the cavities in the stationary or slowly rotating drum. By maintaining the vacuum, the drum is rotated for 180° above the target, the vacuum is turned off, and the powder slug is ejected by means of a short pulse of compressed air. Finally, the drum is rotated to the exhaust channel to remove residual powder adhering to the filter membrane or the walls of the dosing cavity.

There are many new capsule filling machines that work with one or more of the mentioned processes of capsule filling. Figure 30-42 shows a Modular capsule filling machine (Modu-C LS), produced by H Höfliger. Figure 30-43 shows another new capsule filling machine, produced by PharmaLand (PCF 1200 Model).

All capsules, whether filled by hand or by machine, will require cleaning. Small quantities of capsules may be wiped individually with cloth. Larger quantities are generally processed through a capsule deduster/polisher.

Uniformity of Dosage Units

Uniformity of dosage units is defined as the degree of uniformity in the amount of the drug substance among dosage units, USP 34²⁴. The uniformity of dosage units of capsules can be demonstrated by either of two methods: Content Uniformity or Weight Variation (Table 30-4).

Disintegration tests are not usually required for capsules, unless they have been treated to resist solution in gastric fluid (enteric-coated). In this case, they must meet the requirements for disintegration of enteric-coated tablets. For certain capsule dosage forms, a dissolution requirement is part of the monograph. Procedures used are similar to those employed in the case of compressed tablets.

SOFT ELASTIC CAPSULES

Soft elastic capsules, also known as soft gelatin capsules or softgels, differ from hard gelatin capsules in several ways. First, the capsule shells are typically made and filled simultaneously. Commercially filled soft gelatin capsules are available in a wide variety of shapes (e.g., round, oblong, tubular). The gelatin is plasticized by the addition of glycerin, sorbitol, or other suitable material, which gives the shell its characteristic flexibility. Preservatives are often included in the capsule shell formulation to prevent microbial or fungal growth.



Figure 30-42. Automated capsule filling machine, Modu-C LS. (Courtesy of Harro Hofliger Verpackungsmaschinen GmbH.)



Figure 30-43. Automated capsule filling machine, Modu-PCF 1200. (Courtesy of PharmaLand.)

Soft gelatin capsules are generally filled with liquids, where the drug is dissolved or dispersed in a solvent. The oral bioavailability of poorly water soluble drugs can be enhanced when the drug is dissolved in the fill liquid. Selection of the fill liquid (solvent) depends on the physicochemical properties of the drug, presence of other additives, and weight or volume of the content. Product stability is dependent on the properties of the capsule shell, storage conditions, and fill content²⁵. For example, water vapor migration from the atmosphere can migrate through the capsule shell into the fill liquid and potentially cause precipitation of the drug.

Manufacturing Methods

Plate Process—In this method, a set of molds is used. A warm sheet of prepared gelatin is laid over the lower plate, and the liquid is poured on it. A second sheet of gelatin is carefully put in place, followed by the top plate of the mold. The set is placed under a press where pressure is applied to form the capsules. Filled capsules are then washed off with a volatile solvent to remove any traces of oil from the exterior. The sheets of gelatin may have the same or different colors.

Rotary-Die Process—In 1933, the rotary-die process for soft gelatin capsules was perfected by Robert P Scherer.²⁶ The rotary-die machine is a self-contained unit capable of continuously and automatically producing finished capsules from a supply of gelatin mass and filling material, which may be any liquid, semiliquid, or paste that will not dissolve gelatin. Two continuous gelatin ribbons, which the machine forms, are brought into convergence between a pair of revolving dies and an injection

wedge. Accurate filling under pressure and sealing of the capsule wall occur as dual and coincident operations; each is delicately timed against the other. Sealing also severs the completed capsule from the net. The principle of operation is shown in Figure 30-44 (see also Fig. 30-45).

By this process, the content of each capsule is measured individually by a single stroke of a pump, so accurately constructed that a plunger travel of 0.025 in will deliver 1 <minim> (apoth). The Scherer machine contains banks of pumps arranged so that many capsules are formed and filled simultaneously. All pumps are engineered to extremely small mechanical tolerances and to an extremely high degree of precision and similarity. All operations are controlled on a weight basis, by actual periodic checks with a group of analytical balances. Individual net-fill weights of capsules, resulting from large-scale production, vary no more than 3% from theory, depending on the materials used.

The rotary-die process makes it possible to encapsulate viscous materials, such as ointments and pastes. In this manner, solids can be milled with a vehicle and filled into capsules. When it is desirable to have a high degree of accuracy and a hermetically sealed product, this form of enclosure is ideally suited.

The modern and well-equipped capsule plant is completely air-conditioned with a high-quality water source, a practical necessity for capsule production. Its facilities and operations include the availability of carbon dioxide at every exposed point of operation for the protection of oxidizable substances before encapsulation. Special ingredients have also been used in the capsule shell to exclude light wavelengths destructive to certain drugs.

Norton Capsule Machine—This machine produces capsules completely automatically, by leading two films of gelatin between a set of vertical dies. These dies, as they close, open, and close, are in effect a continual vertical plate, forming row after row of pockets across the gelatin film. These are filled with medicament and, as they progress through the dies, are sealed, shaped, and cut out of the film as capsules, which drop into a cooled solvent bath.

Accogel Capsule Machine—Another means of soft gelatin encapsulation uses the Accogel machine, a process developed at Lederle. The Accogel, or Stern machine, uses a system of rotary dies but is unique, in that it is the only machine that can successfully fill dry powder into a soft gelatin capsule. The machine is available to the entire pharmaceutical industry by a lease arrangement and is used in many countries of the world. It is extremely versatile, not only producing capsules with dry powder, but also encapsulating liquids and combinations of liquids and powders. By means of an attachment, slugs or compressed tablets may be enclosed in a gelatin film. The capsules can be made in a variety of colors, shapes, and sizes.

Seamless Process (Bubble Method)—The seamless technique produces soft gelatin capsules without requirement of dies. In this process, a molten gelatin stream runs through an outer nozzle of a concentric tube at a constant rate. The medicated liquid formulation is dispensed through the inner orifice by means of a pump. The emerging stream is broken up into an intermittent but steady flow of uniform sized by a pulsating mechanism, leading to the formation of droplets enveloped in molten gelatin. The formed capsules are quickly removed from the nozzle, slowly congealed, and automatically ejected from the system.

Table 30-4. Application of Content Uniformity (CU) and Weight Variation (WV) Tests for Capsule Dosage Forms

Dosage type	Subtype	Dose and ratio drug substance	
		≥25 mg	<25 mg
Hard gelatin capsule	Powder or others	WV	CU
Soft gelatin capsule	Suspension, emulsion or gel Solutions	CU	CU
		WV	WV

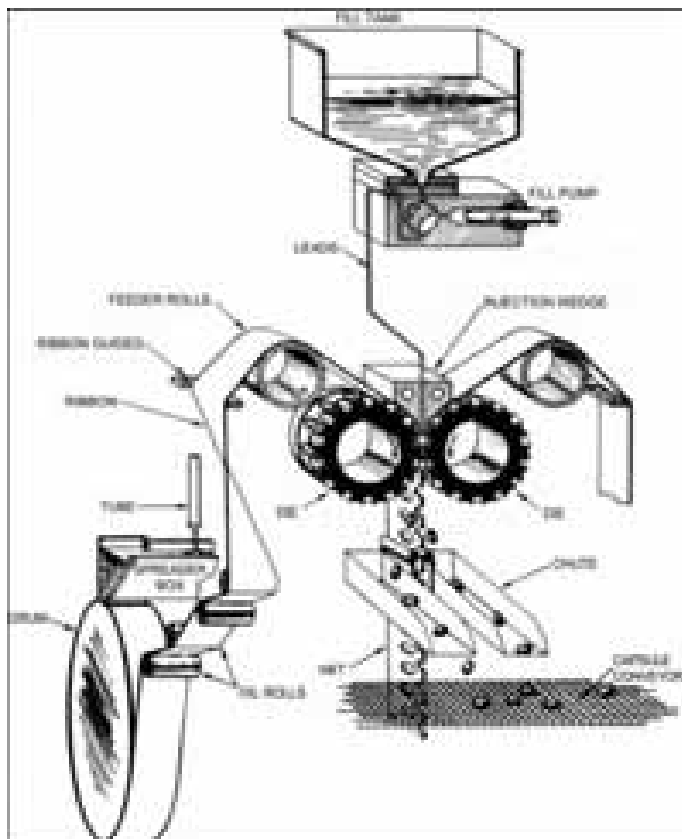


Figure 30-44. Schematic drawing of a rotary-die soft gelatin capsule filler. (Courtesy of R.P. Scherer: Detroit, MI.)

Microencapsulation— Microencapsulation is a process or technique by which thin coatings can be applied reproducibly to small particles of solids, droplets of liquids, or dispersions, thus, forming microcapsules. It can be differentiated readily from other coating methods in the size of the particles involved; these range from several tenths to about 5000 micrometers in size. As a technology, microencapsulation is placed in the section on capsules only because of the relationship in terminology to mechanical encapsulation previously described. The topic is also discussed in Chapter 32.

A number of microencapsulation processes have been disclosed in the literature, some based on chemical processes involving a chemical or phase change and others are mechanical in nature, requiring specialized equipment.^{27,28} A number of coating materials have been used successfully to microencapsulate materials; examples of these include gelatin, polyvinyl alcohol, ethylcellulose, cellulose acetate phthalate, and styrene maleic anhydride. Film thickness can be varied considerably, depending on the surface area of the material coated and other



Figure 30-45. Rotary-die soft gelatin capsule filler. (Courtesy of Pharmagel Engineering SPA: Milan, Italy.)

physical characteristics of the system. The microcapsules may consist of a single particle or clusters of particles. After isolation from the liquid manufacturing vehicle and drying, the material appears as a free-flowing powder. The powder is suitable for formulation as compressed tablets, hard gelatin capsules, suspensions, and other dosage forms. Microencapsulation has applications for masking the taste of bitter drugs, but can also be used as a means of formulating prolonged-action dosage forms, protecting chemicals against moisture or oxidation, and modifying a material's physical characteristics for ease of handling in formulation and manufacture.

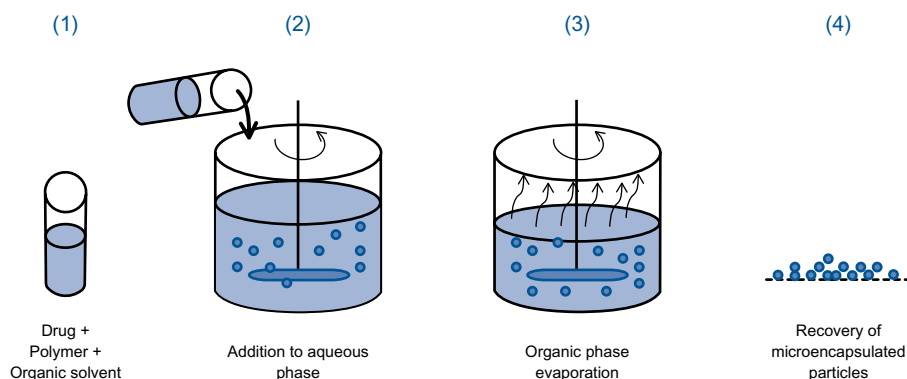


Figure 30-46 Basic steps of microencapsulation by solvent evaporation.

The technique of microencapsulation by solvent evaporation is widely applied in pharmaceutical industries to obtain the controlled release of drug (Fig. 30-46). The basic steps of microencapsulation by solvent evaporation include: 1) dissolving of the hydrophobic drug in an organic solvent containing the polymer; 2) emulsification of this organic phase in an aqueous phase; 3) extraction of the solvent from the organic phase, by the aqueous phase, accompanied by solvent evaporation, transforming droplets of dispersed phase into solid particles; and 4) recovery and drying of microencapsulation particles to eliminate the residual solvent. For insoluble or poorly water-soluble drugs, the oil-in-water (o/w) method is frequently used.

REFERENCES

- Ritschel WA, Kearns GL. *Handbook of Basic Pharmacokinetics: including clinical applications* 6th ed. Washington, DC: American Pharmacists Association, 2004.
- Vogt M *et al.* Cogrounding enhances the oral bioavailability of EMD 57033, a poorly water soluble drug, in dogs. *Eur J Pharm Biopharm.* (2008). 68: 338–345.
- Turkoglu M, Sakr A. Tablet dosage forms. In: Florence AT, Siepmann J, eds. *Modern Pharmaceutics*, Vol. 1. New York: Informa Healthcare, 2009: 481–497.
- Augsburger LL, Hoag SW (Eds.). *Pharmaceutical Dosage Forms: Tablets, vols 1–3*. New York: Informa Healthcare, 2008.
- Buggins TR *et al.* The effects of pharmaceutical excipients on drug disposition. *Advanced Drug Delivery Reviews.* (2007). 59: 1482–1503.
- Chono S *et al.* Enhancement of the dissolution rate and gastrointestinal absorption of Pranlukast as a model poorly water-soluble drug by grinding with gelatin. *Int. J. Pharm.* (2008). 347: 71–78.
- Jackson K *et al.* Drug-excipient interactions and their effect on absorption. *Pharmaceutical Science & Technology Today.* (2000). 3 (10): 336–345.
- Patel H *et al.* The effect of excipients on the stability of levothyroxine sodium pentahydrate tablets. *Int J Pharm.* (2003). 264: 35–43.
- Rowe R, *et al.* (Eds.). *Handbook of Pharmaceutical Excipients*, 6th ed. Washington, DC: APhA/Pharmaceutical Press, 2009.
- Rudnic EM *et al.* Effect of molecular structure variation on the disintegrant action of sodium starch glycolate. *J Pharm Sci.* (1985). 74: 647.
- Khattab I *et al.* Effect of mode of incorporation of disintegrants on the characteristics of fluid-bed wet-granulated tablets. *J Pharm Pharmacol.* (1993). 45: 687–691.
- Khattab I *et al.* A study of the effect of disintegrant. distribution ratio on tablet characteristics using a central composite design. *Eur J Pharm Biopharm.* (1993). 39: 260–263.
- Sakr A *et al.* Comparative effectiveness of super disintegrants on the characteristics of directly compressed triamterene/hydrochlorothiazide tablets. *Pharm Ind.* (1993). 55: 953–957.
- Natoli D. Tooling for Pharmaceutical Processing. In: Augsburger LL, Hoag SW, eds. *Pharmaceutical Dosage Forms: Tablets, vol 3*. New York: Informa Healthcare, 2008: 1–48.
- Watt PR. *Tablet Machine Instrumentation in Pharmaceutics: Principles and Practice*. West Sussex: Ellis Horwood, 1988.
- United States Pharmacopoeia*, 34th ed, 2011.
- Berry IR, Nash RA. *Pharmaceutical Process Validation*. New York: Marcel Dekker, 1993.
- Hussain A. (2002). *PAT Subcommittee of the Advisory Committee for Pharmaceutical Science Meeting Presentation, February 25*. http://www.fda.gov/ohrms/dockets/ac/02/slides/3841s1_01_hussain/sld001.htm.
- Hausman DS *et al.* Application of Raman spectroscopy for on-line monitoring of low dose blend uniformity. *Int J Pharm.* (2005). 298: 80–90.
- Delaney R. Surveying consumer preferences. *Pharm Exec.* (1982). 2(3): 34.
- Ridgway K. *Hard Capsule Development and Technology*. London: Pharmaceutical Press, 1987; 49–60.
- Drakulich A. (2011). *Draft Guidance Issued on Capsule Bead Size, EPT-the Electronic Newsletter of Pharmaceutical Technology, Jan 20, 2011*. <http://pharmtech.findpharma.com/pharmtech/Drug+Delivery/Draft-Guidance-Issued-on-Capsule-Bead-Size/ArticleStandard/Article/detail/703988>.
- Tannenbaum PJ *et al.* The influence of dosage form on the activity of a diuretic agent. *Clin Pharmacol Ther.* (1968). 9: 598–604.
- United States Pharmacopoeia*, 34th edition, 2011.
- Tabibi E, Gupta LG. Soft gelatin capsules development. In: Rong L, ed. *Water-Insoluble Drug Formulation*. CRC Press, 2008: 590–606.
- Ebert WR. Soft elastic gelatin capsules: a unique dosage form. *Pharm Technol.* (1977). 1(10): 44–50.
- Jyothi NV *et al.* Microencapsulation techniques, factors influencing encapsulation efficiency. *J Microencapsul.* (2010). 27(3): 187–197.
- Li M. Microencapsulation by solvent evaporation: state of the art for process engineering approaches. *Int J Pharm.* (2008). 363 (1–2): 26–39.

Chapter 31

Coating of Pharmaceutical Dosage Forms

Stuart C. Porter, PhD

INTRODUCTION 611

EVOLUTION OF THE COATING PROCESS 611

PHARMACEUTICAL COATING PROCESSES 612

FILM-COATING PROBLEMS 616

COATING PROCEDURES AND EQUIPMENT 616

QUALITY CONTROL OF COATED TABLETS 619

STABILITY TESTING OF COATED PRODUCTS 619

RECENT TRENDS IN PHARMACEUTICAL COATING TECHNOLOGY 620

INTRODUCTION

Any introduction to tablet coating must be prefaced by an important question—*Why coat tablets?*²—since coatings are often applied to dosage forms that are already functionally complete. That said, a broad range of pharmaceutical oral solid dosage forms are coated, for a plethora of reasons that include

1. Protecting the drug from its surrounding environment (particularly air, moisture, and light) in order to improve stability.
2. Masking unpleasant taste and odor.
3. Making it easier for the patient to swallow the product.
4. Improving product identity, from the manufacturing plant, through intermediaries, and to the both healthcare workers and patients.
5. Facilitating handling, particularly in high-speed packaging/filling lines, and automated counters in pharmacies, where the coating minimizes cross-contamination due to dust elimination.
6. Improving product appearance, particularly where there are noticeable visible differences in tablet core ingredients from batch to batch.
7. Reducing the risk of interaction between incompatible components. This would be achieved by using coated forms of one or more of the offending ingredients (particularly active compounds).
8. Improving product robustness because coated products generally are more resistant to mishandling (abrasion, attrition, etc).
9. Modifying drug release, as in repeat-action, delayed release (enteric coated) and sustained-release products.

EVOLUTION OF THE COATING PROCESS

Coating of medicinal products is a concept steeped in history. Rhazes (850–932 AD) used the mucilage of psyllium seeds to coat pills that had an offending taste. Subsequently, Avicenna¹ was reported to have used gold and silver for pill coating. Since those early times, a wide range of materials has been used in *tablet coating*. White² mentioned the use of finely divided talc in what was at one time popularly known as *pearl coating*, while Kremers and Urdang³ described the introduction of the gelatin coating of pills by Garot in 1838.

An interesting reference⁴ reports the use of waxes to coat poison tablets. These waxes, being insoluble in all parts of the gastrointestinal tract, were intended to prevent accidental poisoning (the contents could be “activated” by breaking the tablet open prior to administration).

While earlier coated products were produced by individuals working in pharmacies, particularly when extemporaneous

compounding was the common practice, that responsibility now has been assumed by the pharmaceutical industry. Early attempts to apply coatings to pills yielded variable results and usually required the handling of individual pills. Such pills would have been mounted on a needle or held with a pair of forceps and literally dipped into the coating fluid, a procedure that would have to be repeated more than once to ensure that the pill was coated completely. Subsequently, the pills were held at the end of a suction tube, dipped, and then the process repeated for the other side of the pill. Not surprisingly, these techniques often failed to produce a uniformly coated product.⁵

Initially, the first sugar-coated pills seen in the United States were imported from France about 1842⁵; while Warner, a Philadelphia pharmacist, became among the first indigenous manufacturers in 1856.⁶

Pharmaceutical pan-coating processes were initially based on those used in the candy industry, where techniques were highly evolved, even in the Middle Ages. Candy coating processes typically used coating pans made of copper because drying was accomplished by applying an external heat source directly to the outside of the coating pan. Current pharmaceutical coating processes use coating pans, of a broad range of designs, made from stainless steel, where drying of the product being coated is achieved by means of a supply of heated air, and moisture and dust-laden air is removed from the vicinity of the pan by means of an air-extraction system.

Conventional pharmaceutical pan-coating processes, employed essentially for sugar coating, remained unchanged for the first half of the twentieth century. However, since then there have been significant advances made in coating-process technology, mainly as a result of a steady evolution in pan design and associated ancillary equipment, primarily control systems.

Until the early 1950s, pharmaceutical coating was dominated by the sugar coating of tablets. However, at that time, a new form of technology (called film coating) was developed. Recognizing the potential limitations of the sugar-coating process, the pioneers of film coating created significant improvements by employing coating formulations based on polymers dissolved in highly volatile organic solvents, with the result that coating processes that took days to carry out could now be completed in only a few hours.

While the use of organic solvents circumvented the problems associated with the poor drying capabilities of conventional equipment available at the time, this approach had its disadvantages; these organic solvents were highly flammable, potentially toxic, and presented significant handling and environmental problems.

Fortunately advances in equipment design, including the introduction of fluid-bed coating processes and those using side-vented pans coating pans, have resulted in the gradual emergence of coating processes where drying efficiencies have been maximized. The result has been the emergence of aqueous

coating processes as the preferred means, with a few exceptions, of coating pharmaceutical products

While advances in equipment design have resulted in film coating becoming the dominant process for coating pharmaceutical oral solid dosage forms, these improvements (in equipment design) have also benefited the sugar-coating process, creating fully automated processes that can produce a batch in less than one day.

PHARMACEUTICAL COATING PROCESSES

There are essentially four major techniques for applying coatings to pharmaceutical solid dosage forms: (1) sugar coating, (2) film coating, (3) microencapsulation, and (4) compression coating.

Although it could be argued that the use of mucilage of psyllium seed, gelatin, and so on, as already discussed, was an early form of film coating, *sugar coating* is regarded as the oldest method for applying pharmaceutical coatings. This process involves the application of aqueous solutions of sugar (typically sucrose, although sugar alcohols such as sorbitol may also be used). Formerly, sugar coating involved a significant number of individual applications of various coating formulations (resulting in tablet weight increases of 50–100%) and employing techniques that needed highly-skilled process operators. The result was that a batch of tablets could easily take up to 5 days to coat (although recent advances, as mentioned earlier, have now allowed this time to be reduced to less than one day).

Film coating, which involves the deposition of a thin polymeric film (comprising polymers, plasticizers, coloring agents and, sometimes, flavoring agents) onto the dosage form from solutions that were originally organic-solvent-based, but which now rely much more on water as the prime solvent. While the majority of film-coating formulations are intended for immediate-release applications (that is, they are not intended to affect drug release from the final dosage form), film coatings can also be used for modifying drug release from the final dosage form as well. Polymer systems used to modify drug release are not soluble in water and thus these systems are commonly formulated as aqueous-based suspensions.

Microencapsulation is a modified form of film coating, differing only in the size of the particles to be coated and the methods by which this is accomplished. This process is based on either mechanical methods such as pan coating, air-suspension techniques, multi-orifice centrifugal techniques, and modified spray-drying techniques, or physicochemical ones involving coacervation-phase separation, in which the material to be coated is suspended in a solution of the polymer. Phase separation is facilitated by the addition of a nonsolvent, incompatible polymer or inorganic salts or by altering the temperature of the system.

Compression coating involves the use of modified tableting machines that allow the compaction of a dry coating around a tablet core produced on the same machine. Essentially, a tablet core is produced on one side of the machine, is then transferred into a larger die partially filled with coating powder, additional coating powder is added on top of the core tablet, and the powdered coating material is finally compacted into a layer surrounding the core tablet. The main advantage of this type of coating process is that it eliminates the use of any solvent, whether aqueous or organic in nature. However, compression coating is mechanically complex and has not been commonly adopted as a method for coating tablets, although in recent times, modifications of this process have been found useful as a means of applying special coatings for novel drug-delivery applications.

SUGAR COATING OF COMPRESSED TABLETS

While the term *sugar* is somewhat generic and lends itself to describing a range of carbohydrate materials, sugar coating processes commonly use sucrose because it is one of the few materials that produces smooth, high-quality coatings that are essentially dry and tack-free at the end of the process. While the

popularity of sugar coating has certainly declined, this process is still used by many companies that have invested in the complete modernization of the process. In spite of certain inherent disadvantages of sugar coating, products that have been expertly sugar coated still remain among the most elegant available.

Since sugar coating is a multistep process, where esthetics of the final coated product is an important attribute, it has typically been highly dependent on the use of skilled process operators. Sugar coating is commonly a manual process, where coating materials (either in liquid or powder form) are applied by hand to the surface of the rotating bed of tablets. While sugar coating has historically been a lengthy and tedious process to undertake, processing times have been effectively reduced in recent years through process modification, typically involving thin sugar-coating procedures, and process automation (in which case, coating application is accomplished using automated dosing techniques).

The sugar-coating process is commonly divided into six main steps: (1) sealing, (2) subcoating, (3) smoothing, (4) color coating, (5) polishing, and (6) printing.

Sealing

This step involves the application of a specialized polymer coating directly to the tablet core for the purpose of separating tablet ingredients (primarily the drug) from water (which is a major constituent of subsequent processing steps) in order to achieve good product stability. A secondary function is to strengthen the tablet core. Sealing coats usually consist of alcoholic solutions (approximately 10–30% solids) of resins such as shellac, zein, cellulose acetate phthalate, or polyvinyl acetate phthalate.

Historically, shellac has been a popular choice of material for the sealing step; however this natural polymer can undergo further polymerization on storage, causing the seal coat to become completely insoluble to the point where bioavailability of the active pharmaceutical ingredient (API) may be compromised. This problem may be resolved by incorporating a small amount of polyvinylpyrrolidone (PVP) into the shellac-based seal coat formulation.

The quantities of material required to be applied as a sealing coat will depend primarily on tablet and batch size. Another variable is tablet porosity because highly porous tablets will tend to soak up the first application of solution, thus preventing it from spreading uniformly across the surface of every tablet in the batch. In this case, one or more further applications of resin solution may be necessary to ensure that the tablet cores are sealed effectively.

Because most sealing coats develop a degree of tack (stickiness) at some time during the drying process, it is usual to apply a dusting powder to prevent tablets from sticking together or to the pan. A common dusting powder is asbestos-free talc. Excessive use of talc may cause problems, firstly, by imparting a high degree of slip to the tablets, thus preventing them from rolling properly in the pan, and secondly by creating a surface that, at the beginning of the subsequent subcoating stage, is very difficult to wet. Such poor wetting often results in uneven subcoat buildup, particularly on the tablet edges. If there is a tendency for either of these problems to occur, one solution is to replace part or all of the talc with some other material such as terra alba, which will form a slightly rougher surface. Use of talc now is being frowned upon because of its potential health problems.

If it is necessary to prepare a delayed-release (enteric-coated) product, this can be achieved by making additional applications of the seal-coat solution. Under these circumstances, however, it is more preferable to use seal-coating formulations based on synthetic polymers such as polyvinyl acetate phthalate or cellulose acetate phthalate, rather than shellac because these have more suitable delayed-release characteristics.

Subcoating

Subcoating is a step in the sugar-coating process that can have a marked effect on ultimate tablet quality. Possibly 50% of the final coating weight gain is achieved during the subcoating stage.

Historically, subcoating has been accomplished by applying a “glue” (in the form of an aqueous solution of a suitable gum, such as gum acacia, or even elatin) to the sealed tablet cores: once this solution has been distributed uniformly throughout the tablet mass, it is followed by a liberal dusting of powder, which serves to reduce tack and facilitate tablet buildup. This procedure of application of gum solution, spreading, dusting, and drying is continued until the requisite buildup has been achieved. Thus, the final subcoating is a sandwich of alternate layers of gum and powder. Some examples of binder solutions are shown in Table 31-1 and those of dusting powder formulations in Table 31-2.

While this approach has proved to be very effective, particularly where there is difficulty in covering edges, a *lumpy* subcoat can be the result, especially if care is not taken during application of the alternate layers. Additionally, if the amount of dusting powder applied is not matched to the binding capacity of the gum solution, not only will the ultimate coating be brittle, but also dust will collect in the back of the coating pan, and by becoming reattached to the tablet surface during subsequent applications of the gum solution, contribute to excessive roughness. An alternative approach, particularly when using an automated dosing system, involves the application of a suspension subcoat formulation (essentially a coating formulation where the powdered materials responsible for coating buildup are dispersed into the gum-based solution). An example of a suspension subcoating formulation is shown in Table 31-3. Employing suspension subcoating formulations not only allows the solids loading to be matched more closely to the binding capacity of the base solution, but often enables the less-experienced operator to achieve satisfactory results.

Smoothing

Depending on the success achieved during subcoating, especially in terms of achieving a smooth surface, it may be necessary to smooth out the tablet surface further prior to application of the color coating. Smoothing usually can be accomplished

by the application of a simple syrup solution (approximately 60–70% sugar solids).

Often, the smoothing syrups contain a low percentage of titanium dioxide (1–5%) as an opacifier. This can be particularly useful when the subsequent color-coating formulation uses water-soluble dyes as colorants because a white, opaque smoothing coat creates a more reflective surface, resulting in a brighter, cleaner final color.

Color Coating

This stage is often the most critical in sugar coating because it has the greatest impact on the final visual effect, and involves the multiple application of syrup solutions (60–70% sugar solids) containing appropriate coloring materials. The types of coloring materials used can be divided into two categories: water-soluble dyes or water-insoluble pigments. The nature of the colorant selected often defines the type of color-coating procedure to be employed.

When used by a skilled artisan, water-soluble dyes produce the most elegant sugar-coated tablets because it is possible to obtain a cleaner, brighter final color. However because water-soluble dyes are migratory colorants (that is to say, moisture that is removed from the coating on drying will cause migration of the colorant, resulting in a nonuniform appearance), great care must be exercised in their use, particularly when dark color shades are required. Such care can be achieved by applying small quantities of colored syrup that are just sufficient to wet the surface of every tablet in the batch, and then allowing the tablets to dry slowly. It is essential that each application be allowed to dry thoroughly before subsequent applications are made, otherwise moisture may become trapped in the coating and may cause the tablets to *sweat* on standing.

When using water-soluble dyes, the final color obtained may be the result of up to 60 individual applications of colored syrup. This factor, combined with the need to dry each application slowly and thoroughly, results in very long processing times (e.g., assuming 50 applications are made, which can take between 15 and 30 min each, the coloring process can take up to 25 hours to complete). The more recent introduction of preformulated dye-based coloring systems has obviated many of these problems.

Tablet color coating with water-insoluble pigments offers some significant advantages. First, because pigments are insoluble in water, color migration during the drying step is essentially eliminated. In addition, if the pigment is opaque or is combined with an opacifier such as titanium dioxide, the desired color can be developed much more rapidly, thus resulting in a thinner color coat. Since each colored syrup application now can be dried more rapidly, and fewer applications are required, significant reductions can be achieved in both processing times and costs.

Although water-insoluble pigment-based color coatings are by no means foolproof, they will permit more abuse than a dye color-coating process and are easier to use by less-skilled coating operators. Pharmaceutically acceptable pigments can be classified either as inorganic pigments (e.g., titanium dioxide, iron oxides) or certified lakes. Certified lakes are produced from water-soluble dyes using a process known as *laking*, whereby each dye molecule is bonded to the surface of a suitable insoluble substrate (such as alumina hydrate, a material chemically very similar to the aluminum hydroxide used in many antacid formulations).

Certified lakes, particularly when used in conjunction with an opacifier such as titanium dioxide, provide an excellent means of coloring sugar coatings and permit a wide range of shades to be achieved. However, the incorporation of pigments into the syrup solution is not as easy as with water-soluble dyes because it is necessary to ensure that the pigment is wetted completely and dispersed uniformly. Thus, the use of pigment color concentrates, which are commercially available, is usually beneficial.

Table 31-1. Binder Solution Formulations for Subcoating

	A, % W/W	B, % W/W
Gelatin	3.3	6.0
Gum acacia (powdered)	8.7	8.0
Sucrose	55.3	45.0
Water	to 100.0	to 100.0

Table 31-2. Dusting Powder Formulations for Subcoating

	A, % W/W	B, % W/W
Calcium carbonate	40.0	—
Titanium dioxide	5.0	1.0
Talc (asbestos-free)	25.0	61.0
Sucrose (powdered)	28.0	38.0
Gum acacia (powdered)	2.0	—

Table 31-3. Typical Suspension Subcoating Formulation

	% W/W
Distilled water	25.0
Sucrose	40.0
Calcium carbonate	20.0
Talc (asbestos-free)	12.0
Gum acacia (powdered)	2.0
Titanium dioxide	1.0

Polishing

Sugar-coated tablets are, by nature, very dull in appearance, and thus need to be polished to achieve a glossy finish. Polishing is accomplished by applying mixtures of waxes (beeswax, carnauba wax, candelilla wax, or hard paraffin wax) to the tablets in a polishing pan. Such wax mixtures may be applied as powders (usually in a finely milled form) or as solutions/dispersions in various organic solvents.

Printing

To identify sugar-coated tablets (in addition to shape, size, and color) often it is necessary to print them, either before or after polishing, using pharmaceutical branding inks, by means of the process of *offset rotogravure*.

Sugar-Coating Problems

Various problems may be encountered during the sugar coating of tablets. Any process in which tablets are kept tumbling constantly can cause problems if the tablets are not strong enough to withstand the applied stress. Tablets that are too soft or have a tendency to laminate may break up and the fragments adhere to the surface of otherwise good tablets; thus a few bad tablets in the batch can end up spoiling the whole batch.

Traditional sugar-coating pans exhibit inherently poor mixing characteristics. The consequence of this is that if care is not exercised during the application of the various coating formulations, the distribution of coating materials throughout the batch can be nonuniform, creating finished tablets of varying sizes.

Excessive use of dusting powders may result in a coatings being formed in which the quantity of fillers exceeds the binding capacity of the coating formulation, creating soft coatings or those with an increased tendency to crack.

Irregularities in appearance are not uncommon and occur either as the result of color migration during drying when water-soluble dyes are used, or of *washing back* when overdosing of colored syrups causes the previously dried coating layers to be re-dissolved. Rough tablet surfaces will produce a *marbled* appearance during polishing because wax buildup occurs in the small depressions in the tablet surface.

FILM COATING OF SOLID DOSAGE FORMS

Film coating is a process involving the deposition of a thin, uniform coating onto the surface of the substrate. Unlike sugar coating, which is typically only applicable to the coating of tablets, film coating has greater flexibility in allowing a broad range of products (e.g., tablets, powders, granules, nonpareils, capsules) to be coated. Film coatings are typically applied continuously to a moving mass of product, usually by means of a spray atomization technique, although manual application procedures have been used.

Historically, film coating was introduced in the early 1950s to combat the shortcomings of the then predominant sugar-coating process. Film coating has proved successful as a result of the many advantages offered, including

1. Minimal weight increase (typically 2–3% of tablet core weight)
2. Significant reduction in processing times
3. Increased process efficiency and output
4. Increased flexibility in formulations
5. Improved resistance to chipping of the coating.

In the early years of film coating, the major process advantages resulted from the greater volatility of the organic solvents used. However, the use of organic solvents is not without its disadvantages, including:

1. Flammability hazards
2. Toxicity hazards

3. Concerns over environmental pollution
4. Cost (relating to minimizing items 1 to 3 and to the cost of the solvents themselves).

However, subsequent to its initial introduction, film coating has benefited from significant advances in process technology and equipment design, with the result that less volatile, and safer, solvents such as water can be easily accommodated by the process.

Film Coating Raw Materials

The major components in any film-coating formulation consist essentially of a polymer, plasticizer, colorant, and solvent (or vehicle).

Ideal properties for the polymer include solubility in a wide range of solvent systems to promote flexibility in formulation, an ability to produce coatings that have suitable mechanical properties, and appropriate solubility in gastrointestinal fluids so that key objectives with respect to bioavailability can be achieved.

Cellulose-based polymers are commonly used in film coating, particularly hydroxypropyl methylcellulose. Suitable substitutes are hydroxypropyl cellulose, which may produce slightly tackier coatings, and methylcellulose. Alternatives to the cellulose polymers are acrylic copolymers (such as methacrylate and methyl methacrylate copolymers) and vinyl polymers (such as polyvinyl alcohol, polyvinylpyrrolidone-vinyl acetate copolymer, and polyvinyl alcohol-PEG copolymer).

For immediate-release (IR) film-coating applications, where there is no intent to modify drug-release characteristics, polymers are soluble in water and typically used as solutions in either water (preferred) or organic solvents.

Many of the commonly used polymers are available in a range of molecular-weight grades, a factor that also must be considered in the selection process. Molecular weight may have an important influence on various properties of the coating system, such as solution viscosity and mechanical strength and flexibility of the resultant film.

The incorporation of a plasticizer into the formulation improves the flexibility of the coating, reduces the risk of the film cracking, and potentially improves adhesion of the film to the substrate. To ensure that these benefits are achieved, the plasticizer must show a high degree of compatibility with the polymer and be retained permanently in the film, if the properties of the coating are to remain consistent on storage. Examples of typical plasticizers include glycerin, propylene glycol, polyethylene glycols, triacetin, acetylated monoglycerides, citrate esters (e.g., triethyl citrate), or phthalate esters (e.g., diethyl phthalate).

Colorants are usually added to improve product appearance, aid in product identification, and improve coated product stability. As in the case of sugar coating, colorants can be classified as either water-soluble dyes or insoluble pigments.

The use of water-soluble dyes is precluded with organic solvent-based film coating because of the lack of solubility in the solvent system. Thus, the use of pigments, particularly aluminum lakes and inorganic pigments (such as titanium dioxide and iron oxides), provides the most useful means of coloring film-coating systems. Although it may seem obvious to use water-soluble dyes in aqueous formulations, the use of insoluble pigments is preferred because:

1. They can reduce the permeability of the coating to moisture, thus potentially improving product stability.
2. They serve as bulking agents to increase the overall solids content in the coating dispersion without dramatically increasing viscosity.
3. They tend to be more light stable.
4. They are generally opaque, and thus can help to improve the stability of photolabile APIs, such as nifedipine.

The major solvents used in film coating typically belong to one of these classes: alcohols (such as methanol, ethanol, and

isopropyl alcohol), ketones (such as acetone), esters (such as ethyl acetate), chlorinated hydrocarbons (such as dichloromethane), and water. Solvents perform an important function in the film-coating process because they facilitate the application of the coating to the surface of the substrate. Good interaction between solvent and polymer is necessary to ensure that optimal film properties are obtained when the polymer solution dries and a solid film coating is formed. The initial interaction between solvent and polymer will yield maximum polymer chain extension, producing films having the greatest cohesive strength and, thus, the best mechanical properties.

Although it is very difficult to give typical examples of film-coating formulations because these will depend on the properties of the materials used, such formulations usually are based on 5–25% (w/w) coating solids in the requisite vehicle (with the higher concentration range preferred for aqueous formulations), of which 60–70% is polymer, 6–7% is plasticizer, and 20–30% is pigment.

Over the last 30 years, the general trend with respect to film-coating formulations for IR applications is for pharmaceutical companies to buy ready-made coatings (these are usually dry powders, comprising polymers, plasticizers, and colorants that are reconstituted at the point of use by dispersion in water). With a major focus on cost reduction in film coating, many of these coating systems can be used at higher concentrations in water (typically 18–25% w/w solids, compared to the 10–15% w/w solids experienced with regular coating formulations), enabling significant reductions in processing time to be achieved, as suggested by Porter.⁷

MODIFIED-RELEASE FILM COATINGS

Film coatings can be applied to pharmaceutical products to modify drug release. The USP describes two types of modified-release dosage forms, namely those that are *delayed release* and those that are *extended release*. Delayed-release products often are designed to prevent drug release in the upper part of the gastrointestinal (GI) tract. Film coatings used to prepare this type of dosage form are commonly called *enteric coatings*, while film coatings that are required to extend drug release over a long period of time (from 6 to 24 hours) are commonly called *sustained- or controlled-release* film coatings.

Enteric Coatings

Enteric coatings generally remain intact in the stomach but will dissolve and release the contents of the dosage form once it reaches the small intestine. The purpose of an enteric coating is to delay the release of drugs that are inactivated by the stomach contents, (e.g., pancreatin, erythromycin, and substituted benzimidazole compounds, such as proton pump inhibitors) or may cause nausea or bleeding by irritating the gastric mucosa (e.g., aspirin, and many nonsteroidal anti-inflammatory drug substances). In addition, such coatings can be used to give a simple repeat-action; in this case, a tablet core containing part of the dose is enteric coated, and then additional API is applied to the surface of the enteric coated tablet, with the result that the outer-most layer of API will be released immediately on ingestion, while the remainder of the dose will only be released once the enteric coating has dissolved.

The functionality of enteric coatings stems from the way the polymers react to differences in composition of the respective gastric and intestinal environments, especially with regard to pH and enzyme content. Although there have been repeated attempts to produce coatings that are susceptible to enzymatic breakdown, this approach is not popular because such breakdown of the coating can be relatively slow. Thus, most currently used enteric coatings are weak acids that remain undissociated and thus insoluble in the low pH environment of the stomach but readily ionize and dissolve when the pH rises to above 5. The most effective enteric polymers are polyacids having a pK_a of about 5. Coatings that respond to enzymatic breakdown are

now being considered as protective coatings suitable for the controlled delivery of polypeptide drugs.

Historically, the earliest enteric coatings used formalin-treated gelatin, but this approach was unreliable because the polymerization of gelatin could not be controlled accurately and often resulted in failure to release the drug, even in the lower intestinal tract. Another early candidate was shellac, but again the main disadvantage resulted from further polymerization that occurred on storage, often resulting in failure to release the active contents. Pharmaceutical formulators now prefer to use synthetic polymers to prepare more effective enteric coatings.

One of the oldest synthetic polymers used for enteric coating is cellulose acetate phthalate (CAP). However, a pH greater than 6 usually is required to allow the coating to dissolve, and thus a significant delay in drug release may ensue. It also is relatively permeable to moisture and gastric fluid compared to other enteric polymers, and it is very susceptible to hydrolytic decomposition (where phthalic and acetic acids are split off from the polymer chains), resulting in a change in polymer properties, and thus enteric coating performance.

Other useful enteric-coating polymers include polyvinyl acetate phthalate (PVAP, which is less permeable to moisture and gastric fluid, more stable to hydrolysis, and able to ionize at a lower pH); hydroxypropyl methylcellulose phthalate (HPMCP, which has properties similar to PVAP); acrylic copolymers, such as methacrylic acid-methacrylic acid ester copolymers (some of which have a high dissociation constant); cellulose acetate trimellitate (CAT, which has properties similar to CAP); carboxymethyl ethylcellulose (CMEC); and hydroxypropyl methylcellulose acetate succinate (HPMCAS). Today, acrylic copolymers are often preferred (in terms of performance and global acceptability) for designing enteric coating formulations.

Since enteric coating polymers are, by nature, insoluble in water (except at a pH >5), their use in aqueous coating systems is predicated by the use of either liquid polymer dispersions (sometimes called latices or pseudolatices) or dry powder coating systems that can readily be dispersed in water prior to use.

Sustained-Release Coatings

The concept of sustained-release formulations was developed to eliminate the need for multiple dosage regimens, particularly for those drugs requiring reasonably constant blood levels over a long period of time. In addition, it also has been adopted for those drugs that need to be administered in high doses, but where too rapid a release is likely to cause undesirable side effects (e.g., the ulceration that occurs when potassium chloride is released rapidly into the milieu of the gastrointestinal tract).

Formulation methods used to obtain the desired drug release rate from sustained-action dosage forms include

1. Increasing the particle size of the drug.
2. Embedding the drug in a matrix.
3. Coating the drug or dosage form containing the drug.
4. Forming complexes of the drug with materials such as ion-exchange resins.

Only those methods that involve some form of coating fall within the scope of this chapter. A discussion of other controlled release drug delivery systems can be found in Chapter 32 (Oral Modified Release Drug Delivery Systems). The mechanisms of drug release from film-coated products are also provided.

Materials that have been found suitable for producing sustained-release coatings include:

1. Mixtures of waxes (beeswax, carnauba wax, etc.) with glyceryl monostearate, stearic acid, palmitic acid, glyceryl monopalmitate, and cetyl alcohol. These provide coatings that dissolve slowly or are broken down by other means in the GI tract.
2. Shellac and zein. These polymers remain intact until the pH of gastrointestinal contents becomes less acidic.

3. Ethylcellulose, which provides a membrane around the dosage form and remains intact throughout the gastrointestinal tract. However, it does permit water to permeate the film, dissolve the drug, and diffuse out with the dissolved drug.
4. Acrylic resins, which behave similarly to ethylcellulose as a diffusion-controlled drug-release coating material.
5. Cellulose acetate (diacetate and triacetate), which can function as semi-permeable membranes.
6. Silicone elastomers.
7. Polyvinyl acetate.

As with enteric coatings, many of the synthetic polymers suitable for sustained-release film-coating applications are available as aqueous polymer dispersions that can be used in aqueous coating processes.

Various methods have been used to prepare sustained-release products using film-coating techniques. Examples include the application of suitable film coatings to:

1. Dried granules (either irregular or spheronized)⁸
2. Drug-loaded beads (or nonpareils)
3. Drug crystals
4. Drug/ion-exchange-resin complexes
5. Tablets, including mini tablets⁹

In the first four examples, the final coated particles can be either filled into two-piece hard-gelatin capsules or compacted into tablets. Additionally, coated drug/ion-exchange-resin complexes may be dispersed in viscous liquids to create liquid suspensions.

An interesting application of the film-coated, sustained-release tablet is the osmotic pump concept as described by Wong *et al.*¹⁰ In this device, a tablet core (formulated to contain osmotically active ingredients) is film coated with a semi-permeable membrane. This membrane is subsequently *pierced* with a laser to create a delivery orifice. Once such a device is ingested, the infusion of water generates an osmotic pressure within the coated tablet that *pumps* the drug out through the orifice.

With sustained-release products, one must remain aware constantly of the fact that the final dosage forms typically contain drug loadings that are sufficiently high to cause problems if the entire dose is released quickly. This phenomenon, commonly called *dose-dumping*, can be avoided only if:

1. The film coating is mechanically sound and will resist rupture on ingestion of the dosage form.
2. Sufficient coating is applied uniformly across the surface of the material that is to be coated.
3. The dosage form is not chewed or crushed prior to ingestion.

FILM-COATING PROBLEMS

As with sugar coating, problems may occur during, or subsequent to, the film-coating process. Porter¹¹ has provided a comprehensive review of common problems experienced with film-coated products.

The tablets being coated may not be sufficiently robust or may have a tendency to *laminate* while being coated. Since film coats are relatively thin, their ability to hide defects is significantly less than that of sugar coating. Hence, tablets that have poor resistance to abrasion (i.e., they exhibit high friability characteristics) can be problematic because imperfections may readily be apparent after coating. It is very important to identify tablets with suspect properties, whether mechanical or performance related (e.g., poor dissolution), prior to a coating process because subsequent recovery or reworking of tablets may be extremely difficult after a coating has been applied.

Various process-related problems can occur during the application of a film coating. One example is *picking*, which is

a consequence of the fluid delivery rate exceeding the drying capacity of the process, causing tablets to stick together and subsequently break apart. Another example, *orange peel* or *roughness*, is usually the result of premature drying of atomized droplets of solution, or it may be a consequence of spraying too viscous a coating solution such that effective atomization is difficult.

Mottling, or lack of color uniformity, can result from uneven distribution of color in the coating, a problem often related to the use of soluble dyes in aqueous film coating, when color migration can occur, either by evolution of residual solvent in the film or by migration of plasticizer in which the colorant may be soluble. The use of pigments in the film-coating process minimizes the incidence of this latter objection considerably. However, uneven color also can result from inadequate dispersion of the pigments in the coating solution.

Finally, some major problems occur as the result of internal stresses that develop within the film as it dries. One example is *cracking*, which occurs when these stresses exceed the tensile strength of the film. This problem may be compounded by the existence of post compaction strain relaxation (a phenomenon that can occur with certain types of tablet formulations, such as those containing ibuprofen, after ejection from the die during the tableting process), which causes tablets to expand. Another example is *logo-bridging* (i.e., bridging of monograms present in the surface of the tablet core), which occurs when the internal stresses are able to overcome the adhesive bonds formed between the coating and the tablet surface, causing the film to pull away so that legibility of the monogram is lost. An understanding of the properties of the various ingredients used in the film-coating formulation and how these ingredients interact with one another can allow the formulator to avoid many of these internal-stress-related problems.

Porter and Terzian¹² have described a systematic approach that may be used to solve common film-coating problems.

COATING PROCEDURES AND EQUIPMENT

COATING PANS

Sugar coating historically has involved the ladling of the various coating fluids onto a cascading bed of tablets in a conventional coating pan (Fig 31-1) fitted with a means of supplying drying air to the tablets and an exhaust to remove moisture and dust-laden air from the pan.

Typically, after the requisite volume of liquid has been applied, a period of time is allowed for the tablets to mix and the liquid to become fully dispersed throughout the batch. To facilitate the uniform transfer of liquid, the tablets often are *stirred* by hand, or in larger pans, by means of a rake, to overcome mixing problems often associated with *dead spots*, an inherent problem seen with conventional pans. Finally, tablets are dried by blowing air onto the surface of the tablet bed. Thus, sugar coating is a sequential process consisting of consecutive cycles of liquid application, mixing, and drying.

During the early history of film coating, the equipment used was adapted essentially from that already employed for sugar coating. Although ladling of coating liquids during the film-coating process has been practiced, usually the liquid is applied using a spray technique. Spray equipment used is essentially of two types:

1. Airless (or hydraulic) spray, where the coating liquid is pumped under pressure to a spray nozzle with a small orifice and atomization of the liquid occurs as it expands rapidly on emerging from the nozzle. This is analogous to the effect achieved when one places one's finger over the end of a garden hose.
2. Air spray, where liquid is pumped, under little or no pressure, to the nozzle and is subsequently atomized by means of a blast of compressed air that makes contact

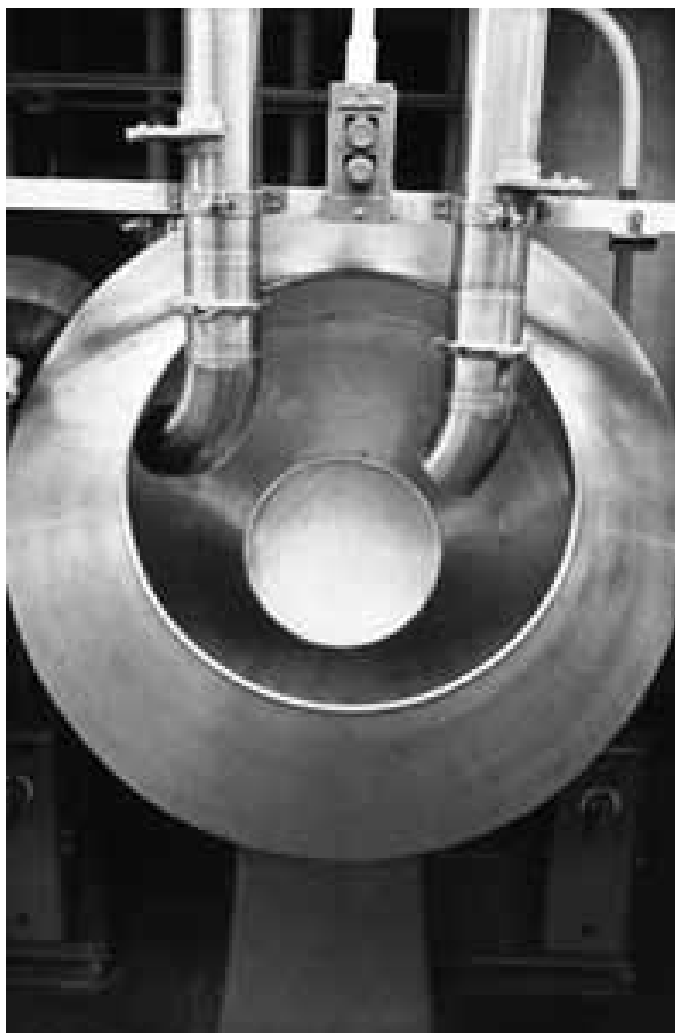


Figure 31-1. Typical equipment setup for conventional sugar coating.

with the stream of liquid as it passes from the nozzle aperture.

Airless-spray techniques typically are used in large-scale film-coating operations employing organic solvents, while air-spray techniques are more effective in both small-scale laboratory set-up or all aqueous film-coating operations.

Spray application enables finely atomized droplets of the coating solution to be delivered across the surface of the moving tablet mass in a manner that achieves uniform coverage while preventing adjacent tablets from sticking together as the coating solution is rapidly dried. While all the events that take place during the spray application process occur continuously and concurrently, the overall picture can be more simply represented as shown in Figure 31-2.

The spray process can be operated either intermittently or continuously. In the early years of film coating, the lack of adequate drying conditions inside the coating apparatus, together with the preference for using airless coating techniques (with their inherently higher delivery rates) with organic solvent-based formulations often required the use of intermittent spray procedures. This technique allowed excess solvent to be removed during the nonspray part of the cycle and thus reduced the risk of *picking* and the tendency for tablets to stick together. However, in later years, improvements in drying capabilities have resulted in the preferred use of continuous spray procedures, as this permits uniform coatings to be applied in a shorter process.

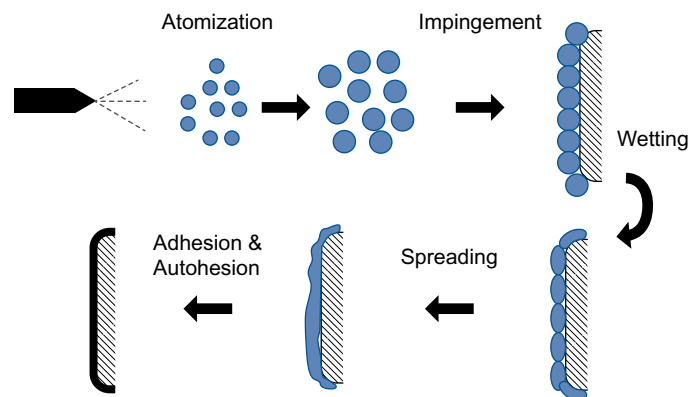


Figure 31-2. Schematic representation of the film-coating process.

As mentioned earlier, pan equipment initially was completely conventional in design and, with the exception of the addition of spray-application equipment, was similar to that used in sugar coating. Fortunately, film-coating formulations were based on relatively volatile organic solvents, which enabled acceptable processing times to be achieved in spite of the relative deficiencies of the air-handling systems. However, such equipment did not produce a completely enclosed system, a fact that made effective solvent containment extremely difficult to achieve. Although conventional pans possessed acceptable properties with regard to mixing of the tablet mass in the sugar-coating process (particularly as this could be augmented by manual stirring of the tablets during processing), they were less suited to meet the more rigorous demands of the film-coating process, even when some simple baffle system was installed.

The introduction of aqueous film coating in the latter part of the twentieth century presented a more serious challenge to the continued use of conventional processing equipment. Limitations in both drying and mixing capabilities potentially resulted in significantly increased processing times while substantially compromising product quality and long-term stability. Fortunately, these problems have been eliminated as coating-pan design has evolved and coating process technologies have improved.

Although conventional, solid-wall coating pans had their limitations, the design shown in Figure 31-3, commonly called the Pellegrini pan and being a somewhat angular pan rotating on a horizontal axis, provided significant improvements for the sugar-coating process and was more easily adaptable to film coating because the presence of an integral baffle system ensures that more uniform mixing is achieved. Additionally, because process drying air and extraction are introduced through the rear opening, the front can potentially be closed off to produce a more-or-less enclosed coating system. Pellegrini pans are available with capacities ranging from the 10-kg laboratory scale-up to 1000 kg for high-output production.

In solid-wall coating pan designs, including that shown in Figure 31-3, drying air is blown only onto the surface of the tablet bed, thus inherently limiting drying capacity. To offset this limitation, several attempts have been made to improve air exchange, particularly within the tablet bed. One such approach, shown schematically in Figure 31-4, allows one of the air plenums (inlet or exhaust) to be immersed into the tablet bed, thus improving drying capacity.

Two such types of equipment, both based on the Pellegrini style of coating pan, are supplied by IMA (called the GS Coater) and Nicomac.

A major advance in pan coating technology occurred with the introduction of the side-vented pan concept, an innovation developed by Eli Lilly. This concept, originally called the Accela-cota, became the basis for a wide range of *side-vented* coating

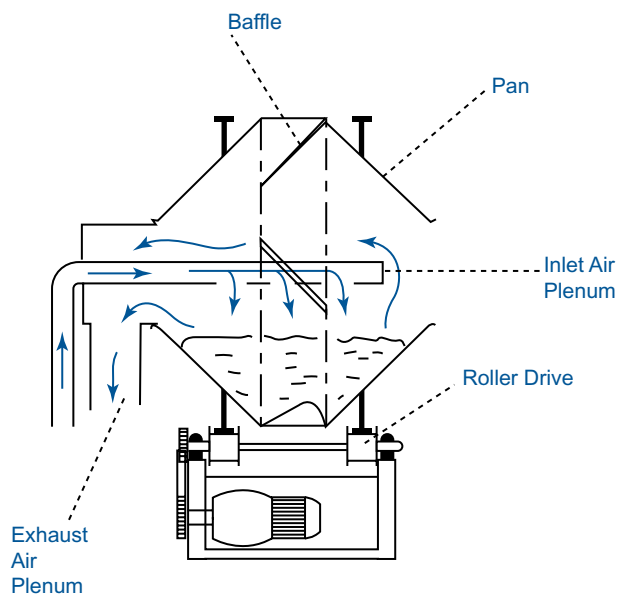


Figure 31-3. Schematic diagram of a Pellegrini coating pan.

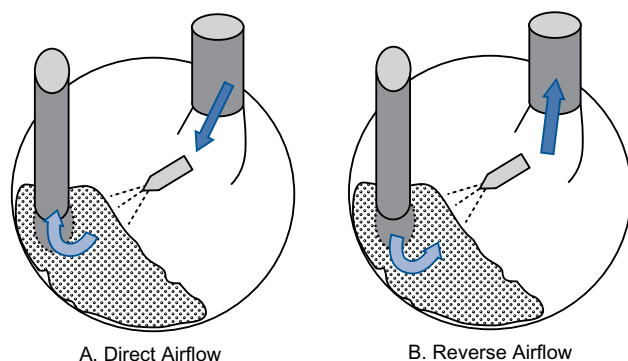


Figure 31-4. Schematics of upgraded conventional coating pans. A: Direct airflow. B: Reverse airflow.

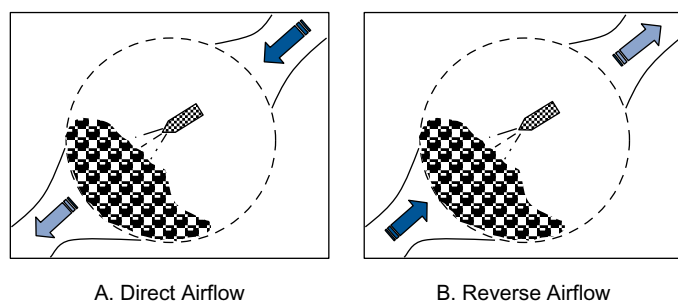


Figure 31-5. Schematics of side-vented coating pans. A: Direct airflow. B: Reverse airflow.

pan designs, a schematic of which is shown in Figure 31-5. The salient features of side-vented coating pans are:

1. An angular pan (fitted with an integral baffle system) that rotates on a horizontal axis.
2. A coating system that is completely enclosed.
3. A perforated pan that allows drying air to be pulled through a cascading bed of tablets while the coating liquid is applied to the tablet surface using a spray-atomization technique.

Side-vented coating pans exhibit dramatically improved drying characteristics, a feature that facilitated the successful adoption of aqueous film-coating technology. Manufacturers of side-vented coating pans include Thomas Engineering, Oystar-Manesty, O'Hara, Glatt, Dumoulin, Vector Freund, Driam, IMA, and Nicomac, to name but a few. A comprehensive overview of coating equipment that can be used for pharmaceutical applications has been provided in a recent publication.¹³

Ongoing trends with side-vented coating pans have produced:

1. Designs that permit multidirectional air flow.
2. Fully automated, computerized coating processes (especially for production-scale coating purposes).
3. Effective clean-in-place (CIP) or wash-in-place (WIP) cleaning systems that facilitate compliance with good manufacturing processes (GMPs).
4. Laboratory-scale coating equipment provided with interchangeable coating pans representing batch capacities in the range of 3–40 kg (depending on product density).
5. Coating pans designed to permit continuous processing (where the product is constantly introduced into one end and flows, fully coated, out the other).

Although improvements in coating-pan design have predominately occurred to improve the aqueous film-coating process, they have also benefited other processes, including sugar coating.

FLUIDIZED-BED COATING EQUIPMENT

Fluid-bed processing technology has long been used in the pharmaceutical industry, initially for drying, and subsequently for granulation. While several attempts have been made to apply this technology to the film-coating process, a significant advance came with the introduction of the Wurster concept (a schematic of which is shown in Figure 31-6) in the 1950s.

At a time when organic-solvent-based coating formulations were still primarily used, the Wurster process was extremely popular for coating a variety of pharmaceutical dosage forms, especially tablets. Although fluid-bed processing inarguably exhibits the most effective drying characteristics of any film-coating process, the introduction of aqueous coating formulations initially created waning interest in using the Wurster process for coating tablets. A major factor in this trend undoubtedly was related to the increased potential (compared with use of coating pans) for tablet breakage in the fluid-bed process. During the last 30 years, however, resurgent interest in the Wurster process has occurred as a result of the growing demand for applying film coatings to pellets, granules, and powders (so-

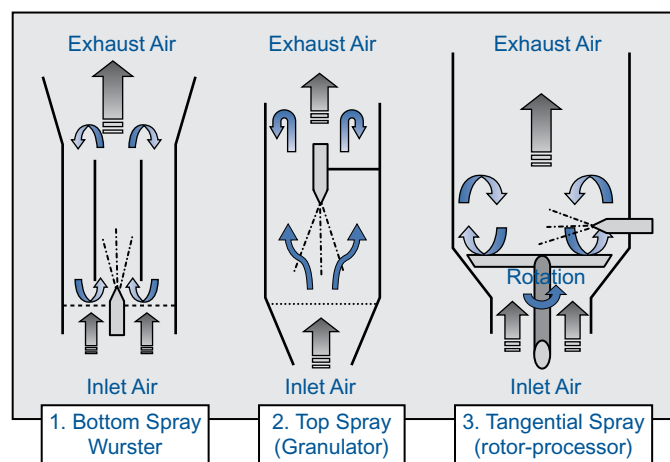


Figure 31-6. Schematic diagrams for three basic processes used for fluid-bed film coating.

called *multiparticulates*) when producing modified-release coated dosage forms.

The suitability of the fluid-bed process for film coating multiparticulates also has generated interest in processes other than the Wurster for this application. In particular, modifications of the spray granulation process (often termed the *top-spray coating process*) and a rotary process (often called the *tangential spray process*) have both been used for the film coating of multiparticulates. Schematics for all these processes also are shown in Figure 31-6.

The major manufacturers of fluid-bed processing equipment (Glatt Air Techniques, Vector Corporation, and GEA) all have adopted a principle in which a basic processing unit is designed to accept modular inserts for each of the three fluid-bed coating processes shown in Figure 31-6. Selection of a particular type of insert often is determined by the nature and intended functionality of the coating applied; for example

1. Granulator Top-Spray Process—preferred when a taste-masking coating is being applied; additionally suitable for the application of hot-melt coatings.
2. Wurster, Bottom-Spray Process—preferred for the application of modified-release coatings to a wide variety of multiparticulates; also suitable for drug layering when the drug dose is in the low-to-medium range.
3. Rotor, Tangential-Spray Process—suitable for the application of modified-release film coatings to a wide range of multiparticulate products; ideal for drug layering when the dose is medium to high; also useful as a spheronizing process for producing spheroidal pellets from powders.

POTENTIAL FOR TOTALLY AUTOMATED COATING SYSTEMS

Over the course of time, the pharmaceutical industry has witnessed a general transition away from manually operated sugar-coating processes, requiring total operator involvement, to film-coating processes, where operator intervention is infrequent. Increasing familiarity with, and understanding of, tablet coating as a unit process, and a desire to ensure compliance with GMPs, ultimately have increased the desire to achieve reproducible and consistent conformity to design specifications for every batch of product made; this is especially the case if the requirements of more recent FDA initiatives, such as Quality by Design (QbD), are to be met. Meeting these requirements is clearly compromised if the idiosyncrasies of individual process operators are allowed to have a less-than-positive impact on final product quality (in its broadest sense).

Employing the tenets of QbD, by designing fully optimized processes, creating an effective design space, and implementing effective Process Analytical Technologies (PAT) all have a positive impact on process control. Process automation, especially when responding to appropriate feed-back and feed-forward directives from a well designed process that is being monitored by suitable PAT capabilities, contributes effectively to the whole concept of QbD.

Since a sugar-coating process always has been highly operator dependent, such dependence can be minimized through automation. However, the natural sequencing of events that are the basis of the sugar-coating process adds a level of complexity when considering the implementation of automation. Nonetheless, implementation of PAT, such as the effective use of NIR sensors, simplifies these challenges when it comes to automation.

QUALITY CONTROL OF COATED TABLETS

The most important aspects of coated tablets that must be assessed from a quality-control standpoint are appearance characteristics and drug availability. From the appearance standpoint, coated tablets must be shown to conform, where applicable, to some color standard, otherwise the dispenser and the consumer may assume that differences have occurred

from previous lots, signifying a changed or substandard product. In addition, because of the physical abuse that tablets, both in their uncoated and coated forms, receive during the coating process, it is essential to check for defects such as chipped edges and picking, and ensure that they do not exceed predetermined limits.

Often, to identify the products, coated tablets may be imprinted (particularly with sugar-coated tablets) or bear a monogram (commonly seen with tablets that are film-coated). The clarity and quality of such identifying features must be assessed. The failure of a batch of coated tablets to comply with such preset standards may result in 100% inspection being required or the need for the batch to be reworked.

Batch-to-batch reproducibility for drug availability is of paramount importance; consequently each batch of product should be submitted to some meaningful test such as a dissolution test. Depending on the characteristics of the tablet core to be coated, tablet coatings can modify the drug release profile, even when not intended (unlike the case of enteric- or controlled-release products). Since this behavior may vary with each batch coated (being dependent, for example, on differences in processing conditions or variability in raw materials used), it is essential that this parameter be assessed, particularly in products that are typically borderline (refer to Chapter 30 Oral Solid Dosage Forms).

A review of various techniques that may be used to assess the quality of coated products, especially film-coated tablets, has been provided by Porter and Felton.¹⁴

STABILITY TESTING OF COATED PRODUCTS

The stability-testing program for coated products will vary depending on the dosage form and its composition. Many stability-testing programs are based on studies that have disclosed the conditions a product may encounter prior to end use. Such conditions usually are referred to as normal and include ranges in temperature, humidity, light, and handling conditions. The conditions to be employed in modern stability-testing programs often conform to the guidelines established by the International Committee on Harmonization (ICH). A more detailed discussion on the stability of pharmaceutical products may be found in Chapter 4.

Limits of acceptability are established for each product for qualities such as color, appearance, availability of drug for absorption, and drug content. The time over which the product retains specified properties, when tested at normal conditions, may be defined as the *shelf life*. The container for the product may be designed to improve the shelf life. For example, if the color in the coating is light-sensitive, the product may be packaged in an amber bottle and/or protected from light by using a paper carton. When the coating is friable, resilient material such as cotton may be incorporated in both the top and bottom of the container, and if the product is affected adversely by moisture, a moisture-resistant closure may be used and/or a desiccant may be placed in the package. The shelf life of the product is determined in the commercial package tested under normal conditions.

The stability of the product also may be tested under exaggerated conditions. This usually is done for the purpose of accelerating changes so that an extrapolation can be made early, concerning the shelf life of the product. Although useful, highly exaggerated conditions of storage can supply misleading data for coated dosage forms. Any change in drug release from the dosage form is measured *in vitro*, but an *in vivo* measurement should be used to confirm that drug availability remains within specified limits over its stated shelf life. This confirmation can be obtained by testing the product initially for *in vivo* availability and then repeating at intervals during storage at normal conditions for its estimated shelf life (or longer).

Interpretation of stability data for coated, modified-release products should be undertaken with extreme care because the diffusion characteristics of polymeric films can change

significantly under exaggerated temperature and/or humidity conditions. This change may be confounding when trying to predict their diffusion characteristics under more moderate conditions and thus can prove misleading when predicting shelf life.

When elevated-temperature stability studies are conducted on products coated with aqueous polymeric dispersions (latices or pseudo-latices), the data obtained might be more indicative of morphological changes that have occurred in the film. Such changes may result from partial destruction of the film when coated material adheres together in the container and subsequently is broken apart; additionally, these changes might result from further coalescence of the coating (which can occur when the coating is not coalesced completely during the coating process).

Stability tests usually are conducted on a product at the time of development, during the pilot phase and on representative lots of the commercial product. Stability testing must continue for the commercial product as long as it remains on the market because subtle changes in a manufacturing process and/or a raw material can have an impact on the shelf life of a product.

RECENT TRENDS IN PHARMACEUTICAL COATING TECHNOLOGY

There is an inherent conservatism expressed by pharmaceutical manufacturers towards accepting major changes in raw materials (i.e., non-active ingredients) and processing technologies. Thus, change tends to be evolutionary rather than revolutionary. Still, some interesting events have occurred over the last decade.

Of particular note is the growing interest in *Process Analytical Technology* (PAT). This has resulted in bringing many analytical procedures out of the laboratory and closer to the manufacturing process with which they may be associated. The desire here is to introduce, ideally as an in-line control function, specific analytical techniques that can be used to enhance the quality of the final coated products. One example is the use of near infra-red techniques that can be used to analyze coated product in such a way that, for example, product moisture contents, drug contents, amounts of coating applied, and even, to some extent, drug release rates can be predicted before that product is discharged from the coating process. The use of terahertz in-line analytical techniques has recently been proposed¹⁵ as a means of determining coating process endpoint.

A major regulatory initiative that is having a significant impact on pharmaceutical manufacturing processes in general is the quest to implement *Quality by Design* (QbD). Certainly, film-coating processes are not exempt from this initiative, but at the same time pose a significant challenge, primarily because in the past scant attention has been paid to critical issues relating to film-coating formulations and processes that ultimately affect product quality. Today, PAT initiatives are considered an integral part of the focus on QbD.

A major change having a significant impact on the way coating of pharmaceutical products is viewed encompasses the growing interest in continuous film coating processes. Initially, continuous coating processes were primarily considered as a useful tool for the manufacture of large-volume consumer healthcare products; in recent years, focus has begun to switch to more mainstream pharmaceutical products, and, in particular, their relevance to in-line continuous pharmaceutical products (where raw materials are fed in at one end, and finished packaged products emerge at the other end). Certainly, such in-line continuous processes facilitate the implementation of PAT and QbD initiatives, but also lead, at least potentially, to significant cost reductions in the manufacturing process.

Generally, continuous coating processes are based on the concept of a stretched side-vented coating pan, where uncoated product is introduced at one end, passes by a whole bank of

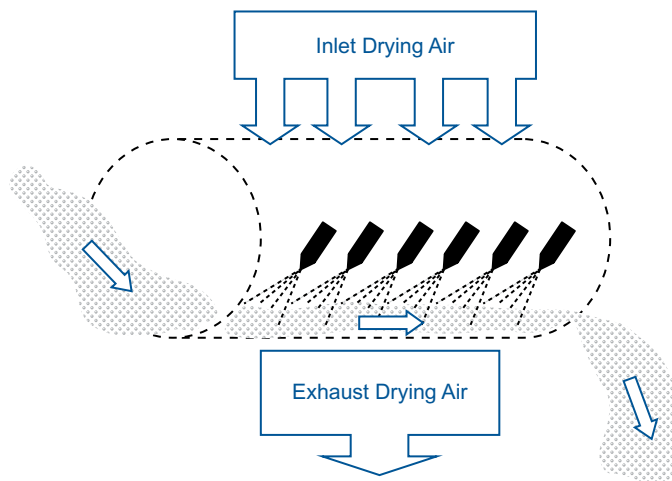


Figure 31-7. Schematic diagram of continuous film-coating process.

spray guns, and emerges from the other end fully coated (see Figure 31-7). The advantages of this type of process include:

1. Increasing output (typical outputs are in the range of 500 to 1000 kg h⁻¹), compared to common batch processes which might coat a 250 kg batch in one to two hours, while a 500 kg batch might be coated in 3–4 hours.
2. Reducing residence time in a process where product is typically exposed to stressful conditions (attrition, high humidities and temperatures) from several hours to about 15 minutes.
3. Improving uniformity of distribution of coating materials.

As stated earlier, continuous coating processes have typically been reserved for coating large-volume products where desired applied coating levels are in the range of 3–4% (based on the tablet core weight). With the growing interest in complete in-line manufacturing processes, where the throughput rate of the coating process needs to be matched to the output of a single tablet press, lower volume output continuous coating processes (with outputs in the range of 25 to 50 kg h⁻¹ up to 100 to 150 kg h⁻¹) are now available. The subject of continuous film-coating processes has been reviewed in detail by Porter¹⁶ and Cunningham *et al.*¹⁷

Currently, most coating processes involve the spray application of liquid coating systems where solidification of the coating is achieved through solvent removal (i.e., drying), and distribution of coating materials is facilitated through constant motion of the material being coated. Recently, a method of electrostatically applying powder coatings, essentially using currently available pharmaceutical coating pans, has been described by Zhu *et al.*¹⁸ This process involves the sequential application of a liquid plasticizer to the tablets being coated, followed by a powder application of the remaining coating materials, and then completing the formation of the coating using a heat-curing step. This technique has been utilized with a range of existing pharmaceutical coating materials, and has application for both immediate-release and modified-release coating applications. Other dry powder coating techniques that do not rely on electrostatic charges have also been investigated.

REFERENCES

1. Urdang G. What's New, 1943, pp 5–14; through *JAPhA* 1945; 34: 135.
2. White RC. The coating of compressed tablets. *JAPhA* 1922; 11: 345–350.
3. Kremers E, Urdang G. *History of Pharmacy*. Philadelphia: Lippincott, 1940: 20, 319.
4. Anon. Current Medical Literature. *JAMA* 1920; 74: 826–838.

5. Wiegand TS. *Am J Pharm* 1902; 74: 33.
6. Warner WR Jr. *Am J Pharm* 1902; 74: 32.
7. Porter SC. The role of high-solids coating systems in reducing process costs. *Tablets & Capsules*. 2010; 8(3): 10–15.
8. Balasubramaniam J *et al.* Preparation of esomeprazole magnesium pellets by extrusion spherization. *24th Annual Meeting & Exposition of AAPS*. 2009.
9. Porter SC *et al.* Modified-release tablets coated with a modified-release film coating: comparison of regular tablets with mini-tablets as an alternative to pellets. *17th Annual Meeting and Exposition of AAPS*. 2002.
10. Wong PSL *et al.* Osmotically controlled tablets in modified-release drug delivery technology. In: Rathbone MJ *et al.* eds. 2003: 101.
11. Porter SC. Troubleshooting appearance-related problems of film-coated tablets. *Tablets & Capsules* 2008; 6(3): 16–22.
12. Porter SC, Terzian L. Predicting coating performance and solving common film-coating problems using a formulation database created using statistical D.o.E. *22nd Annual Meeting and Exposition of AAPS*. 2007.
13. Anon. Supplement to *Tablets & Capsules*: 2008; 16.
14. Porter SC, Felton LA. Techniques to assess film coatings and evaluate film-coated products. *Drug Dev & Ind Pharm* 2010; 36(2): 128–129.
15. May RK *et al.* Terahertz in-line sensor for direct coating thickness measurement of individual tablets during film coating in real-time. *J Pharm Sci* 2011; 100(4): 1535–1544.
16. Porter SC. Continuous film-coating processes: A review. *Tablets & Capsules* 2007; 4: 26–33.
17. Cunningham C *et al.* Evaluation of recent advances in continuous film coating processes. *Drug Dev & Ind Pharm* 2010; 36(2): 227–233.
18. Zhu J *et al.* A method for electrostatically powder coating tablets. *Tablets & Capsules* 2010; 8(3): 18–21.

Chapter 32

Oral Modified-Release Drug Delivery Systems

Ali R. Rajabi-Siahboomi, PhD; Manish S. Rane, PhD and Linda A. Felton, PhD

INTRODUCTION 623

RATIONALE AND DEFINITIONS OF ORAL
MODIFIED-RELEASE TECHNOLOGIES 623

DELAYED-RELEASE SYSTEMS 623

EXTENDED-RELEASE SYSTEMS 624

ORAL EXTENDED-RELEASE TECHNOLOGIES 627

SUMMARY 631

INTRODUCTION

In recent years, the scope of drug delivery technologies has expanded significantly, irrespective of the route of administration, as a result of an advanced understanding of disease, science, and safety associated with pharmaceuticals. The overall purpose of drug delivery systems, however, has remained constant: to provide a therapeutically effective amount of drug to the appropriate site in the body for a desired duration of action. The site at which a drug is delivered (drug targeting) and the rate at which the drug is released (profile) have to be carefully considered during dosage form design and product development. Site-specific drug delivery has been one of the key focus areas for pharmaceutical formulation scientists for many decades. The main purpose of site-specific drug delivery systems is to improve the efficacy and safety of drugs. Pharmaceutical scientists have also focused on the development of formulations that deliver drugs at a desired release rate, the duration of which may span from very fast (a few seconds) to very slow and controlled (days, weeks, and months). Combining site-specific aspects of drug delivery with controlled release rates is highly desirable for patient treatment.

With better understanding of gastrointestinal (GI) tract anatomy, physiological barriers to drug absorption, and the need for different release profiles for different disease conditions, more efficient and advanced drug delivery systems have been developed. In recent years, the purposes of such modified-release drug delivery technologies have evolved to the optimization of drug performance and to the enhancement of patient tolerance (reduction in side-effects). These technologies have not only had a significant impact on the success of developing and commercializing new chemical entities, but the reformulation of marketed drug products for better patient acceptance has allowed pharmaceutical companies to extend the patent life of their products. As a result of the market success and thus opportunities for enhanced drug delivery systems, considerable efforts have been expended on developing appropriate polymeric carriers and sophisticated processing and manufacturing machineries. A tremendous amount of research has been directed toward the development of robust modified-release (MR) oral dosage forms. These efforts have followed the science-based and risk-managed process of drug product development, supported by quality-by-design (QbD) principles. This chapter describes various modified-release technologies used for drug delivery to the oral cavity (oral route of administration) as well as formulation approaches and manufacturing considerations.

RATIONALE AND DEFINITIONS OF ORAL MODIFIED-RELEASE TECHNOLOGIES

The oral route for delivering medications has been the preferred route for most drugs due to patient acceptance, ease of administration, accurate dosing, cost-effective manufacturing methods, and generally the improved shelf-life of the product.

For conventional oral dosage forms, the drug (active pharmaceutical ingredient [API]) is rapidly released after administration and subsequently absorbed into the body from the GI tract. The concentration of drug in the blood peaks shortly after administration as the drug absorption process dominates, then decreases over time as metabolism and/or excretion processes dominate. Conventional immediate-release (IR) dosage forms, however, do not maintain the plasma levels of the drug within the therapeutic range for an extended period of time and thus a short duration of action may be observed. For many drugs and therapeutic indications, multiple dosing of IR formulations provides satisfactory clinical performance with an appropriate balance of efficacy and safety. For example, multidose therapy may be tolerated for short-term treatment, but is not desirable for treating chronic conditions. To reduce dosing frequency and eliminate the fluctuations in blood concentration associated with conventional delivery, extended-release (ER) systems have been and continue to be developed, where the drug is slowly released over an extended timeframe. Delayed-release (DR) technologies exhibit a lag time in drug release (no drug released immediately) to target the drug to a specific site in the body. Both ER and DR systems are broadly referred to as modified-release (MR) dosage forms. The common goal for the development of any MR formulation is to enhance the drug's therapeutic benefits, minimize side-effects, and improve the overall management of the disease. These technologies may be combined with conventional IR delivery or combined with other MR technologies.

DELAYED-RELEASE SYSTEMS

Delayed-release systems are designed to target drug release to a specific site in the body. The delay in drug release is usually achieved by coating an oral solid dosage form with an enteric polymeric film. These polymers, commonly referred to as polyacids, possess acidic functional groups (such as carboxyl groups) and exhibit pH-dependent solubility. The enteric film remains intact in the low pH of the stomach and starts to dissolve at the higher pH of the small intestine when sufficient ionization of these acidic functional groups occurs. According to the US Food and Drug Administration (FDA), a delayed-release dosage form is defined as one "that releases a drug (or drugs) at a time other than promptly after administration. Enteric-coated articles are delayed-release dosage forms."

The general purposes of enteric coatings are:

- to protect the drug in the dosage form from the harmful effects of the gastric environment (i.e., acidic or enzymatic degradation);
- to minimize the irritating effects of certain drugs (e.g., nonsteroidal anti-inflammatory drugs) on the gastric mucosa; and/or
- to deliver the drug to the intestine for local effects.

Table 32-1. Common Enteric Polymers and Their Respective pH Thresholds

General class	Enteric polymer	pH trigger ^a
Cellulosics	Cellulose acetate phthalate (CAP)	6.0
	Cellulose acetate trimellitate (CAT)	4.8
	Hydroxypropyl methylcellulose acetate succinate (HPMCAS)	5.5–6.8
	Hydroxypropyl methylcellulose phthalate (HPMCP)	5.0–5.5
Vinyls	Polyvinyl acetate phthalate (PVAP)	5.0
Acrylics	Poly (methacrylic acid: ethylacrylate) (1 : 1)	5.5
	Poly (methacrylic acid: methylmethacrylate) (1 : 1)	6.0
	Poly (methacrylic acid: methylmethacrylate) (1 : 2)	7.0
Natural polymers	Shellac	7.2
	Rosin	6.0

^aAs indicated by their manufacturers.

Commonly used polymers in enteric coating applications are classified based on their chemical backbone (i.e., cellulose, vinyls, and acrylics) (Table 32-1). Natural polymers such as shellac and rosin have also been used.¹ Some of the enteric polymers are esters that are susceptible to hydrolysis when exposed to elevated humidity. The polymer materials are commercially available as aqueous dispersions or as powders that are dissolved or dispersed in an appropriate solvent prior to application. More details about the application of polymeric films to solid dosage forms can be found in Chapter 30. In addition to the polymer, excipients, such as plasticizers, pigments, and anti-adherents, may be added for processing or functionality. Several fully formulated systems containing any necessary excipients are also commercially available.^{2,3} Note that when formulating an acid-labile drug in an enteric dosage form, an intermediate sealcoat may be necessary to separate the API from the acidic enteric material.

Enteric-coated tablets are tested for their resistance to disintegration in 0.1 N HCl (pH 1.2) media, to simulate their performance in the stomach, a low pH condition when fasted. The enteric-coated system may also be evaluated at a slightly elevated pH (e.g., USP acetate buffer pH 4.5) to simulate fed state (food effects) and/or the administration of multiple doses of proton pump inhibitors.⁴ In addition, the amount of acid taken up by the coated dosage form using gravimetric evaluation after exposure to the acidic media may be determined. Drug dissolution from the enteric-coated dosage forms is evaluated in acidic pH initially (typically 1–2 h, as recommended by the drug product monograph for a specific API) and then in an appropriate phosphate buffer. A typical drug-release profile of an enteric-coated tablet is shown in Figure 32-1. Ideally, no drug is released during exposure to the acidic environment and rapid release is observed when the pH of the test media is increased. It is possible to apply an enteric coat to an ER dosage form. In that case, the enteric coating will cause a delay in drug release (as described above) and, after the enteric layer dissolves, the drug will slowly release from the product over an extended time period. Extended-release products are described in detail below.

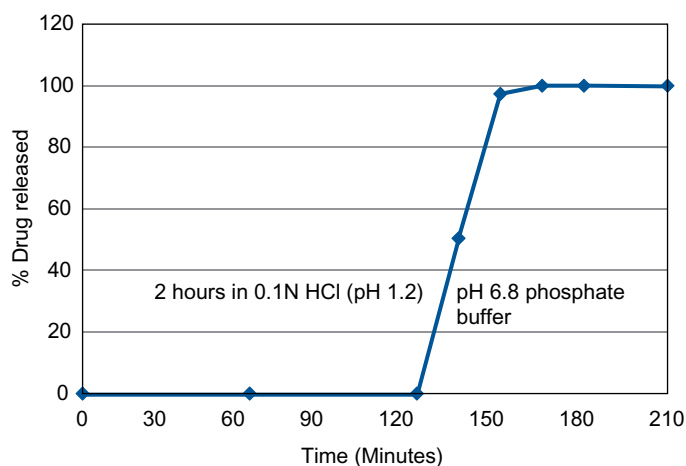


Figure 32-1. Typical release profile for an enteric-coated dosage form.

EXTENDED-RELEASE SYSTEMS

Extended-release systems for oral drug delivery maintain therapeutic blood levels of a drug for a prolonged period. Terms such as extended-release, sustained-release, and slow-release have been used interchangeably in the literature. The term “controlled-release” is a more specific term and may be defined as delivery of a drug at a predetermined rate for a defined time period. Commercially available ER products are often designated by suffixes such as “CR” (controlled-release), “LA” (long-acting), and “ER,” “XL,” and “XR” (extended-release). Figure 32-2 provides an example of a typical drug release profile of an ER product. These systems reduce fluctuations in drug concentration in the blood. Since these products are administered less frequently, patient adherence to the dosing regimen may be improved. Despite the higher unit cost of the delivery system in comparison to conventional IR dosage forms, there is an economic savings benefit for ER formulations. Economic savings may also result from a decrease in hospitalizations and less lost work time, as chronic disease conditions are better managed.

When developing an ER dosage form, physiological and biopharmaceutical factors (e.g., GI tract transit time and pattern, rate and window of absorption, localized metabolism in the GI tract, and potential food effect) and the physicochemical properties of the drug (e.g., chemical nature and physical form of the

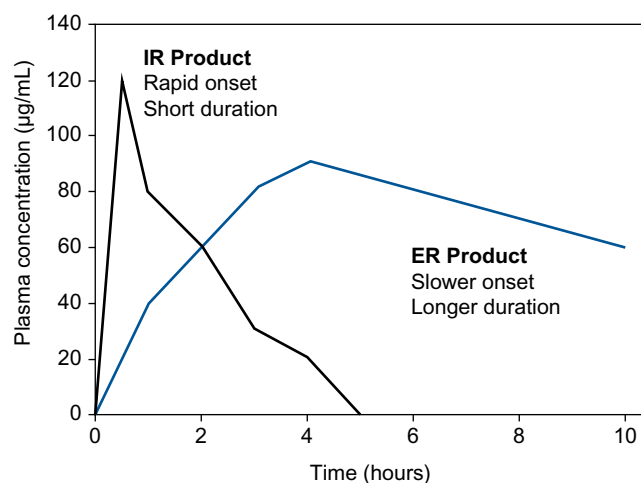


Figure 32-2. Typical release profile for an extended-release (ER) dosage form in comparison to an immediate-release (IR) product.

drug, dose, solubility, and stability of the drug in the GI tract) must be considered, as these may influence the design of the delivery system and the desired release performance of the product. The following sections discuss these parameters as well as safety aspects of ER products. The remainder of the chapter focuses on specific delivery technologies that are used today to achieve extended drug release from solid oral dosage forms.

BIOPHARMACEUTICAL CONSIDERATIONS FOR EXTENDED-RELEASE DRUG PRODUCT FORMULATION AND DEVELOPMENT

The anatomy and physiology of the GI tract are complex. There are differences in the lengths, pH, transit time, and surface area available for absorption. Table 32-2 provides a summary of average anatomical and physiological features found in healthy humans, irrespective of gender or race. The values may differ from person to person, along with the age of the person and also GI health status (e.g., gastric acidity, diarrhea, and colitis). Oral ER formulations generally remain intact and are subjected to these considerably different conditions while traveling through the GI tract.

Knowledge of the GI anatomy and physiology is important when designing suitable ER dosage forms. A drug must first be released from the dosage form, then dissolve in the GI media before being absorbed through GI membranes. Thus, a drug must be soluble in the given GI media for absorption to occur. Most drugs, however, are weak acids and bases and their solubility is affected by the different pH in the GI tract. GI conditions, transit time, and the presence and types of food ingredients can also significantly influence drug absorption.

Some drugs may be preferentially absorbed in a specific region of the GI tract, referred to as an “absorption window.” Knowledge of such narrow regional absorption is critical for designing suitable ER delivery systems.⁵ For this specialized delivery, one formulation strategy would be to create an ER gastroretentive system, where the dosage form is retained and the drug is released in the stomach so the drug is in solution when passing the absorption window.^{6–8}

The most absorptive region of the GI tract is the small intestine. However, the general transit time of oral dosage forms through the small intestine is relatively short. Various studies using noninvasive modalities have assessed the effect of food and the dosage form shape and size on transit time of nondisintegrating tablets through segments of the human GI tract. In addition to the presence of food, other inherent GI motility ailments may further alter the transit time of the solid dosage form in GI tract. Table 32-3 shows a range of transit times reported for ER tablets, caplets, and pellets.^{9,10}

Dietary food consists of different components such as proteins, carbohydrates, fats, fibers, minerals, and other nutrients. The presence of food alters the volume of liquid, pH and viscosity of the GI content, as well as GI transit times. Hence the performance of an ER formulation may differ when administered with food.^{11,12} Food intake can influence the rate of drug release from the dosage form, the rate of drug absorption, the amount of drug absorbed, or all of these parameters simultaneously.¹³

Various physiological mechanisms prevent the absorption of foreign material from the GI tract. Intestinal metabolic enzymes found in abundance in the intestinal epithelium walls may metabolize the drug before absorption can occur. The activity of such enzymes varies in the small intestine, with levels rising slightly from the duodenum to the jejunum and declining in the ileum and colon. The presence, concentration, and distribution of these enzymes are highly variable from person to person and also vary with gender, age, and race. There is also an efflux mechanism involving P-glycoproteins. Many drugs have been shown to possess poor bioavailability due to intestinal enzymes and the efflux pump systems.¹⁴ Components of food can inhibit GI metabolism and efflux of certain drugs, resulting in an increase in absorption and bioavailability. For example, grapefruit juice has been shown to enhance bioavailability of a number of

drugs by inhibiting both the efflux transporter P-glycoprotein and intestinal metabolism.^{15,16} In fact, grapefruit juice has been reported to cause as high as 300 percent increase in the concentration of cyclosporin, midazolam, terfenadine, and felodipine, orally administered medications.¹⁷

SAFETY CONSIDERATIONS AND RISK MANAGEMENT

Oral ER formulations carry higher drug payloads compared to IR formulations. A formulation that fails to sustain drug release and abruptly releases the drug in an IR fashion (called “dose-dumping”) is potentially dangerous.^{18,19} Careful consideration, therefore, must be given to the safety and risk management of these ER formulations. One of the key reasons reported for dose-dumping is related to the food effect, which is compounded by a formulation which is not robust. For example, very low levels of rate-controlling material may make a formulation more sensitive to both food components and the higher mechanical stress imposed by the GI tract.²⁰

Most ER formulations are intended to be used/administered whole. Any damage to the integrity of the dosage form may also result in dose-dumping. Warning labels should be included on the packaging of such ER formulations. The concept of dose-dumping is also applicable for DR products. Note that there are several ER patient-friendly formulations, such as sprinkle capsules or multiparticulate systems, which have been formulated to be subdivided prior to administration without compromising the drug release mechanism.²¹

Another consideration for possible dose-dumping relates to concomitant use of alcohol with ER formulations. Alcohol is known to be a good solvent for many release-controlling materials, and the alcohol content in beverages may dissolve or weaken the rate-controlling system, potentially resulting in dose-dumping in the stomach.²² During drug product development, the potential alcohol effect on the performance of the formulation should be evaluated.^{23–25}

DRUG PROPERTIES RELEVANT TO ER PRODUCT FORMULATION AND DEVELOPMENT

Drug dose and solubility are the most important physicochemical properties to consider when developing an ER product. Extended-release systems designed to reduce dosing frequency inherently carry greater amounts of the drug than their corresponding IR formulations. A combination of safety data, elimination half-life, the therapeutic window, and typical drug levels in conventional formulations provide some indications of the total amount of drug required for an ER preparation. For large-dose drugs, the size of the dose may present significant challenges when formulating oral ER dosage forms. This is further complicated if the high-dose drug is also highly soluble in water.

As mentioned previously, drugs must be in solution in order to be absorbed from the GI tract. The aqueous solubility of a drug affects its dissolution rate, which in turn establishes its concentration in the solution and, hence, the driving force for passive diffusion across the absorptive membranes in the GI tract. The Noyes-Whitney equation, shown below under sink conditions, demonstrates the importance of drug solubility in the rate of mass transfer of solute particles into the continuous phase:

$$\frac{dM}{dt} = \frac{DAC_s}{h}$$

where dM/dt is the rate of mass transfer or rate of dissolution, D is the diffusion coefficient, A is the total surface area of the device, C_s is the aqueous saturated solubility of the drug, and h is the height of the boundary (diffusion) layer.

In general, drugs with extreme aqueous solubilities pose challenges for ER formulation development. For example, any system relying on diffusion of the drug through a polymer barrier membrane (reservoir system) may not be suitable for poorly soluble drugs. In contrast, it may be difficult to sufficiently slow the dissolution of a drug with very high aqueous solubility. Thus,

Table 32-2. Summary of Physiological Considerations of Human Gastrointestinal Tract

GI Region	Length (cm)	Diameter (cm)	pH	Micro-organism (counts/g content)	Proposed mechanisms of absorption						SA (m ²)	Transit time for food (h)	Villi present	Constituents (enzymes, others)
					P	C	A	F	I	E				
Mouth	15–20	10	5.0–6.8	NR	Y	Y	N	N	N	N		Short	N	Amylase, maltase, ptyalin, mucins
Esophagus	25	2.5	5.0–6.0	NR								Very short	N	–
Stomach	20	15.0	1.2–5.0	≤10 ²	Y	Y	Y	N	N	N	0.1	0.25–3.00	N	Hydrochloric acid, pepsin, rennin, lipase, intrinsic factor
Small intestine	700										12.0	3.0–10.0		
Duodenum	25	5.0	4.6–6.0	≤10 ²	Y	Y	Y	Y	Y	Y	0.3	0.3–2.0	Y	Bile, trypsin, chymotrypsin, amylase, maltase, lipase, nuclease, CYP3A4
Jejunum	300	5.0	6.3–7.3	≤10 ²	Y	Y	Y	Y	Y	Y	3.0	1.0–6.0	Y	Amylase, maltase, lactase, sucrose, CYP3A5
Ileum	400	2.5–5.0	7.6	≤10 ⁷	Y	Y	Y	Y	Y	Y	4.0	1.0–6.0	Y	Lipase, nuclease, nucleotidase, enterokinase
Large intestine	150										0.3	1.0–20.0		
Caecum	10–30	7	7.0–8.0		Y	Y	Y	N	N	Y	0.05	Short	Y	–
Colon	150	5	7.9–8.0	≤10 ¹¹	Y	Y	Y	N	N	Y	0.25	4.0–20.0	N	Esterases, sulphatase glycosidases, reductases, amidases
Rectum	15–19	2.5	7.0–8.0		N	N	N	N	N	N	–	Variable	N	–

NR, Not reported; P, passive diffusion; C, convective transport; A, active transport; F, facilitated transport (carrier mediated); I, ion pair transport; E, endocytosis; Y, yes; N, no; SA, surface area available for absorption.

Table 32-3. Reported Ranges of Transit Times of Different Dosage Forms Through Human Gastrointestinal Tract

Dosage form	Transit time in stomach (h)		Transit time in small intestine (h)	
	Fasted	Fed	Fasted	Fed
Extended-release tablets (11 mm round)	0.1–1.1	2.0–4.8	3.9–5.6	2.2–7.7
Extended-release caplets (17 × 4 mm)	0.3–2.0	1.0–9.0	1.1–6.0	1.2–5.8
Extended-release pellets (0.3–1.8 mm diameter, administered in gelatin capsules)	0.1–4.5	0.1–4.5	1.8–7.5	1.2–5.8

the aqueous solubility of a drug is a critical property that influences the technology selected to achieve the extended release.

DRUG RELEASE RATE CONSIDERATIONS

Ideally, an ER system should deliver drug to the desired site at a rate according to the needs of the body (i.e., a self-regulated system based on a feedback control). Although some researchers are investigating possible approaches to develop such an ideal dosage form (e.g., self-regulating insulin delivery pump), this is still a difficult goal to achieve, especially for oral drug administration. The key question for the formulation scientist is what should be the rate of drug release to maintain therapeutic blood levels. Ideally, the rate of absorption of the drug should be equal to the overall rate of elimination. This means that drug release must be independent of the amount of drug in the dosage form and should remain constant over time. This type of release is referred to as zero-order kinetics. Although it may not be necessary to maintain a constant level of drug in the blood for all therapeutic cases, zero-order release has been idealized as the ultimate goal for an ER delivery system. The majority of ER systems follow first-order kinetics, where drug release is dependent on the amount remaining in the dosage form at any given time. Drug release is thus faster initially and slows over time as drug is depleted from the delivery system. As mentioned previously, ER systems may be combined with IR delivery to attain high initial blood concentrations that are then maintained over an extended time period. Delayed-release technologies may also be combined with ER systems, resulting in an initial delay in release followed by prolonged drug release.

There are vast amounts of literature covering the mathematical description of drug release kinetics for ER systems.^{26,27} While a detailed discussion of mathematical modeling of drug release kinetics is beyond the scope of this chapter, the most common mathematical equations used to describe different release profiles are shown in Table 32-4, where Q is the cumulative amount of drug released at time t , K is a rate constant, and n is the diffusional exponent. The Peppas-Sahlin model (Table 32-4) describes systems that have intermediate values of n , where at least two processes contribute to the overall release mechanism. According to the model, K_d is the diffusional rate constant, K_r is the relaxational rate constant, and m is a purely Fickian diffusion exponent for a device of any geometrical shape exhibiting extended release. The value of n has been routinely used to characterize drug release mechanisms, using the calculated ranges shown in Table 32-5.

Table 32-4. Summary of Some Equations Describing Drug Release Kinetics from the Extended-Release Systems

Kinetics	Equation
Zero order	$Q = k t + a$
Higuchi model (1963)	$Q = k (t)^{0.5}$
Korsmeyer–Peppas model (1983)	$Q = k (t)^n$
Peppas–Sahlin model (1989)	$Q = K_d t^n + K_r t^{2n}$

Table 32-5. Diffusional Exponent (n) and The Theoretical Mechanism of Drug Release From Swellable Systems having Spherical Geometry

Diffusional exponent (n)	Drug release mechanism
$0.43 < n < 0.50$	Fickian diffusion (first order)
$0.50 < n < 0.89$	Anomalous release (pseudo-first order)
$0.89 < n < 1.00$	Case II transport (zero order)
$n > 1.00$	Super case II transport (unknown release)

ORAL EXTENDED-RELEASE TECHNOLOGIES

Most currently available ER formulations incorporate one or more of the following general technological approaches:

- **Matrix systems:** Monolithic (*mono* is single and *lith* is a block of material) matrices are the most common version. These systems comprise one or more release-controlling materials with one or more drugs uniformly dispersed in the matrix. Two types of such systems are employed, depending on the nature of the rate-controlling material.
 - **Hydrophilic matrix systems:** Drug particles are dispersed in a hydrophilic polymeric matrix; drug release occurs by dissolution of the drug, diffusion through the gel layer (formed upon exposure to GI fluids), and/or erosion of the matrix.
 - **Insoluble matrix systems:** Drug is dispersed in a matrix of water-insoluble polymers or waxes; drug release occurs as the GI fluids permeate the matrix and dissolve the drug.
- **Reservoir systems:** A drug-containing unit (core) is enclosed by a polymeric barrier coat. Two key reservoir systems have been commonly used:
 - In **simple diffusion systems**, a drug-containing core is surrounded by a water-insoluble polymer coating. Drug release is achieved by diffusion of drug through the coating.
 - **Osmotic systems** contain osmotic agents in the coated drug core. Drug release occurs through an orifice in the coating layer due to an osmotic pressure gradient generated as the GI fluids permeate into the core.

Details about each of these systems is provided below.

Matrix-Based Extended-Release Systems

Wax matrices are prepared by adding drug to molten fat or wax, cooling to congeal the material, then granulating and compressing into tablets. Fatty acids, alcohols, and waxes which are solid at room temperature and do not melt at body temperature are used for this type of delivery system. Examples of materials used for the preparation of wax matrices include carnauba wax, fatty alcohol, glycerol palmitostearate, stearyl alcohol, beeswax, aluminum monostearate, and glycerol monostearate.

These materials may be used alone or in combination with a hydrophilic polymers. Drug release occurs as GI fluids permeate into the pores of the matrix, causing drug dissolution, and the dissolved drug then diffuses out of the system. The presence and levels of wetting and channeling agents, such as soluble sugars or hydrophilic low-viscosity grade polymers, in the matrix are used to further modulate drug release. Lipidic matrix systems have also been developed, where drug release occurs through a combination of erosion of the lipid matrix and diffusion through pores in the matrix.

These waxy materials are derived from natural products and are complex mixtures of multiple components.²⁸ Hence, variability of source or changes in the isolation and purification processes may alter their composition in subtle ways that may influence drug release profiles. In addition, fatty acids and waxes exhibit complex solid state behavior, including polymorphism, leading to phase transition behavior on cooling and storage. These physical changes can also affect drug release rates. If waxes coat the drug particle during processing, drug release may not be complete.

Ion-exchange resins are water-insoluble crosslinked polymeric resins that complex oppositely charged drugs. These systems are formed by prolonged exposure of the drug to the resins. Drug-containing resins are milled to the desired particle size and can be subsequently filled into capsules, tableted, or suspended in liquids. Drug release from these systems occurs, as the name implies, by an ion-exchange mechanism. For example, a cationic drug in an anionic resinate is replaced by protons, sodium, or other electrolyte in the GI fluids. The GI fluids must penetrate into the resin and the dissolved drug has to diffuse out of the matrix. As would be expected, drug release from these systems is dependent on pH and electrolyte concentration. These systems can be coated with rate-controlling membranes to further control drug release.

Inert matrices are prepared by wet granulation of a drug with water-insoluble polymers such as ethylcellulose, methylacrylate, methylmethacrylate, polyvinyl chloride, or polyvinyl acetate.^{29,30} Drug release from these matrices is by simple dissolution of the drug and diffusion through fluid-filled pores. Water ingresses into the matrix, and the solid drug on and adjacent to the surface dissolves into the media, creating pores and channels in the tablet core that allow additional drug to dissolve, and the process continues.

Higuchi's mathematical equation defines drug release from inert matrices as:

$$Q = \left[\left(D \in \frac{C_s}{\tau} \right) (2A - \in C_s) t \right]^{1/2}$$

where Q is the amount of drug released per unit surface area after time t , D is the diffusion coefficient of the drug in the dissolution medium, $\in$ is the porosity of the matrix, τ is the tortuosity of the matrix, C_s is the saturated solubility of the drug in the dissolution medium, and A is the total amount of drug present per unit volume in the matrix. Drug release rates from inert matrices may be altered by changing either the porosity or the tortuosity of the matrix. A high tortuosity means that the effective average diffusion path is large and thus drug release is slower. The porosity term takes into account the space available for drug dissolution, and an increase in porosity results in faster drug release. Water-soluble excipients can be added as pore formers, and these agents alter the tortuosity of the matrix. Inclusion of some water-insoluble excipients may reduce the wettability of the matrix, resulting in a reduction in the penetration of the dissolution media and subsequent slower drug release. Higher compaction forces during tableting in general lead to lower porosity and hence slower drug release.

As mentioned, ethylcellulose is one of the polymers used to prepare inert matrices. It is a water-insoluble cellulose ether polymer available in different molecular weight grades which

exhibit varying viscosities in organic solutions as well as available in different particle size grades. Ethylcellulose is available commercially as a dry powder (e.g., Ethocel) and as an aqueous dispersion (e.g., Surelease). Small particle size grades of ethylcellulose are required to generate a high tortuosity (large diffusion path) in the tablet. However, small particle sizes in the dry state coupled with high concentrations of the polymer will likely lead to a formulation with poor flow properties and static charges so wet granulation methods may be required. Polymethacrylates and polyvinyl acetate have also been used to study inert matrices with ER characteristics. Drug release from these inert matrix formulations is dependent on drug solubility, polymer level, mechanical properties of the tablets, and the type and amount of other excipients used.

Hydrophilic matrices are prepared by direct compression or by wet granulation of a blend of one or more drugs with suitable release-retardant hydrophilic polymers plus other excipients necessary for processing. These matrices are a simple and cost-effective approach for the manufacture of ER systems.³¹ A wide range of natural and semisynthetic polymers have been studied to produce hydrophilic matrices, as shown in Table 32-6. These materials can be used alone or in combination to fabricate hydrophilic matrices.

The most widely utilized hydrophilic matrix formers are the high molecular weight (high viscosity grade) nonionic cellulose ethers, and more specifically, hydroxypropyl methylcellulose (HPMC, also referred to as hypromellose). The popularity of HPMC can be attributed to its flexibility to obtain desired drug release profiles, cost-effectiveness, and broad safety and regulatory acceptance. These systems can be readily prepared with conventional tableting equipment. To fully explore the formulation, manufacturing, and drug release mechanism from hydrophilic matrices, examples of HPMC matrices will be used throughout this section of the chapter.

On exposure to water or gastrointestinal fluids, hydrophilic polymers in the matrix system rapidly hydrate to form a viscous gel layer. The polymer undergoes transition from a "glassy" solid to a "rubbery" gelatinous state. Failure to generate a uniform and coherent gel may result in a more rapid drug release. It is the physicochemical characteristics of the gel layer which control water uptake and drug release from the matrix. Gel growth occurs as water permeates in the matrix to hydrate the solid polymer particles immediately beneath the gel, increasing the thickness of the gel layer. Simultaneously, the outer surface of the matrix becomes hydrated, polymer chains fully relax and disentangle from each other and erode from the matrix surface into the surrounding media. Water continues to penetrate

Table 32-6. Common Polymers Studied in Hydrophilic Matrix Formulations

Cellulosic	Hypromellose (hydroxypropylmethylcellulose (HPMC)) Hydroxypropylcellulose (HPC) Hydroxyethylcellulose (HEC) Sodium carboxymethylcellulose (Na-CMC)
Noncellulosic: gums/polysaccharides	Sodium alginate Xanthan gum Carrageenan Ceratonia (locust bean gum) Chitosan Guar gum
Noncellulosic: others	Poly(ethylene oxide) Homopolymers and copolymers of acrylic acid (e.g., carbomers)

Table 32-7. Formulation and Manufacturing Considerations in the Design of Hydrophilic Matrices

Formulation components	Key considerations
Drug	Solubility, dose, pKa, stability, and particle size
High viscosity grade of polymer	Particle size, type, and level
Filler excipients	Level/type (water soluble/insoluble)
Lubricants and glidants	Level/type (stearates, nonstearates, and fatty acids)
Others	Release rate modifiers, stabilizers, solubilizer, surfactant, and buffering agents
Manufacturing aspects	
Direct compression	Flow and compressibility of the formulation
Dry granulation	Double compression and reworkability
Wet granulation	Solvent (aqueous or hydro-alcoholic), binders, high or low shear, fluid bed methods
Hot melt extrusion	Type, level, and melting point of polymer and drug

toward the core of the tablet at a controlled rate, until all of the tablet hydrates and ultimately erodes completely.

Although a rapid burst release of drug from the external surface area may occur, drug release is controlled either by diffusion through the gel layer or by gradual erosion of the gel. Diffusion is the dominant mechanism controlling the release of water-soluble drugs while water-insoluble drugs are released predominantly by the erosion of the gel layer. Practically, drug release generally occurs by a combination of dissolution, diffusion, and erosion, and follows first-order kinetics. Consistent and reliable release requires that polymer hydration and gel layer formation is rapid to obviate potential matrix disintegration and premature drug release. For this reason, hydrophilic polymers for matrix applications are usually supplied in small and defined particle size ranges (e.g., METHOCEL™ CR grades). HPMC and polyethylene oxide (POLYOX™) are well known for their rapid hydration and gel formation and thus highly suitable as rate-controlling polymers in hydrophilic matrices.³² Polyethylene oxide resins have been shown to be among the fastest hydrating water-soluble polymers, quickly forming a gel layer to regulate drug release.³³ Table 32-7 shows some of the general formulation and manufacturing considerations for hydrophilic matrices.

In order to reduce the initial burst effect, scientists have investigated combinations of faster hydrating polymers to rapidly form the gel layer. The application of an insoluble but permeable film coat on the surface of the matrix is another approach to reducing the initial drug burst. Water ingress through the insoluble membrane and drug dissolution are subsequently slowed, generating a lag time in the release profile. Following sufficient water ingress, the hydrophilic polymer(s) hydrates

and, due to swelling of the matrix, causes rupture of the insoluble membrane.

Modulation of drug release in hydrophilic matrices is achieved by various approaches that regulate gel strength and structure. For example, faster drug release can be obtained when lower polymer concentration, lower viscosity grades of the polymer, and/or highly water-soluble excipients are used in the matrix formulation. All these factors reduce gel strength which in turn increases both the diffusion rate (faster water ingress and faster drug diffusion out of the device) and the rate of polymer disentanglement (erosion). According to the percolation theory, which describes a cluster percolating as a group of adjacent particles of the same component from one side to the other side of a system, polymer concentration must be above a threshold level (the critical point). When formulating above this critical point, a coherent gel layer controlling the drug release rate is formed.³⁴ When the polymer concentration is close to the critical point, the matrix system may be too sensitive to small variations in formulation or manufacturing processes. Based on the percolation theory, the critical concentration for different types of HPMC to obtain robust ER formulations is above 18 percent v/v of polymer.³⁵ Different viscosity grades of polymer are used to modulate faster or slower drug release. For example, a sufficient concentration (above the critical point) of a low-viscosity grade of HPMC (METHOCEL™ K100LV or K4M) is generally used for drugs with poor water solubility. However, for drugs with very high water solubility, a stronger gel layer is required and hence higher viscosity grades of HPMC (METHOCEL™ K15M or K100M) are used.

Another factor affecting drug release from hydrophilic matrices is the excipient(s) used in the formulation. More specifically, excipient solubility and concentration in the formulation have been reported to significantly affect drug release.³⁶ The manufacturing process is also important and may affect release characteristics from the dosage form. Use of polymer blends is yet another approach to refine drug release. For example, HPMC matrices containing other hydrophilic polymers such as polyethylene oxide, carbomer, sodium carboxymethylcellulose (Na CMC), or sodium alginates have been shown to alter drug release.³⁷ These materials create strong hydrogen bonding between the carboxyl and hydroxyl groups of HPMC and the other polymers, leading to stronger gel layers and slower drug release.

Since release from a hydrophilic matrix is directly related to the surface area of the tablet, geometric approaches to modify drug release have been investigated, predominantly for achieving zero-order release profiles. The Geomatrix technology is an example of such an approach. The system consists of a multi-layered tablet, the interior being a simple HPMC matrix containing the drug that is sandwiched between two outer layers that control the exposed surface area during GI transit. These outer layers contain HPMC and other ingredients, but no drug (Figure 32-3). The outer layers hydrate and, as they dissolve, a larger surface area of the core layer is exposed to the dissolution media, ideally leading to a constant, zero-order release profile (www.skyepharma.com).

Reservoir Systems

As mentioned earlier, a simple reservoir ER system is a unit that contains a drug-containing core surrounded by a polymeric film coating. The film coating controls the rate of water uptake into the system and the drug, once dissolved, diffuses out

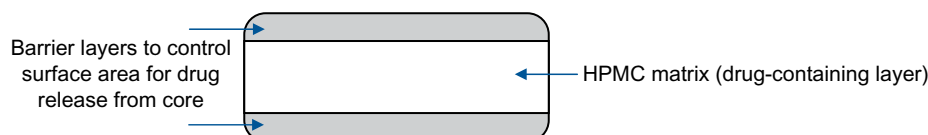


Figure 32-3. Schematic of the Geomatrix technology for zero-order drug release.

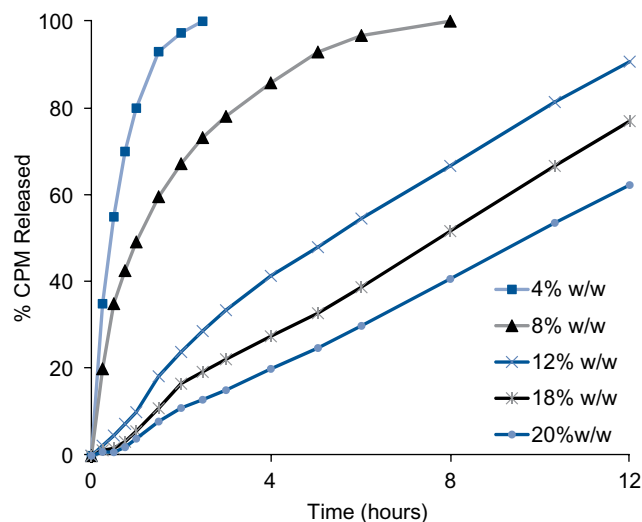


Figure 32-4. Influence of the amount of coating (Surelease) applied on the release of chlorpheniramine maleate (CPM) pellets.

through the membrane. The reader is referred to Chapter 30 for details about the coating process. Drug release across the insoluble membrane from a reservoir-type system is described by Fick's first law of diffusion:

$$\frac{dM}{dt} = \frac{ADK\Delta C}{l}$$

where dM/dt is the dissolution rate, A is the surface area, D is the diffusion coefficient, K is the partition coefficient, ΔC is the concentration gradient, and l is the diffusion path length (film thickness). For a given drug and formulated system, A , D , K , and ΔC remain nearly constant; thus drug release is inversely proportional to the diffusion path length or the coating thickness. Film thickness is not generally determined during manufacturing, although it may be investigated during product development stages. Regardless, film thickness is proportional to the amount of polymer applied, often reported as a percentage weight gain. Figure 32-4 shows that increasing the amount of applied polymer slows drug release. Note that many other variables influence drug release from reservoir systems. For example, the permeability of the coating may be altered by the inclusion of hydrophilic polymers or water-soluble materials. These agents dissolve upon ingestion to create pores in the coating. Commonly used polymers for reservoir systems include ethylcellulose ethers (such as ETHOCEL™), acrylic polymers (such as Eudragit RL 30D and Eudragit RS 30D), and polyvinyl acetate (Kollicoat SR 30D).

Reservoir systems may be formulated in different forms such as tablets, mini-tabs, or spheres/pellets (typically 350–1000 μm). Mini-tabs are small tablets with a diameter of about 2–5 mm. These units can be coated with rate-controlling polymers and filled into hard capsule shells. Mini-tabs can be manufactured by direct compression processes, similar to conventional tablets, with specialized tablet tooling. These systems combine the

advantages of multiparticulate dosage forms (described in the next paragraph) with the more established manufacturing techniques of tableting.³⁸

A great deal of interest has recently been focused on microparticulate reservoir systems. These microparticles can be filled into capsules or compressed with other excipients to produce tablets. These systems can potentially be formulated to contain multiple drugs and/or multiple release profiles by blending different populations of microparticles. After ingestion, the individual coated particles, each containing a portion of the total dose, disperse and move through the GI tract. One of the primary advantages of these systems is their more consistent transit through the GI tract in comparison to larger particles (i.e., tablets). Gastric emptying variations as well as inter- and intra-patient variability are generally reduced. In addition, the risk of dose-dumping is minimized, since it is unlikely that the rate-controlling membrane of all subunits would be compromised, assuming a robust formulation.

There are two general processes used to manufacture reservoir-type multiparticulates, differing predominantly in the structure of the core, as shown in Figure 32-5. A wet granulation or hot melt extrusion process can be used to create drug-containing cores which are then coated with the rate-controlling membrane. Alternatively, drug can be applied to starter seeds or inert carrier material (e.g., Suglets), where a liquid binder is used to adhere drug to the core. Subsequently, the rate-controlling membrane is applied. The selection of the manufacturing process is dependent on the expertise and equipment available within the company.

It is highly desirable to have spheres of uniform and narrow size distributions, smooth surfaces, and low friability in order to achieve reproducible drug release profiles.³⁹ Changes to the particle size and size distribution will influence surface area available for drug release. Figure 32-6 shows how a decrease in particle size from 14–18 mesh (1000–1410 μm) to 30–35 mesh (500–590 μm) significantly increased the release of a model drug when similar amounts of the rate-controlling membrane (ethylcellulose) were applied.³⁹ These results were attributed to a decreased film thickness of the coating, arising from an increase in total surface area of the smaller pellets. While this specific example used drug-layered pellets, the principles also apply to coated pellets prepared by wet granulation or other approaches. A narrow particle size distribution of the microparticles will minimize batch-to-batch variation in the thickness of the rate-controlling membrane and thus in drug release. It is also important that microparticles, irrespective of the manufacturing process, are physically strong enough to withstand the rigors of coating. Friable particles may break or fracture during processing, leading to product loss and, most importantly, change in surface area.

Osmotic Pump Systems

Another option of the coated reservoir system is the osmotic pump. Osmotic agents (osmogens) are included in the tablet core, and a delivery orifice is drilled through the coating membrane. As water diffuses through the coating, the osmotic pressure increases, and it is this osmotic pressure gradient that is

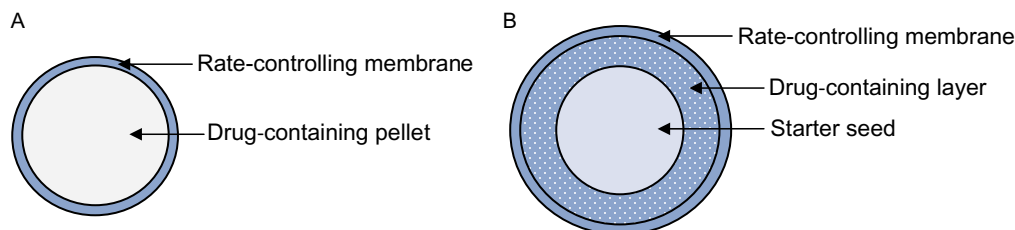


Figure 32-5. Schematics of two reservoir-type microparticulates (A) wet granulation (extrusion) and spheronized drug-containing pellet coated with a rate-controlling membrane; (B) drug-layered pellet coated with a rate-controlling membrane.

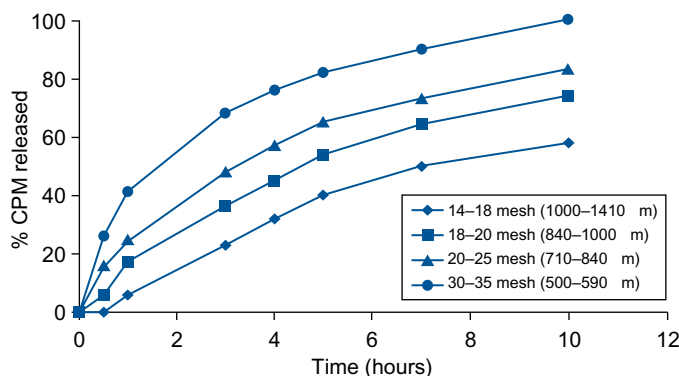


Figure 32-6. The influence of starter-seed size distribution on chlorpheniramine maleate (CPM) release from ethylcellulose barrier membrane coated (10 percent weight gain of surelease) multiparticulate systems

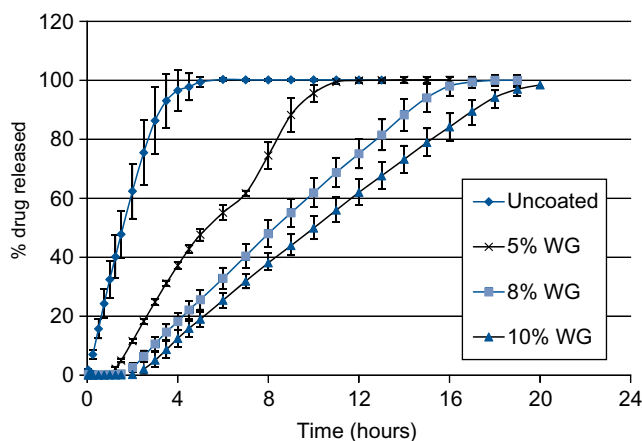


Figure 32-7. Drug release profiles of bilayer osmotic tablets uncoated and coated with varying amounts of a semipermeable cellulose acetate membrane.

the driving force to push drug out of the system through the delivery orifice. The primary advantage of the osmotic system is that only osmotic pressure is required to be effective, and thus drug release from the system is independent of the environment. That is, drug release is not influenced by the pH changes that occur as the product moves through the GI tract. These systems are also less sensitive to food effects.

The most common polymer used for barrier coating of osmotic systems is cellulose acetate. This material has been shown to be mechanically strong enough to withstand core swelling and the hydrostatic pressure created inside the system. Ethylcellulose, another water-insoluble polymer, has also been investigated as a barrier membrane coating for osmotic pumps to a lesser extent.⁴⁰

A key parameter for controlling drug release from the osmotic pump systems is the permeability of the barrier membrane to GI fluids. Figure 32-7 shows how drug release is affected by increasing weight gain (thickness) of a semipermeable membrane. In addition to changing the permeability of the membrane, drug release can also be influenced by surface area of the tablet, as this variable also affects the ingress of the GI fluids.

The first osmotic pump developed was the “elementary osmotic pump.” After an initial lag time, sufficient fluid would permeate through the membrane to generate the osmotic pressure gradient. Drug release would occur at a constant rate, but only until the osmotic pressure gradient diminished. In addition, drug release from these systems was not complete.

A newer version of the osmotic pump was developed to overcome these issues. The “push–pull osmotic pump” consists of a



Figure 32-8. Cross-section of a push–pull osmotic pump (PPOP) tablet. 1. Pull layer (drug layer). 2. push layer, 3. semipermeable membrane, 4. drug delivery orifice.

bilayer tablet core, one layer containing the drug and low-viscosity POLYOX™ (e.g., POLYOX N80), called pull-layer, and the other layer containing a swellable polymer (e.g., POLYOX coagulant and osmogen, called the push-layer). A cross-section of a push–pull osmotic pump is shown in Figure 32-8. The swellable layer maintains the osmotic pressure gradient as the drug (and osmogen) is depleted from the system. The critical formulation variables for the push–pull osmotic include polymer hydration rate and swelling kinetics as well as osmogen levels. Note that both dissolved and dispersed drug can be released from this system. As with the elementary pump, drug release is independent of pH, ionic strength, and other physiological factors. These attributes reduce patient variability. Moreover, since osmotic pressure is maintained over time due to the swelling layer, not only does drug release follow zero-order kinetics, drug release is more complete than with elementary pumps. Push–pull osmotic pump systems have been used to successfully deliver some challenging drugs, making them a huge commercial success in the market.⁴¹

SUMMARY

Advancements in the science and technology of oral modified-release drug delivery systems have resulted in the commercial availability of a number of products. More importantly, these improved delivery systems are impacting the health of patients worldwide. The development and commercialization of modified-release products is no longer considered just a life-cycle management strategy for the manufacturer. Formulation scientists will continue to investigate and develop novel drug delivery platforms to improve the overall effectiveness of drug therapies.

REFERENCES

1. Satturwar PM *et al.* Evaluation of new rosin derivatives for pharmaceutical coating. *Int J Pharm* 2004; 270: 27–36.
2. Cunningham C, Fegely K. One-step aqueous enteric coating systems scale-up evaluation. *Pharm Technol Eur* 2001; 25(11): 36–44.
3. Missaghi S *et al.* Delayed release film coating applications on oral solid dosage forms of proton pump inhibitors: case studies. *Drug Dev Ind Pharm* 2010; 36(2): 180–189.

4. Rohss K *et al.* Esomeprazole 40 mg provides more effective intragastric acid control than lansoprazole 30 mg, omeprazole 20 mg, pantoprazole 40 mg, and rabeprazole 20 mg in patients with gastro-oesophageal reflux symptoms. *Eur J Clin Pharmacol* 2004; 60(8): 531–539.
5. Parrott N *et al.* Predicting pharmacokinetics of drugs using physiologically based modeling—application to food effects. *AAPS J* 2009; 11(1): 45–53.
6. Rosen H, Abribat T. The rise and rise of drug delivery. *Nat Rev Drug Discov* 2006; 4(5): 381–385.
7. Waterman K. A critical review of gastric retentive controlled drug delivery. *Pharm Dev Technol* 2007; 12(1): 1–10.
8. Davis SS. Formulation strategies for absorption windows. *Drug Deliv Technol* 2005; 10(4): 249–256.
9. Coupe AJ, *et al.* Variation in gastrointestinal transit of 718 pharmaceutical dosage forms in healthy subjects. *Pharm Res* 1991; 8: 360–364.
10. Davis SS, Hardy JG, Fara JW. Alimentary tract and pancreas: transit of pharmaceutical dosage forms through the small intestine. *Gut* 1986; 27: 886–892.
11. Singh BN, Malhotra BK. Effects of food on the clinical pharmacokinetics of anticancer agents: underlying mechanisms and implications for oral chemotherapy. *Clin Pharmacokinet* 2004; 43: 1127–1156.
12. Maka DA, Murphy LK. Drug-nutrient interactions: a review. *AACN Clin Issues* 2000; 11: 580–589.
13. Fleisher D *et al.* Drug, meal and formulation interactions influencing drug absorption after oral administration: clinical implications. *Clin Pharmacokinet* 1999; 36: 233–254.
14. Benet LZ, Cummins CL. The drug efflux–metabolism alliance: biochemical aspects. *Adv Drug Deliv Rev* 2001; 50(Suppl 1): S3–S11.
15. Martinez MN, Amidon GL. A mechanistic approach to understanding the factors affecting drug absorption: a review of fundamentals. *J Clin Pharmacol* 2002; 42: 620–643.
16. Deferme S, Augustijns P. The effect of food components on the absorption of P-gp substrates: a review. *J Pharm Pharmacol* 2003; 55: 153–162.
17. Harris RZ *et al.* Dietary effects on drug metabolism and transport. *Clin Pharmacokinet* 2003; 42: 1071–1088.
18. Hendeles L *et al.* Food induced “dose-dumping” from a once-a-day theophylline product as a cause of theophylline toxicity. *Chest* 1985; 87: 758–765.
19. Jonkman JH. Food interactions with sustained-release theophylline preparations: a review. *Clin Pharmacokinet* 1989; 16(3): 162–179.
20. Abrahamsson B *et al.* In-vitro and in-vivo erosion of two different hydrophilic gel matrix tablets. *Eur J Pharm Biopharm* 1998; 46(1): 69–75.
21. Breitzkreutz J, Tuleu C. Pediatric and geriatric pharmaceuticals and formulation. In: Florence AT, Siepmann J, eds. *Modern Pharmaceuticals*. New York: Informa Healthcare, 2009: 221–257.
22. FDA News. FDA asks Purdue Pharma to withdraw Palladone® for safety reasons (July 12, 2005). Available at: <http://www.fda.gov/bbs/topics/news/2005/NEW01205.html> (accessed 2007).
23. Fadd HM, Mohamad MAM, Basit AW. Impairment of the in vitro drug release behaviour of oral modified release preparations in the presence of alcohol. *Int J Pharm* 2008; 360: 171–176.
24. Levina M, Rajabi-Siahboomi AR. The influence of concomitant use of alcoholic beverages on hypromellose matrix tablets. *Pharm Technol* 2008; 20(12).
25. Palmer D *et al.* The influence of hydro-alcoholic media on drug release from polyethylene oxide extended-release matrix tablets. *Pharm Technol* 2011; 35(7): 50–58.
26. Siepmann J, Streubel A, Peppas NA. Understanding and predicting drug delivery from hydrophilic matrix tablets using the “sequential layer” model. *Pharm Res* 2002; 19(3): 306–314.
27. Siepmann J, Peppas NA. Mathematical modeling of controlled drug delivery. *Adv Drug Deliv Rev* 2001; 48(2–3): 137–138.
28. Ozyazici M, Gokce EH, Ertan G. Release and diffusional modeling of metronidazole lipid matrices. *Eur J Pharm Biopharm* 2006; 63(3): 331–339.
29. Chang RK *et al.* Polymethacrylates. In: Rowe RC, Sheskey PJ, Quinn ME, eds. *Handbook of Pharmaceutical Excipients*. London, Chicago: Pharmaceutical Press and American Pharmacists Association, 2009: 525–533.
30. Gallardo D, Skalsky B, Kleinebudde P. Controlled release solid dosage forms using combinations of (meth) acrylate copolymers. *Pharm Dev Technol* 2008; 13(5): 413–423.
31. Tiwari SB, Rajabi-Siahboomi AR. Extended-release oral drug delivery technologies: monolithic matrix systems. *Drug Deliv Syst* 2008; 437: 217–245.
32. Patel N, Madan P, Lin S. Development and evaluation of controlled release ibuprofen matrix tablets by direct compression technique. *Pharm Dev Technol* 2011; 16(1): 1–11.
33. Maggi L, Bruni R, Conte U. High molecular weight polyethylene oxides (PEOs) as an alternative to HPMC in controlled release dosage forms. *Int J Pharm* 2000; 195(1–2): 229–238.
34. Gonçalves-Araújo T, Rajabi-Siahboomi AR, Caraballo I. Application of percolation theory in the study of an extended release verapamil hydrochloride formulation. *Int J Pharm* 2008; 361: 112–117.
35. Aguilar-de-Leyva A *et al.* Study of the critical points and the role of the pores and viscosity in carbamazepine hydrophilic matrix tablets. *Eur J Pharm Biopharm* 2012; 80(1): 136–142.
36. Levina M, Rajabi-Siahboomi AR. The influence of excipients on drug release from hydroxypropyl methylcellulose matrices. *J Pharm Sci* 2004; 93(11): 2746–2754.
37. Tiwari SB, Rajabi-Siahboomi AR. Applications of complementary polymers in HPMC hydrophilic extended release matrices. *Drug Deliv Technol* 2009; 9(7): 20–27.
38. Riis T *et al.* pH-independent drug release of an extremely poorly soluble weakly acidic drug from multiparticulate extended release formulations. *Eur J Pharm Biopharm* 2007; 65: 78–84.
39. Parmar J *et al.* Formulation of extended release multiparticulate systems using ethylcellulose. *Pharma Times* 2010; 42(4): 35–39.
40. Shamblin SL. Controlled release using bilayer osmotic tablet technology: reducing theory to practice. In: Wen H, Park K, eds. *Oral Controlled Release Formulation Design and Drug Delivery: Theory to Practice*. New York: John Wiley & Sons, Inc., 2010: 129–153.
41. Malaterre V *et al.* Oral osmotically driven systems: 30 years of development and clinical use. *Eur J Pharm Biopharm* 2009; 73: 311–323.

Chapter 33

Aerosols

John J. Sciarra, PhD and Christopher J. Sciarra, BS, MS Industrial Pharmacy

INTRODUCTION 633

MODE OF OPERATION 634

PROPELLANTS 639

CONTAINERS 642

VALVES 642

ACTUATORS 643

PACKAGING 643

APPLICATIONS 644

FORMULATION OF MDIS USING HFCS AS THE PROPELLANT 645

NEWER DEVELOPMENTS 649

INTRODUCTION

Inhalation therapy has been used for many years, and there has been a resurgence of interest in delivery of drugs by this route of administration. The number of new drug entities delivered by the inhalation route has increased over the past five to ten years. This type of therapy also has been applied to delivery of drugs through the nasal mucosa, as well as through the oral cavity for buccal absorption. Originally, this type of therapy was used primarily to administer drugs directly to the respiratory system (treatment of asthma); inhalation therapy is now being used for drugs to be delivered to the bloodstream and finally to the desired site of action. Proteins (insulin), steroids, cardiac agents, immunizing agents, etc., are being developed for delivery in this manner.

Drugs administered via the respiratory system (inhalation therapy) can be delivered either orally or nasally. Further, these products can be developed as a

- nebulizer/atomizer
- dry powder inhaler
- nasal inhaler
- metered-dose aerosol inhaler.

Drugs delivered via a nebulizer/atomizer are generally formulated as sterile aqueous solutions (or suspensions) and are inhaled by the patient through an atomizer, nebulizer, or other similar devices. These products are not included in this chapter.

Dry powders have been used for inhalation therapy for over 75 years. The active ingredients were packaged in capsules, representing a single dose of drug. The capsule was punctured, and a small amount of powder fell into a chamber while the patient inhaled. The procedure was repeated until all of the powder was inhaled. While these dry powders were somewhat popular during the early 1940s to 1950s, they fell into disuse with the introduction of the aerosol metered dose inhaler, which became available around 1955. This first generation MDI was formulated with chlorofluorocarbons (CFC), was compact and portable, and contained epinephrine hydrochloride or albuterol as the active ingredient. These MDIs quickly became the dosage form of choice for inhalation therapy, especially for the treatment of asthmatics. With the phase-out of CFCs starting in 1996, dry powders containing about 25–30 to 60 doses of active ingredient were developed and became commercially available from 2000 to 2003. Several dry powder inhalers currently available include salmeterol, fluticasone, and budesonide. Mometasone dry powder inhaler is available in Europe. These dry powder inhalers do not contain a propellant. These consist of active, very potent drugs that are dispensed from a specially designed package. An accurate amount of drug as a dry powder is released

from a small unit dose package while the patient inhales deeply. The dry powder will then travel to the lungs along with the inspired air. Carrier molecules, such as lactose, are often used to reduce agglomeration of the small drug particles as well as facilitate fluidization during the inhalation process. The technology of dry powders is not a part of this chapter.

The nasal metering drug delivery system produces an aqueous spray, consisting of active ingredient and excipients. The drugs used can act locally within the nasal mucosa or systemically by passing through the nasal mucosa and enter the general circulation system. This occurs via numerous capillary vessels present in the mucosa. These nasal sprays can also be formulated similarly to MDIs, using propellants and a nasal actuator. The pressurized type of nasal inhaler is included in this chapter.

The development of the metered-dose inhaler (MDI) in the mid-1950s made possible a convenient dosage form for the delivery of medication to the respiratory system. Atomizers and nebulizers were cumbersome to use and in many instances did not offer convenience of use, so that administration of drugs by atomizers/nebulizers was generally left to hospital or at-home use. While many improvements were made to these nebulizers and atomizers, they lacked the convenience of use especially as to their portability and use outside of a hospital and/or home setting.

Metered-dose inhalers consist of a pressurized container filled with active ingredient, excipients and propellant, and a metered-dose valve. The pressurized container is placed within an oral adapter (mouthpiece), and when the unit is dispensed, an exact amount of drug is expelled in the proper particle size distribution to achieve maximum deposition of drug into the lungs. The aerosol dosage form (MDI) has become the dosage form of choice for delivery of drugs to the lungs. Metered-dose inhalers are formulated as solutions or suspensions of active drug in a mixture of solvents, dispersing agents, and liquefied gas propellants.

Topical aerosol products are becoming more popular because they are easy to administer and have a better feel than ointments and creams. Topical pharmaceutical aerosols can be formulated as a spray, foam, and semisolid. They can be used to deliver therapeutic agents topically (to the skin surface), rectally, and vaginally. They consist of a liquid, emulsion, or semisolid concentrate and liquefied gas or compressed gas propellant. Each of these systems is discussed in later parts of this chapter.

Many therapeutically active ingredients have been administered or applied to the body by means of the aerosol dosage form. This dosage form has been used orally to dispense a variety of agents, such as budesonide, salmeterol xinafoate, fluticasone propionate, fenoterol, epinephrine hydrochloride, albuterol, albuterol sulfate, metaproterenol sulfate, cromolyn sodium, flunisolide hemihydrate, ipratropium bromide, beclomethasone dipropionate, and triamcinolone acetate.

These MDIs were formulated using a CFC propellant. However, these CFC propellants are being phased out.

Oral aerosols have been used mainly for the symptomatic treatment of asthma, as well as for the treatment of several other ailments. These aerosols have been readily accepted by both physician and patient.

ADVANTAGES

One of the main reasons for the rapid and widespread acceptance of the MDI dosage form for the administration of therapeutically active agents is that it affords many distinct advantages to the user. These advantages have been described by various investigators and, for MDIs, include the following:

- Rapid onset of action.
- Circumvention of the first-pass effect and avoidance of degradation in the GI tract.
- Lower dosage that will minimize adverse reactions, especially in the case of steroid therapy, in which most of the steroid reaches the respiratory tract and less is swallowed.
- Dose titration to individual needs and ideal for prn medication.
- Alternative route when therapeutic agent may interact chemically or physically with other medicinals needed concurrently.
- Viable alternative when the drug entity exhibits erratic pharmacokinetics upon oral or parenteral administration.
- Container and valve closure are tamperproof.

The pressure package is convenient and easy to use. Medication is dispensed in a ready-to-use form at the push of a button. There is generally no need for further handling of the medication. Since the medication is sealed in a tamperproof pressure container, there is no danger of contamination of the product with foreign materials, and at the same time, the contents can be protected from the deleterious effects of both air and moisture. Easily decomposed drugs, such as epinephrine, lend themselves to this type of package, for oxygen is excluded from the headspace.

Sterility is always an important consideration with certain pharmaceutical and medicinal preparations. While initial sterility is generally no problem to the manufacturer, there is concern for the maintenance of the sterility of the package during use, for example, with ophthalmic preparations. When necessary, the aerosol package can be prepared under aseptic conditions, and sterility can be maintained throughout the life of the product. For those products requiring regulation of dosage, a metering valve can be used. An accurately measured dose of therapeutically active drug can be administered quickly and, in the case of drugs for inhalation, buccal, or nasal application, in the proper particle-size range.

There are many advantages to the administration of medicinal agents by inhalation, buccally and nasally. Response to drugs administered by inhalation, buccally and nasally, is prompt, often very specific and with minimal side-effects, faster in onset of activity than drugs given orally and, with most drugs, approaching intravenous therapy in rapidity of action. Drugs that normally are decomposed in the GI tract can be administered safely by inhalation, buccally and nasally. The use of the self-pressurized aerosol package makes this type of therapy simple, convenient, and acceptable, compared with the use of atomizers and nebulizers, which are bulky and require cleaning.

DEFINITIONS

The term aerosol is used to denote various systems ranging from those of a colloidal nature to systems consisting of pressurized packages. Aerosols have been defined as colloidal systems consisting of very finely subdivided liquid or solid particles, dispersed in and surrounded by a gas. Originally, the term aerosol referred to liquid or solid particles having a specific size range, but this concept has fallen into disuse.

The present-day definition refers to those products that depend upon the power of a liquefied or compressed gas to dispense the active ingredient(s) in a finely dispersed spray, foam, or semisolid. Pump systems that also dispense the active ingredient(s) in the form of a finely dispersed mist (although of greater particle size) often are classified as aerosols. These pump systems generally are used to dispense medication intranasally.

In 1978, the use of certain chlorofluorocarbons (CFCs) was curtailed by the FDA, EPA, and CPSC. These restrictions applied to the use of Propellants 11, 12, and 114 (CFCs) for use in all aerosol products. Exemptions were granted to MDIs and a few other essential uses. Because of these restrictions, new valve systems and dispensing systems, which allowed greater use of liquefied hydrocarbons and compressed gases, were developed for non MDIs. Individual drugs which have been successfully converted to HFA propellants have had the CFC products phased out. These regulatory requirements are discussed in greater detail in the propellant section of this chapter.

MODE OF OPERATION

LIQUEFIED-GAS SYSTEMS

Liquefied gases have been used widely as propellants for most aerosol products. These compounds are useful for this purpose, since they are gases at room temperature and atmospheric pressure. However, they can be liquefied easily by lowering the temperature (below the boiling point) or by increasing the pressure. The compounds chosen generally have boiling points below 70°F (21°C) and vapor pressures between 14 and 85 psia at 70°F (21°C). When a liquefied-gas propellant is placed into a sealed container, it immediately separates into a liquid and a vapor phase.

Since these materials are liquefied gases, some of the molecules will leave the liquid state and enter the vapor state. As molecules enter the vapor state, a pressure gradually develops. As the number of molecules in the vapor state increases, the pressure also will increase. An equilibrium soon is attained between the numbers of molecules changing from a liquid to a vapor and from a vapor to a liquid. The pressure at this point is referred to as the vapor pressure (expressed as psia) and is characteristic for each propellant at any given temperature. The term psig (pounds/square inch gauge) represents the uncorrected gauge pressure and is to be distinguished from psia (pounds per square inch absolute), which is corrected to include atmospheric pressure (0 psig, which equals 14.7 psia). This vapor pressure is exerted equally in all directions and is independent of the quantity of liquefied gas present.

The pressure exerted against the liquid phase is sufficient to push the latter up a dip tube and against the valve. In cases when there is no dip tube (MDIs), the container is used in the inverted position so that the liquid phase is in direct contact with the valve. When the valve is opened, the liquid phase is emitted and comes into contact with the warm air at atmospheric pressure. The liquid propellant immediately reverts to the vapor state, since its boiling point is substantially below room temperature. As the contents of the container are expelled, the volume within the container occupied by the vaporized propellant increases, causing a temporary fall in pressure. However, as soon as the pressure decreases, a sufficient number of molecules change from the liquid state to the vapor state and restore the original pressure. When a compressed gas is used as the propellant, the relationship is quite different, and there is a drop in pressure as the contents are used.

TWO-PHASE SYSTEM

This is the simplest of all aerosol systems. It consists of a solution or a suspension of active ingredients in liquid propellant or a mixture of liquid propellant and solvent. Both a liquid and a vapor phase are present, and when the valve is depressed,

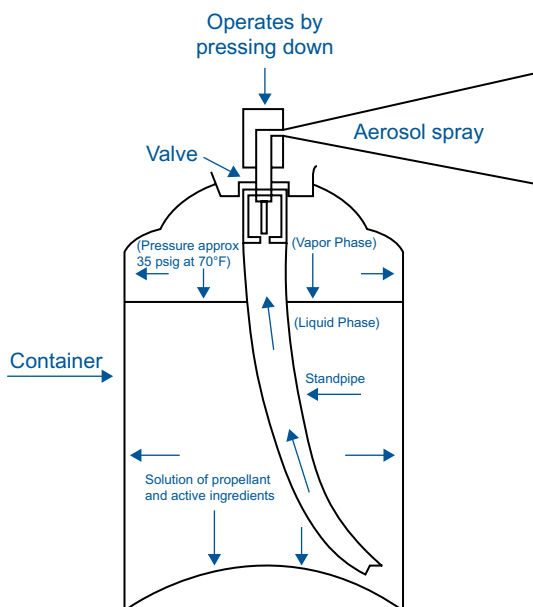


Figure 33-1. Cross section of a typical space or surface coating aerosol spray.

liquid propellant containing dissolved active ingredients and other solvents are released. Depending on the nature of the propellants used, the quantity of propellant present, and the valve mechanism, a fine mist or wet spray is produced because of the large expansion of the propellant at room temperature and atmospheric pressure. This system is used to formulate aerosols for inhalation or nasal application.

Fluorocarbon propellants, primarily trichloromonofluoromethane (P-11), dichlorodifluoromethane (P-12), and dichlorotetrafluoroethane (P-114), are used for MDIs, provided that they currently have been approved under an NDA or ANDA. All new products must use other fluorocarbon propellants that are environmentally acceptable, a hydrocarbon propellant, or a compressed gas. The section dealing with propellants indicates those that are useful for this purpose.

A space spray generally contains from 2 to 20 percent active ingredients and from 80 to 98 percent propellant. While the pressure of space sprays is in the range of 30 to 40 psig, the particles that are produced range from less than 1 to 50 μm . These particles remain suspended in air for relatively long periods of time. Space insecticides, room deodorants, and vaporizer sprays are examples of this type of system. MDIs formulated with a non-CFC propellant will generally have an internal pressure of about 50 to 70 psig.

A surface-coating spray (a relatively wet or coarse spray) can be achieved by decreasing the amount of low-boiling propellants

and increasing the ratio of active ingredients and solvents. The product concentrate can vary from 20 to 75 percent, and the propellant from 25 to 80 percent. Particles are produced ranging in size from 50 to 200 μm . Products such as hair sprays, residual insecticides, perfumes, colognes, paints, protective coatings, and topical sprays are formulated in this manner. The pressure of this system is generally lower than that in the space spray.

Figure 33-1 shows a cross-section of a typical space or surface-coating aerosol spray.

The liquefied gas propellants widely used for MDIs include those shown in Table 33-1. Combinations of these propellants are used to achieve the desired spray characteristics. In certain instances the nature of the product will determine the propellant combination. Dispersion or suspension sprays used for MDIs are similar to space sprays in that they are two-phase systems in which the active ingredients are either dissolved or suspended in the liquid propellant phase. Over the past several years, Propellants 12, 12/11, or 12/114 have been used for these inhalation aerosols and were exempted from the CFC ban. Propellant 134a and 227ea are used to formulate non-CFC MDIs approved by the FDA for use in the United States. Their properties are shown in Table 33-2. Propellant 152a can be used in topical aerosols along with hydrocarbons, dimethyl ether (DME), and compressed gases. The properties of these propellants are included in Tables 33-3 to 33-5.

THREE-PHASE SYSTEM

This system is useful for topical pharmaceutical aerosols in that it allows a greater use of liquid components not miscible with the propellants. Water is not miscible with liquefied-gas propellants and, in many instances, presents a problem, since active ingredients are soluble in water. With the increased emphasis upon the decrease of volatile organic compounds (VOCs) in all products, these systems are finding increased use. These problems have been overcome to a large extent by use of the three-phase system. Depending on the nature of the formulation, one of the following two systems may be employed. Dimethyl ether is most useful for products containing large amounts of water.

Two-layer system

In this system the liquid propellant, the vaporized propellant, and the aqueous solution of active ingredients make up the three phases. Since the liquid propellant and water are not miscible, the liquid propellant will separate as an immiscible layer. When a hydro alcohol mixture is used, the propellant and hydro alcohol solution will mix and form a single layer. When this propellant is of the fluorocarbon type, being denser than water, it will fall to the bottom of the container. Hydrocarbons, on the other hand, are lighter than water and, when used in this manner, will float on top of the aqueous layer. A spray is produced by the mechanical action of an exceedingly small valve orifice through which the liquid and some vaporized propellant are

Table 33-1. Properties of Fluorocarbons (CFCs)

Property		Trichloromono-fluoromethane	Dichlorodi-fluoromethane	Dichlorotetra-fluoroethane
Molecular formula		CCl_3F	CCl_2F_2	$\text{CClF}_2\text{CClF}_2$
Numerical designation		11	12	114
Molecular weight (g/mol)		137.28	120.93	170.93
Boiling point (1 atm)	$^{\circ}\text{F}$	74.7	-21.6	38.39
	$^{\circ}\text{C}$	23.7	-29.8	3.55
Vapor pressure (psia)	70 $^{\circ}\text{F}$	13.4	84.9	27.6
	130 $^{\circ}\text{F}$	39	196	73.5
Liquid density (g/mL)	70 $^{\circ}\text{F}$	1.485	1.325	1.468
	130 $^{\circ}\text{F}$	1.403	1.191	1.36
Solubility in water (weight %)	77 $^{\circ}\text{F}$	0.11	0.028	0.013

Table 33-2. Properties of Hydrofluorocarbons (HFCs)

Property		Tetrafluoroethane	Heptafluoropropane
Molecular formula		CF ₃ CH ₂ F	CF ₃ CHFCF ₃
Numerical designation		134a	227
Molecular weight (g/mol)		102	170
Boiling point (1 atm)	°F	-15.0	-3.2
	°C	-26.2	-16.5
Vapor pressure (psig)	70°F	71.1	43 at (20°)
	130°F	198.7	—
Liquid density (g/mL)	21.1°C	1.22	1.41
Flammability		Nonflammable	Nonflammable
Solubility in water	% w/w	0.15	0.058

Table 33-3. Properties of Hydrochlorofluorocarbons

Property		Difluoroethane
Molecular formula		CH ₃ CHF ₂
Numerical designation		152a
Molecular weight (g/mol)		66.1
Boiling point (1 atm)	°F	-12.0
	°C	-11.0
Vapor pressure (psia)	70°F	63
	130°F	176.3
Liquid density (g/mL)	70°F	0.91
	130°F	—
Solubility in water (weight %)	77°F	<1.0

Table 33-4. Properties of Hydrocarbons and Ethers

Property	Propane	Isobutane	<i>N</i> -butane	Dimethyl ether
Molecular formula	C ₃ H ₈	C ₄ H ₁₀	C ₄ H ₁₀	CH ₃ OCH ₃
Molecular weight (g/mol)	44.1	58.1	58.1	46.1
Boiling point (°F)	-43.7	10.9	31.1	-13
Vapor pressure (psig at 70°F)	110	30.4	16.5	63
Liquid density (g/mL at 70°F)	0.5	0.56	0.58	0.66
Flash point (°F)	-156	-117	-101	—

forced by the vapor pressure of the propellant. The vapor layer is replaced continuously by vaporization of the liquid layer of propellant. This action results in the maintenance of a constant vapor pressure in the headspace.

An important characteristic of this system is that the propellant layer can be adjusted by varying the components so its specific gravity is almost equal to, but does not exceed, that of the hydroalcoholic phase. The propellant floats on top of the hydroalcoholic phase and, when shaken, is dispersed easily. When the valve is depressed, sprays are produced of varying characteristics, depending on the nature of the formulation.

Table 33-5. Properties of Compressed Gases

Property	Carbon dioxide	Nitrous oxide	Nitrogen
Molecular formula	CO ₂	N ₂ O	N ₂
Molecular weight (g/mol)	44	44	28
Boiling point °F	-109 ^a	-127	-320
Vapor pressure, psia, 70°F	852	735	492 ^b
Solubility in water, °c 77°F	0.7	0.5	0.014
Density (gas) g/mL	1.53	1.53	0.96699

^a Sublimes

^b At the critical point (233°F)

^c Volume of gas at atmospheric pressure soluble in one volume of water.

This system is designed to dispense pressurized products efficiently and economically, using relatively small amounts of hydrocarbon, HFA, or HCFC propellants.

The vapor phase of the propellant and the product concentrate enter the mixing chamber in the actuator through separate ducts or channels. The vaporized propellant enters, moving at tremendous velocity, while the product is forced into the actuator by the pressure of the propellant. It is at this point that product and vapor are mixed with violent force, resulting in a uniform, finely dispersed spray. Depending on the configuration of the valve and actuator, either a dry or a wet spray can be obtained.

Water-based aerosols developed for use in this system have the advantage that the chilling effect associated with liquefied-gas systems is eliminated. Since only vaporized propellant is dispensed, less propellant is required in the container. With greater use of water as a solvent for active ingredients, a greater range of products can be developed. Because the use of volatile organic compounds (VOCs) is now being curtailed, water is being used, when possible, as an alternative to some solvents such as alcohol. The use of P-152a and/or DME as a propellant also helps to reduce the VOC content of some aerosols such as hairsprays. Table 33-3 and Table 33-4 illustrate the properties of these propellants.

Foam system

Foam aerosols, which often are classified separately, consist of three-phase systems in which the liquid propellant, which normally does not exceed 10 to 15 percent by weight, is emulsified with the drug-containing liquid. When the valve is depressed, the emulsion is forced through the nozzle, and in the presence of warm air and at atmospheric pressure, the entrapped propellant reverts to a vapor and whips the emulsion into a foam. The use of a dip tube is optional with this type of system, and when present, the container is designed for upright use. For those containers where the dip tube is omitted, the container must be inverted prior to use.

Foam valves have been developed that are applicable to both types of packages. Foam products operate at a pressure of about 30 to 45 psig at 70°F (21°C) and generally contain about 4 to 7 percent propellant, depending upon the nature of the propellant. Hydrocarbon propellants are used at the lower percentage, since their density is much lower than that of their fluorocarbon counterparts. A typical foam-type aerosol can be seen in Figure 33-2. Shave creams, suntan foams, and several topical dermatological aerosols have been formulated as foam aerosols. Generally, a blend of propane/isobutane or propane/isobutane/butane or isobutane alone is used for foam aerosols.

Some foams use P-152a as the propellant, since this propellant will produce a somewhat more stable foam and is less flammable than hydrocarbons. As a replacement for the CFCs, the

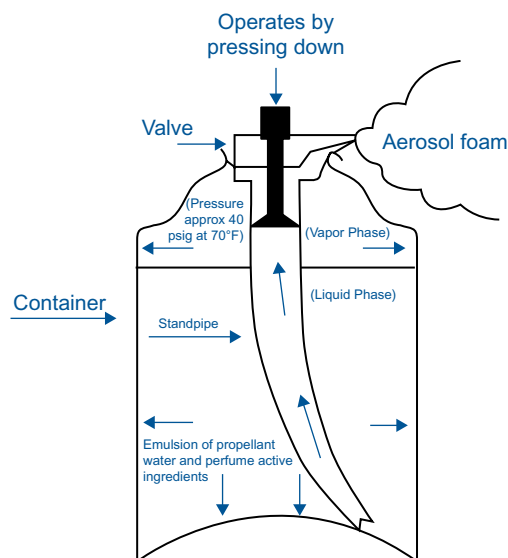


Figure 33-2. Foam type aerosol.

HFA propellant 134a is being currently used for foam formulations, making the product non-flammable. Depending on the formulation, some aerosols use nitrous oxide, carbon dioxide, or a mixture of both as the propellant. Contraceptive foam aerosols are formulated with a hydrocarbon, generally A-31, as the propellant.

COMPRESSED-GAS AEROSOLS

Aerosols using compressed gases as the propellant are finding increased use. These propellants, especially nitrogen, carbon dioxide, and nitrous oxide, are acceptable for use with pharmaceuticals. Compressed gases are used to dispense the product as a solid stream, wet spray, or foam. These aerosol products use an inert gas such as nitrogen, carbon dioxide, or nitrous oxide as the propellant. The gas is compressed in the container, and it is the expansion of the compressed gas that provides the push or the force necessary to expel the contents from the container. As the contents of the container are expelled, the volume of the gas will increase, causing a drop in pressure, according to Boyle's law. This enables one to calculate the drop in pressure as the contents of a compressed-gas aerosol are used. Table 33-5 indicates some of the more important properties of these compressed gases. Depending upon the nature of the formulation and the type of compressed gas used, the product may be dispensed as a semisolid, foam, or spray.

Semisolid dispensing

The concentrate generally is semisolid in nature, and since the gas is insoluble and immiscible with the concentrate, the product is dispensed in its original form. This system is applicable to the dispensing of dental creams, hairdressings, ointments, creams, cosmetic creams, foods, and other products. Compressed-gas aerosols operate at a substantially higher initial pressure of 90 to 100 psig at 70°F (21°C). This pressure is necessary to ensure adequate pressure for the dispensing of most of the contents from the container. The amount of product retained in the unit after exhaustion of the pressure varies with the viscosity of the product and loss of pressure, due to seepage of gas during storage. Since the concentrate generally is semisolid in nature and the dispensing characteristics depend largely on the viscosity of the product and the pressure within the container, the viscosity of the product concentrate must be adjusted accordingly.

Foam dispensing

Soluble compressed gases, such as nitrous oxide and carbon dioxide, can be used to produce a foam when used with emulsion

products. This system is typical for whipped creams and toppings and several pharmaceutical and veterinary products. When this system is used, the gas dissolved in the concentrate will be evolved and cause a whipping of the emulsion into a foam. To facilitate the formation of a foam, this system is shaken prior to use to disperse some of the gas throughout the product concentrate.

Spray dispensing

This system is similar to a space or surface spray except that a compressed gas is used as the propellant. Since these gases do not possess the dispersing power of the liquefied gases, a mechanical breakup actuator is used. The product is dispensed as a wet spray and is applicable to solutions of medicinal agents in aqueous solvents.

Another application for this type of system is found in the contact lens saline solutions. These consist of a normal saline solution packaged in an aluminum aerosol container and pressurized with nitrogen. Since these solutions may come in contact with the eye, they are sterilized using cobalt-60 gamma irradiation.

BARRIER-TYPE SYSTEMS

The barrier-type aerosol system is gaining popularity in the pharmaceutical prescription and over-the-counter markets. Currently many of the sunscreen products are formulated in a barrier system into a spray, lotion or foam. These packaging systems are desirable because they are compact, portable, and ease to rub in. These systems separate the propellant from the product itself. The pressure on the outside of the barrier serves to push the contents from the container. The following types are available.

Piston type

Since it is difficult to empty the contents of a semisolid from an aerosol container completely, a piston-type aerosol system has been developed. This uses a polyethylene piston fitted into an aluminum container. The product is placed into the upper portion of the container. The pressure from nitrogen (about 90 to 100 psig) or a liquefied gas pushes against the other side of the piston, and when the valve is opened, the product is dispensed. The piston scrapes against the sides of the container and dispenses most of the product concentrate. The piston-type aerosol system is shown in Figure 33-3. This system has been used successfully to package cheese spreads, cake decorating icings, and some ointments and creams. Since the products that use this system are semisolid and viscous, they are dispensed as a lazy stream rather than as a foam or spray. This system is limited to viscous materials since limpid liquids, such as water or alcohol, will pass between the wall of the container and the piston. The piston type system has also been used to formulate post-foaming type gels.

Plastic-bag and bag-in-bag type

This system consists of a collapsible plastic bag fitted into a standard, three-piece tinplate or aluminum container, as shown in Figure 33-4. The product is placed within the bag, and the propellant is added through the bottom of the container. Since the product is placed into a plastic bag, there is no contact between the product and the container wall except for any product that may escape by permeation through the plastic bag. A variation of this system is shown in Figure 33-5. The valve and a collapsed inner bag are inserted into a container. A compressed or liquefied gas is added at the same time as the valve/bag is inserted, and then the valve is crimped. The product is forced through the valve and into the bag. The bag expands and will compress the propellant resulting in an increase in pressure. As the valve is opened, the product will be dispensed. Ointments, creams, and gels can be packaged in this system.

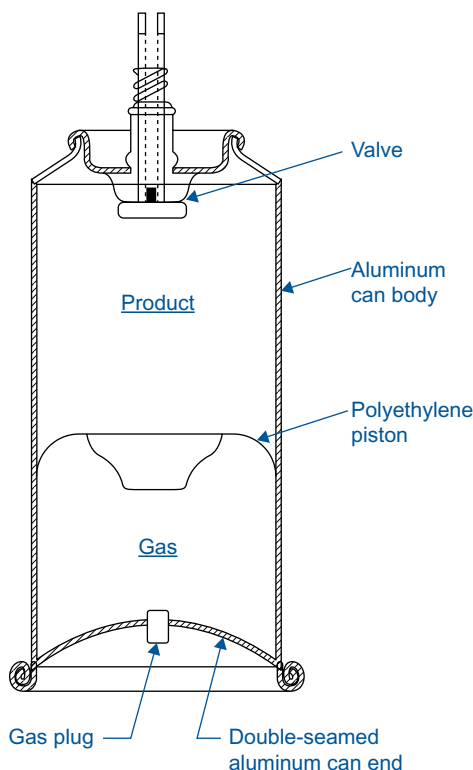


Figure 33-3. Free piston aerosol system.

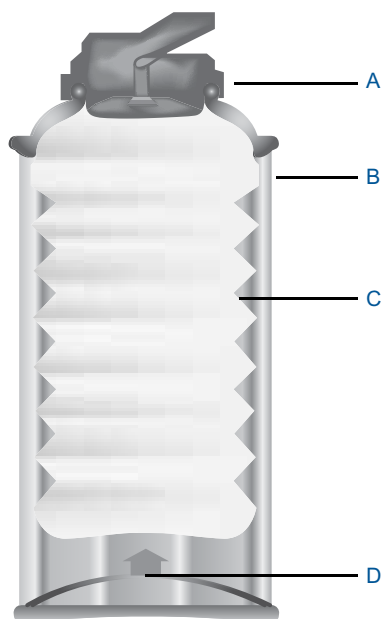


Figure 33-4. Plastic-bag aerosol system.

Limpid liquids, such as water, can be dispensed as either a stream or a fine mist, depending on the type of valve used, while semisolid substances are dispensed as a stream. To prevent the gas from pinching the bag and preventing the dispensing of product, the inner plastic bag is accordion-pleated. This system can be used for a variety of different pharmaceutical and nonpharmaceutical systems, including topical pharmaceutical products as a cream, ointment, or gel.

A modification of the barrier system dispenses the product as a gel that will then foam. By dissolving a low-boiling liquid such as isopentane or pentane in the product, a foam will result when the product is placed on the hands, and the warmth of



Figure 33-5. Bag-in-valve container system. (Courtesy of CCI container.)

the hands will cause vaporization of the solvent. This system, as well as the piston system, is used in post-foaming shave gels.

Can-in-can systems

Figure 33-6 illustrates a system consisting of an aluminum can into which a second aluminum thin-walled can has been inserted. This inner can is glued to the outer can at the neck and forms a gas-tight seal. Then the neck of the can is fabricated. The propellant (liquid or compressed) is added through a small opening in the bottom of the can that is sealed with a rubber plug. A recent addition to this system includes replacement of the inner aluminum pouch with an inner plastic bag made of organic polymers. Sufficient space remains between this bag and the walls and the bottom of the outer container to accommodate sufficient propellant to evacuate the product completely. Systems illustrated by Figure 33-6 can be used with a continuous or metered-dose valve to dispense medicated solutions, gels, creams, and lotions.



Figure 33-6. Cross-section of the Lechner barrier pack. It consists of a rigid or flexible inner bag that can be evacuated more than 95 percent, depending upon the viscosity of the product. (Courtesy of Lechner GMBH.)

PROPELLANTS

The propellant generally is regarded as the heart of the aerosol package. In addition to supplying the necessary force to expel the product, the propellant must also act as a solvent and diluent and has much to do with determining the characteristics of the product as it leaves the container. Various chemical compounds have been used as aerosol propellants.

Compounds useful as propellants can be classified as the following:

- Liquefied gases
- Chlorofluorocarbons (CFC)
- Hydrochlorofluorocarbons (HCFC)
- Hydrofluorocarbons (HFC) or Hydrofluoroalkanes (HFA)
- Hydrocarbons (HC)
- Hydrocarbon ethers
- Compressed gases.

LIQUEFIED GASES

The liquefied-gas compounds have widespread use as propellants, since they are extremely effective in dispersing the active ingredients into a fine mist or foam, depending on the form desired. In addition, they are relatively inert and non-toxic. They have the added advantage that the pressure within the container remains constant. Two types of liquefied gases are used. The chlorofluorocarbons (CFCs) and hydrofluorocarbons (HFCs) find greater use since they are non-flammable in contrast to the flammable hydrocarbons. The hydrocarbons are advantageous since they are less expensive than any of the fluorocarbons and generally are environmentally acceptable.

Chlorofluorocarbons (CFCs)

These compounds have been implicated in causing a depletion of the ozone layer and for responsibility for the greenhouse or for the global warming effect (increase in earth's temperature, rising sea levels, and altered rainfall patterns). Depletion of the ozone layer is also alleged to have resulted in an increase in the incidence of skin cancer. This is due to a greater penetration of the ozone layer by the skin-cancer-causing UV radiation from the sun (the ozone layer will prevent these rays from penetrating the earth's atmosphere). In 1974, the Environmental Protection Agency (EPA), Consumer Product Safety Commission (CPSC), and the FDA promulgated a ban on the use of chlorofluorocarbons, namely Propellants 11, 12, and 114, in most aerosol products. Certain pharmaceutical aerosols for inhalation use (MDIs) were exempted from this ban. According to the Montreal Agreement reached in 1988, beginning in 1989 the production of these propellants was restricted worldwide. Starting January 1, 1996, worldwide production of CFCs was reduced to only the amount needed for certain exempted uses that included MDIs for the treatment of asthma and chronic obstructive pulmonary disease (COPD). CFC propellants are used in all of the metered-dose inhalers in the United States. MDIs have been classified by the Environmental Protection Agency and Food and Drug Administration as "essential use" and as such are exempted from the ban on the use of CFCs. The EPA granted allocation of CFC 11, 12, and 114 to those manufacturers of MDI inhalers currently being sold so that these products can continue to be manufactured and available in the US marketplace. Currently EPA regulations prohibit the granting of a CFC allocation for any MDI classified as essential use and not approved by the FDA prior to December 31, 2000. Essentially this ruling prevents the development and sale of a generic version of these CFC containing MDIs. MDIs containing albuterol are the only MDI products available in generic form. All other CFC propelled MDIs can continue to be sold in the marketplace free from generic product competition.

Generic MDI products containing albuterol and those containing epinephrine have been developed since the late 1990s and are currently marketed by generic pharmaceutical companies. No other generic version of an MDI product is available in the United States.

In the latter part of the 1990s, FDA issued a ruling to encourage the development of ozone-friendly propellants (HFAs) so that essential use designation could be removed. This ruling allowed for the removal of essential use designation from existing MDIs containing CFCs if the following conditions are met:

- At least one non-CFC product with the same active drug is marked with the same route of administration, for the same indication, and with approximately the same level of convenience of use as the CFC product containing that active moiety (while these alternatives are not required to be MDIs, the presumption is that HFA MDIs would most easily fit criteria compared to, for example, dry powder inhalers).
- Supplies and production capacity for the non-CFC product(s) exist or will exist at levels sufficient to meet patient need.
- Adequate US post marketing use data are available for the non-CFC product(s).
- Patients who medically required the CFC product are adequately served by the non-ODS product(s) containing that active moiety and other available products.

The FDA Pulmonary-Allergy Drugs Advisory Committee met on June 10, 2004 to consider removing "Essential Use" designation for albuterol MDIs containing CFCs. The committee was in agreement that albuterol MDI met the first three criteria but was not in agreement as to the last criteria. Concern was raised as to the effect of removal of albuterol MDIs containing CFCs from the market place and what effect their replacement with an HFA albuterol would have upon the cost to the consumer. albuterol MDI is available in the US as a generic product while the HFA albuterol will not be available as a generic until 2015. This has caused concern to the committee members. The FDA has phased out albuterol CFC MDI as of December 31, 2005, after which date no albuterol CFC MDIs can be legally marketed in the United States. Epinephrine (Primatene Mist) containing CFC will be banned starting December 31, 2011.

While MDIs are currently being formulated with a HFC propellant in place of the CFC, only two drug entities have received FDA approval for marketing in the US. These include albuterol and beclomethasone propionate. In contrast to Europe and the rest of the world, almost every CFC metered-dose inhaler has been replaced with a HFC propellant. These replacements will be discussed in greater detail in another section of this chapter.

Since the early 1990s topical aerosol products have been reformulated using a hydrocarbon or a compressed gas propellant in place of the CFC. Compressed gasses, hydrocarbon, and HFC propellants are the suggested alternative propellant. Other than some of the specific properties of the CFCs and the HFCs, the principles of use in dosage form development remain essentially the same.

Liquefied gases provide a nearly constant pressure during packaging operations and have a large expansion ratio. Several of the fluorinated hydrocarbons have an expansion ratio of about 240, that is, 1 mL of liquefied gas will occupy a volume of approximately 240 mL if allowed to vaporize. Dimethyl ether has a value over 350. On the other hand, compressed gases expand only to the extent of 3 to 10 times the original volume.

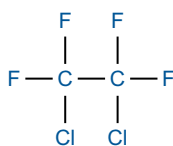
The physicochemical properties of these compounds are of prime importance in the formulation and manufacture of aerosol products. The solvent power, stability, and lack of reactivity

of the propellants have made them extremely useful for this purpose.

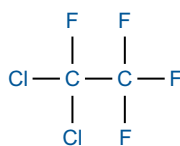
Nomenclature

To refer easily to the fluorinated hydrocarbons, a relatively simple system of nomenclature was developed some time ago by the refrigeration industry. A numerical designation is used to identify each propellant.

- All propellants are designated by three digits (000). When the first digit is zero, the propellant is designated by the last two digits, and zero is assumed to be the first digit (e.g., Propellant 011 is Propellant 11).
- The first digit is one fewer than the number of carbon atoms in the compound. When there are only two digits, (0) is understood to be the first digit and indicates a methane derivative. When this first digit is (1), the propellant is an ethane derivative; when (2), it is propane; and when (3), it is a butane derivative.
- The second digit is one more than the number of hydrogen atoms in the compound.
- The last digit represents the number of fluorine atoms.
- The number of chlorine atoms (for CFCs) in the compound is found by subtracting the sum of the fluorine and the hydrogen atoms from the total number of atoms that can be added to saturate the carbon chain.
- In the case of isomers, each has the same number, and the most symmetrical one is indicated by the number alone. As the isomers become more and more asymmetrical, the letter a, b, c, etc., follows the number.
- For cyclic compounds, a C is used before the number. The use of this system can be exemplified as follows: CFC 114–Dichlorotetrafluoroethane.
- Propellant 114 is an ethane derivative, has no hydrogens, and contains 4 fluorine atoms.
- Since 6 atoms are required to saturate the carbon chain, of necessity there must be 2 chlorine atoms. These can be arranged in two different ways; however, since there is no letter following the numerical designation, the symmetrical structure refers to Propellant 114.



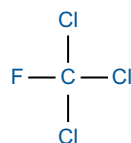
Propellant 114



Propellant 114a

For CFC 11–Trichloromonofluoromethane

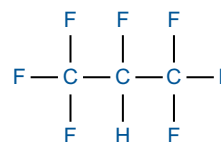
- The designation is 0 for methane (first digit)
- 1 for number of fluorine atoms (third digit)
- 1 for one more than number of hydrogen atoms (second digit)
- 3 chlorine atoms required to saturate molecule.



Propellant 11

For HFC-227–Heptafluoropropane

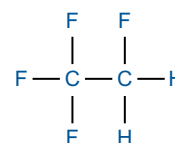
- The designation is 2 for propane (first digit)
- 7 for number of fluorine atoms (third digit)
- 2 for one more than number of hydrogen atoms (second digit)
- Since there is no letter following the third digit, the one H atom must be on the #2 carbon as this is the most symmetrical configuration.



Propellant 227

For HFC-134a–Tetrafluoroethane

- The designation is 1 for ethane (first digit)
 - 4 for number of fluorine atoms (third digit)
 - 3 for one more than number of hydrogen atoms (second digit)
- There are two possibilities for the 2 hydrogen atoms (one on each carbon or two on one carbon). Since this designation has an “a” following the last digit, it is the asymmetrical compound.



Propellant 134a

Physical properties

Table 33-1 shows some of the more useful physicochemical properties of CFC propellants. Propellants 11, 12, and 114 are included in the latest issue of the USP/NF and the *British Pharmacopoeia*. Specifications for these propellants, hydrocarbons, as well as the HFCs, HCFCs, and compressed gases, can be found in the *Handbook of Pharmaceutical Excipients*, Sixth Edition.

From a solubility standpoint, the CFC, HFC, and HCFC propellants, which are non-polar, are miscible with most non-polar solvents over a wide range of temperature. They also are capable of dissolving many substances. For the most part the propellants are not miscible with water, although the degree of miscibility depends on the individual propellants. A cosolvent such as ethanol, 2-propanol, or DME, must be used when water is present to produce a clear solution. However, when one considers that these propellants are used for metered-dose aerosols, the choice of cosolvent is extremely limited and, in many cases, to the use of ethyl alcohol. The alternative is to form an emulsion for topical aerosol pharmaceuticals.

One of the most important physicochemical properties of a propellant is its vapor pressure, which may be defined as the pressure exerted by a liquid in equilibrium with its vapor. When the vapor pressure exceeds atmospheric pressure, boiling and vaporization take place. However, if the vaporized molecules are prevented from leaving the container (by placing the propellant into a sealed container), they will fill the head space and eventually cause an increase in pressure. The pressure developed at equilibrium is the vapor pressure. The vapor pressure of a liquefied gas is independent of the quantity used but is influenced by temperature changes. Assuming ideal behavior for the liquefied gas, the effect of temperature on the vapor pressure can be calculated from

$$\log P = -\frac{\Delta H_{\text{vap}}}{2.303 RT}$$

where P is the vapor pressure, H is the heat of vaporization, R is the gas constant (generally $1.987 \text{ cal deg}^{-1} \text{ mole}^{-1}$), and T is the absolute temperature.

Since

$$\ln P = -\frac{\Delta H_{\text{vap}}}{RT} + C$$

a plot of $\log P$ versus $1/T$ should yield a straight line, and from this the heat of vaporization may be calculated.

$$\Delta H_{vap}(\text{cal mole}^{-1}) = -(\text{slope})(2.303R)$$

These equations can be used to predict the behavior of pure propellants at elevated temperatures. When one considers that an aerosol preparation consists of a propellant and solvents or mixtures of these, the vapor pressure considerations are somewhat different. By mixing various propellants, such as Propellants 11 and 12 or Propellants 12 and 114, a range of vapor pressures is obtained. This is not possible when the HFCs are used, since the range in pressure between P-134a and P-227 is relatively small (about 26 psig, compared with about 70 psig between P-11 and P-12). The vapor pressure of a mixture of propellants may be calculated from Raoult's law, which states that the vapor pressure of a solution is dependent upon the vapor pressure of the individual components. For ideal solutions, the vapor pressure is equal to the sum of the mole fractions of each component present times the vapor pressure of the pure compound at the desired temperature.

Mathematically, this law may be expressed as

$$p_A = \frac{n_A}{n_A + n_B} p_A^\circ = N_A p_A^\circ$$

where p_A = partial vapor pressure of Component A, p_A° = vapor pressure of pure Component A, n_A = mols of Component A, n_B = moles of Component B, and N_A = mol fraction of Component A.

$$p_B = \frac{n_B}{n_B + n_A} p_B^\circ = N_B p_B^\circ$$

The total vapor pressure of the system is obtained by

$$P_{\text{total}} = p_A + p_B$$

When the mole fraction of one component is large, the other component has a small mole fraction, and as such, it does not appreciably affect the vapor pressure. This system approaches ideal behavior.

When the components are of similar physical and chemical nature, the experimentally determined values and the calculated values are approximately the same. In the case of the fluorinated hydrocarbons, the deviation from ideal behavior is not great, and the results are approximately equal or within 5 percent. When other solvents are present, such as alcohols, the vapor pressures can be calculated in a similar manner.

Chemical properties

The fluorinated hydrocarbons have been widely used as aerosol propellants because they generally are considered to be chemically inert. From the standpoint of formulation, the only chemical property that need be considered is hydrolysis, in regard to Propellant 11. While addition of fluorine to a carbon atom generally increases stability, a propellant such as trichloromono-fluoromethane may undergo hydrolysis with the formation of hydrochloric acid. Propellant 11 is not used with aqueous products, as hydrolysis will occur; Propellant 114 generally is used instead. For topical and cosmetic aerosols, hydrocarbons or hydrochlorofluorocarbons are used (Propellants 152b or DME). Propellants 134a and 227 have properties similar to those of P-12 except for their solubility characteristics.

Hydrocarbons

Hydrocarbon propellants have replaced CFCs for topical pharmaceutical aerosols. Their low-order toxicity makes them suitable, while their flammability tends to limit their use. With the development of newer types of dispensing valves, the flammability hazard has been reduced considerably. The advantage of hydrocarbons is their greater range of solubility and a lower cost than CFCs. To date they represent a readily available replacement for CFCs as propellants, provided that the flammability hazard can be reduced.

The HFC propellant used for MDIs is also applicable to topical aerosol pharmaceuticals, with the added advantage of non-flammability.

In addition to having the proper vapor pressure, hydrocarbons have other properties that make them useful as propellants. Their density of less than 1 and their immiscibility with water make them useful in the formulation of three-phase (two-layer) aerosols. Being lighter than water, the hydrocarbon remains on top of the aqueous layer and serves to push the contents out of the container. Not being halogenated, hydrocarbons generally possess better solubility characteristics than the fluorinated hydrocarbons.

As with CFCs, a range of pressures can be obtained by mixing various hydrocarbons in varying proportions. As the composition of the hydrocarbons is likely to vary somewhat, depending on their source, blending of hydrocarbons must be based on the final pressure desired and not on the basis of a stated proportion of each component, whose pressure will depend on its purity. Table 33-6 lists some commonly used blends that are commercially available.

Finally, it should be indicated that the hydrocarbons are characterized further by their extreme chemical stability. They are not subject to hydrolysis, making them useful with water-based aerosols. They will react with the halogens but only under severe conditions.

ALTERNATIVE PROPELLANTS (HCFCS AND HFCS)

Many pharmaceutical aerosols were developed originally using chlorofluorocarbons (CFCs) 11, 12, and 114. These propellants have found widespread use because of their inertness, non-flammability, and non-toxicity. Unfortunately, the CFCs have been implicated in depleting the ozone layer, and their use as aerosol propellants has practically been eliminated, except for exempted medical uses, which included MDIs.

Topical pharmaceutical aerosols have been successfully reformulated with Propellants 152a, DME, hydrocarbons, and compressed gases. Suitable valves are available that, together with modifications in formulation and propellant blends, produce topical aerosol pharmaceuticals that are satisfactory and acceptable.

Several new liquefied-gas materials have been developed to replace the CFCs as refrigerants and foaming agents and in other non-pharmaceutical uses. Propellant 134a and Propellant 227 have been developed as substitutes for Propellant 12 in MDIs and have survived many of the short- and long-term toxicity studies. To date, no suitable replacement has been found for Propellants 11 and 114. Propellant 114 is not essential for use with MDIs, but most of the present suspension formulations require a minimum amount of Propellant 11. Propellant 11 is used to form a slurry with the active ingredient and dispensing agent. This is impossible to accomplish with Propellants 134a and P-227 (unless these propellants are chilled well below their boiling point and handled as a cold fill). Propellant 11 also has

Table 33-6. Commonly Used Hydrocarbon Blends

Designation ^a	Pressure (psig at 70°F)	Composition (mol percent)		
		<i>N</i> -Butane	Propane	Isobutane
A-108	108 ± 4	Traces	99	1
A-31	31 ± 2	3	1	96
A-17	17 ± 2	98	Traces	2
A-24	24 ± 2	49.2	0.6	50
A-40	40 ± 2	2	12	86
A-46	46 ± 2	2	20	78
A-52	52 ± 2	2	28	70
A-70	70 ± 2	1	51	48

^a Designations used by Phillips Chemical Co, Bartlesville, OK. Other designations include: Aeron—Diversified CPC International, Inc., Channahon, IL.

been used to dissolve the surfactants that have been used with CFC MDIs. The HFCs are extremely poor solvents and will not dissolve a sufficient amount of the currently used FDA-approved surfactants (oleic acid, sorbitan trioleate, and soya lecithin).

It also has been noted that some of the currently used valves are not compatible with these newer HFC propellants. The gaskets and sealing compounds used in metered-dose valves may present compatibility problems to the formulator; however, other gasket materials (EPDM) have been developed and found to be satisfactory. Several of the critical properties of these newer propellants are shown in Table 33-2. Additional details about their use in formulation of MDIs is included in a later part of this chapter.

COMPRESSED GASES

The compressed gases such as nitrogen, nitrous oxide, and carbon dioxide have been used as aerosol propellants. Depending on the nature of the formulation and the valve design, the product can be dispensed as a fine mist, foam, or semisolid. However, unlike the liquefied gases, the compressed gases possess little, if any, expansion power and will produce a fairly wet spray and foams that are not as stable as liquefied-gas foams. This system has been used for the most part to dispense food products and, for non-foods, to dispense the product in its original form as a semisolid. Compressed gases have been used in products such as dental creams, hair preparations, ointments, and aqueous antiseptic and germicidal aerosols and are extremely useful in contact lens cleaner saline solution and barrier systems.

CONTAINERS

METAL

Tin-plated steel

To produce an aerosol container that was light and relatively inexpensive, tin-plated steel was used for aerosol containers. This resulted in the large-scale production of aerosol containers. For certain products the tin affords sufficient protection, so that no further treatment is necessary. For additional protection to either the drug product or container, a coating, usually organic in nature, may consist of an oleoresin, phenolic, vinyl, or epoxy coating. The liner (single or double coat) is added to the container prior to fabrication; that is, it is applied to the flat sheets of tin plate.

Aluminum

Many MDIs and pharmaceutical aerosols use an aluminum container. Aluminum containers (sometimes referred to as canisters) are produced by an impact extrusion process so that the container is seamless. These containers are extremely strong and will withstand relatively high pressures. A variety of different aluminum aerosol containers, ranging in size from 10 mL to 1260 mL, is available. While aluminum is less reactive than other metals used in can manufacture, added resistance can be obtained by coating the inside of the container with organic materials such as epoxy, vinyl, phenolic, or polyamide resins. Many of the MDIs will use an anodized or non-anodized internal surface, with or without an organic inner coating. Most of the aluminum containers for MDIs formulated with an HFC propellant are coated with an organic inner coating.

Glass

For pharmaceuticals and medicinals, glass is preferred because of the absence of incompatibilities, as well as for its aesthetic value. The use of glass containers is limited to those products having a lower pressure and lower percentage of propellant. While glass is basically stronger than most metallic containers, a potential hazard is present if, and when, the container is dropped with subsequent breakage. Two types of glass aerosol containers are available. The uncoated glass container has the advantage of decreased cost and high clarity. The contents can

be viewed at all times. The plastic-coated glass containers are protected by a plastic coating that prevents the glass from shattering in the event of breakage. The plastic coated glass container is used for some topical and MDI aerosols.

VALVES

Probably the most basic part of any aerosol or pressurized package is the valve mechanism through which the contents of the package are emitted. Together with the formulation, the valve determines the performance of a pressurized package. The interaction of these two components is such that one cannot readily be discussed without reference to the other.

The primary purpose of the valve is to regulate the flow of product from the container. It provides a means of discharging the desired amount when needed and prevents loss at other times. The valve also exerts a major effect on the character of the dispensed product. For example, a product formulated to produce a foam can be dispensed as a spray or as a wet stream by the use of different actuators or push buttons on the valve. The selection of proper propellants also governs whether a foam, spray, or wet stream will be produced.

CONTINUOUS-SPRAY VALVES

Figure 33-7 illustrates the basic subcomponents used in aerosol valves. A fully assembled valve is shown in Figure 33-8.

A small hole about 0.013 to 0.020 in diameter sometimes is placed in the valve body, as seen in Figure 33-7. This allows the escape of a small quantity of vaporized propellant along with the product. This gives a greater degree of dispersion to the emitted spray, as well as cleaning the valve orifices following discharge. However, since a greater amount of propellant is used than with non-vapor tap systems, care must be exercised during formulation of the product to take this into account. One may also note a change in spray pattern from start to finish because of the change in propellant composition that takes place as the contents are used. Vapor-tap valves are used with powder aerosols, water-based aerosols, aerosols containing suspended materials, and other agents that would tend to clog the valve. They currently are used with hydrocarbon aerosols since the flame extension of the spray can be reduced substantially through use of a vapor-tap valve. This is accomplished by balancing the size of the vapor-tap opening and the valve orifice.

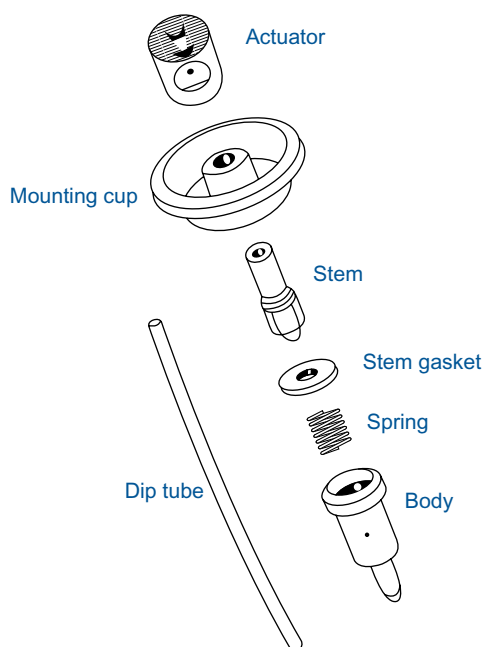


Figure 33-7. Continuous-spray aerosol valve, showing subcomponents used for sprays, foams, and semisolids.

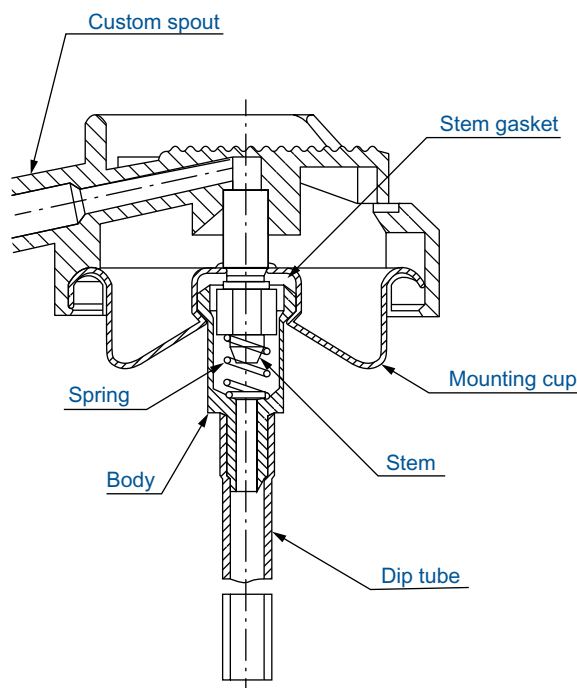


Figure 33-8. Assembled continuous-spray valve.

FOAM VALVES

Valves for foam or aerated products usually have only one expansion orifice, the one at the seat. Following this is a single expansion chamber that serves as a delivery nozzle or applicator. It is sufficiently large in volume to permit immediate expansion of the pressurized product to form the familiar ball of foam. As demonstrated earlier, the same formulation will be discharged as a solid stream when dispensed with a valve and actuator having small orifices and expansion chambers. Under these latter conditions, the ball of foam will begin to develop where the stream impinges on a surface. This rather interesting performance is used in some pressurized surgical soaps on the market. These products are preferred for use by surgeons and other operating room personnel since when applied the foam breaks down easily and is rubbed into large areas of the hands and arms. Another similar product contains ethyl alcohol and is made into a similar foam and used as a skin disinfectant or sprayed onto instruments, etc.

Because of their large openings, foam valves may lend themselves to use with viscous materials such as syrups, creams, and ointments. Foam valves also have been used to dispense rectal and vaginal foams. Metered valves are discussed later in this chapter.

ACTUATORS

The actuator provides a rapid and convenient means for releasing the contents from a pressurized container. It provides the additional functional use in allowing the product to be dispensed in the desired form, that is, a fine mist, wet spray, foam, or solid stream. Mechanical breakup actuators are used for three-phase or compressed-gas aerosols. In addition, special actuators are available for use with pharmaceutical and medicinal aerosols that allow dispensing of products into the mouth, nose, throat, vagina, or eye.

PACKAGING

Two methods have been used to package aerosol products. Unlike non-aerosol products, part of the manufacturing of necessity takes place during the filling operation. The propellant and product concentrate must be brought together in a way that ensures uniformity of product.

Depending on the nature of the product concentrate, the aerosol can be filled by a cold-filling or a pressure-filling process. There are advantages and disadvantages to both methods, and there are many factors that must be considered before deciding which process to use. Since aerosol packaging is a very specialized procedure, many of the pharmaceutical aerosols are manufactured and packaged at commercial contract filling facilities. A typical unit used to fill MDIs in the laboratory is shown in Figure 33-9, while Figure 33-10 illustrates a typical Pamasol aerosol packaging line for MDIs. Figure 33-11 shows

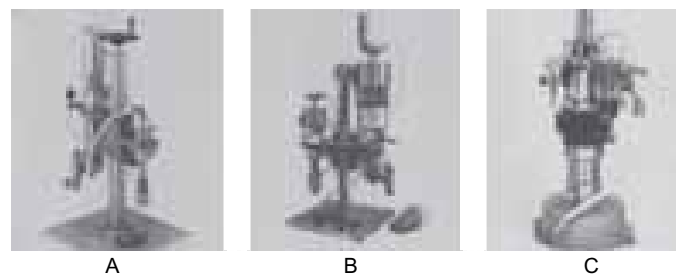


Figure 33-9. Aerosol laboratory and pilot-sized filling equipment. A, Product filler. B, Crimper and pressure filler for propellant. C, Propellant pump. (Courtesy of Pamasol Willi Mader AG.)

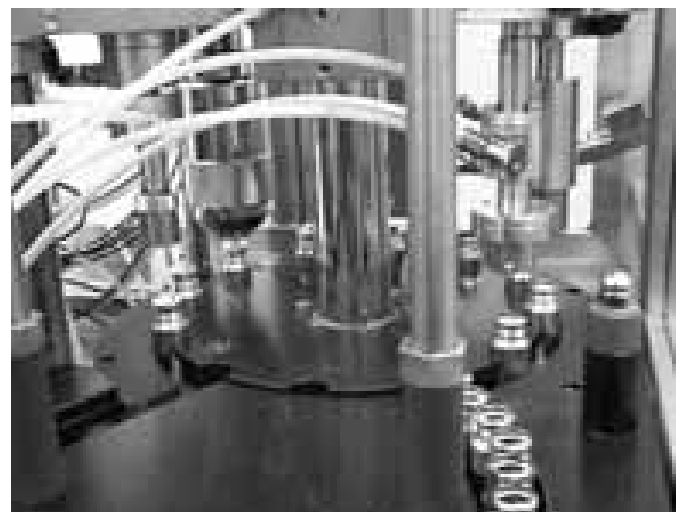


Figure 33-10. (A) Commercial aerosol filling and packaging Pamasol line for MDIs. (B) Detail of crimping and pressure filling head. (Courtesy of Sciarra Laboratories, Inc.)



Figure 33-11. Aerosol filling and packaging Terco line for topical aerosols. (Courtesy of Sciarra Laboratories, Inc.)

a Terco filling and packaging line for topical pharmaceutical aerosols.

APPLICATIONS

Aerosol technology has been applied to the formulation of products containing therapeutically active ingredients. A pharmaceutical aerosol may be defined as an aerosol product containing therapeutically active ingredients dissolved, suspended, or emulsified in a propellant or a mixture of solvent and propellant and intended for oral or topical administration or for administration into the nose, eye, ear, rectum, or vagina.

MDIs are intended for administration as fine, solid particles or as liquid mists via the respiratory system or nasal passages.

Table 33-7. Metered-Dose Inhalants (Solution and Suspensions): Prototype Formulation

Solution (CFC, HFC)^a

Active ingredient(s): solubilized antioxidants: ascorbic acid
Solvent blends: water, ethanol, glycols
Propellants: 12/11, 12/114 or 12 alone; 134a, 227, 134a/227
Suspensions (CFC)

Active ingredient(s): micronized and suspended
Dispersing agent(s): sorbitan trioleate, oleyl alcohol, oleic acid, lecithin, etc.

Propellants: 12/11, 12/114, 12 or 12/114/11 suspensions (HFC)^a

Active ingredient(s): micronized and suspended solvent: ethanol
Dispersing agent(s): sorbitan trioleate, oleyl alcohol, oleic acid, lecithin, etc.

Propellants: 134a, 227, 134a/227

or

Active ingredient(s): micronized and suspended propellants: 134a, 227, 134a/227

^a The reader is directed to the patent literature to ensure that the formulations are not covered by a patent.

They are used for their local action in the nasal areas, throat, and lungs, as well as for prompt systemic effect when absorbed from the lungs into the bloodstream (inhalation therapy). The particle size must be considerably below 10 μm and, in most instances, should be between 3 and 6 μm for maximum therapeutic response.

An alternative definition of this dosage form includes the following: Pharmaceutical aerosols are products that are packaged under pressure and contain therapeutically active ingredients that are released upon actuation of an appropriate valve system. They are intended for topical application to the skin as well as local application into the nose (nasal aerosols), mouth (lingual aerosols) or lungs (inhalation aerosols).

AEROSOL FORMULATION

Topical pharmaceuticals may be formulated as aerosols using solutions, suspensions, emulsions, powders, and semisolid preparations, while MDIs are formulated as solutions or suspensions. Table 33-7 illustrates the basic formulation of aerosol products for use as metered-dose inhalants.

Solution aerosols

Topical aerosols consist of a solution of active ingredients in pure propellant or a mixture of propellant and solvents. The solvent is used to dissolve the active ingredients and/or retard the evaporation of the propellant. Solution aerosols are relatively easy to formulate, provided the ingredients are soluble in the propellant. However, the propellants are non-polar in nature and in most cases are poor solvents for some of the commonly used medicinal ingredients. Through use of a solvent that is miscible with the propellant, one can achieve varying degrees of solubility of the active ingredient. For topicals (isopropyl alcohol, isopropyl myristate, polyethylene glycols, etc.), ethyl alcohol has found the greatest use, although some other solvents may be of limited value. For those substances that are insoluble in the propellant or propellant/solvent system, a dispersion or suspension can be produced. In this case the drug must be micronized so that the particles are less than 10 micrometers in average diameter.

The usual fluorocarbon propellants used in currently available MDIs are blended as indicated in Table 33-7 or used alone when appropriate. Propellant 11 is used often when solubility of the drug and solvents presents a problem, for it is a better solvent than either Propellant 12, 114, 134a, or 227. Additionally, Propellant 11 may be required to prepare a suitable slurry when preparing a dispersion aerosol. Generally the propellant represents upward of 60 weight-percent of the final formulation and, in most cases, may be as high as 99.9 percent. Propellant 12 may be used alone or in combination, as indicated. The proportion of each propellant is varied to obtain the desired pressure within the container and the proper particle-size distribution.

Topical pharmaceutical solutions are formulated using the hydrocarbon propellants, butane, isobutene, and propane. Although butane and isobutene can be used individually, the hydrocarbons are generally used as a blend, as shown in Table 33-6. Other non-MDIs are formulated as aqueous solutions (eye care, etc.) and utilize nitrogen as the propellant. These are packaged as a conventional aerosol or utilize a barrier system.

Dispersions or po(powder aerosols)

These aerosols are similar to solution aerosols except that the active ingredients are suspended or dispersed throughout the propellant or propellant and solvent phase. This system is useful with antibiotics, steroids, and other poorly soluble compounds. Problems associated with the formulation of this system include agglomeration, caking, particle size growth, and valve clogging. Some of these problems have been overcome through use of lubricants such as isopropyl myristate, sorbitan trioleate, oleic acid, or other substances that provide slippage between particles of the compound as well as lubricating component parts of the valve. Surfactants also have been used to disperse the

particles. The use of dispersing agents such as sorbitan trioleate, oleic acid, or lecithin is useful in keeping the suspended particles from agglomerating. Thought also should be given to both the particle size and the moisture content of the powder. The moisture content should be kept between 100 and 300 ppm or less, depending upon the type of product, and the propellants and solvents must be dried by passing them through a drying agent. The particle size for metered-dose inhalants should remain in the micrometer range and should be between 2 and 8 micrometers or less, with a mass median diameter of between 3 and 6 micrometers.

FORMULATION OF MDIS USING HFCS AS THE PROPELLANT

Several new environmentally acceptable propellants are available worldwide as replacements for CFCs in MDI inhalers. Among the alternatives is tetrafluoroethane (HFC 134a), which is available as Dymel 134a/P from DuPont Fluoroproducts; Solkane 134a pharma from Solvay; and Zephex 134a from Ineos Fluor.

Tetrafluoroethane (P-134a) is a hydrofluorocarbon (HFC) or hydrofluoroalkane (HFA) aerosol propellant (contains hydrogen, fluorine, and carbon) as contrasted to a CFC (contains chlorine, fluorine, and carbon). The lack of chlorine in the molecule and the presence of hydrogen reduce the ozone depletion activity to practically zero; therefore, tetrafluoroethane can be considered an alternative to CFCs in the formulation of metered-dose inhalers. It has replaced CFC-12 as a refrigerant since it has essentially the same vapor pressure. Its very low Kauributanol value and solubility parameter indicate that it is not a good solvent for the commonly used surfactants for MDIs. Sorbitan trioleate, sorbitan sesquioleate, oleic acid, and soya lecithin show limited solubility in tetrafluoroethane, and the amount of surfactant that actually dissolves may not be sufficient to keep a drug readily dispersed. Tetrafluoroethane has been used as a replacement for CFCs in MDIs containing albuterol. Two such products containing albuterol sulfate and one product containing beclomethasone dipropionate are currently available in the United States. Outside the US, this propellant has found greater use in Europe and the rest of the world, where many MDIs have been developed using tetrafluoroethane.

The use of tetrafluoroethane as a propellant for MDIs has been the subject of numerous patents throughout the world. These patents cover the formulation of MDIs, use of specific surfactants, cosolvents, etc. Many of these formulation patents are no longer valid in many countries of the world. The US patents have not been challenged to date.

US Patent No. 5605674 claims a self-propelling aerosol formulation which may be free from CFCs which comprises a medicament, 1,1,1,2-tetrafluoroethane, a surface active agent, and at least one compound having a higher polarity than 1,1,1,2-tetrafluoroethane.

The formulator is referred to the patent literature prior to formulating an MDI with tetrafluoroethane and/or heptafluoropropane (P-227) as the propellant. The use of an HFC as the propellant may also requires a change in manufacturing procedure that necessitates a redesign of the filling and packaging machinery for an MDI.

One commercially available MDI is Proventil HFA (Schering), which contains albuterol sulfate suspended in ethanol, oleic acid, and tetrafluoroethane (P-134a). Each actuation delivers 108 µg of albuterol sulfate equivalent to 90 µg of albuterol from the mouthpiece. Similar versions of this product are available in the United Kingdom and the rest of the world. In 1998, 3M released Qvar for sale in the United Kingdom. This MDI contains beclomethasone in solution form. Since the respiratory fraction of the product is substantially higher than the current CFC product, according to the literature, a 200-microgram dose of Qvar achieved a total beclomethasone level comparable to a 400-microgram dose of the CFC-containing beclomethasone formulation. An NDA for Qvar

has since been approved by the FDA and Qvar MDI is now available commercially in the United States.

Another replacement for CFCs in metered-dose inhalers is heptafluoropropane (P-227), which is available as Dymel 227 ea/P from DuPont Fluoroproducts; Solkane 227 pharma from Solvay; and Zephex 227ea.

Heptafluoropropane is classified as a hydrofluorocarbon (HFC) aerosol propellant since the molecule consists only of carbon, fluorine, and hydrogen atoms. It does not contain any chlorine and consequently does not affect the ozone layer, nor does it have an effect upon global warming. It is therefore considered as an alternative propellant to CFCs for metered-dose inhalers. The vapor pressure is somewhat lower than that of tetrafluoroethane and dichlorodifluoromethane but considerably higher than the vapor pressure used to formulate most MDIs. Similar to tetrafluoroethane, heptafluoropropane is not a good solvent for medicinal agents used in the formulation of MDIs. Its use as a propellant is included in US Patent 5605674.

Although there are no MDIs formulated with this propellant currently available in the United States, the rest of the world, including Europe and Asia, have many MDIs that contain P-227. There are several MDIs formulated with P-227 currently under review by the FDA.

Emulsions

An emulsion system is useful for a great variety of topical pharmaceutical products. Since these systems contain a relatively small amount of propellant (4 to 10 percent), there is little, if any, chilling effect. Active ingredients that may be irritating if inhaled can be used as a foam. Depending on the nature of the formulation and the manner in which the product is to be used, the foam is aqueous or nonaqueous and can be stable or quick-breaking.

Emulsions can be dispensed from an aerosol container as a spray, stable foam, or quick-breaking foam, depending on the type of valve used and the formulation. Two types of emulsions can be formulated for use in an aerosol. A W/O emulsion is one in which the water phase is dispersed throughout the oil phase; an O/W emulsion is one in which the water is the continuous phase.

If the product concentrate is dispersed throughout a propellant, the system behaves similarly to a W/O emulsion. However, since the propellant is in the external phase, the product is dispersed as a wet stream rather than as a foam. When the propellant is in the internal phase (O/W), a foam will be produced. The consistency and stability of the foam can be modified by choice of surfactants and solvents used.

Many water-based aerosols are of the W/O type, in which the propellant is in the external phase. Stable shave-cream foams, on the other hand, are produced by keeping the propellant in the internal phase.

The stable foam is similar to a shaving-cream formulation, into which therapeutically active ingredients are incorporated. The foam is dispensed and rubbed into the skin or affected area. By substituting glycols and glycol derivatives for the water in an emulsion, a nonaqueous foam is obtained. The foam stability can be varied by the choice of surfactant, solvent, and propellant. It has been suggested that these foams are applicable to ointment bases, rectal and vaginal medication, and burn preparations.

A quick-breaking foam allows convenient and efficient application of medication. In certain instances the product was dispensed as a foam that quickly collapsed. This was useful in covering large areas with no rubbing necessary to disperse the medication. These quick-breaking foams consist of alcohol, surfactant, water, and propellant.

CONTAINER AND VALVE COMPONENTS

Pharmaceutical containers

Aluminum is used as the material of construction for most metered-dose aerosols. While aluminum can be used without an

internal organic coating for certain aerosol formulations, many containers are available that have been anodized or may have an internal coating made from an epoxy, epoxy phenolic, or polyamide resin.

Aluminum containers are produced with a 20-mm opening so as to receive the standard metered 20 mm and non-metered valves. These canisters are used for MDIs and are fitted with a 20 mm metered-dose valve. albuterol made with a CFC will utilize a non-anodized aluminum canister while other MDIs use an anodized canister or one that has been coated internally with an organic liner. A variety of openings ranging from 13 to 20 mm are available for special and customized applications. Aluminum containers are manufactured from a slug of aluminum and are seamless; therefore, there is virtually no danger of leakage. Figure 33-12 shows a typical aluminum container used for MDIs.

Pharmaceutical valves

A typical metered-dose aerosol delivery system is illustrated in Figure 33-13. Metering valves fitted with a 20-mm ferrule are used with the above containers for all metered-dose inhalation, nasal aerosols, and oral products.

The metering valve delivers a measured amount of product, and the amount delivered is reproducible not only for each dose delivered from the same package but from package to package.

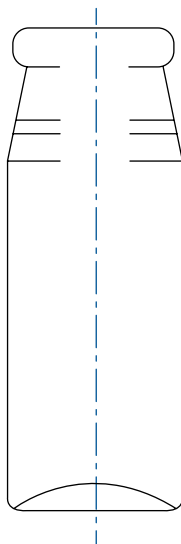


Figure 33-12. Typical aluminum aerosol container, cut-edge type used with an O ring.

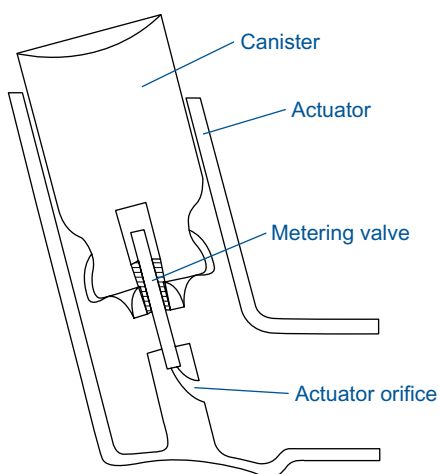
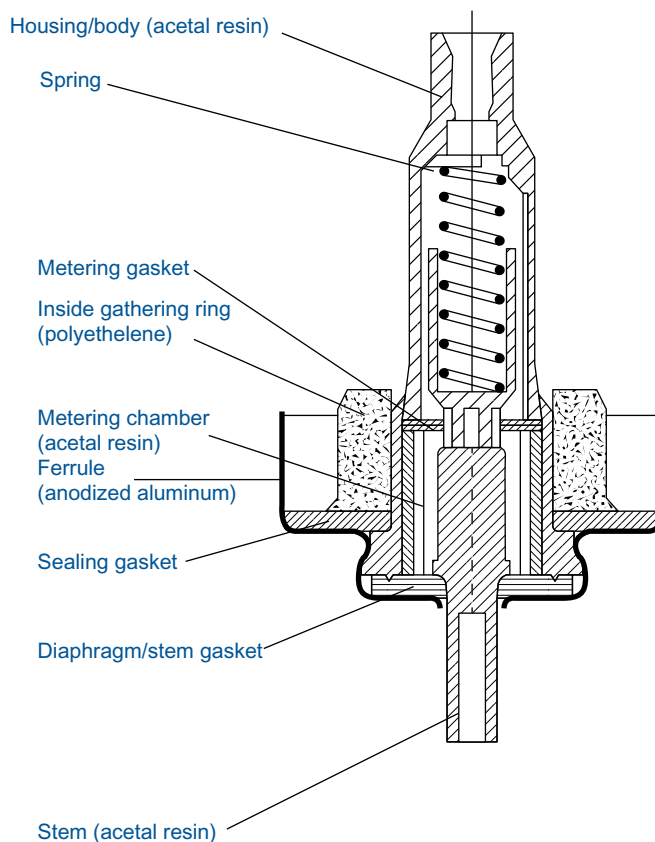
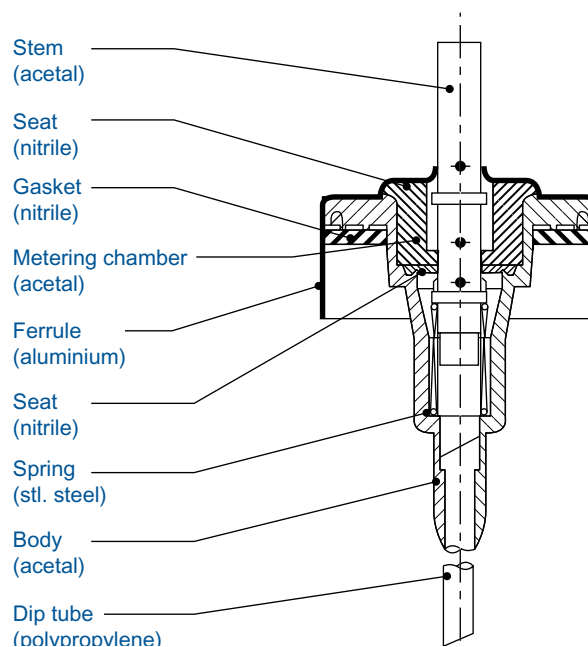


Figure 33-13. Typical metered-dose aerosol delivery system.

Two basic types of metering valves are available, one for inverted use and the other for upright use. Generally, valves for upright use contain a thin capillary dip tube and are used with solution-type aerosols. On the other hand, suspension or dispersion aerosols use a valve for inverted use that does not contain a dip tube. Figures 33-14 and 33-15 illustrate both types of valves and are typical of those commercially available.



Figures 33-14. Metering valve, inverted. (Courtesy of Valois.)



Figures 33-15. Twenty-mm metered-dose valve showing the subcomponent parts and metering chamber. It is used in the upright position. (Courtesy of Bespoke.)

An integral part of these valves is the metering chamber that directly is responsible for the delivery of the desired amount of therapeutic agent. The size of the chamber can be varied, so that from about 25 to 150 microliters of product can be delivered per actuation. Most of the products commercially available use dosages in the range of 25 to 75 microliters. The chamber is sealed via the metering and stem gasket. In the actuated position, the stem gasket will allow the contents of the metering chamber to be dispensed while the metering gasket will seal off any additional product from entering the chamber. In this manner the chamber always is filled and ready to deliver the desired amount of therapeutic agent.

These valves should retain their prime over fairly long periods of time. However, it is possible for the material in the chamber to return slowly to the main body of product in the event the container is stored upright (for those used in the inverted position). The degree to which this can occur varies with the construction of the valve and the length of time between actuations.

Both types of valves currently are used on commercially available oral inhalation aerosols. During the development stage, the compatibility of the valves should be determined with the exact formulation to be used, to determine the accuracy of the metered-dose in regard to doses delivered from the same container of product and from different containers. Additionally, one should ensure that there is no interaction between the various valve subcomponents and the formulation. If distortion or elongation of some of the plastic subcomponents occurs, this may result in leakage, inaccurate dosage, and/or decomposition of the active ingredients.

There also have been instances in which the therapeutic agent was adsorbed or absorbed onto the various plastic components, and a lower than normal dose of the active ingredient was dispensed. For these reasons, one must not only determine the total weight of product dispensed per dose but also the actual amount of active ingredient in each dose. Some test procedures use the results obtained by taking ten doses of material and determining the average amount present in one dose. When possible and when the analytical procedure permits detection of fairly small amounts of active ingredients present per dose, multiple single-dose assays should be performed. Using the average of ten doses may fail to reveal problems of variations in each of the individual doses dispensed.

EVALUATION OF MDIS AND TOPICAL PHARMACEUTICALS

Various tests have been devised to ensure the integrity of the aerosol package. These aerosol products are said to be tamper-proof, since they cannot be opened and closed in the usual manner. Because these products are all under pressure, it is very difficult to add any foreign material to the product once the entire package is assembled. This also makes it rather difficult to obtain suitable samples for an analysis. Special sampling procedures and test methods have been developed and are used to determine the suitability of the product.

Topical pharmaceutical aerosols do not present any special problems other than the sampling procedure. The USP includes several tests under the specific monographs for the topical aerosols. These include delivery rate, leak testing, microbial limit test, and assay. While several of these products are dispensed as sprays, no special emphasis or consideration is given to the particle size of the droplets or particles emitted. The spray may be defined as a fine, dry, or wet spray. Of special interest for topical spray products is the concern that some of the smaller particles may be inhaled by the user.

MDIs require a greater amount of testing since the metered valve, oral adapter, and the formulation are collectively responsible for delivering the therapeutically active ingredient to the appropriate site in the respiratory passages. This assumes that the patient will administer the product properly, so that both the dose and depth of penetration of the medication can be ensured. Unfortunately, this is not always done. Both the physician and the pharmacist

have provided a most valuable service to the patient by taking the time to demonstrate the correct use of these inhalers.

Many of the tests required for the evaluation of MDIs are similar to those used for other dosage forms. These include description, identification, and assay of the active ingredient; microbial limits; moisture content; net weight, degradation products and impurities (if any); extractables; and any other tests deemed appropriate for the active ingredient.

Other tests specific for MDIs include the following:

Dose uniformity over the entire contents—(MDIs only)

This test is described in the USP/NF and determines the amount of active ingredient delivered through the mouthpiece (oral adapter) per a specified number of actuations (dose taken by patient).

Leakage rate

This test is also available in the USP/NF and is used to estimate the weight loss over a one-year period. Since there are several sealing gaskets present in a metered-dose valve, this test determines the integrity of the gaskets, as well as the proper crimping of the valve onto the container.

Total number of discharges per container

This is defined as the number of actuations per container and is not less than the label claim.

Spray pattern and/or plume geometry

This test evaluates the type of spray pattern emitted for the MDI and relates to the characteristics of the metering valve and oral adapter, as well as to the formulation.

It is beyond the scope of this chapter to discuss these tests in greater detail. The reader is referred to the USP/NF for specifics on each test. However, particle-size distribution is covered in greater detail because of its relationship to deposition of particles of drug in the respiratory system.

PARTICLE-SIZE DISTRIBUTION

Particle-size distribution is probably one of the most important characteristics of an MDI. To be effective, the particles emitted from the spray must be below 10 micrometers and, in most cases, between 2 and 8 micrometers in diameter. Several methods are available for the determination of the particle-size distribution for MDIs. A common method includes a cascade impactor that depends upon the principle of carrying particles in a stream of air through a series of consecutively smaller jet openings. The heavier and larger-diameter particles are impacted on a slide under the larger opening, and as the openings

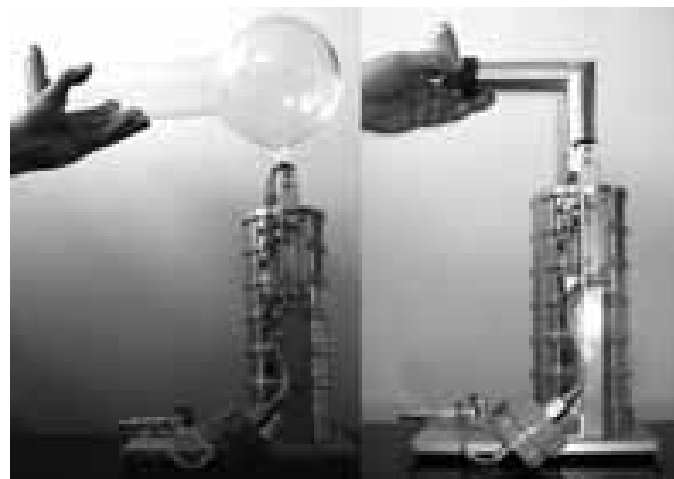


Figure 33-16. Cascade impactor (Impaq AS-6) used for particle-size distribution of MDIs (left-fitted with a 1000 mL sampling throat; right-fitted with USP throat).

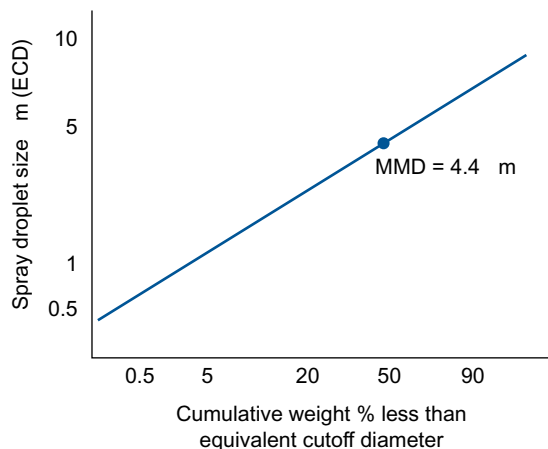


Figure 33-17. Log probability plot of data from the cascade impactor (MMD, mass median diameter).

get smaller, the velocity of the stream increases, and the next larger particles are deposited on the next slides. Figure 33-16 illustrates a cascade impactor that can be used to indicate the particle-size distribution of MDIs.

Figure 33-17 and Table 33-8 give a typical analysis of a suspension MDI. Table 33-9 gives a full analysis of the particle size for a solution MDI. Other methods include the use of a microscope or instrumentation based on the use of laser technology. The reader is referred to the USP/NF for a more comprehensive review of this subject.

Formulation factors

Included among formulation factors are the physicochemical characteristics of the active ingredients, the particle size and shape of the drug, the type and concentration of surface-active agent used, and, to some extent, the vapor pressure and the metered volume of propellants. In terms of physicochemical properties, the lipoidal solubility and pulmonary absorption rates of the active ingredient are of utmost importance. Another physicochemical factor governing the biopharmaceutics of a drug is its dissolution characteristics in pulmonary fluids. Drugs having a rapid dissolution rate in pulmonary fluids predictably produce much more intense and rapid onset of action, having a shorter duration than their less soluble derivatives. Therapeutic agents that exhibit very poor solubility in pulmonary fluids are to be avoided since they are likely to serve as irritants and precipitate bronchial spasms.

The selection of the appropriate surface-active agent (required in most pressurized inhalation suspension aerosols) is another important consideration, since the surfactant will influence droplet evaporation, particle size, and overall hydrophobicity of the particles reaching the respiratory passageways and pulmonary fluids. Solubility of these dispersing agents or surfactants is limited when formulating with an HFC propellant. Ethyl alcohol has been added to increase their solubility.

Table 33-8. Cumulative Particle Size Distribution

Slide no.	Particle size (μm)	Cumulative particle size distribution (percent)
Filter	Less than 0.5	0.55
6	0.5–1	5.35
5	1–2	18.98
4	2–4	61.59
3	4–8	90.2
2	8–16	96.67
1	16–32	100

Table 33-9. Particle Size Distribution of Solution Type MDI

Collection unit	Mass found		Cumulative particle size distribution
	μg	percent	percent
Valve stem	38.7	1.43	—
Mouthpiece	298.3	11.02	—
Collar	27.6	1.02	—
Induction Port	1057.8	39.07	—
Filter—(0.3 μm)	70.2	5.46	5.46
Stage 6—(0.5 μm)	66.3	5.16	10.61
Stage 5—(1.0 μm)	151.4	11.77	22.39
Stage 4—(2.0 μm)	303.8	23.64	46.02
Stage 3—(4.0 μm)	408.8	31.8	77.83
Stage 2—(8.0 μm)	204.4	15.9	93.73
Stage 1—(16.0 μm)	80.6	6.27	100
Total A	2300.7		
Total B	1285.4		
Total C	2707.8		
Total R	1204.7		

Respirable dose (μg) = 120.47

Respirable fraction = 52.36 percent

Mass median aerodynamic diameter (MMAD) (μm) = 1.95

Geometric standard deviation (GSD) = 2.81

Mass balance = 99.73 percent

A = Total mass of drug found on the collar, induction port, filter, and stages 1–6.

B = Total mass of drug found on the filter and stages 1–6.

C = Total mass of drug found on the valve stem, mouthpiece, collar induction port, filter, and stage 1–6.

R = Total mass of drug found on stages 2–6 and filter.

The effects of propellant vapor pressure and the metered volume of propellants on drug deposition in the lungs recently have been studied using rather large, specialized, plastic adapters. Findings in this area have demonstrated that the amount of material deposited in the mouth, tube, and actuator (likely sites of material loss) increased as the vapor pressure was decreased and the metered volume increased.

Component design

Component design, specifically that of the actuator and adapter, also has been shown to alter the particle size and the penetration and deposition of drugs into the lungs. Numerous studies have demonstrated that a complex set of interactions exist between the actuation type, valve dimensions, distance from actuator, and other component variables and that particle size (mass median diameters) could vary up to 40 percent by altering one or more of the aforementioned components.

One component that has undergone enormous modification in the last few years to improve drug delivery is the adapter or mouthpiece. Up to about the mid-1970s, almost all adapters were short and rather simplistic so as to minimize possible holdup of material in the adapter. The holdup in the short-stem adapters averages anywhere from 5 to 20 percent. Recently, however, numerous customized adapters having specific designs and dimensions have entered the marketplace.

Interest in the larger adapters (often referred to as tube spacers) can be attributed to any one or more of the following reasons. The larger adapter designs permit a complete evaporation of propellant, reducing initial droplet velocity and particle size. This reduction of particle size improves depth of drug penetration into the lungs, while a lower initial velocity decreases product impaction to the back of the esophagus (whiplash effect),

an effect common to short-stem adapters. The larger adapter designs also permit a decrease in pressure and increased volume flow that also has been reported to increase penetration of particles into the lungs. It should be pointed out that the larger tube spacers are not without problems. They are inconvenient because of their size, are expensive, and are somewhat difficult to clean. They also present the manufacturer with the problem of assessing product holdup in a rather complex device.

Administration techniques

The metered-inhalation aerosol dosage form, although popular, generally is considered one of the most complicated to use drug-delivery systems currently marketed by the pharmaceutical industry. It is viewed by many as being only slightly simpler to use than an injectable, since inhalation products often require up to 10 to 15 maneuvers by the patient during use. Failure of the patient to perform any one of these maneuvers correctly may alter significantly the deposition of the drug into the appropriate portion of the lungs.

Differences in the directions for use of each inhalation product are a result of the product formulation and actuator design that the manufacturer deemed most appropriate for the particular product. In light of this, it is not surprising to find patients who require two or more aerosol inhalation products or who are constantly changing their medication (such as the asthmatic patient) occasionally experiencing difficulties in complying with the suggested method of application.

These problems present a unique opportunity for the pharmacist to counsel the patient on the correct use of these inhalers. Several manufacturers will provide placebo inhalers for this purpose. Others will provide videotapes that can be used by the pharmacist and other health professionals to teach the correct use of these inhalers. Brown bag and senior citizen programs, health seminars, and other similar programs can provide a suitable audience for group presentations. However, the most successful programs are conducted on a one-to-one basis in the privacy of the pharmacy.

Many attempts have been made to overcome these problems and increase the efficacy of this dosage form. A breath-activated inhaler has been developed by 3M Pharmaceuticals and is used as an integral part of their pirbuterol acetate inhalation aerosol. They found that in a study of 70 patients, the use of the breath-activated inhaler increased the efficient use of these inhalers from 50 to 91 percent. These patients were given both written and verbal instructions. Reading instructions alone increased the efficiency from 39 to 63 percent. Many other devices, including tube spacers, breath activated, and electronic devices have been under study for numerous years. To date none of these units have become commercially available other than one breath-activated (Maxair Autohaler – 3M) and one spacer (Azmecort In-halation Aerosol – Aventis Pharmaceuticals).

Tube spacers also increase the efficiency of drug delivery via MDIs. These spacers permit atomization of the delivered dose in a confined chamber or bag and eliminate the need for the precise synchronization of actuations and deep breathing with inspiration. Triamcinolone acetonide MDI is available as an MDI and is fitted with a tube spacer instead of the conventional short-stem actuators. Patients must check with their physician before using these spacers, since the particle-size distribution of the drug being dispersed through one of these spacers may be substantially different from those emitted from a short-stem actuator. Certainly, the deposition pattern will change, and the efficiency of the delivered dose reaching the proper pulmonary airways will be increased.

NEWER DEVELOPMENTS

At present, there is much interest in developing MDIs for a variety of conditions, including asthma, emphysema, diabetes, AIDS, cancer, heart disease, and cystic fibrosis. Many of these

compounds have been developed using biotechnology processes, and their delivery to the respiratory system via an MDI is an extremely challenging undertaking. With the introduction of newer, alternative propellants, the challenge becomes even greater and presents a unique opportunity for the delivery of these compounds. There is very little interest by manufacturers of currently available MDIs formulated with a CFC to convert the existing therapeutic agents to a HFC product. As indicated previously, albuterol and beclomethasone propionate are the only two drug entities available in the United States with a HFC propellant. It is doubtful if any others will ever be converted to a HFC. Manufacturers are currently discovering newer therapeutic agents for development as an MDI, using the environmentally acceptable HFA propellant. Several “second generation steroids” are currently under development and are formulated with a HFA propellant.

The valve and container suppliers are cooperating with the industry to develop much-needed hardware to accommodate this change. At present, there is no specific, scheduled date when manufacturers can no longer use CFCs for their metered-dose inhalers. As long as CFCs remain available, the change-over will continue at the current slow pace.

BIBLIOGRAPHY

- Burke GP, Poochikian G, Botstein P. Regulatory science of inhalation aerosols. *J Aerosol Med* 1991; 4: 265–8.
- Dalziel SM, Creazzo JA. Pharmaceutical aerosols propellant, particle perspectives for pulmonary drug delivery. *Spray Technology and Marketing* 2003; 13(6): 20–4.
- Hickey AJ. *Pharmaceutical Inhalation Aerosol Technology*. New York: Dekker, 1992.
- Johnsen M. *The Aerosol Handbook*, 2nd edn. Caldwell, NJ: Wayne E Dorland Co., 1982.
- Johnsen MA. Allergies, asthma and aerosols. *Spray Marketing and Technology* 2001; 11(5): 42–8.
- Johnsen MA. Compartmentalized dispensers: A new era for aerosols. *Spray Marketing and Technology* 2001; 11(10): 22–9.
- Johnsen MA. Compartmentalized dispensers. Part II: The piston can. *Spray Marketing and Technology* 2001; 11(11): 36.
- Johnsen MA. Compartmentalized aerosols Part IV, bag-in-can concepts. *Spray Marketing and Technology* 2002; 12(2): 41–4.
- Leskovsek N. Don't hold your breath. *Drug Delivery Technology* 2003; 3(7): 30.
- Purewal TS, Greenleaf DJ. *Medicinal Aerosol Formulations*, US Pat 5,605,674, Feb 25, 1997.
- Robinson G. Propellants for MDIs: Medical propellants meet specialized criteria. *Spray Technology and Marketing* 2003; 13(6): 16–18.
- Sanders P. *Handbook of Aerosol Technology*, 2nd edn. Malabar, FL: Robert E Krieger Publ, 1987.
- Sciarrà CJ, Sciarrà JJ. Monographs on aerosol propellants. In: Rowe RC et al. eds. *Handbook of Pharmaceutical Excipients*, 6th edn. London UK: AphA and the Pharmaceutical Press, 2003.
- Sciarrà JJ et al., edn. *Remington: The Science and Practice of Pharmacy*, 21st edn. Philadelphia: Lippincott, Williams, and Wilkins, 2000; p 1000.
- Sciarrà JJ. Aerosol suspensions and emulsions. In: Liebermann H, Rieger M, Banker G, eds. *Pharmaceutical Dosage Forms: Disperse Systems*, vol 2, 2nd edn. New York: Marcel Dekker Inc., 1996; p 319.
- Sciarrà JJ. Pharmaceutical aerosols. In: Banker GS, Rhodes CT, eds. *Modern Pharmaceutics*, 3rd edn. New York: Marcel Dekker, Inc., 1996; p 547.
- Sciarrà JJ, Sciarrà CJ. “Aerosols”. In: *KirkOthmer Encyclopedia of Chemical Technology*, 2001. Available at <http://jwsedck.interscience.wiley.com:8095/articles/aeroscia.a01/abstract.html>
- Sciarrà JJ. The next generation of metered-dose inhalers. *US Pharm* 1997; 22(7): 37.
- Sciarrà CJ, Sciarrà JJ. Validation of inhalation aerosols. In: Nash RA, Wachter, AH eds. *Pharmaceutical Process Validation*, 3rd edn. New York: Marcel Dekker, Inc., 2003; p 329.
- Sciarrà CJ, Sciarrà JJ. CFC-free MDIs: A mix of factors influences development in the US and Europe. *Spray Technology and Marketing* 2003; 13(6): 25–6.
- Strobach DR. *Aerosol Age* 1988; 33(7): 32
- USP 26/NF21. Physical Tests and Determination: <601>= Aerosols, Metered Dose Inhalers, and Dry Powder Inhalers. 2002; p 2105.

Chapter 34

Biotechnology and Drugs

Ara H. DerMarderosian, PhD and Zhiyu Li, PhD

BACKGROUND 651

BIOTECHNOLOGY DRUGS 653

UNIQUE PHARMACEUTICAL CHALLENGES OF
BIOTECHNOLOGY-DERIVED THERAPEUTICS 658

PHARMACOGNOSTICAL APPLICATIONS 662

PHARMACOLOGICAL APPLICATIONS 664

PHARMACEUTICAL MANUFACTURING
APPLICATIONS 664

ORGANIC SYNTHESIS APPLICATIONS 664

MORAL AND ETHICAL QUESTIONS 664

APPENDIX A 666

The past 40 years have witnessed the emergence, development, and maturation of biotechnology in medicine and pharmaceuticals. Previously rare, or even unattainable, pharmaceuticals can now be produced in useful quantities by harnessing the power of molecular biology. Interestingly, the term *biotechnology* was first coined in 1919 by the Hungarian engineer Károly (Karl) Ereky to describe how products could be produced from raw materials with the aid of living organisms as agriculture began to join forces with industry following World War I.¹ Hence, biotechnology is not a new concept. Humans have been manipulating living organisms over the millennia to solve problems and improve the quality of life. But today, especially in the context of science and health, the term biotechnology is used interchangeably with “genetic engineering.” The concept of DNA manipulation is central to most modern references to biotechnology.

The practical realization of this technology has followed from our ability to now detect, decode, isolate, produce, and characterize the various proteins that coordinate the numerous functions essential to human life and health. Processes that precede or are causative in pathophysiology can not only be identified but also now manipulated in an attempt to restore normal function. This relatively new methodology involves the synergism of discoveries in recombinant DNA methodology, genetic engineering, immunology, genomics, proteomics, and bioinformatics, with advances in automation and data analysis to create a cogent, high-technology industry. Overall, biotechnology has led to the creation of new products for home and industry, improvement of agricultural yields, diagnosis of genetic disorders, and the enhancement of our medical arsenal against disease. The publications in February 2001 of the virtually complete sequence of the human genome^{2,3} will certainly accelerate the application of these technologies. While the close of the past millennium has clearly witnessed the benefits resulting from the proliferation of biotechnology-derived products, new questions have arisen regarding issues of ethics and pharmacoeconomics. Nonetheless, it is clear that the benefits of biotechnology already have far outweighed the drawbacks.

BACKGROUND

As biotechnology-derived pharmaceuticals have become commonplace in healthcare, pharmacy practitioners should have a detailed knowledge of the manufacture and use of these newer agents.^{4,5} As a backdrop to understanding modern biotechnology, it will be instructive to review some of the basic biological milestones that precede it. Table 34-1 provides a compilation of milestones in biotechnology, especially for their connections with pharmaceutical sciences. It is clear that technology is proceeding at a rate that is already threatening to bypass our ability to manage the ethical dilemmas presented by these advances. Fortunately, visionaries such as Nobel laureate James Watson have used their positions to encourage the proper and ethical use of

genetic information and technology. As the initial director of the publically funded Human Genome Project, Watson announced 18 years ago a plan to set aside 3% of the project budget devoted to ethical, legal, and social implications (ELSI) research, a decision he recently deemed, “probably the smartest thing I did.”⁶

Nature has for some 3.5 billion years been conducting what we may call natural genetic experiments. These include mutation (random heredity alteration), crossing-over (breakage and exchange of corresponding segments of homologous chromosomes), and recombination at meiosis (fertilization). These processes all have contributed to the current diversity of life on this planet. In addition, it is well known that humans have been manipulating genetic characteristics of different species for over 10,000 years through inbreeding and cross-breeding experiments. To cite a few examples, one can point to the modern robust strains of wheat or corn, which are a far cry from their puny ancestors. Similarly, the varied breeds of dogs, cats, poultry, and cattle may be mentioned. These manipulative efforts continue, and in less than a lifetime, the development of larger and sweeter oranges, seedless watermelons, and flamboyant ornamental plants has occurred. Also familiar are such hybridizations as the tangelo (crossing the tangerine and the grapefruit) and the mule (crossing a donkey and a horse).

All cell structures and functions begin with proteins, and the code for building the proteins is found in deoxyribonucleic acid (DNA). This is why the discovery of the double-helix structure of DNA by Watson and Crick in 1953 fundamentally began the unraveling of the mystery of cell processes. (The 50th anniversary of publication of their model was celebrated in 2003, with some exceptional retrospective documentation published in print and on the Internet by Cold Spring Harbor Laboratory and the journals, *Science* and *Nature*.) DNA, the genetic blueprint of an organism, is made up of building blocks known as nucleotides (molecules containing a sugar, nitrogen-containing purine or pyrimidine bases, and a phosphate group) that are connected in a very long ladder-like structure. When this rubber-like twisted-ladder structure is coiled tightly, it is referred to as a two-stranded, or double, helix.

There are four different nucleotides (containing the bases adenine, cytosine, guanine, and thymidine) with a total of about 3 billion nucleotide units in the human genome, tightly packed into chromosomes. These include the genetic code for a large number of genes, originally estimated at 100,000 in the human but downgraded to roughly 25,000 as a result of the Human Genome Project, a surprisingly low number compared with other species. Each of these genes controls the synthesis of a protein made up of a long strand of anywhere from 50 to 3000 amino acids. Nirenberg and Matthaei, in 1961, and others later, elucidated how the nucleotide sequence of a gene regulates the particular sequence in which the 20 different amino acids will be united to produce a particular protein. A single codon is made up of units of three adjacent nucleotides; each codon specifies

Table 34-1. Milestones in Biotechnology. The recent explosion of growth in the development and application of biotechnology may be traced to a number of successive, discrete, milestone discoveries and events.

Discoveries and Events	Time, Scientists, and Companies
1. X-ray diffraction data and proposed double-helix model for the 3-dimensional structure of DNA	RE Franklin and MH Wilkins; JD Watson and FH Crick, 1953
2. Site-specific recognition and cleavage of DNA by restriction endonucleases	W Arber, 1962; M Meselson and R Yuan, 1968; HO Smith, 1970; D Nathans, 1971
3. Determination of the genetic code	M Nirenberg, S Ochoa, and P Leder, 1966; HG Khorana, 1966
4. Identification of DNA ligase	M Gellert, 1967
5. Identification of RNA-directed DNA polymerase (reverse transcriptase)	HM Temin and S Mizutani, 1970; D Baltimore, 1970
6. DNA cloning techniques	HW Boyer, S Cohen, and P Berg, 1971–1972
7. Formal discussions on emerging DNA technologies	Gordon Conference on Nucleic Acids, June 1973
8. Self-imposed standards for rDNA research	Asilomar Conference, Feb 1975
9. Hybridoma created	C Milstein and G Kohler, 1975
10. Recombinant Advisory Committee (RAC) issues guidelines	Recombinant Advisory Committee, 1976
11. DNA sequence technologies	F Sanger, 1977; W Gilbert, 1977
12. US Supreme Court ruled that microorganisms are patentable	General Electric superbug, 1980
13. US approval of first diagnostic kit using MAb technology, anti-C3d BioClone	Ortho Diagnostics, 1981
14. US approval of first ethical pharmaceutical produced by using rDNA technologies, (Humulin (human insulin))	Genentech and Eli Lilly & Co, 1982
15. Expression of a foreign gene in plants: bacterial antibiotic resistance gene expressed in tobacco plants	Monsanto Co, Washington University, and Max Planck Institute, 1982
16. FDA approval of first monoclonal murine antibody drug, Orthoclone OKT3, for reversal of acute kidney transplant rejection	Ortho Biotech, 1986
17. The polymerase chain reaction (PCR) methodology enables targeted amplification of DNA sequences	KB Mullis; Cetus Corp., 1983; use of thermostable DNA polymerase, 1988
18. FDA approval of first recombinant vaccine for hepatitis B virus	Chiron Corp., 1986
19. US Patent and Trademarks Office issues first patent for genetically engineered mammal, transgenic mouse	P Leder and Harvard University, 1988
20. Formal launch of Human Genome Project	USA, 1990
21. First human patient received gene therapy for adenosine deaminase deficiency	WF Anderson, 1990
22. Dolly the sheep becomes the first cloned mammal	I Wilmut, 1997
23. RNA interference's gene silencing activity in nematode worm <i>C. elegans</i>	A Fire and C Mello, 1998
24. FDA approval of the first rationally designed and target specific cancer chemotherapy drug, Gleevec, to treat certain types of leukemia	Novartis, 2001
25. Simultaneous publication of human genome sequence by Human Genome Project and Celera Genomics	Human Genome Project and Celera Genomics, 2001
26. Approval of the first gene therapy drug worldwide, Adenovirus-based delivery of p53 tumor suppressor	Gendicine, in China, 2003
27. EMEA approval of the first biosimilar drug, Omnitrope (recombinant human growth hormone)	Sandoz, 2006
28. US district court in New York rejected the breast cancer gene patents	US district court in New York, 2010

one amino acid. The arrangement of codons in the DNA, following transcription into messenger RNA (mRNA), determines the sequence of amino acids that will form a particular protein. The detailed understanding of how these genes and their proteins govern basic cellular processes is the underpinning of molecular biology and biotechnology.

Because each of the major organs of the body (brain, liver, blood, tissue, etc.) has a specified set of tasks to perform, certain specific sets of genes in each organ (collection of specialized cells) must be activated and deactivated, that is, turned *on* and *off* as needed. Following the directions laid down by the genetic code of DNA and mediated by mRNA, each cell type continuously produces a unique and characteristic array of proteins. Each cell type maintains a complement of transcriptional

activating and repressing proteins whose actions balance to create the specific gene expression profile of a particular tissue. Moreover, epigenetic processes such as gene methylation and histone acetylation status also contribute to tissue-specific gene expression. Expressed proteins are then secreted into the extracellular milieu, while many are used within the cell itself. The number of possible biosynthetic permutations is very high if one considers that a typical protein can be made up of some 500 amino acids and, further, that every one of these sites may be occupied by any one of 20 different amino acids. It is likely that over the long periods of evolution of each organism, given the vast array of possible combinations of these amino acids, a multitude of unique proteins with all sorts of optimized functions have developed.

The concept that genetic information flows from DNA to RNA to proteins has become a fundamental milestone of modern biology. Thus, with the discovery of reverse transcriptase (from an RNA virus) by Temin and Baltimore, in 1970, which could convert its own genomic RNA into double-stranded RNA, a second milestone was reached. Molecular biology relies heavily on this enzyme to convert mRNA into DNA for gene cloning, library construction, and gene sequencing and detecting. Examples of cellular catalysts, or enzymes, include those that are involved in the digestion of food and others that produce the chemical building blocks of cell life, such as sugars and lipids, hormones for organism regulation, fuel for energy production, and important molecules such as DNA.

Proteins also make up the cell cytoskeleton providing an organized, three-dimensional structure. They permit directed transport and movement of molecules throughout the cell. They are embedded in the outer cell membrane and pump nutrients and ions across the membranes. They serve as receptor sites for hormones that finitely adjust the functions of the cell according to changing bodily needs. Another group of proteins regulates gene activities by binding to DNA and activating or repressing gene transcription. Still other proteins, and their smaller fragments (peptides), are secreted by cells as neurotransmitters or hormones like insulin. Some serve as carrier molecules, like hemoglobin, the body's oxygen carrier.

As is well recognized, these hormones and various related peptide molecules hold enormous power, and because they can act on numerous specific cell surface receptors, they can influence virtually all bodily functions from the nervous system to the immune system. It is obvious that their selectivity, potency, and often-desired evanescent effects on selective target cells make them enormously attractive candidates as a new generation of drugs in the *magic bullet* concept of Paul Ehrlich. Further, when administered parenterally, hormones have the potential to reach target receptors on the surface of cells, without the need to penetrate membranes. Not unlike the normal bodily processes, they can bind to cell surface receptors and activate the cell's particular function. An example of one such approach is seen with the anticancer drug interferon α , which can stimulate some immune cells to attempt to overcome cancerous cell growth.

The body's specific defense response to invading organisms is due to the immune system. Normally, phagocytes called to a site of inflammation induced by pathogens mount a generalized attack response. Indiscriminately, they engulf cellular debris as well as anything recognized as foreign. Occasionally, however, this is not enough, and illness ensues. At this point, several more focused counterattacks proceed by the three types of white blood cells known as macrophages, T lymphocytes, and B lymphocytes. The key features of the immune system are specificity (the ability to focus on specific pathogens) and memory (the ability to recognize and respond rapidly to previously encountered infections). About 1% of the blood cells are white blood cells. The ones that are central to the immune responses are the following:

B Cells—Lymphocytes that produce antibodies (antibody-mediated immune response).

Macrophages—Phagocytic cells that alert helper T cells to the presence of pathogens.

Helper T Cells—*Master switches* of the immune system that stimulate the rapid division of both killer T cells and B cells.

Suppressor T Cells—Lymphocytes with regulatory functions; i.e., they slow down or prevent immune responses.

Killer T Cells and Natural Killer (NK) Cells—Lymphocytes that directly destroy body cells that already have been infected by pathogens (or cancer cells).

Memory Cells—A group of the T cell and B cell population that was produced during the primary encounter with a pathogen but was not used in the battle. These circulate through the body ready to respond rapidly to later attacks by the same organisms.

As a further refinement in the understanding of the immune system, several key weapons are involved in the process. These include the antibodies, which are circulating freely or membrane-bound receptor molecules that bind specific foreign invaders and thereby tag them for destruction by the complement system or phagocytes. There are the perforin proteins, which are secreted by certain T cells and kill their cellular targets by punching holes in them. Finally, there are the lymphokines and interleukins, which are secretions by which white blood cells communicate with each other. Thus, the immune system has two fighting branches with specificity, and often both are employed against infections and antigens in general. The T cells dominate one part of the system, and when they are activated, it is referred to as a *cell-mediated* response. The B cells dominate the other branch, and events associated with their activation are referred to as *antibody-mediated* response.

Before the broad application of whole genome sequencing, expressed sequence tags (ESTs) of complementary DNA provided shortcuts to uncover a large number of new genes. Similarly, *light-houses* have been developed along the chromosomes to guide the way for sequencing DNA restriction maps. DNA research, using the polymerase chain reaction (PCR), has become a powerful tool in forensic and research applications. Based on the principle of PCR, real-time polymerase chain reaction (RT-PCR) or quantitative real time polymerase chain reaction (Q-PCR) has been developed and applied as another powerful technology to detect or quantify one or multiple gene targets precisely and simultaneously. Non-specific fluorescent dyes or sequence-specific DNA probes can be used as indicators. In the clinic, RT-PCR provides rapid and accurate diagnoses.⁷

New publications have described how modern metabolic engineering has brought intermediary metabolism back to life through techniques involving enhancing copies of a gene at a rate-controlling point, adding a gene to remove a poisonous product, or adding several genes to introduce a new pathway into an organism that stops short of the desired product. This metabolic engineering has had numerous practical results in addition to helping develop new theories. DNA technology has been applied to metabolic pathways so that branch point control problems can be solved. Even the insertion of similar enzymes from different species into the studied organism has introduced new flexibility and better metabolic characteristics into the older organism.

This issue also covered the recent developments of the vaccinia virus, so that it now can serve as a molecular vehicle for carrying foreign genes into other organisms. As a means for research, this recombinant vaccinia vector has served as a vehicle for producing live vaccines that would otherwise be difficult to produce. Monoclonal antibodies have also been used successfully in diagnosis and therapy. The monoclonal antibody OKT3 (Reversal of acute kidney transplant rejection, *Ortho Biotech* 1986), the first monoclonal antibody drug for humans, was approved by the FDA in 1986 for treatment of acute renal allograft rejection. Antibody power has been enhanced by attachment of a biological toxin such as ricin, a cytotoxin such as calicheamycin, or a radioisotope such as an alpha emitter. The latter can be used to damage tissue adjacent to that with which the antibody interacts. These are all good examples of combined basic research followed by rapid practical application.

BIOTECHNOLOGY DRUGS

In terms of drug therapy, the body's ability to remember and identify previous infections already has been exploited through the vaccination technique. This relatively simple procedure involves injecting the person with a weakened or killed pathogen that induces an effective immune response without causing the disease. The procedure gives a lasting immunity through the formation of the memory cells mentioned above. Continued advantage of these natural phenomena (normal immunization through infection) is being taken through the application

of molecular biology and biotechnology. Among the recent and significant developments are the vaccines using genetically engineered pathogen components as antigen to stimulate the body to produce antibodies against the pathogen. For example, Recombivax HB (*Merck*), a hepatitis B vaccine, uses a non-infectious hepatitis B surface antigen (HBsAg)-derived protein fragment produced in yeast cells as the antigen. Figure 34.1 shows the production of a genetically engineered vaccine.

As of 2008, 633 biotechnology medicines were in development for more than 100 diseases, mainly for cancer, infectious diseases, autoimmune diseases, and HIV/AIDS related conditions.⁸ In 2010, there were just over 200 biopharmaceutical products on the market.⁹ The biopharmaceutical industry represents a fast growing portion of pharmaceutical market. In 2009, the total biological sale reached \$99 billion, including \$38 billion antibody-based drugs and \$61 billion other recombinant therapeutic proteins. The accelerated rise of biological generic drugs (biosimilars) is going to account for a large proportion of biological drugs and decrease the cost of healthcare in the current decade.

In the mid-1990s, the fastest growing category of biotechnology products had been gene therapy; however, the death of an 18-year-old Phase I gene therapy trial participant, Jesse Gelsinger, in 1998 led to a sober reassessment of strategies, precautions, and controls in such trials.¹⁰ The currently fastest growing segments are monoclonal antibodies and vaccines. Vaccine targets are infectious diseases, cancer, AIDS, and autoimmune diseases (rheumatoid arthritis and multiple sclerosis). Monoclonal antibodies have also been instrumental in treating various forms of cancer and autoimmune disorders. Product candidates that are nearly approved also are increasing. By the end of 2010, 41 antibody-based therapeutics had been approved in the USA and Europe.⁹ Novel monoclonal antibodies have been entering clinical at a rate of approximately 40 per year since 2007.

All of these agents have been made possible through the biotechnology techniques that allow the isolation, identification, and production of normally minimal amounts of proteinaceous *signal agents* found in the extracellular fluids of the body. Once

the composition and sequence of amino acids is determined for a protein, that protein can be reproduced in the laboratory. Even better, the gene sequence of the protein now can be analyzed readily, the protein sequenced, synthesized, and cloned, allowing another organism to use that part of the code that determines the protein. This has been made possible through the discovery and use by HO Smith of restriction enzymes, which make specific reproducible cuts along DNA strands. Frederick Sanger and co-workers devised procedures for quickly determining the nucleotide sequence of DNA fragments. This allowed the identification of the DNA sequence of complete genes. Now, fully automated and high-throughput sequencing instruments in conjunction with novel sequencing, detection, and analysis methodologies make it possible to sequence the human genome in a short time at low cost.¹¹ It will open a new avenue for personalized medicine. Also, the discovery of reverse transcriptase by Temin and co-workers became important in biotechnology because it allowed the mass production of genes from mRNA, which led to increased production of a desired protein. Through these procedures it became possible to determine the amino acid sequence of entire proteins via inference of the genetic code.

The 1972–1973 landmark experiments of Stanley Cohen, Herbert Boyer, and Paul Berg applied this technology to produce recombinant plasmid DNA molecules for propagation in *E. coli* and are recognized as the first creation of a genetically engineered organism capable of producing proteins from another species. This hybrid plasmid now could be grown in the common and rapidly producing bacterium, *E. coli*. A plasmid is a circular DNA molecule that carries a few genes that the bacterium perpetuates and is replicated in addition to its own normal chromosomes. More than any other technique, this really heralded the birth of hybrid DNA technology (hybrid DNA is produced by joining pieces of DNA from different sources, and is also designated rDNA technology). This permitted, for the first time, rapid isolation of unique proteins and their mass production by rapidly growing microorganisms. In addition, new organisms having specifically inserted and desired characteristics could be engineered for medical, agricultural, and ecological uses.

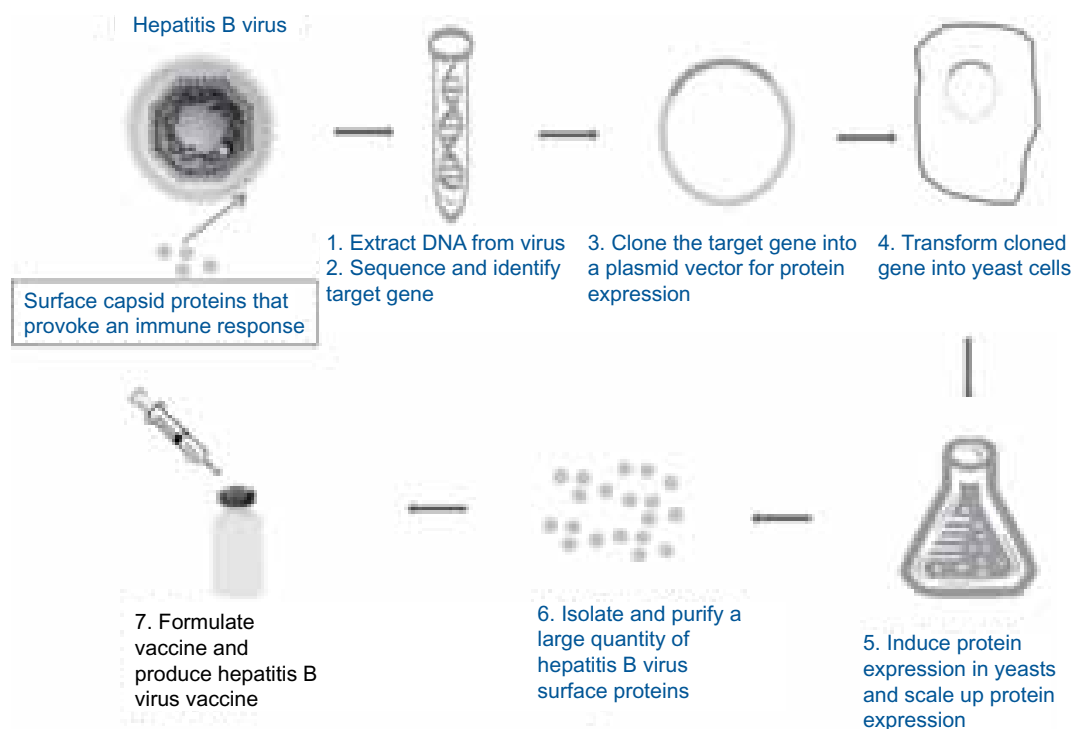


Figure 34-1. Making a genetically engineered vaccine.

Another important aspect of recombinant DNA technology is the use of antibodies in biotechnology, therapy, and diagnosis. Antibodies are produced by plasma cells (B cells) and are made up of four protein chains interconnected by disulfide bonds. The surface of the antibody possesses a highly specific indentation, or *lock* that can recognize the specific foreign particle (*key*) with which it complexes or binds. It long has been known that different antibodies are produced in each individual for their particular immunological experience with antigens. Hence, perhaps millions of different antibodies may be found in any given individual. For a long time it was not known how the B cells were capable of producing this diversity of antibodies that possessed the ability to recognize almost every possible foreign invader. It also was not known whether each B cell secreted a single or many different antibodies.

Fortunately, through the early *clone selection* theory of MacFarlane Burnet in 1957 came the idea that one cell produces only one type of antibody. And in 1975 Georges Kohler and César Milstein devised a method of growing very large numbers of antibody-producing cells from a single B cell. They did this by the ingenious technique of fusing the B cell to a myeloma cancer cell. The resulting *hybridoma* retained two main features from its two parent cells. It could grow indefinitely like the cancer cell yet also produce and secrete antibodies like the B cell. This was the main discovery leading to hybridoma technology and earned Kohler and Milstein (together with Niels Jerne) the 1984 Nobel Prize in Physiology or Medicine.

The antibodies produced by these hybridomas are called monoclonal antibodies (MAbs) because they are derived from a single hybrid cell. Using the ability to identify directly the genes that code for antibodies, it was found that the antibodies are put together from a large number of different gene fragments. When combined in different ways, they can produce a large number of different antibodies. Those portions of the antibody that contain the antigen-binding site are coded by a combination of hundreds of different gene fragments that get reshuffled and permanently fixed in the B cells. Hence, it has become possible to produce MAbs as key reagents in biotechnology procedures as well as exquisite diagnostic tools and specific drugs of great selectivity. Figure 34.2 shows the basic procedures involved in hybridoma technology.

Antibodies are routinely used in biomedical research for affinity protein purification, diagnosis, drug targeting and delivery, and imaging analysis.¹² In addition, hundreds of antibody drugs are under development for the treatment of various diseases.¹³ Muromonab (OKT3) is the first FDA approved monoclonal antibody drug. It is a murine monoclonal antibody against T-lymphocyte CD3 for the reversal of acute kidney transplant rejection. However, using murine antibody in human is associated with the development of human anti-mouse antibody, weak immune response recruitment, and short serum half-life. Genetic and protein engineering methods have been applied to convert murine antibodies into human-like antibodies. Chimeric antibodies are the fusions of murine variable domains (Fv) with human constant domains (Fc). The variable domains of chimeric antibodies can be further modified to adopt human antibody-preferred residues and sequences to create humanized antibodies to minimize immune response and maintain the maximum antigen-binding ability. Furthermore, genetically altered mouse has been used to produce human B cell to create fully human antibodies. In addition, high-throughput peptide display libraries, such as phage display, yeast display, and ribosome display, have been applied to screen Fv domains and construct human monoclonal antibodies.

The phage display approach, described in detail by Rader and Barbas,¹⁴ employs the selectable expression of recombinant antibody molecules on the surface of bacteriophage particles. Generally, random peptide sequences encoding human antibody variable heavy- and light-chain fragments are integrated at a designated position of a surface protein possessed by cells, viruses, or phages through a genetic linkage to the encoding

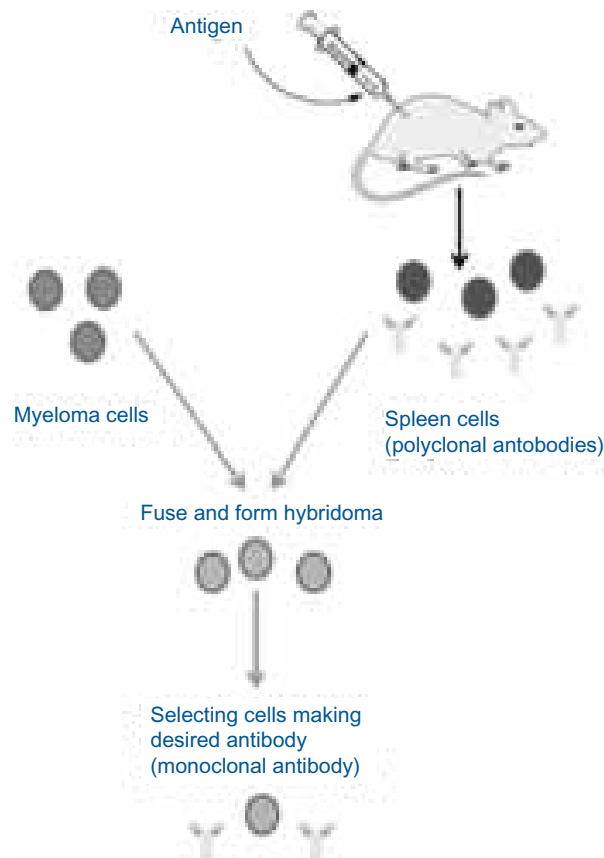


Figure 34-2. Hybridoma production.

nucleic acid. Once antigen-binding cells, viruses, or phages (*in vitro* as well as *in vivo*) have been isolated through a selection process (biopanning), the sequence of this peptide can be derived. The cells to display peptides can be mammalian, yeast, and bacterial cells. The phages can be M13, T7, and P4 phages. Recently, other display technologies, such as ribosome display, mRNA display, and DNA display (all need to establish physical linkage between the displayed peptide and the encoding RNA/DNA through a chemical or protein modification), have been developed to select high binding affinity antibody Fv domain. The screened domain can be fused with human antibody Fc domain through genetic engineering to create human antibody. Then the antibody gene will be integrated into mammalian cells such as Chinese hamster ovary (CHO) cells to produce antibodies.¹²

Unlike vaccination that uses the selected antigens to stimulate the human immune system to generate antibodies for the purpose of prevention or treatment, antibody therapy uses *in vitro* isolated and purified antibodies as drugs. Inside the human body, antibody can recognize and neutralize its target, or acts as a ligand to either activate or inhibit a cellular receptor. In addition, antibody can also accomplish its therapeutic effects through complement-dependent cytotoxicity (CDC), antibody-dependent cellular phagocytosis (ADCP), and antibody-dependent cellular toxicity (ADCC). Antibody is a glycoprotein. Glycosylation of the Fc portion of IgG is essential for antibody activity as well as immune response. The risk of immune reactions to antibodies is the main safety concern for antibody therapy. The high cost of treatment is obviously another limitation for this kind of therapeutic. For example, it costs \$50,000–\$100,000 for one year Herceptin (Trastuzumab for HER2 positive breast cancer) treatment.^{13–15}

It already has become possible to tag monoclonal antibodies with radioisotopes that make possible the detection of very small levels of proteins and peptides in body fluids and tissues.

The limits of detection are often as low as one billionth of a milligram (picogram) in the procedure widely used and known as radioimmunoassay (RIA). This is sufficient to detect low levels of hormones and other protein substances in body fluids. Antibodies can also be conjugated with toxins, cytokines, or enzymes to deliver drugs or prodrugs to the targeted cells. Although these approaches have been proposed and tested for a couple of decades, Mylotarg (anti-CD33 antibody gemtuzumab conjugated with ozogamicin, for acute myeloid leukemia, Wyeth, 2000), the only antibody-drug conjugate on the market, was withdrawn in 2010 due to severe side effects.¹⁶

Another exciting area of biotechnology is that of gene diagnostics and therapy. It is believed that there are as many as 4000 locations in the human genome that are related to different genetic diseases. Of this number some 1200 have been mapped and characterized to various degrees of detail. Some of the abnormalities found on the genes are called point mutations, and they involve cases where a single nucleic acid base in a gene is substituted by a different one. This irregularity results in the exchange of a single amino acid in the encoded protein. Too many changes may result in genetic disorders. For example, in the genes encoding the hemoglobin protein sequences, at least 40 point mutations have been located. Sickle-cell anemia is related to one of these. It is hoped that detailed knowledge of this type at the molecular level will allow the development of agents that can prevent the typical alteration in the shape of blood cells in sickle-cell anemia. Molecular probing or screening at this level also will reveal such disorders prenatally, or early in life, so that appropriate remedial action or preventative measures can be instituted (viz.,

gene therapy). Figure 34.3 shows how genetic defects may be detected.

Pharmacogenomics is another research area that has sprung from pharmacokinetics to examine the key genes or mutations among a subset of the population that confer relative resistance or hypersensitivity to certain drugs. In the pharmacogenomic context, these mutations are more commonly referred to as polymorphisms and are often observed in genes encoding drug metabolizing enzymes, transporters, receptors, or other drug targets. These variations in DNA sequence are often single-nucleotide polymorphisms (SNPs) and can occur in either the coding region of a gene, causing an amino acid change, or in the regulatory region of a gene, thereby causing an alteration in the absolute amount of protein produced. For example, a patient with a polymorphism in one or both copies of a gene encoding a drug-inactivating enzyme would be at risk for exaggerated pharmacological action and potential toxicity for that drug. These pharmacogenetic differences are likely at the heart of most of the heterogeneity of drug responses observed among large populations and have significant impact on the ordinary practice of medicine and in clinical trials. Specific examples of known polymorphisms and the altered effects they confer are delineated in an excellent review article¹⁷ from a premier PharmD pharmacogenomics expert.

Advances in DNA technology also have made this procedure more rapid, particularly when looking for any one of many individual mutations within a single gene that can give rise to disease. This technology, called high-density DNA affinity arrays,¹⁸ or microarrays, is becoming more widespread and financially accessible to even small research institutions. Individual DNA

Sometimes 'misspellings' occurs in the genetic code and errors creep into the chemical sequence. Such defects can be inherited or caused by chemical change, and give rise to disease. Prenatal screening can find faulty genes and identify individuals likely to contract specific ailments

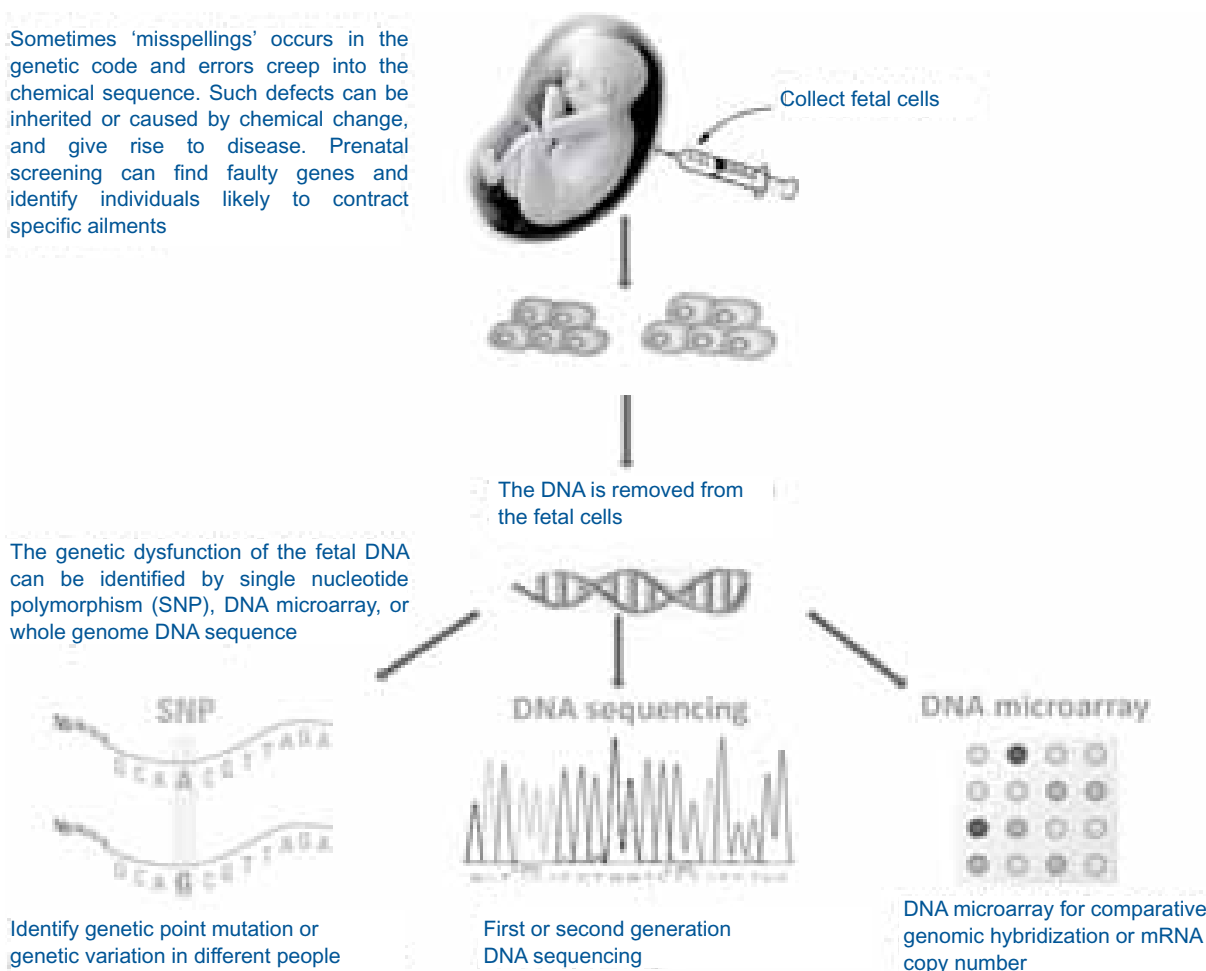


Figure 34-3. Searching for genetic defects.

targets that each contain one of these mutations are adsorbed onto a glass substrate in an array pattern and then hybridized with a fluorescently labeled probe generated from the gene expressed in the individual being tested. Another adaptation of this technology is to investigate the relative expression level of genes implicated in various disease processes. Some diseases are not due to gene mutations but rather to abnormal over- or underproduction of certain regulatory proteins. DNA arrays permit the screening of literally thousands of genes in a single experiment by comparing genes expressed in normal versus diseased individuals. In fact, the expressed mRNAs from the entire human genome can now be screened on two commercially available "chips" and the race is on for the first manufacturer to put the entire expressed human genome on a single chip. Availability of such technology is expected to allow early diagnosis of many diseases that are due to multiple genetic abnormalities, such as cancer. In one specific case of diffuse B-cell lymphoma patients, microarrays have been used to sub-classify the disease to understand which patients will respond best to chemotherapy.¹⁹ In this manner, patients at greatest risk can be given the most aggressive therapy, while those at lower risk can be spared unnecessary toxicity. Microarrays of gene expression can also be applied in drug discovery; a recent application of this technology has been used to predict the general therapeutic classes into which psychotropic medications fall.²⁰

One drawback to microarrays in diagnosis and drug treatment is that gene expression is not always predictive of the amount of protein produced. Moreover, it completely ignores post-translational modification of proteins that can vastly influence biological responses. With advances in mass spectrometry instrumentation, the field of proteomics has emerged as the next means to secure a snapshot of cellular activities and correlate the pattern with specific drug responses or disease processes. Researchers at the US Food and Drug Administration and the National Institutes of Health have recently shown the power of proteomics in early detection of ovarian cancer using simple blood samples.²¹ It was demonstrated that cancerous alterations in organ function cause an altered profile of proteins in serum, allowing this disease to be caught long before patients are symptomatic and with a high degree of accuracy, thereby minimizing the false-positive rate associated with many clinical diagnostic tests. The pharmacy practitioner is wisely advised to keep abreast of developments in proteomic technologies as these advances are quite likely to influence other therapeutic areas with equal magnitude.

As the ability to determine genetic defects that cause a variety of disorders emerges, healthcare will improve. However, society must develop policies governing the use and misuse of this information. Baum has reviewed this problem and provides information on some of genetic screening's serious implications. For instance, in the case of DNA fingerprinting, some observers have questioned the reliability of such an analysis carried out on such a large scale. However, it has yet to be shown to fail definitively when carried out properly. Results may affect civil liberties, insurability, guilt or innocence, etc. The 1996 Health Insurance Portability and Accountability Act (HIPAA) prohibits group health insurance providers from using genetic information for determining eligibility or setting premiums but provides no protection for the millions of Americans with individual insurance policies.²²

As the technique known as restriction fragment length polymorphism (RFLP) analysis develops,² it provides markers throughout the genome that vary among individuals. The identification of particular RFLPs that are tightly associated with genes responsible for certain diseases has become possible. Thus, RFLPs provide markers that identify the chromosome that carries the defective gene. But the presence or absence of the marker does not necessarily indicate disease. RFLP analysis must be performed on both parents or two or more grandparents to determine the status of a disorder. Furthermore it is an expensive and complex analysis.

In the mid-1990s, public attention was raised to the applications of genetic polymorphisms for forensic purposes in criminal

investigations. However, RFLP analysis has considerable drawbacks for this purpose. The need for relatively large amounts of non-degraded DNA and the time necessary for multiple allelic comparisons led to a search for other methods of individual genotyping. The most popular recent technology analyzes inter-individual differences in the variable number of tandem repeats (VTNRs), or *DNA fingerprinting*, as it is known to the lay public.¹⁴ A similar name for the same method is amplified fragment length polymorphisms (AMPFLPs). We each possess in our genome a variable number of repeated DNA sequences that, when measured in combination, provide a unique *fingerprint* of our DNA. These repeated sequences are referred to as minisatellites or microsatellites, depending on their size, and are often made up of repeated sequences of 2 to 70 base-pair-length monomeric sequences. Since the location of many of these polymorphisms is known, the polymerase chain reaction (PCR) can be used to amplify these repeated sequences from even the extremely small amounts of blood, tissue, or semen found at crime scenes. Depending on the number of satellite sequences analyzed, the individual source of the DNA can be identified with near absolute certainty. So useful is this technique that it was employed in 1994 to positively identify the remains of the Romanovs, using mitochondrial DNA isolated from the bones of the Russian royal family who were murdered by the Bolsheviks in 1918.²³

The major diseases of mankind owe much of their origin to heredity, so that it will be exciting in the decades to come to see how molecular biotechnological techniques will allow for early detection, prevention, or even possible cures for many of the maladies of old age such as cardiovascular disease, Alzheimer's disease, diabetes, and cancer. These approaches also are allowing us to understand previous unrecognized causes of diseases. For many years, it has been known that many human leukemias could be diagnosed microscopically on the basis of specific translocations, or rearrangements, of the chromosomes. Today, the abnormal proteins made as a result of these translocations can be identified. Look²⁴ provides a striking review on the function of these aberrant proteins in leukemogenesis. While many of these proteins lead to uncontrolled growth of leukemia cells, some have been shown to cause leukemia by inhibiting the normal cell death that usually occurs in white blood cell populations. This had led to our reclassification of cancer as a disease not necessarily of abnormal cell division but instead, in some cases, of loss of appropriate cell death. Other articles in the same 1997 issue of *Science* deal with more frontiers in cancer research, showing that molecular biology also has led to our understanding of how cancer can be prevented.

It is hoped that continued approaches based on reverse genetics will be fruitful. So far, these methods have allowed researchers to produce a compilation of the exact locations and even the molecular arrangements of several defective genes that are felt to be responsible for Huntington's disease, retinoblastoma, Duchenne's muscular dystrophy, and polycystic kidney disease, among others. From these studies will come specific and reliable tests for the abnormal genes that will allow genetic counselors to make appropriate recommendations for action. Further in the future will come the new therapeutic models and molecules that eventually will translate into clinically effective drugs.

So far, at least two basic approaches in *genetic medicine* can be envisioned. One involves possibly compensating the defective protein by producing it biotechnologically outside the body (e.g., Humulin (diabetes mellitus, Lilly, 1982) is the first FDA approved recombinant human insulin produced in *E. coli*).²⁵ The same can be done with replacing defective or missing enzymes or altering one that should not be produced. The other basic approach is much more complicated and involves compensating or replacing the defective gene. In the case of gene modification that involves germ-line cells (sperm and eggs), it is called germ line gene therapy. Once a gene is introduced stably into the germ line, every cell of the individual will bear this gene. Presently, this type of therapy poses enormous unresolved ethical and scientific questions

and is restricted. If the therapeutic genes are transferred into the somatic cells (specific body organ cells), it is called somatic gene therapy. Dependent on the type of delivery vectors, such as integrated or non-integrated vector, the gene expression can be transient or permanent. However, the transferred genes cannot be inherited and stay with the treated individual. One of the ways to transfer genes into the targeted cells would be to use genetically altered and *harmless* viruses to insert a corrected gene into the human genome. It can be referred to as viral gene therapy. On the other hand, lipid, polymer, protein, and gold particles have been applied for gene delivery. These approaches are described as non-viral gene therapy. Unfortunately, in most cases, it is not yet understood how the inserted gene will be expressed in these tissues, only where and when they are needed. Some ways to circumvent these difficulties have been found, but they are limited. For example, one way to produce tissue specificity is to take out the tissues to be altered, insert the corrected gene, and place the corrected cells back into the body (*ex vivo* gene therapy). Alternatively, tissue-specific genetic control elements, such as transcription factors, have been identified and can be used to restrict expression of the recombinant gene in the targeted cells.

So far, bone marrow cells can be handled in this manner, and they have been the first somatic cell type submitted to clinical testing of this gene-transfer technique. In one example, the goal is the treatment of an inborn disease affecting bone marrow function and involving an enzyme (adenosine deaminase, ADA) deficiency that produces severe weakening of the immune system. In 1990, 4-year-old Ashanthi DeSilva became the first patient to undergo successful gene therapy using retrovirus to insert a healthy ADA gene into white blood cells. However, due to the random integration of retroviral vectors and alteration of oncogenes, four of ten patients receiving the similar therapy in clinical trials developed leukemia in 2007. In 2003, China's (State Food and Drug Administration (SFDA) approved Gendicine, the first gene therapy drug worldwide. It is an adenovirus-based DNA vector designed to deliver cancer suppressor gene p53 for the treatment of head and neck cancer. So far, no gene therapy has been approved by FDA and EMEA. Overall, gene delivery is still the major obstacle to gene therapy.

One of the goals in biochemistry and pharmacology is to understand the molecular features of cell receptors. These are the *locks* into which the *keys* (drugs) fit that alter or control the function of the cell. While now there exists the capability to determine the basic genetic code and thus learn the sequence of the amino acids that make up the protein, its spatial configuration still is not known. This is called the tertiary structure of the protein, the functional form of the protein after the proper folding of a simple straight-chain of amino acids.

The novel combination of x-ray crystallography, molecular mechanics, calculations, and supercomputers are brought to bear to reveal the folded or three-dimensional arrangement. From this 3-D picture of the *lock*, researchers can design specifically shaped drugs that fit the active sites of the folded protein. This computational chemistry methodology is a leap forward and truly a rational approach in drug design. In like manner, the possibility of cloning *receptor sites* of specific design and function provides the pharmacologist with excellent *in vitro* test systems for pharmacological screening and understanding of the mechanisms of actions of drugs.²⁶ For example, the solution of the x-ray crystal structure of the HIV protease has led to the rapid development of four FDA-approved drugs that have revolutionized our treatment of AIDS. Saquinavir (*Hoffmann-La Roche*, 1995) is the first FDA approved HIV protease inhibitor. Another example of successful rational drug design is Gleevec (imatinib, *Novartis*, 2001), which inhibits the tyrosine kinase domain of the gene product (BCR-ABL) of a fusion gene of abl (Abelson kinase) with bcr (breakpoint cluster region) associated with chronic myelogenous leukemia (CML).²⁷

In the past few years, several publications have dealt specifically with the pharmacist's role in implementing pharmaceutical care with regard to biotechnology agents. Tami and Evens²⁸

recently presented an overview on the evaluation of biotechnology products. The authors first discuss the manufacturing, pharmacokinetics, and stability issues that we present in greater detail below. They specifically made the point that biotechnology product costs often can account for 10% of a pharmacy's total budget. In addition, they discuss pharmacoeconomic studies, availability of alternative agents, concomitant drug costs, special pharmacy procedures, reimbursement, and manufacturer's support. Pharmacists are a particularly important link in assisting patients with reimbursement issues for these expensive drugs.

UNIQUE PHARMACEUTICAL CHALLENGES OF BIOTECHNOLOGY-DERIVED THERAPEUTICS

The transition of biotechnology from theory to pharmaceutical practicality has posed a whole new series of challenges to those involved in drug development. With classical small molecules possessing formula weights of less than 1000, a series of chemical compounds normally are screened for a particular pharmacological activity and assessed for specificity. The results of these findings then guide fine-tuning of the chemical entity. As advances in combinatorial chemistry are now generating thousands of compounds in a given class, the rate-limiting step in drug development has shifted to high-throughput screening technologies. Natural products are also regaining prominence in drug discovery because they often possess greater molecular diversity than can be obtained with combinatorial compound libraries. It is now rare that pharmaceutical formulation and drug delivery problems limit the success of small molecules.

In contrast, macromolecular agents (e.g., recombinant proteins and vaccines, antisense DNA, gene therapy constructs) already have the advantage, in theory, of possessing inherent selectivity for a particular biological process.²⁵ By and large, the limitation to the utility of these agents rests with problems related to drug delivery and stability. In fact, it is no surprise that most biotechnology drugs currently approved in the US act at extracellular sites and/or in compartments that are easily accessible, such as the blood-forming elements. In addition, each type of biotechnology agent also is subjected to unique considerations based on our emerging biological understanding of each system. The discussion that follows attempts to address the obstacles to successful therapeutic use of biotechnology products.

Recombinant proteins almost exclusively constitute the currently approved list of biotechnology-derived agents in the US. These proteins usually have resulted from a search for endogenous agents acting by a newly identified physiological mechanism (such as the stimulation of red blood cell production by erythropoietin manufactured in the kidney). Since the therapeutic administration of a recombinant molecule that mimics an endogenous protein carries with it a naturally inherent specificity, it is no surprise that development times for these agents have been considerably shorter than for most conventional small molecules. The probability of regulatory success with recombinant proteins also has been more favorable.²⁹ While a new, small, organic molecule may have a 10% chance of achieving NDA status, this percentage is near 40% for recombinant proteins. When compared with conventional small molecules, recombinant products reaching the clinical trials in Phase I (25% versus 71%) or Phase III (66% versus 93%) are much more likely to become successful therapeutic agents. Nonetheless, Cho and Julian³⁰ state, "The main challenge encountered in development is not so much identifying a bioactive molecule but, rather, how to maintain a therapeutically meaningful concentration of the macromolecule in the vicinity of its target for the desired period of time."

Recombinant protein drugs are produced in various host cells from carefully designed expression systems. For monoclonal antibodies, production is enabled by the genetically engineered mammalian expression systems using Chinese hamster ovary

(CHO) or murine lymphoid cell lines (e.g., NS0, Sp2/0-Ag14).¹² For the bulk of the other recombinant protein drugs, the exploitation of any one of these protein *factories* begins in a similar fashion. As addressed earlier in the section on gene splicing, the complementary DNA encoding a particular protein product is sub-cloned into a DNA vector. This recombinant vector contains gene regulatory sequences that enable highly efficient transcription and translation of the recombinant gene once the construct is introduced into the appropriate host.

The choice of host system (bacteria, yeast, or mammalian cells) depends highly on the biological requirements of the protein. One major consideration is whether the protein product requires glycosylation, the specific addition of certain sugars, for biological activity.²⁰ Bacteria like *E. coli* are unable to conjugate such carbohydrates onto recombinant proteins, but yeast possesses a limited ability for glycosylation. However, mammalian cell systems such as CHO cells have the full complement of glycosyltransferase enzymes. For molecules such as the interferons and filgrastim (G-CSF), glycosylation is not necessary for biological activity; therefore these proteins can be produced in less expensive *E. coli* systems. However, erythropoietin requires mammalian glycosylation and must be produced in the more costly CHO cell system. Hamilton and colleagues³¹ have recently reported the heroic engineering of the yeast *Pichia pastoris* to encode the entire complement of human N-glycosylation enzymes while deleting the yeast's own glycosylation pathways. Such an organism might represent a more cost-effective host for future manufacture of human therapeutic proteins requiring glycosylation.

These modern protein expression systems have several advantages over trying to isolate the corresponding protein from the organs or tissues of other mammals. First, immune reactivity to a nonhuman protein (human insulin versus porcine insulin) is largely obviated (with exceptions described later). Also, protein drugs can be produced that could never be made by conventional methods (interferons, Granulocyte-Colony Stimulating Factor (G-CSF)) or in quantities previously only available in limited amounts (insulin, growth hormone). Finally, the protein is inherently free of potentially pathogenic human viruses (factor VIII or hepatitis vaccines), although treatments to destroy any zoonotic pathogens are often employed in subsequent processing.

As the biotechnology drug sector has matured, manufacturers have begun to follow conventional small-molecule pharmaceuticals in the production of second-generation agents with improved bioavailability, retention, and/or safety. While the cynic might argue that these innovations are merely spurred by pending patent expiration of first-in-class agents, it seems that most second-generation biotech drugs represent true advances. For example, the treatment of anemia due to kidney disease or cancer chemotherapy has been improved by increasing the bioavailability of erythropoietin (EPO) (e.g., Epogen, anemia, *Amgen*, 1989) in the second-generation drug, darbepoietin (e.g., Aranesp, long-acting rEPO, *Amgen*, 2001). Recognizing that N-linked oligosaccharides are essential to the activity of EPO, scientists altered two amino acid residues to accept two additional sialic acid chains. The resulting darbepoietin molecule possesses a three-fold improvement in serum half-life (26.3 h relative to 8.5 h for erythropoietin). This alteration reduces the number of injections necessary for patients to maintain therapeutic hemoglobin levels. While regimens vary by patient and indication, the general rule has been that if a patient had been stabilized previously with three EPO injections per week, only one weekly injection of darbepoietin is now required. From the standpoint of patient quality of life and burden on pharmacy and nursing staff, less frequent dosing has a number of significant advantages.

While scale-up of recombinant protein expression and purification is becoming more routine, these drugs present other challenges not seen with small molecule agents. Issues of proper protein folding, formulation, and stability are proving

as labor-intensive as the initial cloning of the gene itself. Beta-seron (human recombinant interferon β 1b, multiple sclerosis, Bayer, 1995) required modification of one amino acid to enhance the yield of properly disulfide-linked protein after refolding processing.²⁰ This processing modification reflects another advantage of recombinant protein expression that can be employed quite easily, so long as the pharmacological activity of the protein is not compromised. In fact, optimizing the cDNA sequence (and the resulting encoded amino acids) of a recombinant product has resulted recently in the approval of consensus interferon, a single molecule that incorporates the combined activities of multiple interferons.

Manipulation of the expressed gene also can involve the deletion of regions dispensable for biological activity in order to optimize therapeutic utility. Human tissue plasminogen activator (tPA) has been used since 1987 for thrombolytic therapy following myocardial infarction. However, tPA is poorly soluble and must be administered in relatively high concentrations because it is cleared rather rapidly from plasma. Structure-function analysis of individual tPA protein domains allowed the construction of a smaller molecule (reteplase or recombinant plasminogen activator, rPA) that possesses superior solubility and also can be manufactured in *E. coli* (e.g., Rapilysin, acute myocardial infarction, *Roche*, 1996).

The effects of changes in protein formulation or amino acid substitutions can now be assessed rapidly as a result of advances in protein analytical methodology. Protein secondary structure can be monitored quickly and accurately by such techniques as circular dichroism (CD) spectroscopy and Fourier-transformed infrared spectroscopy (FTIR). This technology has been quite useful in that the structural fidelity of the protein drug can be ensured prior to the initiation of more costly pharmacological evaluation. FTIR also has the advantage of being able to detect protein structure in the lyophilized state, greatly facilitating the optimization of formulations capable of maximal stability.

Pharmacokinetic evaluation of recombinant proteins is also an emerging field of significant relevance to pharmacy.³² Since most protein drugs cannot be given orally, the impact of other routes of administration must be assessed. It also must be appreciated that for most biotechnologically derived agents, there is a preexisting and nonconstant concentration of the corresponding endogenous molecule present in plasma. Bioanalytical techniques for monitoring concentrations of the agent require optimization for specificity. Unfortunately the specificity of any one method often depends on the particular matrix in which analysis is performed (blood, urine, or the initial formulation).

In addition, the prediction that recombinant human molecules would not be immunogenic has not proven to be the case.³² Antibodies to several recombinant drugs now are known, but not all neutralize the pharmacological activity of the agent and, in some cases, they can decrease clearance of the agent. These factors obviously complicate the interpretation of pharmacokinetic data. The immunogenicity of a particular protein also may depend on the route of administration. Protein aggregation, known to occur after subcutaneous or intramuscular injection, leads to a greater antigenic response than soluble protein. Finally, the influence of the lymphatic system on protein pharmacokinetics should not be underestimated. After subcutaneous administration, absorption via the lymphatics becomes quantitatively more important than that of blood capillaries as the molecular weight of the drug increases. Because several recombinant protein drugs act primarily through the lymphatics (interferons and interleukins), blood concentrations may be irrelevant to pharmacological activity.

Obviously, there remain immunologic concerns in any instance when a recombinant analog of an endogenous protein is fashioned into a drug product. Known as neoantigenicity, altered protein structure due to amino acid changes, chimeric products (i.e., humanized mouse monoclonal antibodies), or fusion with other constituents could trigger antibody-mediated inactivation of the therapeutic molecule as well as its endogenous

counterpart. In some instances, the very same modification in one context can result in the opposite effect in another. For example, *E. coli* asparaginase had been used to treat certain leukemias based on the observation of their need for the amino acid asparagine. However, some patients experienced immunological reactions to the bacterial protein, and a new product was produced, pegaspargase, where the enzyme was modified by covalent attachment of polyethylene glycol (PEG) molecules. This approach, now known as PEGylation, is also known to improve the efficacy of a number of proteins by reducing their renal clearance.³³ PEGylated therapeutic proteins available in market are: Oncaspar (PEG-asparaginase), Adagen (PEG-adenosine deaminase), Pegasys (PEG-interferon $\alpha 2a$), PEG-Intron (PEG-interferon $\alpha 2b$), Neulasta (PEG-Granulocyte-Colony Stimulating Factor), and Somavert (PEG-growth hormone receptor antagonist). Except for using N-terminal specific PEGylation for Neulasta, random PEGylation on protein surface residues, such as lysine, histidine, and tyrosine, etc., has been applied to create these PEGylated therapeutics.³⁴ It will inevitably yield heterogeneous products.

The same PEGylation strategy used to reduce antigenicity in one context recently caused an immunogenic response with another product, megakaryocyte derived growth factor or thrombopoietin. This product had been in development to treat bleeding disorders due to platelet loss in some disease and with chemotherapy, and a PEGylated form being tested in normal volunteers caused inactivating antibodies to be made to their own endogenous thrombopoietin.³³ This example should serve as a reminder that protein modifications do not always have the intended consequences.

In addition to immunogenicity, another obstacle for protein drug development is delivery. Due to the limitations of big molecular weight and bioavailability, subcutaneous and intravenous injection is still the main route of protein drug delivery. Inhalation formulation is indeed another feasible way for macromolecule drug delivery. The first inhaled insulin, Exubera (by Pfizer), was approved by the FDA in September 2006. It opened a new avenue for non-invasive protein drug delivery and spiked a heatwave for protein inhalation formulation. But the surge faded soon after the withdrawal of Exubera in October 2007.³⁵ Among seven companies developing inhaled insulin, only MannKind still continued its lonely journey. The withdrawal was claimed not to be related with the safety and efficacy issues. The failure of Exubera is attributed to the marketing and result of the products have been analyzed by all news. Exubera inhalation only partially replaced the daily dose of insulin injection, and patients preferred to stay with conventional needle injection. Moreover, the long term use of high-concentration insulin through the lung does trigger concerns of lung cancer.

There are about 200 biological drugs, including around 40 antibodies, available in the market. Patents protecting the IP of these drugs have expired or are close to expiration. The copycat versions of these complicated biological drugs are called biosimilars.³⁶ Biosimilars are also referred as "similar biological medicinal products" by the European Medicines Agency (EMA) and "follow-on biologics" by the FDA. The globe-wide market for biosimilars (biological generic drugs) has emerged and is growing fast. There is a fundamental difference between biosimilar and chemical generics. Due to the unique structure-function relationship of proteins and complicated post-translational modifications, protein expression system, manufacturing processes, formulation, storage, and product handling will all exert some kinds of effects on the therapeutic efficacy of protein therapeutics. The inherent differences between a biosimilar drug and its reference biological product are reflected in the preclinical development strategies and clinical efficacy. It also determines the case-by-case evaluation styles for all biosimilars. No standard guideline has been formed for biosimilar comparisons in pre-clinical and clinical studies in the US. The first approved biosimilar product Omnitrope (human growth hormone), in 2005 by Australia and in 2006 by Europe, represents

the start of a new sector of biopharmaceutical industry. So far, 14 biosimilar products based on 4 reference products (human growth hormone, granulocyte colony-stimulating factor, erythropoietin, and follitropin beta) have been approved by EMA. The driving force for biosimilars is to significantly trim the high cost of biological drugs. It is estimated that biosimilars would cut spending on prescription drugs in the United States by \$25 billion between 2009 and 2018.

The future of recombinant biological agents continues to evolve. To date, most approved agents have been recombinantly expressed, naturally occurring proteins. To a limited extent, some approved agents represent our venture into *protein engineering*, in which chimeric or fusion proteins (immunotoxin conjugates) or mutated or deleted proteins (consensus interferon and rPA) have been developed as a result of our experiences with first-generation agents. A number of other biotechnology approaches are currently under investigation, often in human clinical trials, and their success remains to be fully realized. These agents include antisense RNA and DNA, small interfering RNA (siRNA), ribozymes, aptamers, and gene therapy.

In 1978, the first *in vitro* experiments in which *antisense DNA* was used for specific repression of gene expression opened the door to the opportunity to block disease-causing genes selectively. The antisense nucleotide approach is based on the use of short oligonucleotides (13–25 base pairs) complementary to a specific mRNA whose protein product is implicated in a disease such as cancer or viral infection.³⁷ Binding of the antisense molecule to its target mRNA is then believed to inhibit protein translation by interfering with ribosomal function or inducing mRNA degradation through either cytoplasmic nuclease RNAase H or nuclear RNAase P. This approach meets the two crucial criteria in the design of any successful therapeutic agent: the identification of a target implicated in disease etiology and the ability to find a molecule capable of high affinity and selectivity for the target. Although problems of stability, delivery, and specificity have somewhat dulled the initial promise of these agents, several second or third-generation chemically modified antisense molecules (e.g., 2'-O-methoxyethyl-RNA, peptide nucleic acid, morpholino phosphoroamidate) with improved nuclear resistance and distribution profile are in late clinical trials.³⁸ The first antisense drug, Vitravene (to prevent blindness due to cytomegalovirus infections in HIV-positive patients), was approved in 1998 in the US and subsequently in 1999 in the EU. However, it was voluntarily withdrawn in the EU in 2002 for commercial reasons.

Another encouraging adaptation of antisense therapies is an approach that takes advantage of the Nobel prize-winning discovery of Thomas Cech and Sidney Altman, who demonstrated that certain RNA molecules possess an enzymatic RNA-degrading activity. The RNAs, called *ribozymes*, are directed toward a specific disease-causing RNA by the sequence homology used by antisense molecules. However, the promise of ribozymes relates to their catalytic nature; one ribozyme molecule can lead to the destruction of thousands of target RNA molecules. In addition, ribozyme technology appears to possess greater specificity over some antisense nucleotides. Nonetheless, the instability of RNAs as therapeutic agents is one hurdle that must be overcome before the promise of ribozymes can be fully realized.

Other oligonucleotide technologies also being investigated take advantage of the three-dimensional nature of DNA/RNA molecules. Aptamers are oligonucleic acids specifically selected for high affinity to certain molecular targets. *In vitro* selectivity of these molecules is quite impressive; affinity-selected aptamers can distinguish between theophylline and caffeine, hypoxanthine molecules that differ by only one methyl group.³⁹ The hope is that such selectivity can be applied to closely related enzymes (i.e., COX-1 and COX-2) or receptor family subtypes. The first FDA approved aptamer is Macugen (Pfizer, 2004), a PEGylated oligonucleotide, which binds specifically to vascular endothelial growth factor for the treatment of age-related macular degeneration.

The most recent, highly-hyped technology to emerge in the late 1990s was the use of small, double-stranded RNA molecules to specifically turn off expression of genes involved in pathogenesis.⁴⁰ In 1998, this pathway was first shown to exist in an invertebrate model dear to geneticists, the worm *C. elegans*. Shortly thereafter, it was recognized that mammalian cells possess a similar regulatory mechanism for silencing endogenous genes called RNA interference (RNAi). RNAi is a ubiquitous and fundamental cellular mechanism for gene silencing. While the biological mechanism and consequence of RNAi are still under investigation, RNAi related technologies, such as synthetic small interfering RNA (siRNA) and synthetic or expressed small hairpin RNA (shRNA), are rapidly being developed as therapeutic drugs to specifically and efficiently inhibit disease-related genes. However, converting siRNA or shRNA into a potent therapeutic drug requires addressing problems that are still haunting oligonucleotide drugs (antisense oligonucleotide, ribozyme and aptamer) and gene therapy, such as *in vivo* stability, tissue and cell specificity, intracellular durability, immunogenicity, and toxicity.

siRNA is the functional precursor of RNAi.⁴¹ It is composed of 19–23 base pairs (bp) of double-stranded RNA with 2 bp overhangs at each 3' end. In the cytoplasm, siRNA is processed by Dicer (a RNase III family enzyme) and forms a nuclease complex known as RNA-induced silencing complex (RISC). Subsequently, the RISC endonuclease Ago2 cleaves the sense strand and activates the RISC. Directed by the single-stranded antisense strand, the activated RISC is capable of performing multiple rounds of target mRNA recognition and cleavage. The enzyme-mediated multiple rounds of mRNA cleavage are the key difference between siRNA and antisense oligonucleotide methodologies. siRNA and shRNA are successfully and broadly used to silence target genes for functional validation *in vitro*. Although siRNA is 10–100 times more efficient in gene silencing than antisense oligonucleotide, siRNA and antisense oligonucleotide confront the similar obstacles for therapeutic application *in vivo*. Chemical modifications are usually incorporated into the phosphodiester linkages, the hydroxyl groups of the ribose, the nucleobases and the 3' or 5' terminus of siRNA to improve cellular targeting, mRNA binding, and especially nuclease resistance. All of these modifications have to balance the siRNA activity, cytotoxicity, and pharmacological properties to achieve an optimal efficiency *in vivo*.

Another problem is specificity and the potential for “off-target” gene silencing.⁴² The off-target can be due to (1) the short complementary sequence (8 nucleotides); (2) the saturation of endogenous RNAi machinery and subsequent interference with the systemic microRNA and siRNA regulation and function; and (3) the non-specific siRNA binding to Toll-like cellular receptors and subsequent immune stimulation. A number of companies have sprung up hoping to exploit this technology in treating viral diseases, especially HIV/AIDS, and other disorders caused by over-activity of an enzyme or protein. Most of these studies are in Phase II or I clinical trials. As with other revolutionary technologies that propose to realize Ehrlich's dream of the magic bullet, only time and experience will determine the true clinical utility of these agents.

Until the unexpected death of Jesse Gelsinger in a Phase I recombinant adenovirus trial in 1998, the next wave of products seemed destined to be directed toward gene therapy.⁴³ Many different approaches constitute gene therapy, but the most common is to attempt to replace a nonfunctioning or mutated gene product by the directed expression of a new, non-mutated copy of that gene. In other cases, genes are being introduced to make drug therapy more effective (e.g., HSV thymidine kinase, p53) or gene therapy is combined with other aforementioned approaches (e.g., intracellular expression of antisense RNA to the *ras* oncogene). In general, the DNA encoding these new genes is encoded on a plasmid molecule or is part of a viral vector that can infect cells with the appropriate desirable gene without causing viral disease. (The Gelsinger case points out that the recombinant virus might still be lethal via other mechanisms.)

Delivery methods for these gene sources usually either exploit the DNA delivery tactic of the virus itself or employ cationic liposomal complexes with the DNA to mask the plasmid's negative charge. Obviously, there is substantial concern over the use of modified retroviruses or adeno-associated viruses as gene delivery systems for fear that important host cell genes might be disrupted if the viral DNA is integrated into the host genome. Cationic liposomal strategies have made substantial leaps in the last several years, but their efficiency of gene delivery pales in comparison with viral delivery. Nonetheless, liposome/DNA complexes are amenable to lyophilization and reconstitution, and advances are being made in maximizing the efficiency of these stabilized preparations.⁴⁴

Stem cell therapy is used to introduce new cells to replace the damaged cells in order to repair, recover, and restore the normal function of damaged or diseased tissues or organs and has potential application in the treatment of brain damage, spinal injury, heart damage, blindness and vision impairment, wound healing, and diabetes.⁴⁵ It has been the hot topic of public media, political debate, and medical research for the past decade. Stem cells are characterized by their ability of self-renewal and potency to differentiate. Based on differentiation potential, two kinds of stem cells, pluripotent embryonic stem cells and multipotent somatic stem cells, are commonly used in therapy. Embryonic stem cells can be isolated from the inner cell mass of blastocyst-stage embryos obtained either from *in vitro* fertilization or created by somatic cell nuclear transfer. The immune rejection of embryonic stem cells in clinical treatment is a potential problem. Using somatic nuclear transfer to produce embryonic stem cells may avoid this problem, but it is similar to human cloning, and the technique has not been well developed so far. In addition, due to the differentiation potency of stem cells, the development of teratoma from transplanted cells is another valid concern for stem cell therapy. The first clinical trial using embryonic stem cells to repair spinal injury was approved by the FDA in 2009.

In 2007, Dr. Shinya Yamanaka's group at Kyoto University and Dr. James Thomson's group at University of Wisconsin-Madison have published their methods to genetically reprogram somatic cells (e.g. epithelial cells) with transcription factors to convert them into pluripotent stem cells. It bypasses the ethical issue of destroying embryos and provides a brand new opportunity for stem cell therapy. However, it is uncertain whether these induced pluripotent stem cells are completely exchangeable with embryonic stem cells and are safe to be used in humans.

The combination of nanotechnology and biology has resulted in a new and fast growing scientific discipline, nanobiotechnology. Nanotechnology deals with materials in the nano range (1–100 nm) that are supposed to possess unique physical and chemical properties. Application of nanotechnology in biology and medicine may bring about innovative drugs, biomaterials, and medical devices. However, the main efforts and achievements of nanobiotechnology in drug discovery still remain in the development of nanoparticles using all kinds of materials (gold, quantum dot, biopolymers, and polypeptides) and processings. The biological characteristics and their significant consequences of the interfaces between these nanoparticles and biomolecules are not clearly understood.⁴⁶ Furthermore, the long term benefits and side effects of nanoparticles on human and environment have not been extensively evaluated. Abraxane (*Abraxis Bioscience*), approved by the FDA in 2005, is claimed as the first nanoparticles drug. It takes advantage of the albumin binding ability of a chemotherapeutic, paclitaxel, to make albumin and paclitaxel nanoparticle complexes. Therefore, it solves the poor solubility problem of paclitaxel and avoids the hyperallergic side effect caused by Cremophor EL in regular paclitaxel formulation. Abraxane has been approved for the treatment of breast cancer and is in clinical trial for other cancer treatment.

Finally, a new area has emerged for the potential application of biotechnology products against bioterrorism. Following the

September 11, 2001, terrorist attacks in New York, Washington, DC, and Pennsylvania and the subsequent anthrax scare, Congress earmarked \$37 billion toward homeland defense of which \$6 billion is targeted to developing biodefense strategies. A little-known provision of new FDA guidelines are that counter-terrorism agents are eligible for expedited review, as seen now for drugs to treat AIDS or cancer. Biotechnological approaches have already been proposed for both the detection and treatment for biological agents most likely to be used in terrorist attacks.

By the end of the first decade of the twenty-first century, merging among big pharmaceutical companies had become a trend. Pfizer merged with Wythe, Merck merged with Schering-Plough, and Roche merged with Genentech. Wythe, Schering Plough, and Genentech all carried successful biological products and possessed rich product pipelines. These mergers undoubtedly secured the biological product pipeline for these big pharma. Meanwhile, because of increased R&D cost and budget cuts, big pharma also like to acquire small or medium biotechnology companies with promising products to continue late phase clinical trials.

PHARMACOGNOSTICAL APPLICATIONS

In regard to the applications of biotechnology in pharmacognosy (drugs from natural sources), Cordell⁴⁷ reports that major efforts in this field are underway in Germany, Japan, and the People's Republic of China. These countries are attempting to use manipulated plant-cell cultures to produce otherwise difficult to extract natural products. In Germany, which has a substantial number of efficacious prescription natural products not available in the US, the pharmaceutical industry and government have joined together to form an institute designed specifically to produce natural products commercially through cell-free systems and gene technology of medicinal plants. One company in Japan reported that it can produce ginseng extract, identical in chemical composition with that from mature 6-year-old root, in 75 000 liter fermenters.

Davies summarized the application of genetic engineering to the production of pharmaceuticals. He pointed out that new drugs traditionally have come from natural product sources, usually followed by improved growing techniques or chemical synthesis. However, while the plant or microorganism strain-improvement procedures have resulted in up to 1000-fold increases in yield, the techniques largely have been empirical. Furthermore, these methods have virtually no genetic or biochemical pedigree for successive improvements.

The organisms currently used to produce antibiotics (e.g. penicillin and tetracycline) on a commercial scale are the same species as those originally collected from the natural state and really have been modified only through forced genetic manipulations (by virtue of the media, etc.), based on strain improvement. Of paramount importance is the fact that while gene cloning and recombinant DNA techniques have been successful for proteins and peptides, the higher plants and microbes producing antibiotics have not been manipulated similarly. This is because an entirely different situation exists when more complex genetic and biochemical processes need to be manipulated. As discussed later, there are also some regulatory assurances that remain to be met, particularly in using transgenic plants as a source of pharmaceuticals.

Antibiotics and alkaloids usually are biosynthesized through multistep pathways possessing complex biogenesis and regulatory circuits. The numerous genes involved in the synthesis of a simple antibiotic are not necessarily present as a single genetic linkage group. Thus, cloning the genes needed would require numerous operations, often without the advantage of selective procedures to detect the presence of the cloned genes. Similarly, the same technical problems associated with multiple components of antibiotic synthesis apply to the genetic engineering of improved yields of many secondary metabolites

like alkaloids from plants. Further complicating the issue is that suitable host-vector systems need to be developed fully for the application of recombinant DNA techniques to plants. However, if there is a single rate-limiting step in the pathway of antibiotic biosynthesis, it might be possible to clone the gene for this step by selecting for increased antibiotic production by *shotgun* cloning back into the producing organism.

An alternative way to achieve increased levels of antibiotic production might be to clone appropriate genes on multicopy or high-expression vectors. Another idea would be to engineer antibiotic-producing organisms to produce hybrid or specifically modified antibiotics. This would result in several model compounds that might have more desirable or better properties than the parent antibiotic. For example, Hailes⁴⁸ outlined the directed biotransformation of opiate analogs including morphine and hydromorphone, using genetically engineered bacteria, and discussed the application of this method toward biosynthesis of novel alkaloids. Marsden⁴⁹ reported the engineering of broader substrate specificity into a macrolide antibiotic-producing polyketide synthase from the erythromycin producer, *Saccharopolyspora erythraea*. These investigators modified the carboxylic acid acceptor unit of this multicomponent enzyme complex to utilize over 40 alternative branched-chain starter carboxylic acids (as opposed to the two straight chain acids utilized normally). In doing so, a vast number of novel, macrolide antibiotics can now be synthesized.

In the case of plant tissue culture, many compounds such as secondary metabolites have been produced with yields that are equal to or greater than that of parent plants. Staba⁵⁰ reports on at least 30 natural products that have been generated through plant-cell culture. Among these are included several well-known, but still difficult to obtain, drugs such as diosgenin-derived steroid hormone precursors, the opium alkaloids, digitalis glycosides, several different essential oils, and the *Catharanthus* alkaloids, vincristine, and vinblastine. However, it has been pointed out that these methods still are not economical when compared with the traditional methods used currently, viz., direct extraction from whole plant materials.

Thus far, only one Asiatic drug, known as shikonin (from *Lithospermum erythrorhizon*), has been produced through plant-cell culture methodology in greater quantities and with substantially lower costs than usual extraction procedures.⁵¹ Certainly, continued efforts expended in the biotechnological manipulation of plant genes will prove more successful as research continues in this area. Table 34.2,⁵² shows established hairy root cultures (the result of genetic transformation of the plant by *Agrobacterium rhizogenes*) and examples of secondary product formation.

Finally, even though many efforts are oriented toward the cheap, controlled, pharmaceutical production of secondary plant products, it should be remembered that this new approach can offer a valuable system for basic plant biosynthesis. These techniques offer a means quite suitable for physiological studies as well as for genetic manipulation. For once, it will be possible to have a powerful tool for the study of the control of gene expression at both the cellular and whole plant or organ levels. Beyond this, the increased efficiency with which biosynthetic pathways for desired compounds may be expressed makes plant DNA promising material as a source of mRNA for cloning operations directed at the transfer of specific plant enzymes into microorganisms.⁵²

Awad⁵³ has provided an overview on plant biotechnology that he feels is a field fertile for pharmaceutical research. In addition to providing two tables listing both microorganisms and plants that have been used in agricultural, horticultural, and pharmaceutical research in biotechnology, he shows the major groups of compounds of commercial importance that are derived from plants. These include pharmaceuticals (alkaloids, steroids, anthraquinones), enzymes (papain), latex (rubber), waxes (jojoba, caruba), pigments (food dyes), oils (olive oil, corn oil, etc.), agrochemicals (pyrethrins), cosmetic substances

Table 34-2. Established Hairy Root Cultures with Examples of Secondary Product Formation

Family	Species	Major Secondary Products
Solanaceae	<i>Atropa belladonna</i>	Atropine
	<i>Datura stramonium</i>	Hyoscyamine
	<i>Hyoscyamus muticus</i>	Hyoscyamine
	<i>Nicotiana rustica</i>	Nicotine, anatabine
	<i>N. tabacum</i>	Nicotine, anatabine
	<i>N. hesperis</i>	Nicotine, anabasine
	<i>N. cavicola</i>	Nicotine, nornicotine
	<i>Scopolia japonica</i>	Hyoscyamine
	<i>Solanum laciniatum</i>	Steroidal alkaloids
Apocynaceae	<i>Catharanthus roseus</i>	Ajmalicine, serpentine, vindoline, catharanthine
Chenopodiaceae	<i>Beta vulgaris</i>	Betacyanin, betaxanthin
Polygonaceae	<i>Polygonum hydropiper</i>	
Boraginaceae	<i>Lithospermum erythrorhizon</i>	Shikonin
Compositae	<i>Tagetes patula</i>	Thiophenes
Rubiaceae	<i>Cinchona ledgeriana</i>	Quinoline alkaloids

From Hamill J, Parr A, Robins R *et al.* New Routes to Plant Secondary Products. *Nature Biotechnology* 1987; 5: 800-4.

(essential oils and perfumes), food additives (flavor compounds, non-nutritive sweeteners), and gums (gum acacia and tragacanth). He perceives the major trends in plant biotechnology to include plant-microbe interactions, gene delivery and manipulation, diversity of gene engineering, and microbial and plant secondary metabolites.

Constabel⁵⁴ reviewed medicinal plant biotechnology as a revolutionary methodology useful in enhancing the formation and accumulation of desirable natural products and a possible product-modification method. He describes advances in micropropagation that involve plant regeneration from *in vitro* cultured cells. Here the multiplication factor can be high and of great advantage to speed up slow-growing important medicinal plants. Recent studies with ipecac (*Cephaelis ipecacuanha*) yielded 100 plantlets per shoot-tip explant per year or 600 plantlets per axenic shoot. Similarly, *Digitalis lanata* cultivars with high cardenolide content were obtained by inbreeding and subsequent crossing of selected genotypes. The isolation and *in vitro* culture of shoot tips led to the formation of adventitious shoots. Following short- and long-term culture, rooting of these shoots on solid medium established plantlets that were transferred to soil. The cardenolide yields equaled those of the parent plants.

Similar studies showed that axenic shoot-tip cultures also can be stored for long periods of time and even be cryopreserved for gene banks. Some studies also are focusing on overcoming somaclonal variation, so that stable, high-yield chemovariants can be developed. In terms of enhancement of productivity, gene technology has allowed true transgenic plants. Their uniqueness is shown as a change (i.e., enhancement) in plant performance (insect resistance, herbicide) and productivity (storage proteins, pigmentation). Transgenic cell cultures of drug plants with modified or increased productivity and microorganisms producing phytochemicals are conceivable and may further increase the feasibility of phytochemical production by *in vitro* methods. Already key enzymes for biosynthetic pathways in plants have been identified and related to isolated DNA clones and genes. Schell described plant biotechnology as a powerful tool to use plant resources and to improve the environmental impact of agriculture. He described how transgenic plants can be developed to promote insect tolerance, virus resistance, and tolerance to fungal diseases in crop plants. Critics of plant biotechnology raise legitimate concerns over the potential immunogenicity of foreign proteins expressed in foodstuffs. The EU nations have been particularly opposed to the marketing of these so-called "Frankenfoods," but these genetically engineered food crops have now grown to represent more than

40% of the corn and soybeans harvested annually. Similarly reviewed are the enhancement of stress tolerance, development of hybrid seeds, and improvement of nutritional-quality plants. For example, Miller and Ackerman⁵⁵ also provided new perspectives on food biotechnology, describing tomatoes that are rot-resistant and animal foods being modified by genetic means.

Plant biotechnology has also been used to omit products that are undesirable in a particular crop. Of greatest pharmaceutical relevance is the presence of caffeine in the coffee plant.⁵⁶ Those who wish to avoid the hypertensive and CNS-stimulating effects of this methylxanthine must resort to coffee products that have been decaffeinated by chemical processes that may employ organic solvents, which also remove some components of flavor and aroma. However, Japanese researchers⁵⁷ have now succeeded in using RNA interference technology to downregulate one of the three *N*-methyltransferase enzymes required for caffeine biosynthesis in the non-commercial plant, *Coffea canephora*. The resulting transgenic plants exhibited a 50–70% decrease in caffeine content. The remaining challenge is to cross these plants with the more economically important *Coffea arabica* and achieve a 97% reduction in caffeine content as required in the US for "decaffeinated" labeling.

The use of transgenic plants or animals in producing recombinant proteins for pharmaceutical use is called "pharming," a term accepted in the 2001 release of the *The Columbia Encyclopedia*. A 2003 survey by the Pew Initiative on Food and Biotechnology revealed that while 81% of Americans favored tinkering with plant genes to produce medicines, only half felt comfortable making transgenic animals for the same purpose. Proponents of transgenic plants as pharmaceutical sources point out that production of recombinant proteins requires a specialized facility costing \$300–\$500 million and roughly 5 years of set-up. In contrast, transgenic plant or animal scale-up can be done on the scale of tens of millions of dollars and in half the time. However, no recombinant product produced in plants or animals has yet made it to market, and only three plant-produced products are currently in clinical trials. The industry has suffered numerous setbacks,⁵⁸ not the least of which relate to concerns on the contamination of food crops with plant-made pharmaceuticals, particularly since some companies are using plants to produce orally-active vaccines. In late 2002, a Texas plant biotechnology company was cited for twice contaminating soybean crops in Nebraska and Iowa with transgenic corn harboring either a pig vaccine or a human protease inhibitor. Plants clearly hold great promise for producing biotechnology drugs, but several hurdles must be overcome before this idea is practically implemented.

PHARMACOLOGICAL APPLICATIONS

Pharmacologically directed biotechnological methodology also holds much promise for the medical field. Even before completion of the Human Genome Project, genes had been isolated for literally dozens of neurochemical receptor drug targets, ion channels, and transporters. With these sequences in hand, target proteins could be produced *in vitro* or in cellular systems for high-throughput drug screening. Not only do these expressed proteins provide tools for identifying drug molecules active against a certain system, but the availability of other similar gene sequences enables determination of molecular specificity (e.g., dopamine D2 antagonism versus dopamine D1 antagonism). Moreover, these potent drug discovery technologies are no longer exclusive to pharmaceutical companies. Academic researchers in the neuropharmacology area now have access to recombinant protein screening services of the National Institute for Mental Health (NIMH) for novel CNS-active agents. Through an NIMH contract, Drs. Linda Brady and Bryan Roth at Case Western Reserve University maintain cellular systems for assaying drug activity against more than 70 neurochemical receptors and transporters. In fact, this system has recently been used in an attempt to resolve the controversy over the active antidepressant component(s) of extracts from the herbal remedy, St. John's wort.⁵⁹

It should be mentioned that long before their testing as therapeutic agents, antisense oligonucleotides were used experimentally to advance our understanding of physiology and pharmacology. As an example, Pasternak and Standifer⁶⁰ have outlined their exploitation of antisense molecules in elucidating the functional biology of opioid receptor subtypes. Using antisense molecules to downregulate certain opioid receptor subtypes (or mRNA-splicing variants of the same receptors), these investigators have been able to differentiate among receptors responsible for morphine-induced spinal analgesia, morphine-induced supraspinal analgesia, and the supraspinal analgesia induced by the active glucuronide conjugate of morphine. A similar approach has been used to investigate dopamine, muscarinic acetylcholine, and NMDA receptors.

Huber⁶¹ provides an excellent review on the therapeutic opportunities involving cellular oncogenes because of the novel approaches fostered by biotechnology. Oncogenes can serve as novel therapeutic targets for cancer diagnosis, prognosis, and treatment. Marx⁶² provides another perspective on oncogenes by detailing advances made in our understanding of tumor suppressor genes whose function is inactivated in carcinogenic regression. Tumor suppressor genes are obviously excellent candidates for reintroduction strategies in treating cancer.

PHARMACEUTICAL MANUFACTURING APPLICATIONS

The use of genetically modified bacteria also has played a part in increasing the efficiency of producing certain pharmaceutically important synthetic organic compounds. For example, researchers have isolated a gene from a species of *Corynebacterium* that codes for 2,5-diketo-D-glucose reductase. This particular enzyme catalyzes the conversion of 2,5-diketo-D-glucose to 2-keto-L-glucose, which is an intermediate in ascorbic acid (vitamin C) production. The scientists successfully have inserted the gene into *Erwinia herbicola* bacteria, via a plasmid. Unchanged *E. herbicola* has the ability to ferment D-glucose into 2,5-diketo-D-glucose. Inserting the new gene provides the changed bacterium with the ability to ferment D-glucose to 2-keto-L-glucose in one step. The second step involved in the production of vitamin C is base- or acid-catalyzed cyclization of 2-keto-L-glucose to the acid. Thus, the genetic engineering process saves at least four steps, compared to the previous procedure. The older Keichstein Grussner process involves six steps in the synthesis of ascorbic acid.⁶³

ORGANIC SYNTHESIS APPLICATIONS

In an interesting synthesis application, Iverson and Lerner⁶⁴ report on sequence-specific peptide cleavage catalyzed by an antibody. The monoclonal antibodies necessary for this procedure were produced by immunizing with a Co(III) triethylenetetramine (trien)-peptide hapten capable of catalyzing specific hydrolysis of the GlyPhe bond of peptide substrates at a neutral pH with a metal complex cofactor. As a group, these antibodies are able to bind trien complexes of not only Co(III), but also numerous other metals. At least six peptides were studied as possible substrates with these antibodies as well as various metal complexes. The results of these studies demonstrate the feasibility of using cofactor-assisted catalysis in an antibody-binding site to achieve successful results in difficult chemical transformations.

MORAL AND ETHICAL QUESTIONS

On the matter of moral and ethical questions of biotechnology applications in medicine, numerous articles appear periodically^{10,22,65} to debate the issue. Francis Collins, Director of the National Human Genome Research Institute (NHGRI) has written, "It is estimated that all of us carry dozens of glitches in our DNA—so establishing principles of fair use of this information is important for all of us."

There are many questions, such as the following:

- Does genetic testing constitute invasion of privacy?
- Will there be an increase in abortions that discriminate against the *genetically unfit*?
- Should those destined to be stricken with a fatal genetic disease be informed of their fate,
- especially if there is no remedy available?
- Will these decisions become mandated legally and ultimately demean humans or create a new underclass of the genetically less-fortunate?
- Should gene therapy be used only for treating disease or also for *improving* an individual's genetic legacy?

Currently, most protections on the use of genetic information are regulated at the state level, resulting in uneven application across the population. Senators Jeffords and Daschle recently outlined their respective views on the passage of federal laws that protect the collection and use of genetic information, particularly relating to employment and health insurance.¹⁵ The Human Genome Project's ELSI program (Ethical, Legal, and Social Implications of Human Genetics Research) is now charged with addressing issues that appear as daunting as the sequencing of the genome itself. The benefits of biotechnology in disease prevention and treatment are numerous, and the decoding of the human genome will continue to produce new opportunities to improve our quality of life. But as with all technological advances, safeguards are required to prevent discriminatory and unethical use of this new information.

REFERENCES

1. Bud R. History of "biotechnology". *Nature* 1989; 337: 10.
2. Lander ES, *et al.* Initial sequencing and analysis of the human genome. *Nature* 2001; 409: 860–921.
3. Venter JC, *et al.* The sequence of the human genome. *Science* 2001; 291: 1304–51.
4. Broder S, Caplan A, Evans MJ. Therapeutic horizons—the human genome. *Am Pharm Assoc (Wash)* 2002; 42(5 Suppl 1): S22–3.
5. Roth RI, Fleischer NM. Gene therapy: applications to pharmacy practice. *J Am Pharm Assoc (Wash)* 2002; 42(692): 694–8.
6. Maher B. Postcard from the party. *The Scientist* 2003; 4: 3.
7. Bonetta L. Prime time for real-time PCR. *Nat Methods* 2005: 305–312.
8. Holmer A. Pharmaceutical Research and Manufacturers of America (PhRMA). Washington, DC; 2008.

9. Walsh G. Biopharmaceutical benchmarks 2010. *Nat Biotechnol* 2010; 28: 917–24.
10. Somia N, Verma IM. Gene therapy: trials and tribulations. *Nat Rev Genet* 2000; 1: 91–9.
11. Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol* 2008; 26: 1135–45.
12. Birch JR, Racher AJ. Antibody production. *Adv Drug Deliv Rev* 2006; 58: 671–85.
13. Samaranyake H, Wirth T, Schenkwein D, Raty JK, Yla-Herttuala S. Challenges in monoclonal antibody-based therapies. *Ann Med* 2009; 41: 322–31.
14. Rader C, Barbas CF, 3rd. Phage display of combinatorial antibody libraries. *Curr Opin Biotechnol* 1997; 8: 503–8.
15. Adams GP, Weiner LM. Monoclonal antibody therapy of cancer. *Nat Biotechnol* 2005; 23: 1147–57.
16. Hughes B. Antibody-drug conjugates for cancer. *Nat Rev Drug Discov* 2010; 9: 665–7.
17. Evans WE, Relling MV. Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* 1999; 286: 487–91.
18. Chee M, Yang R, Hubbell E, Berno A, Huang XC, Stern D, Winkler J, Lockhart DJ, Morris MS, Fodor SP. Accessing genetic information with high-density DNA arrays. *Science* 1996; 274: 610–4.
19. Rosenwald A, Wright G, Chan WC, *et al.* The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* 2002; 346: 1937–47.
20. Gunther EC, Stone DJ, Gerwien RW, Bento P, Heyes MP. Prediction of clinical drug efficacy by classification of drug-induced genomic expression profiles in vitro. *Proc Natl Acad Sci U S A* 2003; 100: 9608–13.
21. Petricoin EF, Ardekani AM, Hitt BA, *et al.* Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 2002; 359: 572–7.
22. Jeffords JM, Daschle T. Policy issues. Political issues in the genome era. *Science* 2001; 291: 1249–51.
23. Gill P, Ivanov PL, Kimpton C, Piercy R, Benson N, Tully G, Evett I, Hagelberg E, Sullivan K. Identification of the remains of the Romanov family by DNA analysis. *Nat Genet* 1994; 6: 130–5.
24. Look AT. Oncogenic transcription factors in the human acute leukemias. *Science* 1997; 278: 1059–64.
25. Leader B, Baca QJ, Golan DE. Protein therapeutics: a summary and pharmacological classification. *Nat Rev Drug Discov* 2008; 7: 21–39.
26. Jackson RC. Contributions of protein structure-based drug design to cancer chemotherapy. *Semin Oncol* 1997; 24: 164–72.
27. Capdeville R, Buchdunger E, Zimmermann J, Matter A. Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. *Nat Rev Drug Discov* 2002; 1: 493–502.
28. Tami J, Evens RP. Evaluation of biotechnology products. *Pharmacotherapy* 1996; 16: 527–36.
29. Buckel P. Recombinant proteins for therapy. *Trends Pharmacol Sci* 1996; 17: 450–6.
30. Cho MJ, Juliano R. Macromolecular versus small-molecule therapeutics: drug discovery, development and clinical considerations. *Trends Biotechnol* 1996; 14: 153–8.
31. Hamilton SR, Bobrowicz P, Bobrowicz B, *et al.* Production of complex human glycoproteins in yeast. *Science* 2003; 301: 1244–6.
32. Toon S. The relevance of pharmacokinetics in the development of biotechnology products. *Eur J Drug Metab Pharmacokinet* 1996; 21: 93–103.
33. Sheffield WP. Modification of clearance of therapeutic and potentially therapeutic proteins. *Curr Drug Targets Cardiovasc Haematol Disord* 2001; 1: 1–22.
34. Harris JM, Chess RB. Effect of pegylation on pharmaceuticals. *Nat Rev Drug Discov* 2003; 2: 214–21.
35. Mack GS. Pfizer dumps Exubera. *Nat Biotechnol* 2007; 25: 1331–2.
36. Misra A. Are biosimilars really generics? *Expert Opin Biol Ther* 2010; 10: 489–94.
37. Agrawal S. Antisense oligonucleotides: towards clinical trials. *Trends Biotechnol* 1996; 14: 376–87.
38. Stein CA. Phosphorothioate antisense oligodeoxynucleotides: questions of specificity. *Trends Biotechnol* 1996; 14: 147–9.
39. Gold L, Polisky B, Uhlenbeck O, Yarus M. Diversity of oligonucleotide functions. *Annu Rev Biochem* 1995; 64: 763–97.
40. McManus MT, Sharp PA. Gene silencing in mammals by small interfering RNAs. *Nat Rev Genet* 2002; 3: 737–47.
41. Chi JT, Chang HY, Wang NN, Chang DS, Dunphy N, Brown PO. Genomewide view of gene silencing by small interfering RNAs. *Proc Natl Acad Sci U S A* 2003; 100: 6343–6.
42. Jackson AL, Bartz SR, Schelter J, Kobayashi SV, Burchard J, Mao M, Li B, Cavet G, Linsley PS. Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol* 2003; 21: 635–7.
43. Blau HM, Springer ML. Gene therapy—a novel form of drug delivery. *N Engl J Med* 1995; 333: 1204–7.
44. Anchordoquy TJ, Carpenter JF, Kroll DJ. Maintenance of transfection rates and physical characterization of lipid/DNA complexes after freeze-drying and rehydration. *Arch Biochem Biophys* 1997; 348: 199–206.
45. Singec I, Jandial R, Crain A, Nikkhah G, Snyder EY. The leading edge of stem cell therapeutics. *Annu Rev Med* 2007; 58: 313–28.
46. Park S, Hamad-Schifferli K. Nanoscale interfaces to biology. *Curr Opin Chem Biol* 2010; 14: 616–22.
47. Cordell G. Pharmacognosy: far from dead. *Am Druggist* 1987; 96–8.
48. Hailes AM, French CE, Rathbone DA, Bruce NC. Engineering pathways in *Escherichia coli* for the synthesis of morphine alkaloid analgesics and antitussives. *Ann N Y Acad Sci* 1996; 799: 391–6.
49. Marsden AF, Caffrey P, Aparicio JF, Loughran MS, Staunton J, Leadlay PF. Stereospecific acyl transfers on the erythromycin-producing polyketide synthase. *Science* 1994; 263: 378–80.
50. Staba J. Milestones in plant tissue culture system for the production of secondary products. *J Nat Prod* 1995; 4: 203–9.
51. Fujita Y, Y. H, Suga C. Production of shikonin derivatives by cell suspension cultures of *Lithospermum erythrorhizon*. *Plant Cell Rep* 1981; 1: 59–60.
52. Hamill JD, Parr AJ, Rhodes MJC, Robins RJ, Walton NJ. New routes to plant secondary products. *BioTechnology* 1987; 5: 800–804.
53. Awad A. *Plant Biotechnology: A Fertile Field for Pharmacy*. Ava, OH: Ohio Northern University Press. 1987.
54. Constabel F. Medicinal plant biotechnology. *Planta Med* 1990; 56: 421–5.
55. Miller HI. When is a drug 'safe'? *Nature* 1990; 344(698).
56. Ashihara H, Crozier A. Caffeine: a well known but little mentioned compound in plant science. *Trends Plant Sci* 2001; 6: 407–13.
57. Ogita S, Uefuji H, Yamaguchi Y, Koizumi N, Sano H. Producing decaffeinated coffee plants. *Nature* 2003; 423(823):.
58. Whitehouse D. Doubts over “pharming” technology. BBC Science and Technology News; 2002.
59. Butterweck V, Nahrstedt A, Evans J, *et al.* In vitro receptor screening of pure constituents of St. John's wort reveals novel interactions with a number of GPCRs. *Psychopharmacology (Berl)* 2002; 162: 193–202.
60. Pasternak GW, Standifer KM. Mapping of opioid receptors using antisense oligodeoxynucleotides: correlating their molecular biology and pharmacology. *Trends Pharmacol Sci* 1995; 16: 344–50.
61. Huber BE. Therapeutic opportunities involving cellular oncogenes: novel approaches fostered by biotechnology. *FASEB J* 1989; 3: 5–13.
62. Marx J. Oncogenes reach a milestone. *Science* 1994; 266: 1942–4.
63. Anonymous. Vitamin C production to use chemical engineering. *Chem Eng News* 1985; 63: 6a.
64. Iverson BL, Lerner RA. Sequence-specific peptide cleavage catalyzed by an antibody. *Science* 1989; 243: 1184–8.
65. Roberts L. Ethical questions haunt new genetic technologies. *Science* 1989; 243: 1134–6.

BIBLIOGRAPHY

GENERAL

- Antebi E, Fishlock D. *Biotechnology, Strategies for Life*. Cambridge MA: MIT Press, 1986: 239.
- Collins FS, Green ED, Guttamacher AE, et al. A vision for the future of genomics research. *Nature* 2003; 22: 835–47.
- Drews J. Intent and coincidence in pharmaceutical discovery. *The impact of biotechnology*. *Arzneimittelforsch* 1995; 45(8): 934–9.
- Schlumberger HD, Stadler P. Modern pharmaceutical biotechnology. Situation worldwide and in Germany 1995. *Arzneimittelforsch* 1997; 47(1): 106–10.
- Schonfeld E. *Drugs of choice*. *Fortune* 1997; 11: 293.
- Starr C, Taggart R. *Biology, the Unity and Diversity of Life*, 5th ed. Belmont, CA: Wadsworth, 1989.
- Wiseman A, ed. *Principles of Biotechnology*, 2nd ed. New York, NY: Surrey University Press (Chapman & Hall), 1988.

DRUGS, MEDICINE AND PHARMACY

- Baldwin JN. Curricular guidelines for pharmacy education: substance abuse and addictive disease. *Am J Pharm Educ* 1991; 55: 311–6.
- Brenner MK. Gene transfer to hematopoietic cells. *N Engl J Med* 1996; 35: 337–9.
- Davies JE. In Krongsgaard P, Brøgger Christensen S, Kofod H, eds. *Natural Products and Drug Development* (Alfred Benzon Symp 20). Copenhagen: 1984, p. 65.
- Fix JA. Oral controlled release technology for peptides: status and future prospects. *Pharm Res* 1996; 13 (12): 1760–4.

- Galbraith WM. Safety evaluation of biotechnology derived products. *Prog Clin Biol Res* 1987; 235: 3–14.
- Illum L, Davis SS. Drug delivery. *Curr Opin Biotechnol* 1991; 2: 254–9.
- Szkrybalo W. Emerging trends in biotechnology: a perspective from the pharmaceutical industry. *Pharm Res* 1987; 4: 361–3.
- Vermeij P, Blok D. New peptide and protein drugs. *Pharma World Sci* 1996; 18 (3): 87–93.

PLANTS

- Baum R. *Agricultural biotechnology advances toward commercialization*. *Chem Eng News* 1987 (Aug 10).
- Pareilliux A. [Technology and strategies for production of mass cultures of plant cells.]. *Ann Pharm Fr* 1987; 45: 155–62.
- Powell K. Barnyard biotech—lame duck or golden goose? *Nat Biotechnol* 2003; 21: 965–7.
- Schell J. In Baba S, et al., eds. *National Resources and Human Health—Plants of Medicinal and Nutritional Value* (Proc 1st WHO Symp on Plants and Health for All: Sci Advance, Kobe, Japan, 26–28 Aug 1991). New York: Elsevier, 1992: 49.
- Yoxen E. *The Gene Business. Who Should Control Biotechnology*. New York: Oxford University Press, 1983, p 230. [Note: This reference provides a detailed analytical study of the corporate agenda for biotechnology. It is a readable account of how genetic engineering evolved from a pure science into a profitable business. The structure and function of the multinational gene business, the effects to date, and the economic, scientific, social, and political implications are examined in detail.]

APPENDIX A TOP 10 BESTSELLING BIOLOGICAL DRUGS IN 2009.

Brands	Companies	Indications	Sales (\$ billion)
Enbrel	Amgen, Wyeth Takeda	Rheumatoid arthritis	8.2
Remicade	J&J, Schering Plough, Mitsubishi Tanabe	Crohn's disease	6.5
Humira	Abbott	Rheumatoid arthritis	5.6
Avastin	Roche, Genentech	Metastatic colon cancer glioblastoma, metastatic renal carcinoma	5.4
Rituxan	Genentech, Biogen	IDEC Non-Hodgkin's lymphoma	5.3
Herceptin	Roche, Genentech	HER2 positive metastatic breast cancer	5.2
Lantus	Sanofi Aventis	Diabetes	4.8
Lovenox	Sanofi Aventis	Anticoagulant DVT	4.6
Aranesp	Amgen	Anemia	3.3
Gardasil	Merck	Human papilloma virus vaccine	3.1

The data do not intend to promote the sale of listed representative drugs. It is only for the purpose to show the immense market and profit of biologic drugs.

(Data from Maggon K. Biotechnology Market Review & Top Ten Biologics 2009. Web. <http://knol.google.com/k/biotechnology-market-review-top-ten-biologics-2009#>. Accessed 16 December 2011)

Chapter 35

Pharmaceutical Packaging

C. Jeanne Taborsky, BSChem and Kathleen Deiss, RN

INTRODUCTION 667

CONTAINER SELECTION BY DOSAGE FORM 668

QUALITY CONTROL 678

OVERVIEW OF TEST METHODS 678

SUMMARY 679

INTRODUCTION

A container closure system must be designed to protect the drug during actual conditions of storage, shipment, and use and be able to deliver the correct amount of product at the time of use. It must not interact with the product over its shelf life to the extent that it renders the drug ineffective or unacceptable for use. As defined in the USP, “a container, including the closure, does not interact physically or chemically with the pharmaceutical preparation in any manner to alter the strength, quality, or purity beyond the official requirements under ordinary or customary condition of handling, shipment, storage, sale or use.”¹

The general requirements for containers for foods, including dietary supplements, drugs, cosmetics, and medical devices, are provided in the Food Drug and Cosmetic Act as amended.² The regulations are provided in *Code of Federal Regulations* Title 21. The FDACenter for Food Safety and Nutrition (CF-SAN) is responsible for the regulations that apply to foods and dietary supplements. The Center for Drug Evaluation (CDER) and the USP/NF are responsible for the regulations that apply to drug packaging. The Center for Biologics Evaluation and Research (CBER) and USP are responsible for the regulations that apply to biologics. The Center for Veterinary Medicine (CVM) relies on CDER and USP/NF for its packaging requirements. The Center for Devices and Radiological Health (CDRH) uses 21 CFR Subchapter H—Medical Devices, ISO, and ASTM International standards. If a liquid medication dispenser is purchased by a pharmacist, independent of the drug, it is regulated by CDRH. If it is sold with the drug, it is regulated by CDER. If an injection syringe is sold by itself, it is regulated by CDRH, and, if it contains a drug (prefilled syringe), it is regulated by CDER.

The standards for drug packaging in USP/NF include General Notices, General Chapters <87>, <88> Biological Reactivity Testing, <381> Elastomeric Closures for Injection, <660> Containers-Glass, <661> Containers- Plastic, <670> Auxiliary Packaging Components, <671> Containers- Performance Testing, and <698> Deliverable Volume. In addition to these requirements, one must be cognizant of the EPA requirements for monomer content in Polyvinyl Chloride (PVC), waste disposal policies, CONEG (Coalition of Northeastern Governors), and Toxics in Packaging Clearing House (TPCH) requirements. For certain drugs, such as most tablets and capsules, the regulations of the Poison Prevention Packaging Act (PPPA) apply. There are also regulations in 21 CFR §201.25 Barcoding, which require barcoding the National Drug Code (NDC) number on the drug labels, including blisters. On May 2, 2011, the ASTM issued a new standard for Water Vapor Permeation titled D7709—Standard Test Methods for Measuring Water Vapor Transmission Rate (WVTR) of Pharmaceutical Bottles and Blisters.³ Although these standards are current at the time of publication, it is important to note that the standards in USP and ASTM are evolving, and it is important to refer to the current versions of these standards, which are reviewed and revised routinely. In addition to the law, regulations, ASTM and USP standards, FDA CDER provides guidance in the amount and type of testing required for drug containers.⁴

The amount of effort required to develop or select the appropriate container system is based on a risk assessment considering the route of administration, class of drug product, therapeutic range, and the chemical and physical stability of the drug substance, drug product matrix, packaging components, and the conditions of storage and use. The route of administration determines the degree of testing necessary to determine the acceptability of a container. In some cases, a component, such as a delivery system, will function at one temperature but not another. Therefore, all evaluations must include storage and use temperatures and packaging conditions. The following is a list of dosage forms from highest to lowest risk for use by patients: injectables, implants, inhalation drugs, ophthalmics and otics, oral liquids, oral solids, and transdermals and topical liquids. The concerns about containers have been fostered by actual events of either interaction or the failure to perform. Some examples of those instances are as follows:

- Migration of nitrosamines from rubber;
- Degradation of components from sterilization;
- Reactions to latex allergens in natural rubber;
- Concentration of drug in pouches and PE vials, due to excessive water loss;
- Migration of inks, adhesives, and chemicals in paper into inhalation solutions;
- Migration of vanillin for labels and cartons into inhalation solutions;
- Migration of chemicals from heat seals, plastics, and foils into drugs;
- Presences of vinyl monomer in polyvinyl chloride (toxic);
- Ions and glass in injection solutions;
- Migration of drug to plastic or coating glass;
- Migration of preservatives;
- Migration of plastic additives to drugs;
- Migration of environmental estrogens to drugs;
- Cross-linking of gelatin capsule shells from furfural in Rayon Coil;
- Interaction of residual bleach and dyes in Cotton Coil;
- Failure of a delivery system to deliver the correct dose;
- Low potency or dissolution properties from excess moisture permeation; and
- Degradation from light.

Container designs and materials have evolved to meet new demands and more sophisticated requirements. Due to a tightening of limits on drug product impurities, degradants, and migrants, the selection of containers has become more complicated. Pharmaceutical companies must conduct compendial tests on the containers⁵ and compatibility tests⁶ on the containers with the actual product, using validated methods. The test methods used by the drug companies to test for impurities and migrants are specific to the pharmaceutical formulation and the packaging material.

Container components must also perform, as needed, to protect the products from physical change, loss of contents, and exposure to moisture, oxygen, and light. To develop the appropriate container closure system, the scientist must know the

critical attributes of the specific drug product and chemical and physical properties of the container options. The importance of selecting the appropriate materials of composition cannot be underestimated. This critical step is often not considered until problematic issues arise. It is essential to consider, first, the container's requirements, and then select materials that fulfill those needs. To determine the appropriate container closure system, one should consider the following:

- Preformulation information on the drug substance (sensitivity to heat, oxygen, humidity, light, glass, metals, pH, reactivity with plastics, metals, or types of glass);
- The nature of the dosage form (solid oral, liquids—aqueous, alcohol, oil, suspensions, inhalation powders, creams, ointments, biodegradable polymers, implants, transdermals, etc.);
- Drug product information (open pan studies on impact of humidity, light, heat, pH, and compatibility studies with the components);
- Container material qualification (CAS number, 21 CFR qualification, USP qualification, etc.);
- Container/product compatibility;
- Impact of the packaging operation on the drug product;
- Ruggedness of the container system under actual conditions of packaging, shipping, storage and use;
- Performance; and
- Cosmetic features.

The drug manufacturer has access to all of the data listed, and the information is reviewed by the FDA in support of their submission. Container manufacturers file Type III Drug Master Files (DMFs) with the FDA, so the information can be reviewed in conjunction with the drug application and FDA can determine if the container is compatible and the information is sufficient for approval of the drug product. The challenge for the repackager, pharmacist, or pharmacy compounder is selecting the appropriate container, when the drug manufacturer does not provide repackaging instructions and the container supplier does not provide chemical or performance information. In these cases, the most conservative approach may be the best approach.

Selecting materials that CDER already approved for use for similar products reduces the risk of component failure. Certain chemicals have become associated with concern regarding their safety and appropriateness of use in packaging drug products. These include, but are not limited to, nitrosamines, BHT, certain carbon blacks, and some antistatic agents. It is best to eliminate these concerns by selecting components manufactured without these chemicals. If possible, the packager should know which 21 CFR references, Food Contact Numbers, and USP or EU requirements the component or material meets. It is advisable to choose materials available commercially. Medical grade materials are preferred, but are often more expensive. The classification of medical grade does not mean the materials are of a particular composition; it means the supplier complies with a change control program. Although selecting medical grade materials is preferable, the materials should, at least, come from a reliable source and be cGMP manufactured. One should know the details of the suppliers change control program. When suppliers are not willing to report changes to the packager or pharmacist or control their changes, the quality of every lot is in question and may require full extensive testing, particularly if it is used to package a high-risk product.

CONTAINER SELECTION BY DOSAGE FORM

TABLETS AND CAPSULES

The containers for solid dosage forms are divided into bulk (not for distribution), multiple-unit and unit-dose containers.⁷ In the late 1960s, drug manufacturers packaged tablets and capsules in

multiple-unit glass or polystyrene (PS) bottles with continuous thread metal screw caps with pulp liners. Pharmacy containers consisted of yellow to orange polystyrene prescription vials with white polyethylene (PE) snap caps. Unit-dose containers were not as prevalent at this time. They were used mostly by institutional pharmacies, in the form of strips of plastic or foil sealed together or blister packs, plastic sealed with a card.

The protection afforded by a container is a combination of the permeation of the bottle and the efficacy of the seal with the closure. In 1970, the Poison Prevention Packaging Act (PPPA) was passed, requiring child-resistant/adult-friendly packaging for most oral prescription drugs. After PPPA was passed, prescription tablets and capsules, packaged in the smallest commercial size container for dispensing, were required to be packaged in child resistant containers. Once this requirement was implemented, there was a wave of complaints about efficacy and the failure of drugs to perform. The product complaints included discoloration, increased hardness, not dissolving in the digestive track, friability, increased softness, low potency, and increased side effects. This phenomenon was first reported to the FDA National Center for Drug Analysis Laboratory in Washington, DC, and was later investigated at the USP/NF Drug Standards Laboratory in Washington, DC. The results of those studies forever changed the materials of construction, test standards, and design of container closure systems.

MULTIPLE-UNIT BOTTLES AND PRESCRIPTION VIALS

Prior to 1970, the USP contained chemical standards for glass and plastic, but not performance tests. These standards were general, not based upon end use. Bottles, vials, and prescription vials used by drug manufacturers and pharmacists were composed of either glass or polystyrene. In response to questions about product failures, the Drug Standards Laboratory modified the existing ASTM WVTR method to evaluate the prescription vials and bottles used to package tablets and capsules. Containers on the market were tested, and the results of those studies are the basis for the standards used to qualify containers today (Fig. 35-1).

Observations of the containers supported and explained the permeation results. The glass bottles did not perform as well as anticipated. Although no moisture permeated the glass bottle, the closure and liner did not provide a tight seal. The glass is not compressible and the cap liner was not effective. The polystyrene (PS) containers were observed to have trace cracks in the walls and a defect in the top. The top of the bottles had a small "tail" from the molding process, which cut into the cap, when it was opened and closed. When there was no liner in the cap, or the cap did not fit tightly, the seal was ineffective. Snap caps made of Low Density Polyethylene (LDPE) stretched on repeat use, whereas those made of Polypropylene (PP) were more rigid and better maintained their shape and fit on the vials and bottles.

When the results of the collaborative study were available, the USP/NF Drug Standards Laboratory met with manufacturers of commercial drug bottles and prescription vials. Based on fabrication properties and the ASTM permeation properties, the container manufacturers committed to change to High Density Polyethylene (HDPE) for bottles and PP resin for vials. Table 35-1—Comparison of the ASTM Permeation of Bottle Resins provides a comparison of the properties of the most common plastic bottles for pharmaceuticals, vitamins, and nutraceuticals. These results were determined by the ASTM test methods, not USP methods. The results for containers manufactured of PP are similar to those of HDPE.

With the implementation of the ICH, the stability test conditions were changed. Many drug companies added foil inner seals, to reduce permeation and increase stability during controlled studies. When implementing this change, manufacturers did not consider the stability of the product once the seal was removed, but were only concerned with the conditions impacting on the reported stability testing of the drug on storage. To

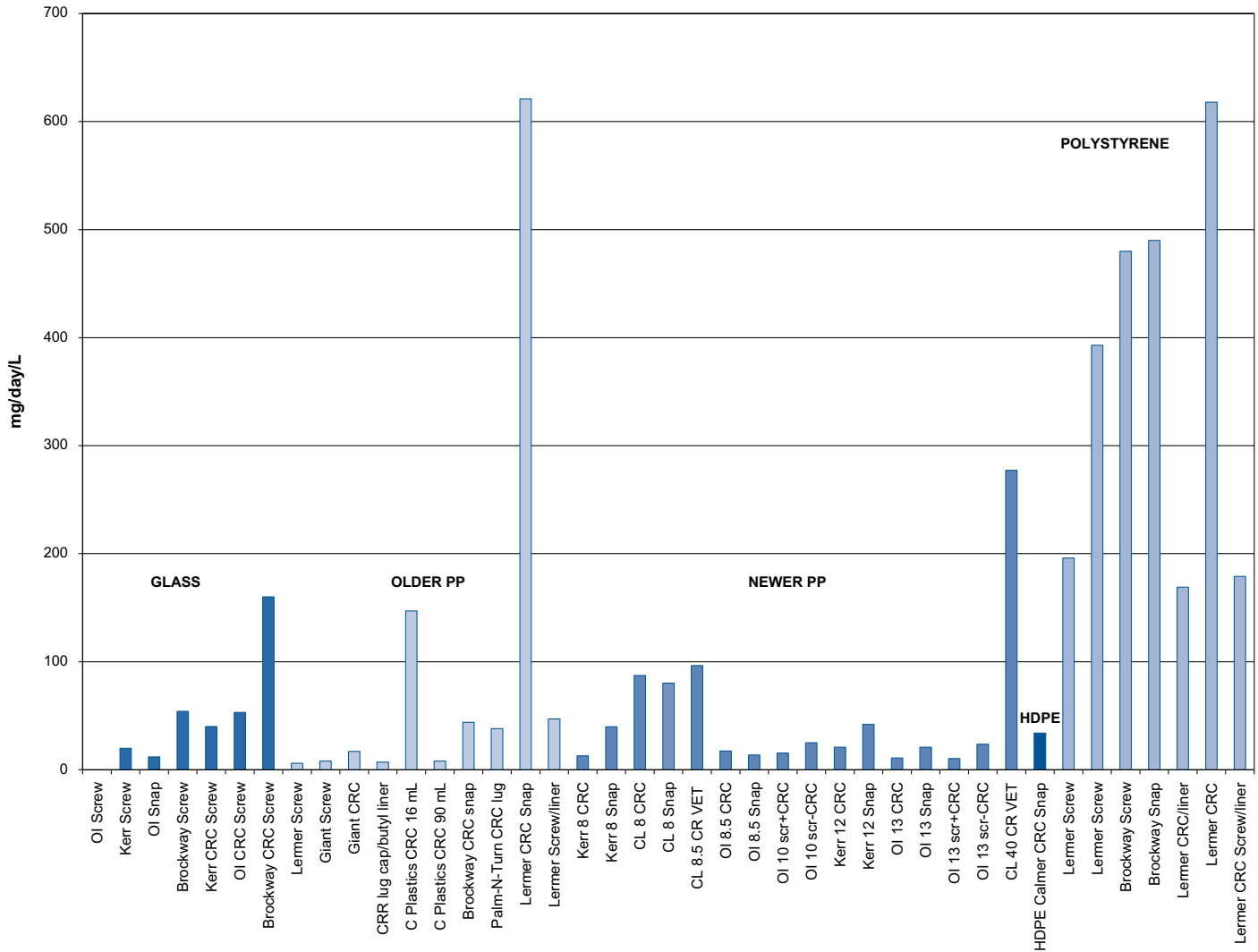


Figure 35-1. USP <671> Moisture Permeation—Bottles and Vials.

determine if the test conditions should be changed to match ICH stability test conditions, the actual concentration of moisture in the air for each condition was calculated and is presented in Table 35-2—Concentration of Water in the Air. Permeation testing was conducted using the ICH conditions for accelerated conditions, controlled room temperature conditions, and USP <671> conditions. The data demonstrated that the USP test mimicked the ICH conditions for controlled room temperature. Since the USP test used a common salt solution to generate the humidity, rather than an expensive temperature/humidity cabinet, it was decided that the test should remain unchanged and a salt solution could be used to generate the test conditions. Keeping the test simple and inexpensive enabled pharmacies and repackagers to do the test in-house.

The critical physical attributes of the bottles include the weight, wall thickness, neck finish, and ovality. These measurements are listed on the bottle drawing and should be part of the routine incoming release testing conducted, before a bottle is used to package a drug product. Additionally, prior to packaging

product into containers, they must be inspected and cleaned, due to the potential for grease, fibers, dust, and insects to be present. Often, the bottles are inverted on the filing line to aid in the removal of debris. Plastic shrink-wrap can be used to minimize contamination. Bottles can also be cleaned by blowing filtered, compressed air into them prior to filling. The air should be filtered, because unfiltered, compressed air often contains oil vapors.

Multiple unit bottles used by pharmaceutical manufacturers and commercial repackagers are universally constructed of white HDPE to fit screw closures. The bottle materials of construction must meet 21 CFR requirements for indirect food additives, such as 21 CFR §177.1520. White colorant is added to protect the product from light. Bottles are fabricated by extrusion or injection blow molding. The resin can be homopolymer or copolymer. The most common copolymer is hexane. The white bottle colorant consists of titanium dioxide extruded in LDPE. The bottle may contain zinc or calcium stearate. The stearate is available of animal or vegetable origin. If the stearate

Table 35-1. Comparison of the ASTM Permeation of Bottle Resins

Property	Polystyrene (PS)	High Density Polyethylene (HDPE)	Low Density Polyethylene (LDPE)	Polyethylene Terephthalate (PET)
WVTR, g/mil/100 in ² /day @ 100 EF, 90% R.H.	5.0	0.3–0.65	1.2	1.3
Oxygen TR, cm ² /mil/100 in ² /day/atm @ 77EF, 0 % R.H.	100–200	30–250	250–840	5

Table 35-2. Concentration of Water in the Air

Temperature °C	% R.H.	Specific Humidity grains/lb dry air
25	40	56
25	60	86
25	75	108
30	60	106
40	15	48
40	75	260

is of animal origin, Bovine Spongiform Encephalopathy (BSE) certification is required. Although some suppliers may say their process would kill the bovine spongiform, there are no test methods that can confirm this and the process cannot be validated. At the present time, most countries, except Brazil, have reported the presence of the disease, and, for this reason, the FDA may ask for the location and name of the farm from which the material originated to approve the source. (See Table 35-3—Information on Typical Materials of Construction for Bottles.)

For product sales within the United States and member countries, the testing requirements used to qualify the HDPE bottles are located in USP <661> Containers-Plastics.⁸ These specifications include tests and limits and are the result of the extensive collaborative study sponsored by the Drug Standards Laboratory, which included the USP, FDA, and industry participation.⁹ A second collaborative study was conducted in 1997 by the FDA Center for Drug Evaluation and Research (CDER), which confirmed the original results, while further investigating the impact of injection versus blow molding and the use of regrind.^{10–14}

USP tests are not required on all shipments of components for tablets; however, if tested, they must pass. The USP <661> physicochemical tests are used to qualify the containers for use in contact with tablets and capsules, and need to be repeated when a change occurs, or to comply with an internal requirement for retesting. The USP tests include two tests for identification: measurement of the infrared spectrum and thermal analysis by differential scanning calorimeter against the USP reference standard. The USP testing also includes the measurement of extractables in hexane, alcohol, and water.

USP <671> requires Moisture Permeation and Light Transmission testing be conducted on all plastic containers, regardless of the resin type or dosage form to be packaged. The USP <671> currently specifies several methods for Moisture Permeation based not only on dosage form, but also on the packaging size and type of configuration (e.g., multiple-unit). The USP Moisture Permeation test was developed for and applies to polyethylene, polypropylene, PET/PETG and glass bottles, and commercial and prescription containers. These test methods for commercial container/closures systems are to be performed with and without seals (Fig. 35-2).

USP <671> Light Transmission test involves cutting a portion of the bottle, mounting it in a holder in a spectrophotometer, and measuring the amount of light that penetrates through the wall of the bottle. The amount of light transmission measured from 290 to 450 nm (ultraviolet/visible) cannot exceed 10%.

Most white HDPE bottles pass this test requirement, although, when the standard was first implemented for HDPE bottles, some did fail.

A study was conducted on white HDPE bottles, PP vials, and PET bottles of various colors to determine if they would pass USP Light Transmission requirements.¹⁵ The bottles included those promoted for use for drugs, nutraceuticals, and prescription vials. All of the white HDPE bottles and PP prescription vials passed USP limits. The results are provided in Figures 35-3 through 35-5. The same could not be said for the various colors of PET bottles, where many failed 10% and one bottle exceeded 60%. The results confirm the premise that, just because the bottles are tinted or contain a colorant, does not guarantee they are light resistant.

Prescription vials and caps are sold together by the manufacturer. Prescription vials are manufactured of amber polypropylene and available with a white polypropylene child resistant closure or a white low density polyethylene snap cap. After the passage of PPPA, the market was flooded with CRC prescription vials, many of which although protecting the child, did not protect the drugs. The amount of moisture that permeated the system varied greatly. Most of the PS prescription vials were removed from the marketplace, since they would not pass USP “tight” limits (see Fig. 35-1). Later testing on marketed prescription vials showed a definite improvement in the protective properties. In most cases, the child resistant closure protects the product better than the non-child resistant closure, since the cap does not get stretched from repeat use.

Closures for bottles are manufactured from white polypropylene and are child resistant (CRC) or standard continuous thread. The smallest container sizes intended for dispensing to the consumer must meet CRC regulations.¹⁶ The CRC is two parts: the outer cap is white polypropylene, and the inner cap is clear polypropylene. The closures are fitted with liners and, in most cases, with heat seals. The liner may be held in place with glue. The cap seal is heat-sealed at the time the bottle is filled and removed by the consumer at the time of use. Thereafter, the cap liner is affecting the seal (see Fig. 35-2).

Cap liners can be manufactured of pulp board, foamed polystyrene, or foamed polyethylene. Pulp board can also be coated with wax. The choice of liner is critical to the protection of the drug from moisture. Although drug stability studies are conducted by the manufacturer with the heat seal in place, once the bottle is opened and the seal removed, it is the cap liner performance that is critical. The cap liner should be able to contract and re-expand after multiple openings and closings. Some cap liners, once compressed, will not re-expand, and the protection from moisture may be compromised. USP <661> Extractables testing, when applied to cap liners, have been known to fail, but these components have not been evaluated by the USP, and specifications relating directly to liners and inner seals have not been listed. The importance of choosing the correct liner and heat seal cannot be overemphasized.

The container closure system design and integration of the bottle, closure, liner, and seal have a major impact on permeation. The repeated opening and closing of a container can create wear on the closures and liners. Removing a heat seal can leave deposits on the container that impact the seal integrity of the closure and liner with the bottle. The permeation results with the seal in place and with the seal removed and closure on the bottle were different. The standards were, therefore, revised

Table 35-3. Information on Typical Materials of Construction for Bottles

Material	Chemical Name	CAS	21 CFR § Reference
HHM 5502 BN	High density polyethylene ethylene hexene copolymer	9002-88-4	177.1520 (c) 3.2a, use conditions B through H
Ampacet white Master Batch	Titanium dioxide extended in low density polyethylene	9002-88-4 1317-80-2	178.3297 177.1520 177.2010

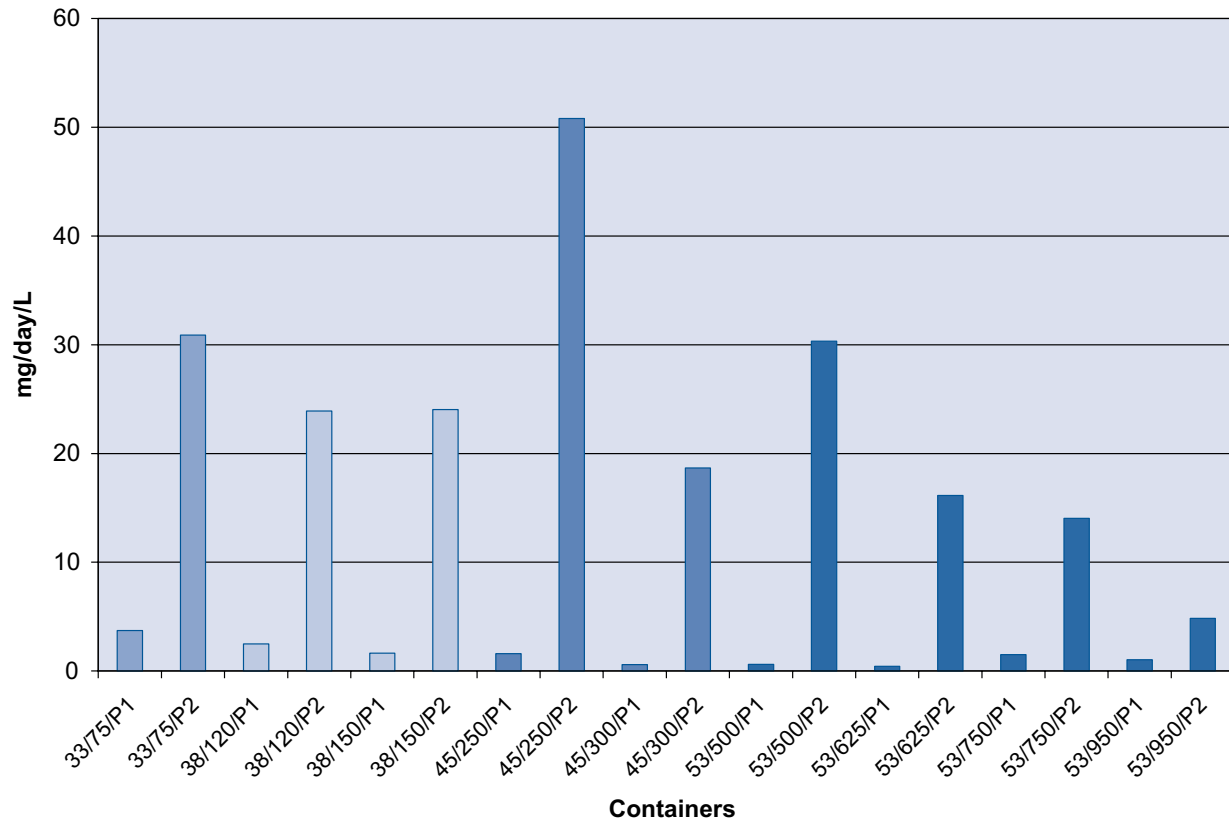


Figure 35-2. USP <671> Moisture Permeation—Bottles with seals (P1) and with caps with liners (P2).

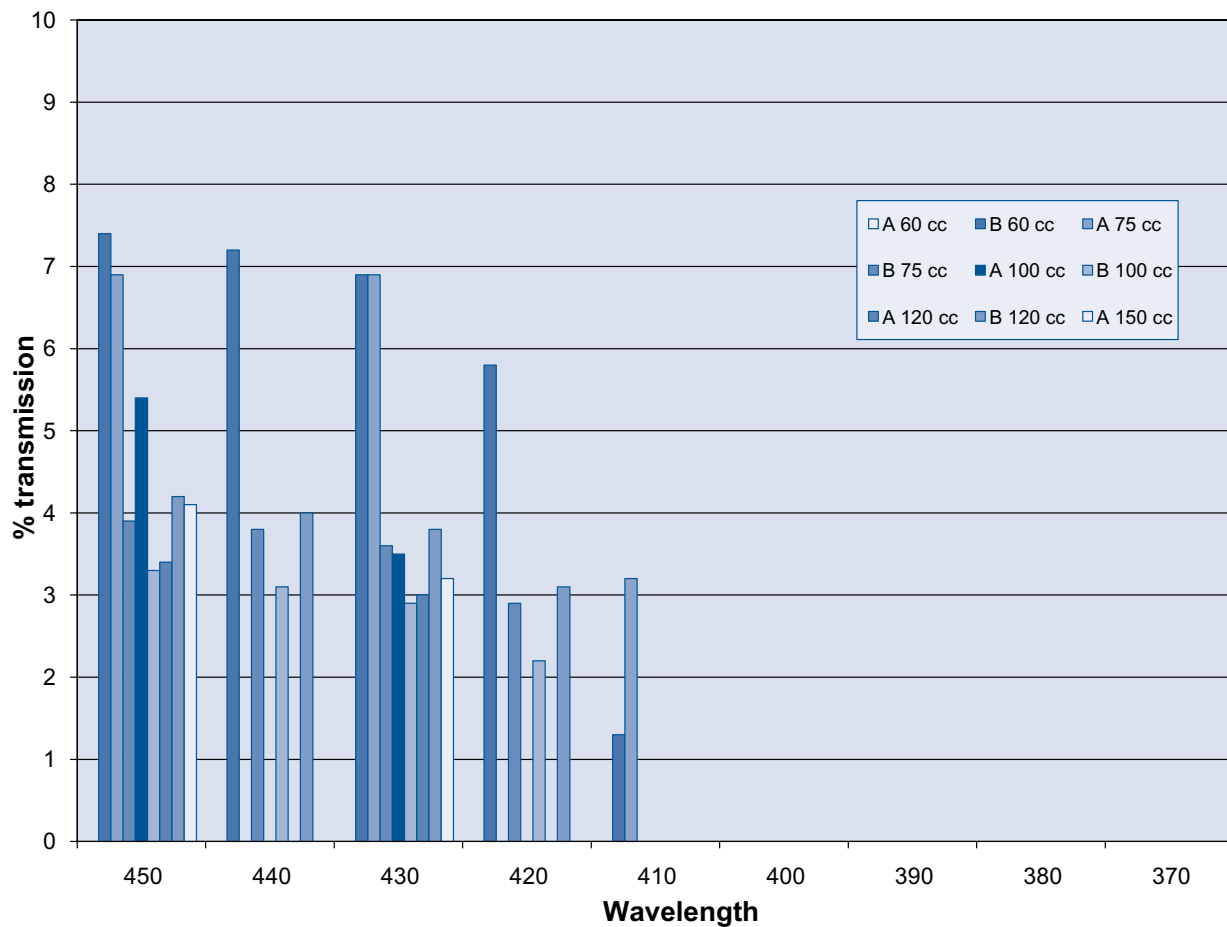


Figure 35-3. Light Transmission of HDPE Bottles. USP limit: nmt 10%.

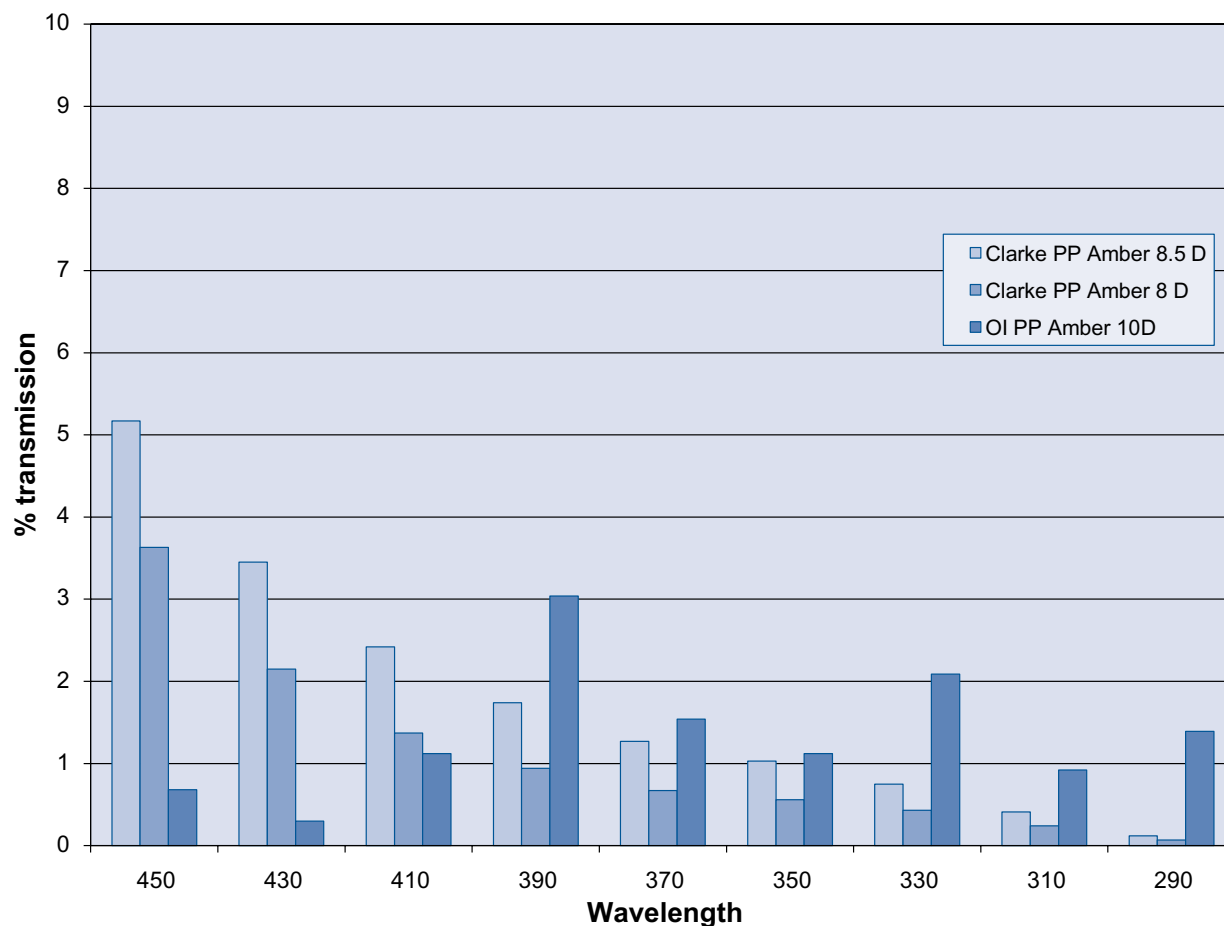


Figure 35-4. USP <661> Light Transmission Prescription Vials. USP limit: nmt 10%.

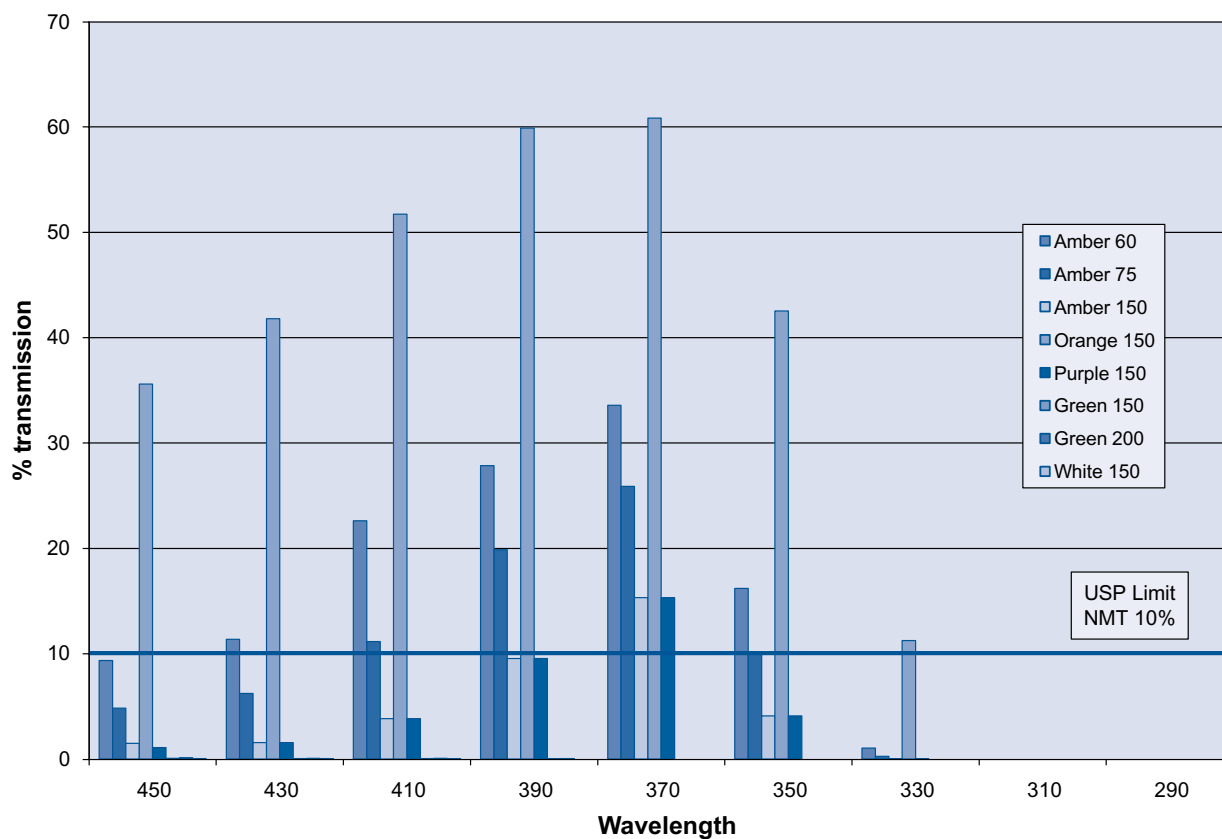


Figure 35-5. Light Transmission PET Colored Bottles. USP limit: nmt 10%.

to provide for testing under both conditions. Testing of the condition with the seals removed is similar to that for a prescription vial (see Fig. 35-2).

It is important to note that the largest and smallest containers are not really bracketing the permeation data, and, therefore, testing the largest and smallest container on stability will not always incorporate the worst case results. It is more meaningful to test the containers, select the worst case, and conduct stability studies on product packaged in this container. As an added benefit, testing on the worst case scenario would significantly reduce the time and cost of conducting stability studies.

Auxiliary Components for Solid Oral Dosage Forms

In addition to bottles and closures, solid dosage forms may be packaged with pharmaceutical filler or desiccant.

Cotton, Rayon, and Polyester Coil are used in bottles of solid oral dosage forms to prevent tablets from breaking during shipping and handling. Filler, or pharmaceutical coil, is available as cotton, rayon, or polyester. Being of plant origin, the Cotton Pharmaceutical Coil is bleached and treated to remove plant residue and other impurities.¹⁷ Rayon Pharmaceutical Coil is essentially rayon fibers, a fibrous form of bleached, regenerated cellulose. Polyester Pharmaceutical Coil is a synthetic polymer that may be treated with various chemicals. It is difficult to identify which coil is used, since all pharmaceutical coils closely resemble each other. There is an identification test that can be conducted, where a drop of dye turns a different color and polyester will burn differently from cotton and rayon. The standards for these materials are provided in USP <670> Auxiliary Packaging Components.

There are advantages and disadvantages of each pharmaceutical coil. Purified Cotton should not be used, if it contains a bluing agent or if there are traces of bleach, which can react with the product. Rayon Pharmaceutical Coil has been found to be a potential source of dissolution problems for gelatin capsules or gelatin-coated tablets, resulting from gelatin cross-linking, due to the production of furfural. Both Purified Cotton and Rayon contain moisture, which can migrate to the product. Polyester contains essentially no moisture. Most of the reactive chemicals have been removed. Polyester filler should be provided by a reputable supplier, where it is known to never have been intended for use as carpet fiber and treated with other chemicals. The finish on fibers used for processing should comply with FDA food contact regulations. (See Table 35-4—Comparison of Pharmaceutical Coils.)

Desiccants are used to reduce the amount of moisture in the head space of the bottle. Desiccants should be removed, once the bottle is open, or they can over absorb moisture and desorb moisture to the product. Only desiccants manufactured for contact with drugs should be used. The shape of the desiccant should

be clearly distinguishable from the shape of the medication to prevent accidental consumption, particularly by the visually impaired. The most common bottle desiccant is filled in a small HDPE canister, which easily fits in the bottle. The amount of desiccant depends on the size of the bottle and is measured in grams.

Desiccants should not be placed in the bottle under the tablets. This practice is detrimental, as moisture is pulled through the tablets, and the tablets act as desiccants for the desiccant. The desiccant should be placed in the bottle on top of the tablets, where it can perform its intended task and be easily discarded when the closure is removed. Desiccants should be stored in a clean, dry place and need checked to determine they have not been exposed to excessive moisture prior to use. An alternative to using a separate desiccant is to use a multi-layered bottle, where the desiccant is incorporated into the inner layer of the bottle.

Oxygen Scavengers are used to remove oxygen molecules from the head space containers of products sensitive to oxidation. Oxygen scavengers can be added to the container closure system or can be incorporated into the container itself. These auxiliary components are available as packets, canisters, or films. Oxygen is removed from the container through irreversible bonding of the oxygen molecules with the oxygen scavenging chemicals. Oxygen scavengers help extend the efficacy and shelf life of oxygen sensitive drug products. Oxygen scavengers and desiccants are also available as a combination product capable of removing both oxygen and moisture.

COMMERCIAL CONTAINERS VERSUS PRESCRIPTION VIALS AND STABILITY

Drug companies and some repackagers conduct stability studies on the product stored in the market containers. The studies conducted by the drug company are used to establish the expiration date in those containers. For an NDA, 6 months of accelerated testing are provided, and, for an ANDA, 3 months of accelerated testing are provided. If the product passes the accelerated testing, it is approved with a 2-year expiration date for the labeled conditions of storage. As long as a repackager uses the same container as used for the stability studies, they can transfer the manufacturers' expiration date.¹⁸ When the standards in USP <671> were first crafted, they were based on a usual dose for 30 days, and the limits for the commercial container with the heat seal and the limits for the drug stored for 30 days without a seal were established. Since that time, dispensing practice has changed, and many prescriptions are now dispensed with 90 tablets for a 3-months' supply, with a discard date of 1 year, not to exceed the expiration date.

UNIT-DOSE CONTAINERS

Unit-dose containers can be configured as blister packs, pouches, strip seals, cups, or blister cards. Commercially, drug

Table 35-4. Comparison of Pharmaceutical Coils

Test	Cotton	Rayon	Polyester
Residual Hydrogen Peroxide Concentration	NMT 50 ppm is found using either method	Not applicable for rayon	Not applicable for synthetic coil
Loss on Drying <731>	Dry 5.00 g of fibers in an oven at 105° to constant weight: it loses NMT 8.0% of its weight	Dry 5.00 g of fibers in an oven at 105° to constant weight: it loses NMT 11.0% of its weight	Dry 5.00 g of fibers in an oven at 105° to constant weight: it loses NMT 0.5% of its weight
Residue on Ignition <281>	NMT 0.20% on a 5.0 g sample	NMT 1.50% on a 5.0 g sample	NMT 0.5% on a 5.0 g sample
Water-Soluble Substances	The residue weighs NMT 0.35%	The residue weighs NMT 1.0%	Not applicable for synthetic coil
Fatty Matter	The weight of the residue of a 10.00 g sample does not exceed 0.7%	The weight of residue of a 5.00 g sample does not exceed 0.5%	Not applicable for synthetic coil
Possible additives of chemicals which could react	Bleach and dyes	Furfural	Antistatic Agents

companies and repackagers use unit-dose blisters or small unit dose foil pouches. The materials used by drug companies and commercial repackagers are a printable foil/polyester lidstock. The blister can be constructed of polyvinyl chloride (PVC), PVC/polyvinylidene chloride (PVDC), PVC/Aclar, or cold form foil, with PVC being the most permeable and cold form being the least permeable. The commercial small foil pouches are created not from strips, but from wider sections of foil, and the units are cut from the web after they are sealed. The best protections for the product, from light and moisture, the “tightest” containers are the foil-to-foil pouches and the cold-form foil-to-foil blisters. The disadvantage of these configurations is that you cannot see the product to verify its appearance and identity. As part of the development of the USP <671> for unit-dose containers, hospitals and pharmacies in the area were visited and unit dose containers from a wide range of configurations were selected for testing.¹⁹ In some cases, the permeation was so high that the testing had to be suspended before the test period ended. A polystyrene clam shell, polyethylene bag, and PVC cup with a paper lidstock exceeded 20 mg/day.²⁰ In other cases, such as the foil-to-foil strip seals, the testing was comparable to the tightest commercial packages. Based on the results, some of the container manufacturers made substantial improvements in the materials or designs (Fig. 35-6).

Blister cards are sometimes used in the pharmacy setting, where they are packaged on demand, pursuant to a prescription, for assisted living facilities. They contain a dosing regimen of 30 tablets packaged in a PVC blister sealed to a card. Standard heat seal cards, where the blister adheres to the card are available in 30-day, 31-day, 32-day, 60-day, 62-day, and 90-day formats. The PVC layer can be various thicknesses; however, for best protection, it should be 10-mil thickness. The usual failure of these systems is due to the inadequate thickness of the film or an insufficient seal. The blister is set in a template, the tablets are hand filled and the pharmacist places the lidstock on top and sets it in the sealing apparatus. The amount of time it is heated is subjective. The process is not usually validated, but the sealing can be checked by sealing an empty card,

letting it cool, and pulling it open. The lidstock should adhere to the blister. The USP <671> Containers—Performance Testing, Permeation studies can be performed in a pharmacy setting to verify the performance. It is also important in a pharmacy setting to store the packaging materials in a cool dry place so the materials will not pick up moisture and pass it to the drug during and after packaging. Since small quantities are packaged at a time, the impact of permeation under long term storage is lessened. The advantage of these systems is patient compliance, which can outweigh other considerations (Fig. 35-7).

Strip seal pouches and cups are more commonly used in hospital settings. Strip seals (unit-dose pouches) are formed by two rolls of material fed between two sealing wheels. The backing of the strip seal is paper coated foil, but it can be plastic as well. There are a large number of materials that have been used in the past. The sealing temperature, wheel alignment, and the cleanliness of the sealing wheels are the most critical aspects. This machine requires more attention than the blister card heat sealing machine. Again, the materials should be stored in a clean dry place, and the packaging operation should be done in a dry area.

Unit-dose cups are used in hospitals for both liquids and solids. The cups can be manufactured of a number of materials and the lidstock can be paper or foil. The lidstock is attached with an adhesive that is coated on the lidstock. These containers have improved, since they were first launched, and have the advantage that they can easily be filled on demand in the hospital pharmacy and transported to the patient's room without the chance of product “mix-up” or loss.

NON-STERILE LIQUIDS

Non-sterile liquids encompasses a wide range of products, including, but not limited to, oral liquids, eye drops, nasal and throat sprays, oral powders for reconstitution, and topical preparations. Aqueous liquid products include suspensions (including magmas or milks), elixirs, emulsions, syrups, solutions, and powders for reconstitution. Liquids are often the preferred oral dosage form for pediatric and geriatric consumers and those

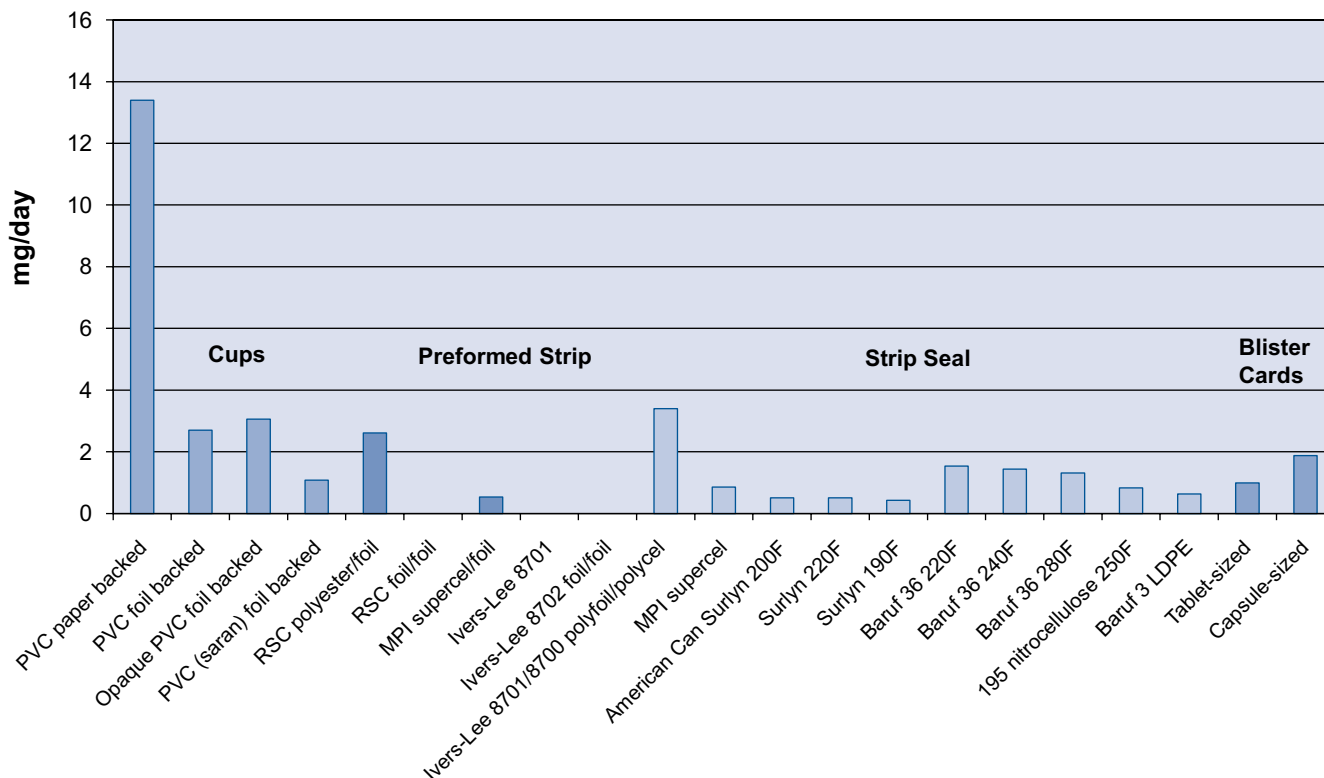


Figure 35-6. USP <671> Moisture Permeation of Pharmacy Unit-Dose. Class A limit: 0.5 mg/day. Class B: 5 mg/day.

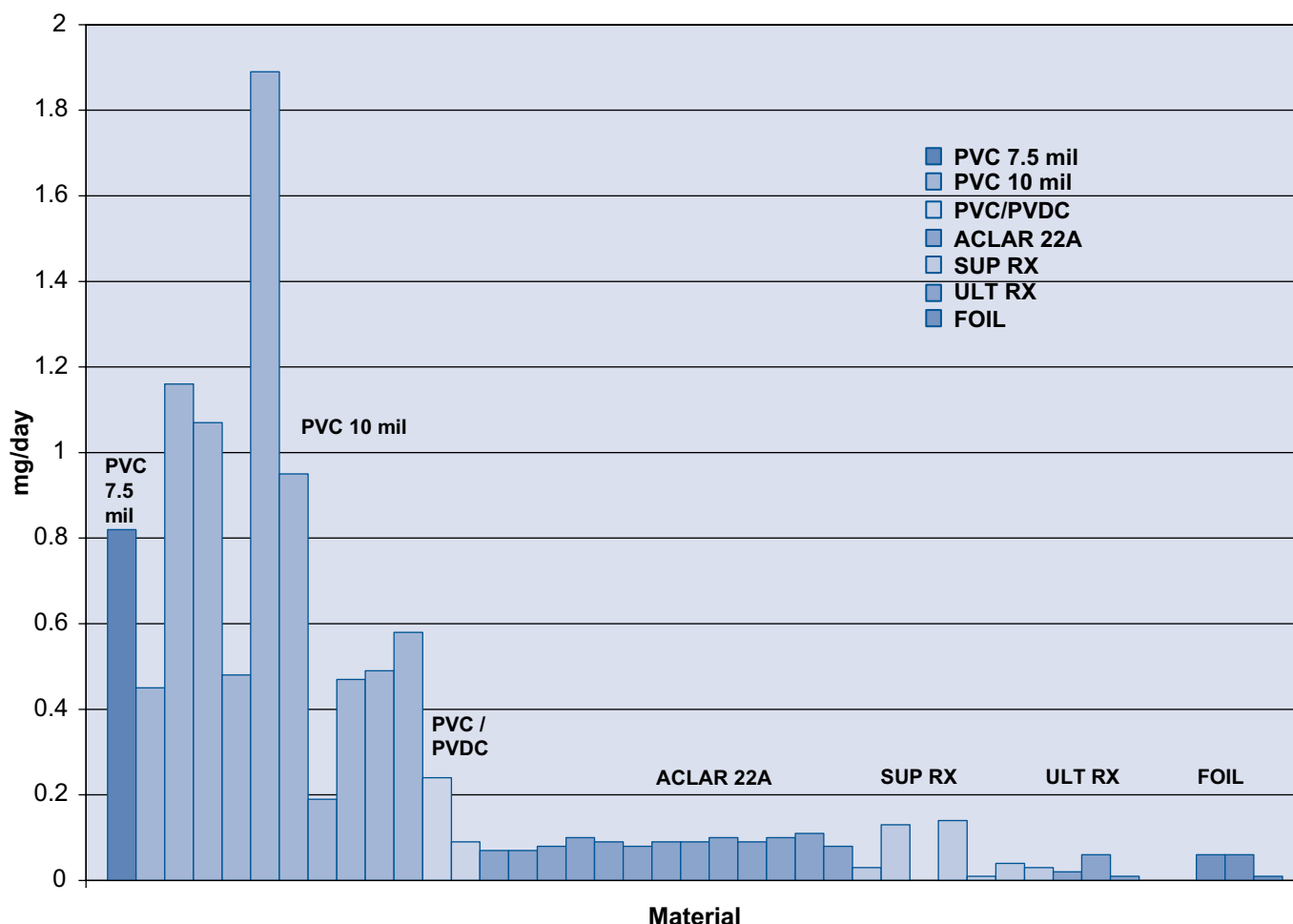


Figure 35-7. USP <671> Moisture Permeation in Blister Packs.

who struggle with swallowing solid oral products. Pursuant to a prescription, pharmacies have had to compound and package oral liquids from tablets and capsules.

The development of suitable and compatible packaging for liquid dosage forms requires greater planning than for solids. The potential for interaction between the packing materials and product are greater, but the health risk is less than that for inhalation and injection products. Challenges encountered with the packaging of liquid products involve the fill of unit dose products, water weight loss, stability, potency, and protection from microbiological contamination. To ensure the liquid product meets its label claim, testing to confirm uniformity of fill and unit dose sampling is required. In many cases, these products do not require a CRC. The fit of the cap is essential to prevent leaking. The thread area for liquids is higher than the SP400 used for tablets. It is important that the top or the bottle fit snugly against the cap liner to affect a suitable seal to prevent leakage.

The main performance concern with aqueous liquids is water loss from the drug and container. The containers should pass USP <661> Containers—Plastic for the type of plastic used. The Buffering Capacity test should also be performed. In most cases, this testing demonstrates the suitability of the container from a physicochemical point-of-view. Polyethylene and polypropylene are common plastics for these containers. The container closure system for a liquid dosage form must be developed based on the product's sensitivities, including, but not limited to, moisture, light, gases, ions, and oxygen.

Additionally, the component composition, risk, and rate of migration under storage and shipment conditions must be considered. Although the primary source of migrants is the imme-

diate container, potential sources of migrants in liquids include inks from labels, additives from adhesives on labels, additives to plastic coatings, such as the inner layer of foil pouches, solvents, plastic additives, vanillin from paper and cartons, and phthalates from shrink sealed bands. It is possible for the product to fail specifications, pose a health risk, and necessitate a recall, if the container components have not been properly tested and the product is not protected (see Fig. 35-8).

For liquid products, the permeation limit is based on the upper limit for assay of the product. Moisture permeation testing conducted on containers intended to package liquid products calculates the moisture permeability of the container as a percentage of water weight loss per year. If the product were to maintain its stability, the potency or assay could exceed the upper limit, if the weight loss of the liquid exceeded 5.0% per year for 2 years, resulting in increased potency through loss of water. If the product degrades while the container loses water, the results give a misleading appearance of stability.

NON-STERILE LIQUIDS—CONTAINING ALCOHOL

Some liquid drug products contain alcohol. Bottles, closures, inner seals, liners, and glue can all interact with alcohol containing products. For products that contain alcohol, the potential for leachables and the risk of interaction with the container are slightly greater than for those that do not contain alcohol. This should come as no surprise, since alcohol is one of the solvents used to test for extractables to determine the suitability of polyethylene and polypropylene.²¹ Testing by USP <661> and <671> are applicable, but compatibility testing is required as well. The results for aqueous and alcohol extractables may demonstrate that a HDPE or LDPE bottle is not suitable. PET

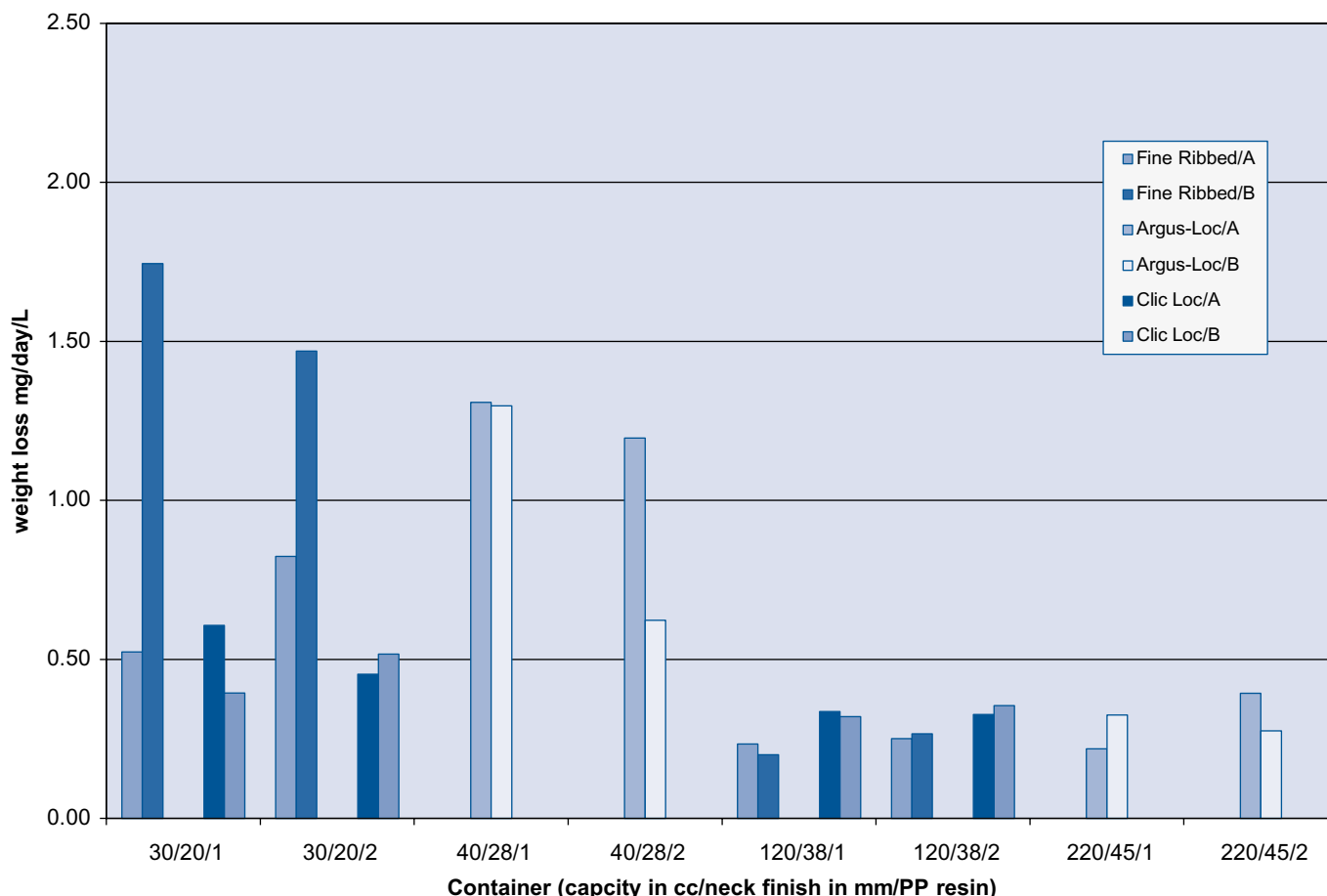


Figure 35-8. USP <671> Containers—Performance Testing. Moisture Permeation Water Weight Loss Percent per Year in Liquid Products.

and PETG bottles are used more often for alcohol containing products, because they are less prone to interact. Compatibility testing with the cap, liner, and glue should be conducted as well. Qualification of the packaging components for these products would require testing and evaluating the extract with alcohol or alcohol/water mixtures at the level at which alcohol would be in the product.

NON-STERILE LIQUIDS—OIL-BASED

Bottles, closures, inner seals, liners, and glue can all interact with an oil based product. Mineral oil products are notorious for interacting with polyethylene, polypropylene, pulp liners, and glue. Use of these materials, even in the manufacturing process, can result in contaminants and loss of drug from transfer lines and holding tanks. Testing by USP <661> Extractables for liquids does not provide any assurance that the components manufactured of these materials will perform satisfactorily with oil based products. Actual contact of the container component with the drug product and placebo (or primary vehicle) under the anticipated conditions are the only meaningful studies that can be used to qualify the components. USP <671> Permeation studies are not applicable.

For these products, the least reactive container is glass or PET. This material is also used to package oil-based food products, such as vegetable oil and salad dressing. The potential migrants are substantially less than those in polyethylene or polypropylene. Polypropylene closures with reduced additives are the only option at present. Most standard closures contain too many additives that can migrate, even if they are not in direct contact with the product. The foamed polyethylene liners have been demonstrated compatible, but, again, testing is necessary to qualify these components for use with a particular product.

Auxiliary Components for Oral Liquids

Dosage delivery devices can include components, such as bulbs, glass and plastic pipettes, closures, glass and plastic droppers, dropper tip caps, optical droppers, eye cups, spoons, blisters, and oral syringes. These components are referred to as liquid medication dispensers, when they are sold independent of a drug product (21 CFR §880.6430). When they are sold as the packaging components for the drug product, they are referred to as dosage delivery systems. The devices are used to dispense liquid drugs and are in contact with the drug during that time. The devices have limited contact with the drug product, except when they are part of the immediate container closure system, such as dropper assemblies. Each component should meet the appropriate USP test based on its composition.²² Suitability and compatibility testing is the responsibility of the end user, not the device manufacturer.

Development and calibration of the dosage delivery systems is a critical aspect of the packaging configuration. The delivery system must assure the proper dosage is delivered to the consumer. The properties of the drug and degree of accuracy required for dosing should be considered, when developing the delivery system. Dosage delivery devices sold with the drug product are custom-made and calibrated for the particular drug product. Pipettes and oral syringes are branded with inks or tapes to indicate calibration lines. Oral syringes provide the more accurate dosing than spoons or cups.

CREAMS, LOTIONS, AND OINTMENTS

Cream, lotion, and ointment products are packaged in glass, PP or PE jars, or plastic-coated aluminum or PE tubes. These products pose a packaging problem, as they tend to shear and

lose viscosity during the packaging operation and tend to migrate during storage. Creams and lotions are aqueous with surfactants and may contain parabens, perfumes, or essential oils, which also tend to migrate, especially into polyethylene. The assay of the parabens tends to drop within 24-hours of packaging but stabilizes thereafter. Creams may set-up differently and have different stability in glass versus plastic; these factors should be considered when selecting packaging materials. A variety of closures can be selected for use with jars. However, tubes are sold as a set with the cap in place. The tube neck finish and cap can be of various designs. Tubes are provided with various closure fitments. In some cases, the tubes are provided with a standard screw cap, with or without a heat seal. In other cases, the tubes are provided with a metal seal, which is punctured by the cap when it is opened by the consumer. The contact layer of the container can impact the product stability and should be evaluated and controlled, as well as influence the choice of materials.

As stated previously, USP testing <660> or <661> needs performed for glass and plastic, respectively. If the product is aqueous-based, USP <661> Buffering capacity testing needs performed. This testing is not required for ointments, because they do not experience any weight loss. The micro challenge test is required for all aqueous-based lotions and creams.

TRANSDERMAL PRODUCTS

FDA classifies transdermal products as topical delivery systems that are self-contained, discrete forms, designed to deliver drug through intact skin or body surfaces into systemic circulation.¹ Transdermal products originally developed through the course of clinical practice, when liquid or ointment drug products were applied directly to gauze or dressings and then affixed to the patient's skin for absorption. For transdermals, it is sometime hard to tell where the product stops and the packaging starts. These products are typically packaged in a continuous process at the time they are manufactured, and repacking is not common. Transdermal packaging includes a protective liner, a contact adhesive, a drug reservoir, which may or may not include a rate controlling membrane, and an outer barrier.² Transdermal products are packaged individually into foil or foil-lined pouches, but thermoformed trays have also been used. Ethylene vinyl acetate is a common choice for the construction of the dose controlling membrane. The protective barrier is often composed of PVDC, aluminum foil, or metallized film. The protective liner is constructed from LDPE and is removed immediately prior to the application to the consumer's skin. Polyurethane monomers can be used effectively, due to their high permeability enabling the diffusion of the drug product across the membrane into the consumer's skin.

Extraction and toxicology studies should be conducted early in the packaging development phase. Transdermal products have direct contact with the consumer's skin, often for a considerable length of time. Extractable studies are needed to determine which chemicals have the potential to migrate from the packaging materials to the drug product and, subsequently, to the consumer. Toxicology studies are then conducted on the identified migrants to determine if the levels of exposure are safe for consumers. Additionally, studies are needed to determine if the proposed packaging configuration is capable of protecting the product from losses.

INHALATION AND NASAL SPRAYS

These products are considered high risk, due to the route of administration and the sensitivity of the patient population. Therefore, the container requires more stringent physicochemical and performance testing. Unlike the reduced testing requirements for oral solids and liquids, these components may require testing of every lot, until the company can demonstrate the components are being manufactured under tight controls. The degree of testing is based on which components have drug or patient contact. The components should all meet or exceed

the requirements listed in USP <381> Elastomeric Closures, <661> Plastics, and the requirements of USP <87>, <88> Biological Reactivity Testing for drug contact components. Since different routes of administration can target different organs and have different rates of absorption, impurities and migrants are of greater concern with these products. Migrants that may be acceptable for an oral liquid may be unacceptable at the same level for nasal sprays or inhalation products.

Spray devices are complex and may contain multiple components, including, but not limited to, rubber, stainless steel, PE, PP, and, in some cases, polycarbonate and styrene butadiene. Some of these materials may not have Drug Master Files and may not be medical grade. In those cases, extra vigilance is required on the part of the drug manufacturer to control the critical attributes of these materials. Testing and knowledge of the inks, dyes, resins, rubber, and thermoform resin compounds, stainless steel quality, and the impact of changes must be understood and incorporated into the packaging design. Components in direct contact with the drug must pass suitability and compatibility testing. Testing of the pump components in contact with the drug product and vehicle under stress conditions will demonstrate the potential for migration.

The chemical nature of extractables may need determined, and a toxicological evaluation be performed. Those materials with no contact may not require extractables testing. Those with patient contact should meet indirect food additives requirements and be latex free. In addition to extractables, it is also important to consider and set limits for particulates that can arise from the rubber and plastic components.

The dose is delivered by the device, and delivered amount, spray patterns, and particle size play a part in qualifying the device for use with the individual. The performance of the dosing system must be evaluated over the predicted expiration period under the conditions of use. To ensure delivery of the prescribed dose of medication, testing needs to be performed on the efficacy of the device, including pump delivery, spray pattern and plume geometry, spray content uniformity, and droplet size distribution. Detailed information can be located in the FDA Guidance for Nasal Spray and Inhalation Solution, Suspension, and Spray Drug Products.²³

Some products for inhalation are packaged in LDPE nebulers. These containers are designed for used with a nebulizer. Nebulers are automatically packaged using an aseptic form-fill-seal (blow molding) system for rigid containers or a seal-fill-seal system for pouches. These containers are permeable and very prone to weight loss and absorption; therefore, they must have low levels of extractables in the resin and be stored in the appropriate conditions. They have been shown to absorb chemicals from paper, ink, glue, and cartons. In the United States, they are packaged in a foil overwrap that contains the label information. The inner layer of the foil must be free of chemicals that can migrate, such as LDPE that is additive free. Additional information can be imposed on the bottom of the nebulers. There are Mylar-backed labels available that help reduce the migration of inks and paper chemicals. They must be tested on the component to verify they are safe and effective for use. These containers are susceptible to product misidentification. Proper precautions need taken to maintain identification, strength, lot number, and expiration date.

LARGE AND SMALL VOLUME PARENTERALS

Before the switch to plastics in the 1970s, glass was the most commonly used packaging material for drugs. It was used to fabricate containers for oral liquids, solids, parenteral products, cartridges (prefilled syringes), ampules, and droppers. Glass is still the preferred material used to manufacture containers for many droppers and injectable drugs. Glass is easily cleaned and sterilized. It is transparent, relatively inert, resistant to most liquids, and impermeable to gases. There is no moisture permeation through the container; however, the integrity of the closure's seal to the container can be a source of loss. Stress

cracks can occur, and, sometimes, glass can react with products and result in flaking into the product.

The physicochemical standards for glass are provided in USP <660> Containers—Glass. Glass is typed based on the hydrolytic chemical durability and composition. Conventional inorganic glass consists of a mixture of crystalline oxides, carbonates, and others melted by heating to a viscous liquid state. Glass consists mainly of silicone dioxide, but the composition varies widely within each type. Medical glass can be composed of either borosilicate glass or soda-lime glass. Properties, processing, and melting temperatures vary, depending on the composition of the glass. Glass vendors and their respective formulations are not interchangeable. Formulations can meet the same Type I glass standards, but they are chemically different. Durability is determined by the ability of the glass to resist the release of soluble mineral substances into water. Sometimes, the inner surface of glass containers can interact with neutral parenteral preparation. Some glasses are available with a silicon dioxide coating on the inner surface, which reduces the risk of interaction and migration, thereby increasing the chemical resistance of the glass. Factors, such as composition, treatments, forming, or secondary processes, can impact the surface quality of glass components. The products can be rendered unacceptable for use, when small amounts of alkali and other ions leach into the product from the container. The intensity of the chemical attack in comparable conditions depends on the nature of the product, the composition of the glass type, the surface treatment, and the method of manufacture.

All three types of glass are available in clear and amber light resistant styles. Testing for light resistance is provided in USP <671> Light Transmission. Type I glass is preferred for use with sterile injectable products, due to its high chemical resistance. Type II glass is made from a high-grade commercial soda-lime glass. The inner surface is treated with sulfur dioxide to dealkalinize the glass. Type II glass can be used for most neutral aqueous or acidic preparations, which can include parenteral products. Type III glass has a higher level of resistance to chemical attack than common commercial glass. It can be used for solid or oral liquid dosage forms and some parenteral products. Some glasses are available with a silicon dioxide coating on the inner surface, which reduces the risk of interaction and migration, which increases the chemical resistance of the glass. Containers with this coating are used to package products of high sensitivity, aggressive buffer systems, high pH, radiopharmaceuticals, and complex agents or those products with a tendency to adsorb to glass surfaces. Glass can degrade as the glass absorbs products. This type of degradation is known as sloughing or weathering. These physical attacks on the glass by the products they package can result in sloughing and delamination, which produce glass particles in the product. These potential events can be controlled by the manufacturer by their selection of composition, secondary processes, and forming processes. Additional information on glass is provided in USP <660>.

Most small volume parenterals are still packaged primarily in glass vials, meeting the standards provided. These vials are sealed, using rubber stoppers and aluminum seals. Although natural rubber was used successfully for many years, it fell out of favor, due to the concern about latex allergy. Even though hard latex rubber is processed to the point at which the allergen was not detected, companies changed to other forms of rubber, such as butyl rubber.

There are different formulations of rubber, and it is best to consult with the manufacturer to determine the best form to use based on the pH of the product and the processing, such as terminal sterilization. Rubber components must meet the 21 CFR requirements, as well as those of USP <381> Elastomer Closures. Rubber is also available with treatments, such as siliconization. It is important to determine if these treatments would cause problems if the chemicals migrated to the drug product. Stoppers are also available with different styles, including those designed for lyophilization. USP also provides

requirements for Biological Reactivity. The colors added to the rubber must meet 21 CFR indirect food additive standards. In some cases, the method of manufacture may be an issue, such as for carbon black. Some methods of manufacture for carbon black preclude it from contact with drug products. If the material meets the limits for *in vitro*, then *in vivo* testing may not be required. There are various colors and styles of aluminum seals. The style of the vial must be considered when the seal is selected, so it can crimp properly on the vial. Cosmetic considerations may apply. For example, a certain color may indicate a certain drug, or a certain design may be preferred for ease of use in different environments, such as an operating room.

Plastic bags and vials have been used successfully to package injectable products or implants. The bags may be manufactured of HDPE, PVC, or LDPE; the vials have been manufactured of PP or multiple layers; and PET trays have been used to package implants. The materials must meet 21 CFR requirements, the applicable sections of <661>, and Biological Reactivity Testing USP <87>, <88>. The bags are prone to the same problems discussed for liquids in permeable plastics. Permeation and excessive weight loss and migration of chemicals into the product can be significant. If the bags are stored unprotected at an elevated temperature, the problem is exacerbated. The more permeable containers are packaged in a foil overwrap. Again, as discussed, the foil overwrap should not contain chemicals that can migrate into the product. If the water weight loss is too great in an IV bag and a drug is injected into it, it is possible for the excipients or drug to fall out of solution. The solutions become super saturated, and crystals form. In some climatic conditions, it is still, therefore, preferred to use glass IV bottles in which the weight loss is not problematic. The actual product in the proposed package configuration should be evaluated under actual conditions of use, including shipping and storage, to ensure product integrity throughout its shelf life.

QUALITY CONTROL

Characterization analysis should be conducted to provide assurances that the proper polymer type is used and that the physical parameters, which could impact the product efficacy and the container configuration suitability for use, are considered. For drug product manufacturers and commercial repackagers, techniques, such as infrared spectrometric analysis, density, melt-flow, and thermal and rheological tests, can assist in providing the necessary assurances of identity and performance. For pharmacists, the quality of the incoming components can be established through comparisons with control samples, weight, physical measurements, fit tests of caps with bottles, and verifying purchase information. Plastic parts or packages should be inspected routinely, on an incoming basis, for dimensional and attribute variability against statistically accepted sampling plans.

OVERVIEW OF TEST METHODS

IDENTITY

Though this may seem obvious, a basic identity test is critical to verify the material received corresponds to the material requested in the purchase order. The identity test can be the most important test performed. There have been cases in which the incorrect material has been shipped. Critical inspection of the materials and accompanying certificates of analysis from the supplier is essential to verify the receipt of the correct material. Errors can occur with the cap liner and heat seal, the neck finish, bottle weight, dropper assembly, and pharmaceutical coil. A fitness test should be performed for closures and dropper assemblies with the bottles. Material treatments must be verified, for example, whether a rubber stopper has been siliconized or not or whether a blister foil has the correct plastic coating. USP <661> contains identity tests based on materials

of construction. It is important to note that there is currently no reference standard for linear low density polyethylene (LLDPE). LLDPE passes infrared spectroscopy but fails differential scanning calorimetry.

EXTRACTABLES (MIGRANTS)

Extractables, migrants, and leachables are all interchangeable terms. The risk of drug failure from migrants in an HDPE bottle and a PP closure are extremely low, if the source of the bottles is from the United States or India. Since the implementation of USP <661> Plastics, most of these bottles, if tested, would pass. These same bottles may perform well with aqueous liquids, if they pass USP <661> and buffering capacity. Many of the bottle manufacturers conduct this testing and can provide the results at the point of sale. When contacting a closure supplier, it is important to inform them of the end use, so they can properly advise regarding material selection. The end use is an important determining factor in the selection of closure and liner materials based on the supplier's knowledge of the additives in their materials.

Oil based products require extra precautions and actual testing with the placebo and product to ensure compatibility and stability. The USP <661> Extractable testing is not applicable for oil-based products, because the extracting solvents do not replicate product contact. Oil-based products are notorious for leaching large amounts of chemical additives from the plastic and pulp. PET or glass is a better choice for the containers for oil-based products than polyethylene, polypropylene, and PVC.

For inhalation, injection, and implant products, the extractables testing is more critical, and the requirements are more stringent. It is important that suppliers understand the end use of the materials. Selecting materials with a minimum of leachable additives helps provide superior protection to the product. It is important to bear in mind that all packaging components in direct contact with the drug or device have the potential for chemical migration and should be carefully selected.

If the nebulas (LDPE form-fill-seal unit dose for liquids) are packaged with a foil overwrap, the inner layer of the foil should be free of additives. LDPE is available to coat the foil, to prevent additives migration. There are many different formulations for the types of glass. Some are more reactive than others, and it is important to learn the difference and use this information, when selecting appropriate containers for products.

LIGHT TRANSMISSION

When trying to protect the drug from light, the container providing optimum protection is a sealed foil pouch or cold-form-blister. The next best protection is afforded by containers composed of amber glass, white HDPE, or amber PP. Many amber blisters, amber PS bottles, or PET bottles of various colors will not meet the USP <671> Light transmission limits of not more than 10%. Clear blisters are considered, by FDA, to be light resistant, when they are stored in cartons. Some drugs are highly light sensitive and require tighter limits than the current USP limits of not more than 10%. These requirements should be considered, when selecting suitable containers for these products.

PERMEATION

Moisture permeation data has been extensively provided in this chapter. For both tablets and aqueous liquids, permeation is a critical issue. Moisture permeation is measured two ways, based on dosage form. For solid dosage forms, it is measured as moisture moving into the container from humidity in the air or from the container itself. For aqueous liquids (calculated as weight loss), it is measured as water migrating from the product to the container's exterior and into the atmosphere. Glass, HDPE, and PP provide sufficient protection against weight loss. Protecting against weight loss is critical to avoid overconcentration of the drug product. Unfortunately, many liquids, such as ophthalmic, otics, eye drops, and nebulas for inhalation products, are all packaged in LDPE. LDPE is considered a semipermeable plas-

tic. The integrity of the container system and wall thickness must be sufficient to protect the drug from excessive water loss over the shelf life.

SUMMARY

When selecting a container closure system, it is important to know the regulations, the product's critical attributes, and the ability of the container to meet the efficacy, stability, storage conditions, and delivery requirements of the drug product. For each product class, there are certain known critical attributes. It is also critical to know the properties of the materials of construction. (Refer to Table 35-5—Summary of Information on Materials of Construction.) For tablets and powders, there are standard HDPE bottles and PP closures that can meet the requirements for protection from light and permeation. There are multiple sources of closure heat seals and liners, which can meet the product requirements. Based on the product, pharmaceutical coil can be added to reduce breakage of friable tablets, and desiccant can be added (on top of the tablets) to remove humidity in the headspace of the bottles. If pharmaceutical coil is needed for highly moisture sensitive products, Polyester Coil is recommended, because it has essentially no moisture. Once the container is opened, the desiccants and coil should be discarded. If the product is repackaged, the manufacturer's expiration date is only meaningful, if the protection afforded by the new container is the same. For prescription vials, the child resistant closures perform better than the LDPE snap caps under repeat use. There are prescription vials on the market that can protect from light and perform as well or better than the manufacturer's container with the heat seals removed.

Containers manufactured of semipermeable plastics, such as PVC, PET, and LDPE, are at greater risk. These issues are less problematic, if HDPE or PP can be used, but that is not always the case. Squeeze bottles, dropper bottles, nebulas, and pouches may all be constructed of these materials. The wall thickness or film thickness of these containers must be such that they meet the permeation limits. In some cases, the permeation can be reduced by the use of an over wrap. If used, the testing must be done with the wrap in place to demonstrate its effectiveness. As with permeation, migration of chemicals into these containers also requires additional attention. Ink from the carton and chemicals from the paper and the glue have been detected in LDPE containers.

The contact surface of glass containers is a critical determining factor to the stability of the medical products packaged within them. Chemical and physical durability indicate the degree to which the product and container interact. Chemicals from the glass can migrate or leach into the product from the container. In unbuffered solutions, the migrating ions can cause a shift in the pH. Migrants can also cause aggregation, precipitation, and degradation in the product, thereby rendering it unacceptable for use.

Different types of unit dose packaging that lend themselves to the sophistication of the packaging operation are available. Drug manufacturers and commercial repackagers have the ability to choose from PVC, PVC/PVDC, PVC/Aclar, and cold form films for blister films and a number of foil lidstocks. The manufacturers and repackagers have the design and manufacturing support required for this sophisticated type of packaging configuration. They also have the equipment and validation support necessary to form-fill and seal the blisters in such a way as to demonstrate the efficacy of the system. The manufacturers and repackagers have the ability to print the bar codes for the NDC number on the back of each blister unit of the blister pack.

Hospitals and neighborhood pharmacies have less sophisticated designs and materials at their disposal. In these cases, the products are stored for shorter periods of time and, in many cases, are packaged pursuant to a prescription. Hospitals may use multiple systems, such as strip seals, blister cards, and cups. The strip machines have a number of materials available.

Table 35-5. Summary of Information on Materials of Construction

Common name	Abbrev.	CAS #	Structure	Uses	Products
Aclar/with PVC	NA	NA	NA	Blister film	Oral solids, tablets, capsules
Aluminum foil	Al	NA	Al	Tubes	Ointments, creams, lotions (treated)
				Foil-lid stock	Solids and liquids unit dose blister, pouches
Glass		NA	Type I	Vials, syringes Drug delivery systems Droppers, Bottles	Most products, including parenteral
			Type II		Most acidic and neutral aqueous preparations, including some parenterals
			Type III		Non-aqueous parenterals, powders, excluding freeze-dried preparations, and non parenterals
Polyethylene—high density	PE, HDPE	9002-88-4	$(CH_2CH_2)_n$ branched Homopolymer or hexane copolymer	Bottles	solid oral and aqueous liquids
				Bags	Bulk pouches for tablets
				Tubes	Aqueous based lotions, ointments
				IV bags	IV solutions
Polyethylene—low density	PE LDPE	9002-88-4	$(CH_2)_n$ linear	Nebules	Liquids for inhalation
				Flexible containers	Eye drops, squeeze bottles,
				Bottles, caps, tubes, dosage delivery systems	Nasal and throat sprays, topical liquids
				Film	
Polyethylene—linear low density	PE, LLDPE	9002-88-4	$(CH_2CH_2)_n$ linear	Film, coatings	Bulk bags for tablets
				Heat seals	Inside caps for tablets, capsules and aqueous liquids, pouches
Polypropylene	PP	9003-07-0	$(CH_2(CH_3)CH)_n$ Homopolymer, Random copolymer	closures	Tablets, capsules, non-sterile liquids, powders for reconstitution
				Tubes, jars	Creams, ointments
				Prescription vials	Tablets and capsules
Polyvinyl chloride	PVC	9002-86-2	$(CH_2CHClCH_2CHCl)_n$	IV bags	IV solutions and drugs
				Blister film	Solid oral dosage forms
				bottles	Some liquids
Polyvinyl pyrrolidone with PVC	PVDE	9003-39-8	$(C_6H_9NO)_n$	Blister film	Tablets, capsules
Polystyrene	PS	9003-53-6	$(CH(C_6H_5)CH_2)_n$	Blister packs bottles	
Polyethylene terephthalate	PET	25038-59-9	$(C_{10}H_8O_4)_n$	Calm shells trays	Implant drugs
				bottles	
Polyethylene terephthalate Glycolate	PETG	25038-59-9	$(C_{10}H_8O_4)_n$	Calm shells, trays	Implant drugs
				Bottles	Alcohol and oil based liquids
Rubber	Rubber	NA	NA	stoppers	Injections
				O rings	For drums Powders, tablets, capsules-
				Bulbs for dosage delivery systems	Non-sterile liquids

Foil-to-foil strips, if properly sealed, afford the maximum degree of protection. Pharmacies packaging for rehabilitation centers or assisted living facilities may use blister cards. There are methods for sealing and barrier materials that can protect the products for the shorter times they are in use.

For all aqueous liquid products, water weight loss is a critical attribute. Unlike solids, such as tablets and powders in which the critical amount of moisture permeation is different for each product, for liquids, the limits are based on the upper limits for potency. The usual limits of 90–110% mean that, if a product

does not degrade and the expiration date is 24 months, the container must not lose more than 10% of the net contents through its walls. If the potency limits are tighter than 110%, then the limits for moisture permeation must likewise be tighter.

The packaging or repackaging of non-aqueous liquids requires more attention and knowledge of the container components. Permeation of an oil-based product is not normally an issue. Oil-based products extract chemicals from components, such as HDPE, PVC, and PP bottles, pulp liners, and LDPE and PP caps. The safest choice for bottles is PET or glass. The closures

should be manufactured with lower amounts of additives than standard closures. Liners should be tested, since they are a primary contact surface. PET and glass bottles that meet light resistance standards can be purchased.

Selecting appropriate liquid medication dispensers depends on the therapeutic dosing requirements and the accuracy of the dispenser. It is important that the dosage delivery system be calibrated to deliver the prescribed dosage of the drug product. Not all syringes deliver the amount indicated by the calibration line, not all droppers have the same drop size, and not all cups and spoons deliver the prescribed dosage of the drug product. There should be administration instructions provided for the drug product.

As with all packaging components, it is, ultimately, the responsibility of the packager to determine a container's compatibility and suitability for use in the packaging of any given product.

ACKNOWLEDGMENTS: We would like to acknowledge Ed McKinley and Mary Foster for their research contribution and collaboration over the years. We would like to acknowledge Brian Reamer for his research on light transmission of bottles, and Elizabeth Deiss for her editorial review.

REFERENCES

1. USP/NF. 10.20. *Containers, Preservation, Packaging, Storage, and Labeling, General Notices and Requirements*. USP, Rockville, MD 20852.
2. Federal Food Drug and Cosmetic Act, Food Drug and Related Law as Amended thorough December 31, 2000, May 2001.
3. ASTM D7709 - 11 *Standard Test Methods for Measuring Water Vapor Transmission Rate (WVTR) of Pharmaceutical Bottles and Blisters*.
4. U.S. Department of Health and Human Services: FDA, CDER, CBER. (1999). *Guidance for Industry Container Closure Systems for Packaging Human Drug and Biologics, Chemistry, Manufacturing and Controls Documentation*. <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm070551.pdf>. (accessed May 2, 2011).
5. USP/NF. <660> *Containers—Glass* and <661> *Containers—Plastics*. General Chapters USP, Rockville, MD 20852.
6. USP/NF. <661> *Containers—Plastics* and <671> *Containers—Performance Testing*. General Chapters USP, Rockville, MD 20852.
7. USP/NF. 10.20 *Containers, Preservation, Packaging, Storage, and Labeling, General Notices and Requirements*. USP, Rockville, MD 20852.
8. USP/NF. <661> *Containers—Plastics*. General Chapters USP, Rockville, MD 20852.
9. Reamer JT. *Project Report 73-956B: Tablet Containers, Collaborative Study*, NF/USP Drug Standards Laboratory, APHA Foundation.
10. Taborsky CJ *et al.* *Project Report—An Evaluation of High Density Polyethylene Containers for Solid Oral Dosage Forms and A determination of the Impact of Using Re grind—A Collaborative Study*. FDA.
11. Unknown. CDER completes HDPE bottle study. *Pharmaceutical and Medical Packaging News* 1997; Sept: 82–83.
12. Taborsky JT *et al.* An evaluation of high-density polyethylene containers used to package solid oral dosage forms and a determination of the impact of using regrind. *Pharm Tech.* (1997). 21(9): 116–130.
13. Okeke C, Taborsky CJ. Evaluation of USP General Chapters on Containers<661> and Containers- Permeation <671> based on high-density polyethylene study results. *Pharmacopeial Forum.* (1997). 23: (5) 4941–4949
14. Capersen G *et al.* Multiple-unit packaging for solid oral dosage forms. *Pharma Technol.* (2000). 24(3): 44–54.
15. Reamer B. *Internal Project Report: USP <661> Light Transmission Study Data*. Sponsored by SciRegs International, Inc., 1999.
16. Poison Prevention Packaging Act, Codified 15 U.S.C.§§ 1471-1477, Public Law 91-601, 84 Stat. 1670, Amended October 14, 2008.
17. USP/NF. <670> *Auxiliary Packaging Components*. General Chapters USP, Rockville, MD 20852.
18. U.S. Department of Health and Human Services: FDA, CDER. *Guidance Expiration Dating of Unit-Dose Repackaging Drugs: Compliance Policy Guide*. May 2005.
19. Reamer JT *et al.* Moisture permeation of typical unit dose repackaging materials. *Am J Hosp Pharm.* (1977). 34: 35–42.
20. Reamer JT *et al.* Moisture permeation of newer unit dose repackaging materials. *Am J Hosp Pharm.* (1978). 35: 787–793.
21. Taborsky J, Sheinin EB. A Critical Approach to the Evaluation of Packaging Components and the Regulatory and Scientific Considerations in Developing a Testing Strategy. *Am Pharma Rev.* (2006). 9(4): 10–15.
22. U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER). *Guidance for Industry Dosage Delivery Devices for Orally Ingested OTC Liquid Drug Products*. May 2011.
23. U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER). *Guidance for Industry Nasal Spray and Inhalation Solution, Suspension, and Spray Drug Products*. July 2002.

BIBLIOGRAPHY

- Ash M, Ash I. *Food Packaging Chemicals & Materials Electronic Handbook*, 2nd ed. New York: Synapse Information Resources, Inc., 2008.
- ASTM-D 1238—American Society of Testing and Materials, 1916 Race St. Philadelphia, PA 19103.
- ASTM-D 1505—Standard Test Method for the Density of Plastics by the Density-Gradient Technique, American Society of Testing and Materials, 1916 Race St. Philadelphia, PA 19103.
- ASTM-D 7709—11 *Standard Test Methods for Measuring Water Vapor Transmission Rate (WVTR) of Pharmaceutical Bottles and Blisters*, 1916 Race St. Philadelphia, PA 19103.
- Babbitt RO, ed. *The Vanderbilt Rubber Handbook*. Norwalk, CT: The Vanderbilt Co., Inc., 1978.
- Capersen G *et al.* Multiple-unit packaging for solid oral dosage forms. *Pharma Technol.* (2000). 24: (3) 44–54.
- Deiss K *et al.* Rethinking stability testing for tablets and capsules in containers. *Tablets & Capsules* 2008; April.
- Federal Food Drug and Cosmetic Act, Food Drug and Related Law as Amended thorough December 31, 2000, May 2001.
- Foster M *et al.* Permeation of unit-dose blister market containers under usp and ich conditions. *Pharma Technol.* (2000). (8) 38–42.
- ICH. *International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH)*. www.ich.org.
- Jenkins WA, Osborn KR. *Packaging Drugs and Pharmaceuticals*. Lancaster, PA: Technomic Publishing, 1993.
- McKinley E *et al.* Proposed revisions to USP Standards for Containers—Glass. *Pharmacopeial Forum.* (2006). 32: (4) 1366–1370.
- Office of the Federal Register National Archives and Records Administration. *Code of Federal Regulations Part 175 and Part 177 Indirect Food Additives*. <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm>.
- Office of the Federal Register, National Archives and Records Administration. *Drug Product containers and closures*. Code of Federal Regulations Title 21 § 211.94.
- Office of the Federal Register, National Archives and Records Administration. *Testing and appraisal or rejection of components, drug products containers, and closures*. Code of Federal Regulations Title 21 § 211.84.
- Okeke C, Taborsky CJ. Evaluation of USP General Chapters on Containers <661> and Containers—Permeation <671> based on high-density polyethylenestudy results. *Pharmacopeial Forum.* (1997). 23(5): 4941–4949.
- Poison Prevention Packaging Act, Codified 15 U.S.C.§ 1471-1477, Public Law 91-601, 84 Stat. 1670, Amended October 14, 2008.
- Reamer JT. *Project Report 73-956B: Tablet Containers, Collaborative Study, NF/USP*. Drug Standards Laboratory, APHA Foundation, August 6 1973.
- Reamer JT *et al.* Moisture permeation of typical unit dose repackaging materials. *Am J Hosp Pharm.* (1977). 34: 35–42.

- Reamer JT *et al.* Moisture permeation of newer unit dose repackaging materials. *Am J Hosp Pharm.* (1978). 35: 787–793.
- Subchapter H-Medical Devices, Code of Federal Regulations Title 21 § 800-1299, Office of the Federal Register, National Archives and Records Administration.
- Taborsky CJ *et al.* *Project Report: An Evaluation of High-Density Polyethylene Bottles Used to Package Solid Oral Dosage Forms and a Determination of the Impact of Using Regrind; A Collaborative Study.* CDER Docket, 1997.
- Taborsky CJ *et al.* Permeation of unit-dose blister market containers under usp and ich conditions. *Pharma Technol.* (2002). 24 (8): 38–42.
- Taborsky CJ, Grady LT. Multiple-unit dietary supplement containers-water vapor permeation in polyethylene terephthalate and high-density polyethylene containers for solid oral dosage forms. *USP Pharm Forum.* (2001). 27 (2): 2410–2416.
- Taborsky CJ, Grady LT. Water vapor permeation of prescription vials. *Pharmacopeial Forum.* (2001). 27 (2): 2417–2427.
- Taborsky CJ, McKinley E. Manufacturer's market containers and closures: Proposed Revision to Containers-Permeation <671>. *Pharmacopeial Forum.* (2003). 29 (1): 326–334.
- Taborsky CJ, Mehta U. Manufacturers' market containers and closures study II: Proposed Revision to Containers-Permeation <671>. *Pharmacopeial Forum.* (2004). 30 (4): 1508–1512.
- Taborsky J, Sheinin EB. A Critical Approach to the Evaluation of Packaging Components and the Regulatory and Scientific Considerations in Developing a Testing Strategy. *Am Pharma Rev.* (2006). 9 (4): 10–15.
- Taborsky JT *et al.* An evaluation of high-density polyethylene containers used to package solid oral dosage forms and a determination of the impact of using regrind. *Pharma Technol.* (1997). 21 (9): 116–130.
- The Toxics in Packaging Clearinghouse (TPCH), Coalition of Northeastern Governors. www.CONEG.org.
- Unknown. *CDER completes HDPE bottle study.* *Pharmaceutical and Medical Packaging News* 1997; Sept: 82–83.
- US Consumer Product Safety Commission. Washington, DC 20207. www.cpsc.gov.
- US Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER). *Guidance Expiration Dating of Unit-Dose Repackaging Drugs: Compliance Policy Guide.* May 2005.
- US Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER). *Guidance for Industry Container Closure Systems for Packaging Human Drug and Biologics, Chemistry, Manufacturing and Controls Documentation.* May 1999.
- US Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER). *Guidance for Industry Dosage Delivery Devices for Orally Ingested OTC Liquid Drug Products.* May 2011.
- US Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER). *Guidance for Industry Nasal Spray and Inhalation Solution, Suspension, and Spray Drug Products—Chemistry, Manufacturing, and Controls Documentation.* July 2002.
- USP/NF. *General Notices and Requirements and General Chapters.* Rockville, Maryland. 2011.

Chapter 36

Pharmaceutical Excipients

William J. Reilly Jr., MBA

INTRODUCTION 683

MATERIAL SUMMARIES 686

INTRODUCTION

The practice of pharmacy is an ever-evolving profession. A non-scientific survey of community pharmacies has revealed that extemporaneously compounding of prescriptions increasingly occurs. Whether this is a result of the physician not writing for compounded medicines or the pharmacy not interested in preparing such medicines is a topic for discussion elsewhere. It is, however, undoubtedly driven by the pharmaceutical industry, addressing many disease treatment and prevention efforts by commercially manufacturing products for mass distribution by the healthcare system in the United States.

The Food and Drug Administration has levied large fines on manufacturers who have failed to comply with what are known as current Good Manufacturing Practices (cGMP). Several companies have been forced to operate under what is known as a "Consent Decree," due to significant deficiencies in cGMPs. Additionally, the FDA's approval of new products is trending downward, this being a result of companies focusing research and development activities on the next "blockbuster" compound and the Agency frequently issuing "not approvable" letters, requiring the sponsor company to conduct additional studies. At the community level, more independent pharmacies are closing their doors or selling their patient lists to national or regional chains. Hospital settings are seeing a greater degree of mergers, so economics are more favorable, and an influx of patients using the facility as a clinic, rather than an acute care provider.

In addition to the profession of pharmacy and pharmacy education has undergone dramatic changes in the United States, since the last edition of the *Remington* was published. The PharmD degree has been the entry-level degree for anyone intending to practice pharmacy for well over a decade. Many practicing pharmacists pursue certification in sub-specialties, such as diabetes or cancer. The focus of pharmacy education has become even more clinically oriented on outcomes and patient interactions. This focus comes at the expense of basic pharmaceuticals, and in some instances, due to course loads, electives, such as industrial pharmacy courses, are not considered by students. The pharmaceutical industry used to be able to hire graduates with pharmacy degrees for positions in production, quality control, and dosage-form development, due to the breadth of understanding the graduate had of pharmaceutical processes. Unfortunately, gaining this knowledge as a PharmD has become increasingly difficult, unless the student pursues an advanced degree in industrial pharmacy or pharmaceuticals. However, many schools of pharmacy have introduced new degree programs focusing on the pharmaceutical sciences. These programs provide a foundation in the sciences that the industry requires of new hires: pre-formulation, analytical chemistry, formulation development, and manufacturing science.

It is imperative that pharmacists in all practice settings know it is their obligation to understand what is used to prepare a medication, whether by commercial means or by extemporaneously compounding it in a practice setting. This chapter does not address the legal aspects of community compounding by a pharmacist, nor does it explain all the specifics of formulating a product for commercial manufacturing. The information contained in these material summaries is intended to provide the practicing pharmacist sufficient background to address questions from an ever more interested patient about what some

of the ingredients in the products they are prescribed or purchasing over the counter are. A more detailed review of these excipients and their commercial applicability to dosage form development can be found in the *Handbook of Pharmaceutical Excipients* (Rowe, Sheskey, and Quinn, eds.), as well as other chapters in this edition of *Remington*.

These excipients, ingredients other than the active pharmaceutical ingredient, in a dosage form have different functionalities, depending on how they are used in a formulation or the type of formulation they are incorporated into. For a more in-depth discussion addressing dosage form development and manufacturing concerns around excipient choice and use, refer to the chapter on Industrial Pharmacy in this edition of *Remington's* or the current edition of the *USP-NF* General Information Chapter <1059>. Every effort has been made to obtain structural forms of the excipients discussed in this chapter. In some instances, the structures have been obtained from references, such as the *Merck Index*, the *Handbook of Pharmaceutical Excipients*, past editions of *Remington's*, or other recognized references. The categories of excipients and the names of the chemicals that fit into each category are subsequently listed:

ANTIOXIDANTS

Such materials, when added to a product, prevent the active from degrading in the presence of oxygen or peroxides.

Excipients

- Butylated Hydroxyanisole (BHA)
- Butylated Hydroxytoluene (BHT)
- Citric acid
- Sodium Metabisulfite

BINDERS

These can be incorporated into a formulation either as a wet material, such as a starch paste or povidone solution, for wet granulation processes, or as a dry powder, such as microcrystalline cellulose, for direct compression or dry granulation processes.

Excipients

- Alginate
- Sodium alginate
- Carboxymethyl cellulose sodium (CMC)
- Microcrystalline cellulose (MCC)
- Powdered cellulose
- Confectioner's sugar
- Dextrin
- Dextrose
- Ethylcellulose
- Guar gum
- Hydroxypropyl cellulose (HPC)
- Hypromellose (HPMC)
- Lactose
- Maltodextrin
- Methylcellulose
- Povidone
- Starch
- Tragacanth
- Zein

COATING AGENTS

Similar to taste masking technology, the first coatings were solutions of sucrose (12–20%). Once polymers were discovered and understood, they were applied in low concentrations (1–3%) using organic solvents. Current technology uses polymers manufactured as latex or pseudo-latex dispersions to avoid the use of organic solvents. Depending on the application, aesthetic or functional coating, the use levels can range from 2 to 20% of the weight of the material being coated.

Excipients

- Carboxymethyl cellulose sodium (CMC)
- Carnauba Wax
- Cellulose acetate phthalate (CAP)
- Ethylcellulose
- Gelatin
- Hydroxypropyl cellulose (HPC)
- Hypromellose (HPMC)
- Maltodextrin
- Methylcellulose
- Starch
- Sucrose
- Zein

COLORS AND PIGMENTS

There are two basic types of coloring agents for pharmaceutical products: dyes and lakes. Dyes are soluble forms of a particular color. They go into solution and can result in very deep, vibrant colors. Lakes are dyes that undergo a processing step that adheres them onto insoluble substrates, such as aluminum or calcium salts. Dyes are usually used in liquid products, whereas lakes are used in chewable products and coating solutions for tablets.

DILUENTS/FILLERS

When considering solid dosage forms, such as tablets and capsules, these ingredients function as a bulking agent for low dose actives, facilitating compression or encapsulation. Diluents are also used in lyophilized products, serving the same purpose of providing bulk to the product, but might be different materials.

Excipients

- Calcium carbonate
- Calcium sulfate
- Microcrystalline cellulose (MCC)
- Powdered cellulose
- Dextrates
- Dextrin
- Dextrose
- Kaolin
- Lactose
- Maltodextrin
- Mannitol
- Starch
- Sucrose

DISINTEGRANTS

These materials facilitate the breakdown of oral solid dosage forms in the gastrointestinal tract. Tablets and capsules must breakdown from their original form to large particles and then to small particles, increasing surface area from which drug can dissolve. Early disintegrants included different types of starch, as a dry addition to a formulation, due to the ability of the starch to swell. Significant improvement in disintegrant performance was achieved with the introduction of the first super-disintegrant, sodium starch glycolate. There are also newer, polymeric disintegrants, such as croscarmellose sodium. These super-disintegrants exhibit significantly superior swelling and water wicking capability, compared to starch, and can be used at much lower concentrations.

Excipients

- Alginic Acid
- Sodium alginate
- Microcrystalline cellulose (MCC)
- Croscarmellose sodium
- Crospovidone
- Guar gum
- Polyacrilin Potassium
- Sodium Starch Glycolate

EMOLLIENTS

These materials are usually used in conjunction with emulsifying agents in the preparation of topical products. An emollient is intended to impart a soft, supple feeling to the skin, helping the skin to retain moisture and natural flexibility.

Excipients

- Glycerin
- Glyceryl monostearate
- Isopropyl Myristate
- Petrolatum
- Polyethylene Glycols

EMULSIFIERS

These materials typically exhibit some surfactant characteristics but also facilitate micelle formation, which allows for stable uniform products consisting of oil-in-water or water-in-oil biphasic emulsions or tri-phasic emulsions.

Excipients

- Carbomer
- Carrageenan
- Lanolin
- Lecithin
- Mineral Oil
- Oleic Acid
- Oleyl alcohol
- Pectin
- Poloxamer
- Polyoxyethylene Sorbitan Fatty Acid Esters
- Sorbitan Esters
- Triethanolamine

FLAVORS

These materials were originally natural products; however, many companies have been created to develop synthetic versions of different flavors. Most manufacturers of finished products provide these Flavor Houses with samples of in-process material for flavoring. By using a Flavor House, unique flavor notes can be imparted to the product that will facilitate brand recognition.

GLIDANTS/ANTIADHERENTS

These products facilitate the movement of the powder or granulation prior to compaction, compression, or encapsulation. Some of these materials possess moisture scavenging capabilities similar to desiccants. By facilitating flow of the material, there is less weight variability of the dosage form, resulting in more consistent dosing of the active ingredient(s). Additionally, by controlling the amount of moisture available to interact with the active ingredient(s) and other excipients, the product's stability is enhanced.

Excipients

- Colloidal silicon dioxide
- Talc

HUMECTANTS

These materials promote the retention of moisture. They are necessary in some semi-solid products (creams and gels) to prevent the dosage form from drying out. They also help prevent

a phenomenon known as cap-locking. Cap-locking involves liquid products that recrystallized at the bottle-cap interface and makes opening the bottle difficult after prolonged periods of non-use. These materials are hygroscopic and should be stored in well closed containers prior to use.

Excipients

Glycerin
Propylene Glycol
Sorbitol
Triethanolamine

LUBRICANTS

These materials facilitate compression/compaction of tablets and consolidation of capsule plugs. Lubricants are used in small quantities in the formulation. One could think of a particle of lubricant as a deck of cards. When the deck is undisturbed, it has a very small coverage area; however, when the individual cards are randomly tossed on a table the coverage area is greatly increased. That is how lubricants, like Magnesium, stearate function. Other lubricants, such as Stearic acid, melt during compaction/compression and exert their lubricant effects in that manner. Many lubricants are hydrophobic—water repellent. If used at too high a concentration or blended for too long, tablet hardness and/or dissolution could be negatively affected.

Excipients

Calcium stearate
Glyceryl monostearate
Isopropyl Myristate
Magnesium stearate
Polyvinyl Alcohol
Sodium Stearyl Fumarate
Stearic Acid
Talc

PLASTICIZERS

Plasticizers lower the glass transition temperature of polymers, specifically polymers used for coatings. By lowering the glass transition temperature, the polymers can partially melt, and the droplets can coalesce to form a completely sealed coating on a tablet or spheroid at a lower temperature, thus, preventing the active from being exposed to elevated processing temperatures. This affords better, more consistent release of the active.

Excipients

Glycerin
Propylene Glycol
Triacetin
Triethanolamine

PRESERVATIVES

If these materials prevent the initiation and growth of microorganisms in products, they are known as bactericidal preservatives. Some preservatives, either by concentration or activity, may only maintain the bacteria level in the product at the time of manufacture and are referred to as bacteriostatic preservatives.

Excipients

Alcohol
Benzalkonium chloride
Boric acid
Butylated Hydroxyanisole (BHA)
Butylated Hydroxytoluene (BHT)
Butylparaben
Methylparaben
Phenol
Phenethyl Alcohol
Potassium Sorbate

Propylene Glycol
Propylparaben
Sorbic Acid

PROPELLANTS

These materials are found exclusively in aerosolized delivery systems. Early propellants were CFCs—chlorinated fluoro-carbons—now many non-CFC products have reached the market and are known as HFCs.

Excipients

Difluoroethane
Nitrogen

PROTECTIVE COLLOIDS

These materials are added to suspensions to protect the suspending agent and other colloidal sized particles from settling. These work in conjunction with the suspending agents or viscosifiers.

Excipients

Hydroxypropyl cellulose (HPC)
Hypromellose (HPMC)
Methylcellulose

SIALAGOGUES

These materials are usually acidic, and their purpose in a formulation is to stimulate the production of saliva. They are mostly found in chewable products for children or in gum-based products. By facilitating the production of saliva, the masticated product is swallowed more readily, limiting the objectionable flavor of the active from lingering too long.

Excipients

Citric acid
Fumaric acid
Tartaric acid

SURFACTANTS

Surfactants function by decreasing the surface tension of a material, coating agent, and a substrate, a tablet. By reducing the surface tension, the coating can more uniformly cover the tablet surface, resulting in a more aesthetically pleasing product. When used in suspension, the surfactant facilitates the wetting of the drug particle, facilitating its ability to go into solution.

Excipients

Polyethylene Glycols
Polyoxyethylene Sorbitan Fatty Acid Esters
Sodium Lauryl Sulfate
Sorbitan Esters

SUSPENDING AGENTS

Suspending agents are similar to viscosifiers in a formulation. They function by keeping small particles of active, and possibly other excipients, suspended during the shelf life of the product.

Excipients

Acacia
Agar
Carbomer
Carboxymethyl cellulose sodium (CMC)
Carrageenan
Microcrystalline cellulose and Sodium Carboxymethyl cellulose Co-processed
Colloidal silicon dioxide
Dextrin
Guar gum
Hydroxypropyl cellulose (HPC)
Hypromellose (HPMC)

Kaolin
Methylcellulose
Pectin
Polyvinyl Alcohol
Povidone
Tragacanth

SWEETENERS

Sweeteners are materials used in both solid and liquid dosage forms that impart sweetness to the product. Sucrose is the standard against which all sweeteners, both natural and artificial, are measured. In addition to being many more times sweeter than sucrose, artificial sweeteners have the advantage of not impacting blood sugars of diabetic or pre-diabetic patients, and they are considered non-cariogenic.

Excipients

Acesulfame Potassium
Aspartame
Confectioner's sugar
Dextrates
Dextrose
Fructose
Mannitol
Saccharin
Sorbitol
Sucralose
Sucrose
Xylitol

TASTE-MASKING AGENTS

The original taste maskers were very high concentrations of sweeteners. The next generation of taste-masking technology involved polymers that coated particles of active ingredients. Newer taste-masking technology involves enveloping/cloaking polymers that hide the particle of active from the taste buds.

Excipients

Confectioner's sugar
Polyacrilin Potassium
Poloxamer

VISCOSIFIERS

Viscosifiers (thickening agents) are similar in function to suspending agents; both impart thickness to liquid products. Some viscosifiers go into solution, such as certain cellulose based polymers, resulting in a Newtonian-type viscosity. Other materials increase viscosity, while imparting a yield stress to the suspension. This type of product is known as thixotrope.

Excipients

Acacia
Agar
Sodium alginate
Bentonite
Carbomer
Carboxymethyl cellulose sodium (CMC)
Guar gum
Hydroxypropyl cellulose (HPC)
Hypromellose (HPMC)
Methylcellulose
Pectin

MATERIAL SUMMARIES

OFFICIAL NAME: ACACIA

INN: Acacia
Synonyms: Gum Arabic
Chemical Name & CAS Number: Acacia (9000-01-5)

Molecular Formula: A complex containing principally calcium, magnesium, and potassium salts of the polysaccharide *arabic acid*, which on acid hydrolysis yields L-arabinose, L-rhamnose, D-galactose, and an aldobionic acid containing D-glucuronic acid and D-galactose.

Structure: A very complex gum, which has no specific structural formula

Description: White to yellowish white, angular microscopic fragments, spheroidal tears or powder

Properties: Soluble 1 in 20 of glycerin and propylene glycol, Insoluble in alcohol (ethanol 95%) and 1 in approximately 2.7 of water at room temperature; 5% aqueous solution pH 4.5–5.0; moisture content 8–13%; specific gravity 1.35–1.49; Solution viscosity varies, depending upon the source of the material, presence of salts (calcium, sodium, etc.), the pH of the solution, and storage conditions of the acacia prior to use.

Incompatibilities: Alcohol or alcoholic solutions precipitate acacia as a stringy mass, when the alcohol amounts to more than about 35% of the total volume. Solution is affected by dilution with water. The mucilage is destroyed through precipitation of the acacia by heavy metals. Borax also causes a precipitation that is prevented by glycerin. It contains calcium and, therefore, possesses the incompatibilities of this ion.

Uses: As a suspending agent for insoluble substances in water, in the preparation of emulsions.

OFFICIAL NAME: ACESULFAME POTASSIUM

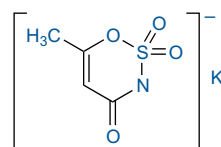
INN: Acesulfame Kalicum

Synonyms: Ace-K, Acesulfame-K

Chemical Name & CAS Number: 6-Methyl-1,2,3-oxathiazin-4(3H)-one-2-dioxide potassium salt (55589-62-3)

Molecular Formula: C₄H₄KNO₄S

Structure:



Description: A white colored powder

Properties: Solubility approximately 1 in 3.7 of water at 20°C, 1 in 1000 of ethanol; exhibiting 200 times the sweetening ability of sucrose

Incompatibilities: No known incompatibilities

Use: Sweetening Agent

OFFICIAL NAME: AGAR

INN: Agar-Agar

Synonyms: Agar, Vegetable Gelatin; Gelosa; Chinese or Japanese Gelatin

Chemical Name & CAS Number:

Molecular Formula: Varies, but the commonality is a 1–3 linked β-D-galactopyranosyl unit joined by a 1–4 linkage to a 3,6-anhydro α-D-galactopyranosyl unit

Structure: Varies greatly, depending on the source of the material

Description: A white to off-white hygroscopic powder

Properties: Insoluble in cold water and alcohol; slowly solubilizes in hot water to approximately 1–1.5% w/w. When cooled, the solution produces a stiff gel that can melt at 60–100°C, depending on composition.

Incompatibilities: Salt solutions can impact the gel formation of the polymer

Use: Substitute for gelatin, gelling agent

OFFICIAL NAME: ALCOHOL

INN: Alcohol; Ethanol; Ethanolum

Synonyms: Ethanol; Ethyl alcohol, grain alcohol

Chemical Name CAS Number: Ethanol (64-17-5)

Molecular Formula: C₂H₆O

Description: A clear, volatile liquid with a characteristic odor and burning taste

Properties: Miscible with water and glycerin; boiling point 78°C

Incompatibilities: Acacia solutions (precipitation of the acacia gum), oxidizing materials while in acidic conditions

Use: As a preservative in concentration greater than 10%; as a disinfectant in concentrations of 60–90%; as a co-solvent for certain active pharmaceutical ingredients or in film coating and tablet printing unit operations; and as a vehicle for both oral and topical finished products

OFFICIAL NAME: ALGINIC ACID

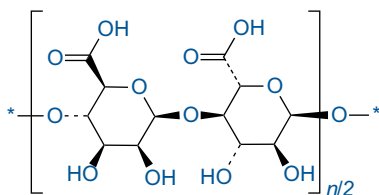
INN: Acidum alginicum

Synonyms: L-gulo-D-mannoglycuronan; Kelaid, polymannuronic acid; Satialgine

Chemical Name & CAS Number: Alginic acid [9005-32-7]

Molecular Formula: Mixtures of β -(1-4)-D-mannosyluronic acid and α -(1-4)-L-gulosyluronic acid. The general formula is $[C_6H_8O]_n$.

Structure:



Description: A fibrous powder that is white to yellow-white that is odorless and tasteless

Properties: pH of 1.5–3.5 of a 3% aqueous dispersion. The powder has a density of approximately 1.6 g/cc with a moisture content of 7%. The powder is soluble in alkali solutions; practically insoluble in ethanol and other organic solvents. The powder does not dissolve in water, but swells by absorbing 200–300 its weight in water.

Incompatibilities: Strong oxidizing agents and alkaline earth metals; in the presence of di- and tri-valent ions, forms a gelatinous white precipitate.

Uses: In solid dosage forms, as either a binder (as a solution) or as a disintegrant (in dry form), as a viscosifying agent in oral liquids, and a stabilizer for emulsions

OFFICIAL NAME: SODIUM ALGINATE

INN: Alginic acid; Natrii Alginas

Synonyms: Algin, Kelcosol, sodium polymannuronate

Chemical Name & CAS Number: Sodium alginate [9005-38-3]

Molecular formula: The purified carbohydrate product extracted from brown seaweeds. It consists chiefly of the sodium salt of alginic acid, a polyuronic acid composed of beta-D-mannuronic acid residues, linked so the carboxyl group of each unit is free, while a glycosidic linkage shields the aldehyde group.

Description: An odorless, tasteless, white to yellow-white powder that is hygroscopic

Properties: A 1% solution has an approximate pH of 7.2; the material is soluble in ethanol and other organic solvents and aqueous solutions with a pH of less than 3. The material is slowly soluble in water, forming a viscous solution at concentration up to 1%

Incompatibilities: As for Alginic acid

Uses: As a binder or disintegrant in oral solid dosage forms; as a thickening agent for topical dosage forms; as an adhesive/reservoir agent for transdermal patches; and as an encapsulating agent for novel drug delivery

OFFICIAL NAME: ASPARTAME

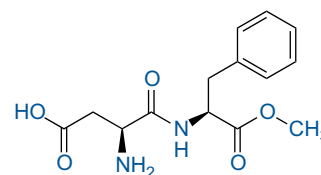
INN: Aspartame; Aspartamum

Synonyms: NutraSweet; Equal; Tri-Sweet

Chemical name: N- α -L-Aspartyl-L-phenylalanine 1-methyl ester (22839-47-0)

Molecular Formula: $C_{14}H_{18}N_2O_5$

Structure:



Description: White to off-white crystalline powder that is approximately 200 times more sweet than sucrose

Properties: Density is between 0.2 and 0.7 g/cc, depending on the grade; only slightly soluble in ethanol and sparingly soluble in room temperature water

Incompatibilities: Some reactions have been recorded with sugar alcohols (sorbitol and mannitol), magnesium stearate, and dicalcium phosphate. Aspartame should not be used when a solution requires heating as degradation does occur.

Uses: As a sweetening agent

OFFICIAL NAME: BENTONITE

INN: Bentonite; Bentonitum; Bentonite

Synonyms: Wilkinite; Soap Clay; Mineral Soap; Veeëgum HS

Chemical Name & CAS Number: Bentonite [1302-78-9]

Molecular Formula: $Al_2O_3 \cdot 4SiO_2 \cdot H_2O$

Description: An odorless, very fine crystalline, clay-like material gray colored with an almost bi-modal particle size distribution of 1–2 μ and 50–150 μ . Bentonite has a slight earthy taste.

Properties: Very hygroscopic material, with a moisture content of 5–12% at ambient relative humidity but capable of increasing its moisture content to 20% at 90% relative humidity. Insoluble in all known solvents, swells to about 12 times its original volume in water. A 2% aqueous dispersion has a pH of $\approx$ 10. Viscosity increases with increasing concentration.

Incompatibilities: Bentonite suspensions are precipitated by acids. Additionally, mixing alcohol into a bentonite suspension will result in precipitation.

Uses: Used in topical formulations to form the gel or suspension. Prior to more effective methods, it was used for taste masking oral preparations.

OFFICIAL NAME: BENZALKONIUM CHLORIDE

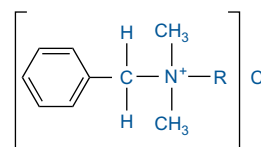
INN: Benzalkonium chloride; Benzalkonii chloridum

Synonyms: Alkylbenzyltrimethylammonium chloride; BKC; BAC

Chemical Name & CAS Number: alkyltrimethyl(phenylmethyl) ammonium chloride (8001-54-5)

Molecular Formula: $[C_6H_5CH_2N(CH_3)_2R]^+Cl^-$, where R can be $-C_8H_{17}$, $C_{12}H_{25}$, $C_{14}H_{29}$ and $C_{16}H_{33}$

Structure:



Description: A white to yellow-white powder, flake, or gel

Properties: 10% solution pH 5–8; hygroscopic; soapy to touch; density 0.98g/cc; very soluble in water and alcohol

Incompatibilities: Aluminum salts, anionic surfactants, citrates, iodides, hypromellose, hydrogen peroxide, lanolin, nitrates, and nonionic surfactants in high concentrations, proteins, permanganates, salicylates, and tartrates

Use: Preservative

OFFICIAL NAME: BORIC ACID

INN: Boric acid; Acidum boricum

Synonyms: Boracic Acid; Orthoboric Acid; Boracic acid; boron trihydroxide

Chemical Name & CAS Number: Boric acid (10043-35-3)

Molecular Formula: H_3BO_3

Description: Hygroscopic white crystalline powder

Properties: Soluble 1 g in 18 mL water, 18 mL alcohol; 5% solution pH 3.5-4.1; density 1.435 g/cc

Incompatibilities: Strong bases, alkali, reacts violently with potassium and acid anhydrides and complexes with glycerin

Uses: As a buffer and preservative for eye drops

OFFICIAL NAME: BUTYLATED HYDROXYANISOLE

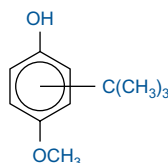
INN: BHA; Butylhydroxyanisolum; Butylated hydroxyanisole

Synonyms: BHA; Sustane; Tenox

Chemical Name: 2-tert-Butyl-4-methoxyphenol (25013-16-5)

Molecular Formula: $\text{C}_{11}\text{H}_{16}\text{O}_2$

Structure:



Description: White or slightly yellow, waxy solid, having a faint, characteristic odor

Properties: Insoluble in water; 1 g in 4 mL alcohol, 2 mL chloroform or 1.2 mL ether; antimicrobial activity against molds and gram positive bacteria.

Storage: In a tightly sealed container, which will protect the material from light. Exposure to light causes discoloration and decreased activity.

Incompatibilities: Avoid use with peroxides and permanganates; contact with oxidizing agents may result in spontaneous combustion. Avoid iron salts, as these reduce effectiveness.

Uses: As an antioxidant and antimicrobial agent in liquid and semi-solid preparations

OFFICIAL NAME: BUTYLATED HYDROXYTOLUENE

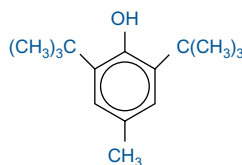
INN: Butylhydroxytoluenum; Butylated hydroxytoluene

Synonyms: Agidol, BHT

Chemical Name & CAS Number: 2,6-Di-tert-butyl-4-methylphenol (128-37-0)

Molecular Formula: $\text{C}_{15}\text{H}_{24}\text{O}$

Structure:



Description: White, tasteless crystals with a mild odor; stable in light or air

Properties: Melts at 70°C; insoluble in water; 1 g in 4 mL alcohol, 1.1 mL chloroform, or 1.1 mL ether

Incompatibilities: Avoid use with peroxides and permanganates; contact with oxidizing agents may result in spontaneous combustion. Avoid iron salts, as these reduce effectiveness.

Uses: As an antioxidant

OFFICIAL NAME: BUTYLPARABEN

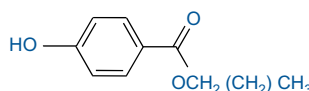
INN: Butyl Parahydroxybenzoate; butylparaben

Synonyms: Butoben, 4-hydroxybenzoic acid butyl ester

Chemical Name & CAS Number: Butyl-4-hydroxybenzoate (94-26-8)

Molecular Formula: $\text{C}_{11}\text{H}_{14}\text{O}_3$

Structure:



Description: White to off-white crystalline powder that is odorless and tasteless

Properties: Very slightly soluble in water or glycerin, freely soluble in propylene glycol, melting range 68-72°C

Incompatibilities: Avoid nonionic surfactants, as these reduce antimicrobial effectiveness, due to micellization. Possible absorption of butylparaben by plastic bottles is under investigation.

Use: Antimicrobial preservative for liquid and semi-solid preparations

OFFICIAL NAME: CALCIUM CARBONATE

INN: Carbonic acid, calcium salt; Calcium carbonate

Synonyms: Cal-Carb; precipitated calcium carbonate

Chemical Name & CAS Number: Carbonic acid calcium salt (471-34-1)

Molecular Formula: CaCO_3

Description: Odorless and tasteless white powder, crystals, or granules

Properties: Moisture content less than 1%, insoluble in all pharmaceutical solvents; particle size can vary depending upon grade. Reacts with acids to release carbon dioxide.

Incompatibilities: All acids and ammonium salts

Uses: As a diluent in tablet formulations and as an opacifier in tablet coatings. It does have therapeutic uses as a nutritional supplement and as an antacid.

OFFICIAL NAME: CALCIUM STEARATE

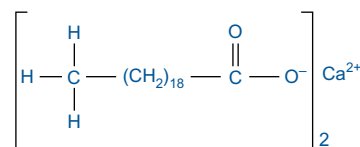
INN: Octadecanoic acid, calcium salt; Calcium stearate

Synonyms: Calcium distearate; stearic acid, calcium salt

Chemical Name & CAS Number: Calcium stearate [1592-23-0]

Molecular Formula: $\text{C}_{36}\text{H}_{70}\text{CaO}_4$

Structure:



Description: A very fine white to off-white powder that is tasteless, but does have a slight characteristic odor

Properties: Practically insoluble in water and ethanol, an approximate surface area of 4.7-8.0 m²/g and a moisture content of approximately 3%.

Incompatibilities: No known incompatibilities

Use: As a tablet and capsule lubricant at levels of less than 3%

OFFICIAL NAME: CALCIUM SULFATE

INN: Calcium sulphate dihydrate; Calcium sulfate

Synonyms: Sulfuric acid, calcium salt (1:1); Gypsum; Terra Alba

Chemical Name & CAS Number: Calcium sulfate [7778-18-9]; CaSO_4 dihydrate [10101-41-4]

Molecular Formula: CaSO_4 and $\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$

Description: Fine, white to slightly yellow-white, odorless powder or granule

Properties: Dissolves in diluted HCl; slightly soluble in water; free flowing powder/granule with a density of 2.3 g/cc

Incompatibilities: Forms complexes with amines and amino acids, when excessive moisture is present, certain antibiotic preparations (i.e., tetracycline, erythromycin, which complex with calcium containing products)

Uses: As a diluent in the manufacture of compressed tablets

OFFICIAL NAME: CARBOMER

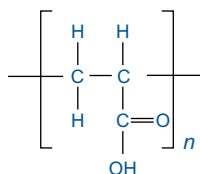
INN: Carbomer; Carbomera

Synonyms: Acrylic acid polymer; carboxyvinyl polymer; carboxy polymethylene; polyacrylic acid

Molecular Formula: A synthetic high-molecular-weight cross-linked polymer of acrylic acid; contains 56-68% of carboxylic

acid (COOH) groups. The viscosity of a neutralized preparation of carbomer 934P (2.5 g/500 mL water) is 30,000–40,000 mPa.s.

Structure:



Description—White, fluffy powder with a slight, characteristic odor; hygroscopic

Properties: Dissolves in water, alcohol, or glycerin, when neutralized with alkali hydroxides or amines; 0.5% dispersion pH 3; specific gravity about 1.41; moisture content 2%; viscosity varies depending upon grade, salt type and concentration, and pH of solution

Incompatibilities: Resorcinol, phenol, cationic polymers, strong acids, and high levels of electrolytes

Uses: As a thickening, suspending, dispersing, emulsifying agent and gel former

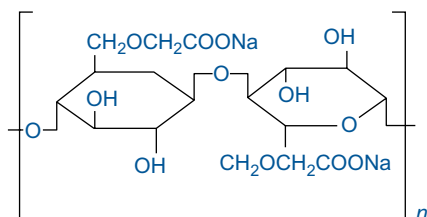
OFFICIAL NAME: CARBOXYMETHYLCELLULOSE SODIUM

INN: Carmellose sodium

Synonyms: Cellulose gum, CMC sodium

Chemical Name & CAS Number: Cellulose, carboxymethyl ether, sodium salt [9004-32-4]

Structure:



Description: Depending on the grade, CMC is a white to off-white odorless powder of varying particle size

Properties: This hygroscopic material should contain less than 10% water, density of 0.7 g/cc (tapped), viscosities of a 1%w/w aqueous solution can range from 10 to 8000 mPa.s, depending on the grade being evaluated.

Incompatibilities: Strongly acidic solutions and soluble salts of iron, aluminum, zinc, and mercury. Complexes can be formed with gelatin and pectin.

Uses: As a suspending agent for liquid formulations, as a binder for tablets or capsules (in solution) and as a coating agent on tablets

OFFICIAL NAME: CARNAUBA WAX

INN: Carnauba wax; Cera wax; Cera carnauba

Synonyms: Brazil wax, caranda wax

Chemical Name & CAS Number: Carnauba wax (8015-86-9)

Description: Light-brown to pale-yellow, moderately coarse powder; characteristic bland odor;

Properties: Insoluble in water; specific gravity about 0.99; melts at about 84°C

Uses: As a polishing agent in the manufacture of coated tablets

OFFICIAL NAME: CARRAGEENAN

INN: Carrageenan

Synonyms: Chondrus crispus, Irish moss, pearl moss

Chemical Name & CAS Number: λ-Carrageenan (9064-57-7)

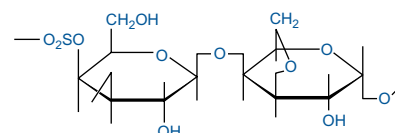
κ-Carrageenan (11114-20-8)

ι-Carrageenan (9000-07-1)

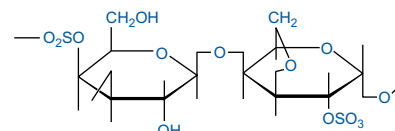
Molecular formula: It is a variable mixture of potassium, sodium, calcium, magnesium, and ammonium sulfate esters of galactose and 3,6-an-hydrogalactose copolymers, the hexoses being alternately linked at the α-1,3 and β-1,4 sites in the polymer chain.

Structure:

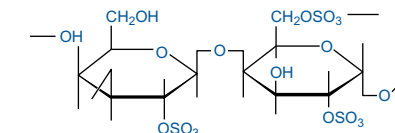
Kappa



Iota



Lambda



Description: Yellow-brown to white, coarse to fine powder; odorless; tasteless. The three main types of copolymers present are *kappa*-carrageenan, *iota*-carrageenan, and *lambda*-carrageenan, which differ in the composition and manner of linkage of monomeric units and the degree of sulfation (the ester sulfate content for carrageenans varies from 18% to 40%). *Kappa*-carrageenan and *iota*-carrageenan are the gelling fractions; *lambda*-carrageenan is the non-gelling fraction.

Properties: All carrageenans hydrate rapidly in cold water, but only *lambda*-carrageenan does not require the presents of cations to initiate functionality. Gelling carrageenans require heating to about 80°C for complete solution when potassium and calcium ions are present.

Incompatibilities: All carrageenans are reactive with cationic materials, resulting in complexation.

Uses: As an emulsifying, suspending, and gelling agent

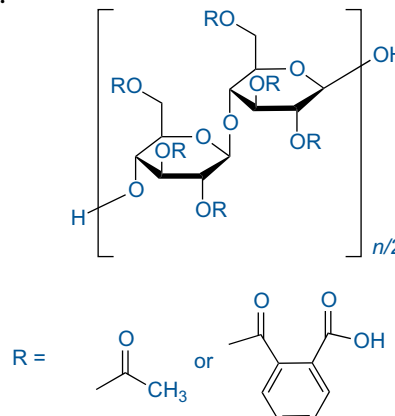
OFFICIAL NAME: CELLULOSE ACETATE PHTHALATE

INN: Cellacefate, cellacephate; Cellulose acetate phthalate

Synonyms: Acetyl phthalyl cellulose; CAP; cellulose acetate hydrogen 1,2-benzenedicarboxylate

Chemical Name & CAS Number: Cellulose, acetate, 1,2-benzenedicarboxylate (9004-38-0)

Structure:



Description: Free-flowing, white powder; may have a slight odor of acetic acid

Properties: Insoluble in water or alcohol; soluble in acetone; moisture content of approximately 2.2%, a bulk density of approximately 0.26g/cc

Uses: As an enteric tablet coating material

OFFICIAL NAME: MICROCRYSTALLINE CELLULOSE

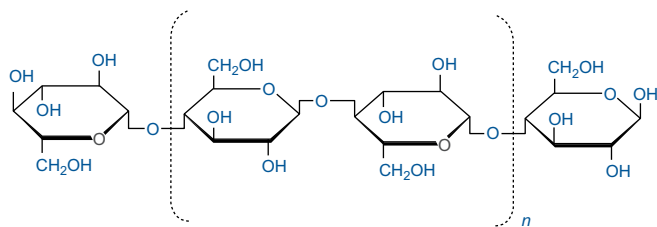
INN: Microcrystalline cellulose; Cellulosum microcrystallinum

Synonyms: Cellulose gel; crystalline cellulose; Avicel PH

Chemical Name & CAS Number: Cellulose [9004-34-6]

Molecular Formula: $(C_6H_{10}O_5)_n$, where $n = 220$

Structure:



Description: Fine, white, odorless, crystalline powder; consists of free-flowing, non-fibrous particles

Properties: Insoluble in water, dilute acids, or most organic solvents; slightly soluble in NaOH solution (1 in 20); density approximately 1.5120-1.668 g/cc; moisture content (grade specific) ranges from 1.5-5.0%; particle size varies by grade from <20-200 μ .

Incompatibilities: Strong oxidizing agents

Uses: As a tablet diluent, disintegrant, and dry binder

OFFICIAL NAME: MICROCRYSTALLINE CELLULOSE AND SODIUM CARBOXYMETHYLCELLULOSE CO-PROCESSED

Synonyms: Cellulose gel

Description: Tasteless, odorless, white to off-white, coarse to fine powder

Properties: pH (dispersion) 6-8; swells in water, producing, when dispersed, a white, opaque dispersion; at concentrations above 2%, can form a thixotropic gel; insoluble in water, organic solvents, and dilute acids; soluble in strong acids. Temperature of liquid at time of dispersion will affect viscosity.

Incompatibilities: High concentrations of dissolved salts

Uses: As a thixotropic suspending agent

OFFICIAL NAME: POWDERED CELLULOSE

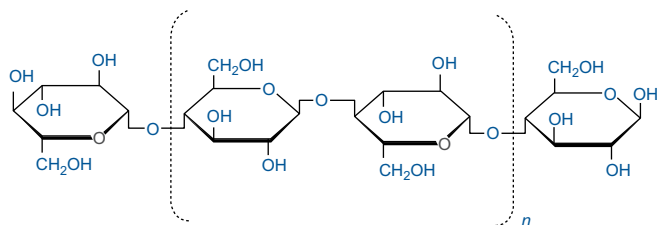
INN: Powdered cellulose

Synonyms: Elcema, Solka-Floc

Chemical Name & CAS Number: Cellulose [9004-34-6]

Molecular Formula: $(C_6H_{10}O_5)_n$

Structure:



Description: White, odorless substance, consisting of fibrous particles, exhibiting degrees of fineness, ranging from a free-flowing dense powder to a coarse, fluffy, non-flowing material

Properties: Insoluble in water, dilute acids, and nearly all organic solvents; slightly soluble in NaOH solution (1 in 20); poorly flowing, due to its fibrous nature, but compressible; particle size is dependent upon the grade

Incompatibilities: No known incompatibilities

Uses: As a tablet diluent, binder, or adsorbent

OFFICIAL NAME: CETYL ALCOHOL

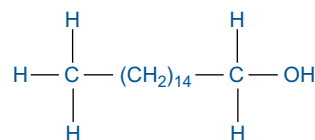
INN: Cetanol

Synonyms: Cetostearyl Alcohol; Palmityl Alcohol; Aldol 52

Chemical Name & CAS Number: 1-Hexadecanol (36653-82-4)

Molecular Formula: $C_{16}H_{34}O$

Structure:



Description: Unctuous, white flakes, granules, cubes, or castings; faint characteristic odor and a bland, mild taste; consisting of a mixture of not less than 90% of cetyl alcohol, the remainder being mostly stearyl alcohol

Properties: Melts at 45-52°C; freely soluble in ethanol, practically insoluble in water; miscible with other melted waxes

Solubility: Insoluble in water; soluble in alcohol, chloroform, ether, or vegetable oils

Uses: As a base for suppositories, creams, and lotions; as a stabilizer in water-in-oil emulsions, due to its high water absorption characteristics and emollient

OFFICIAL NAME: CITRIC ACID

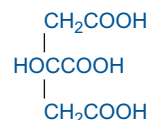
INN: Citric acid

Synonyms: 2-Hydroxypropane-1,2,3-tricarboxylic acid monohydrate

Chemical Name & CAS Number: 1,2,3-Propanetricarboxylic acid, 2-hydroxy-, Citric acid [77-92-9] $C_6H_8O_7$; monohydrate $C_6H_8O_7 \cdot H_2O$ [5949-29-1]

Molecular Formula: Citric acid $C_6H_8O_7$; monohydrate $C_6H_8O_7 \cdot H_2O$

Structure: Citric acid



Description: A white, granular to fine crystalline powder; odorless; strongly acid taste

Properties: 1 g in 0.5 mL water, 2 mL alcohol; density 1.5 g/cc

Uses: As an acidifier, antioxidant, buffering agent, and sialogogue

OFFICIAL NAME: COCOA BUTTER

INN: Cacao Butter

Synonyms: Cacao Butter; Theobroma Oil; Oil of Theobroma

Chemical Name & CAS Number: Theobroma Oil

Molecular Formula: It is a mixture of stearin, palmitin, olein, laurin, linolein, and traces of other glycerides.

Description: Yellowish, white solid; faint, agreeable odor; bland (if obtained by extraction) or chocolate-like (if obtained by pressing) taste; obtained from the roasted seed of *Theobroma cacao* Linné (Fam *Sterculiaceae*)

Properties: Usually brittle below 25°C; specific gravity 0.858-0.864 at 25°C; refractive index 1.454-1.458 at 40°C; slightly soluble in alcohol; not soluble in water

Uses: As a base for suppositories, creams, and lotions; essentially replaced by less variable semi-synthetic materials today

OFFICIAL NAME: COLLOIDAL SILICON DIOXIDE

INN: Colloidal anhydrous silica; Silica colloidalis anhydrica; Colloidal silicon dioxide

Synonyms: Colloidal silica; fumed silica; light anhydrous silicic acid; silicic anhydride

Chemical Name & CAS Number: Silica (7631-86-9)

Structure: SiO_2 (60.08)

Description: Light, white, non-gritty powder

Properties: Insoluble in water or acids (except hydrofluoric); dissolved by hot solutions of alkali hydroxides; particle size approximately 15 nm; 4% solution pH 3.5-4.4

Incompatibilities: Diethylstilbestrol (DES)

Uses: As a moisture scavenger, glidant, or anti-caking agent in tablets and powders; as a suspending and thickening agent in liquid and semi-solid preparations

OFFICIAL NAME: CONFECTIONER'S SUGAR

INN: None

Synonyms: Icing sugar, powdered sugar

Chemical Name & CAS Number: A mixture of sucrose and corn starch, such that not less than 95% of the mixture is sucrose

Description: Fine, white, odorless powder; sweet taste; stable in air

Solubility: The sucrose portion is soluble in cold water; this is entirely soluble in boiling water.

Uses: As a binder in tablets, carrier in taste masking, a sweetening agent

OFFICIAL NAME: COTTONSEED OIL

INN: Cottonseed oil

Synonyms: Cotton Seed Oil; Cotton Oil

Chemical Name & CAS Number: Cottonseed Oil (8001-29-4)

Molecular formula: A mixture of mono-, di- and tri-glyceride fatty acid esters, comprising 39.3% linolenic acid, 33.1% oleic acid, 19.1% palmitic acid, 1.9% stearic acid, 0.6% arachidic acid, and 0.3% myristic acid

Description: Pale yellow, oily liquid with a bland taste; odorless or nearly so

Properties: Slightly soluble in alcohol; miscible with ether, chloroform, solvent hexane, or carbon disulfide; particles of solid fat may separate below 10°C; solidifies at about 0 to -5°C; specific gravity 0.915-0.921; viscosity approximately 70 mPa s

Use: As a solvent and vehicle for injections, emollient

OFFICIAL NAME: CROSCARMELLOSE SODIUM

INN: Carmellosum natricum conexum

Synonyms: Crosslinked Carboxymethylcellulose sodium; modified cellulose gum; Ac-Di-Sol

Chemical Name & CAS Number: Cellulose, carboxymethyl ether, sodium salt, crosslinked (74811-65-7)

Structure: Crosslinked polymer of carboxymethylcellulose sodium

Description: An odorless white to off-white powder

Properties: Density approximately 0.53 g/cc; particle size varies, depending on manufacturer; insoluble in water, though rapidly swells to 4–8 times its original size

Incompatibilities: Complexes with amine containing drugs

Use: A tablet/capsule disintegrant (super disintegrant due to its low use level (<2%))

OFFICIAL NAME: CROSPVIDONE

INN: Crospovidone, Crospovidonum

Synonyms: Crosslinked povidone, crosslinked PVP, polyvinylpyrrolidone, PVPP; Polyplasdone XL; Kollidon CR

Chemical Name & CAS Number: 1-ethylene-2-pyrrolidinone homopolymer (9003-39-8)

Molecular formula: $(C_6H_9NO)_n$ $n > 1,000,000$

Description: White to off-white, free flowing, tasteless/odorless, hygroscopic powder

Properties: Insoluble in water, although rapidly swells and wicks fluid into the center of a dosage form facilitating disintegration; particle size is dependent on the grade and manufacturer.

Incompatibilities: No known incompatibilities

Uses: As a tablet/capsule disintegrant (super disintegrant due to its low, <2% use level)

OFFICIAL NAME: CYCLODEXTRIN

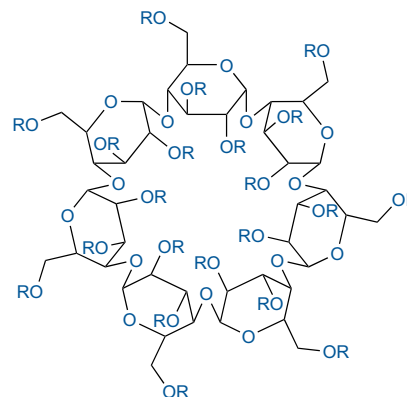
INN: β -cyclodextrin, betadex

Synonyms: cycloamylose; cycloglucan

Chemical Name & CAS Number: β -cyclodextrin (7585-39-9)

Molecular Formula: β -cyclodextrin consists of 7 glucose units connected in a circle

Structure:



Description: Cyclodextrin is a cyclic oligosaccharine occurring as a white, almost odorless crystalline powder, having a slightly sweet taste

Properties: solubility 1 in 50 parts of water, 1 in 200 parts of propylene glycol, practically insoluble in ethanol; 13–15% moisture; particle size ranges from 7 to 45 μ

Incompatibilities: Cyclodextrin can impact some antimicrobial agents, when both are in solution

Uses: As a solubility enhancer by way of inclusion complexation

OFFICIAL NAME: DEXTRATES

INN: Dextrate

Synonyms: Emdex, Corn Syrup Solids

Chemical Name & CAS Number: Dextrates (39404-33-6)

Molecular Formula: Purified mixture of saccharides (hydrated or anhydrous), with a dextrose equivalent of NLT 93% and NMT 99%

Description: A white, free-flowing, odorless powder that is sweet tasting. Dextrates are comprised of dextrose and 3–5% maltose and lesser amounts of higher dextrose oligomers (maltotriose, maltotetraose, etc.)

Properties: 1 in 1 parts in water, insoluble in ethanol; mean particle size of approximately 200 μ ; flow 9.3 g/sec

Incompatibilities: Incompatible with oxidizing agents and under the proper conditions primary and secondary amines

Uses: As a tablet diluent and sweetening agent for chewable tablets

OFFICIAL NAME: DEXTRIN

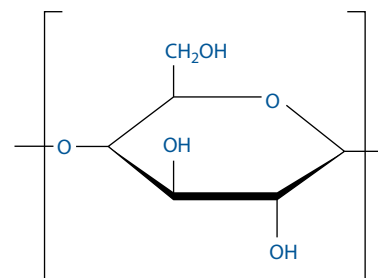
INN: Dextrin

Synonyms: British Gum; Starch Gum

Chemical Name & CAS Number: Dextrin (9004-53-9)

Molecular formula: $(C_6H_{10}O_5)_n$

Structure:



Description: White or yellow, amorphous powder (white: practically odorless; yellow: characteristic odor); dextrorotatory; does not reduce Fehling's solution; gives a reddish color with iodine

Properties: Soluble in boiling water (forming a gummy solution); less soluble in cold water; moisture content of 5%; density approximately 1.5 g/cc

Incompatibilities: Strong oxidizing agents

Uses: As a tablet diluent and binder; as a suspending agent

OFFICIAL NAME: DEXTROSE

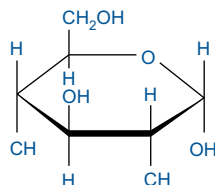
INN: Glucose; Glucose monohydrate; Dextrose; Glucosum monohydricum

Synonyms: Blood Sugar; Anhydrous Dextrose; Dextrose Monohydrate; D-(+)-Glucopyranose monohydrate; Medicinal Glucose; Purified Glucose; Grape Sugar; Bread Sugar; Cere-lose; Starch Sugar; Corn Sugar

Chemical Name & CAS Number: D-Glucose monohydrate (5996-10-1); anhydrous (50-99-7)

Molecular Formula: Monohydrate: $C_6H_{12}O_6 \cdot H_2O$

Structure:



Description: White, odorless, sweet tasting crystalline or granular powder

Properties: Solubility 1 g in 1 mL of water or 100 mL of alcohol; more soluble in boiling water or boiling alcohol; density of 1.5 g/cc; melting point 83°C

Incompatibilities: Aldehyde form interacts with amines, amides, amino acids, peptides, and proteins. Novobiocin sodium, warfarin sodium, and kanamycin sulfate

Uses: As a filler/binder in solid dosage forms; sweetening agent for liquids and chewable tablets; and a tonicity agent and vehicle in parenteral formulations

OFFICIAL NAME: DIFLUOROETHANE

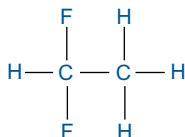
INN: None

Synonyms: HFC 152A; halocarbon 152A

Chemical Name & CAS Number: 1,1-Difluoroethane (75-37-6)

Molecular Formula: $C_2H_4F_2$

Structure:



Description: Liquified gas at room temperature under its own vapor pressure. The liquid is odorless and tasteless.

Properties: Boiling point -24.7°C; critical temperature 113.5°C; surface tension 11.25 dynes/cm; viscosity 0.243 mPa.s

Uses: As an aerosol propellant

OFFICIAL NAME: EDETIC ACID

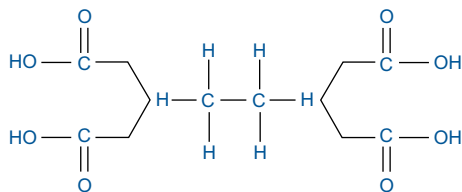
INN: Edetic Acid; Acidum edeticum

Synonyms: EDTA; Ethylenediaminetetraacetic acid

Chemical Name & CAS Number: N,N'-1,2-ethanediylbis[N-(carboxymethyl)glycine] (60-00-4)

Molecular Formula: $(HOOCCH_2)_2NCH_2CH_2N(CH_2COOH)_2$

Structure:



Description: White, crystalline powder

Properties: Very slightly soluble in water; soluble in solutions of alkali hydroxides; pH of a 0.2% solution 2.2

Incompatibilities: With polyvalent ions, alkali containing products, strong oxidizers

Uses: As a metal complexing agent

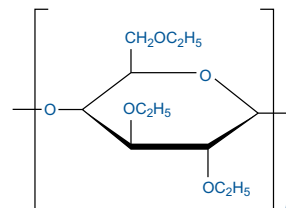
OFFICIAL NAME: ETHYLCELLULOSE

INN: Ethylcellulose

Chemical Name & CAS Number: Cellulose ethyl ether (9004-57-3)

Molecular Formula: $C_{12}H_{23}O_6(C_{12}H_{22}O_5)_n C_{12}H_{23}O_5$

Structure:



Description: Free-flowing, tasteless white to light tan powder

Properties: Practically insoluble in water, glycerin and propylene glycol; density 0.4g/cc; t_g 129–133°C

Uses: As a tablet binder for modified release tablets, film-coating tablets, and taste-masking barrier for chewable products

OFFICIAL NAME: ETHYL OLEATE

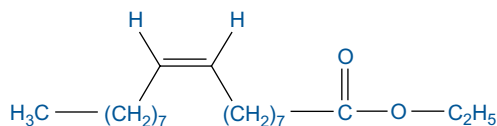
INN: Ethyl oleate; Ethylis oleas

Synonyms: Oleic acid, ethyle ester, Ethyl 9-octadecenoate

Chemical Name & CAS Number: (Z)-9-Octadecenoic acid, ethyl ester (111-62-6)

Molecular formula: $C_{20}H_{38}O_2$

Structure:



Description: Mobile, practically colorless liquid, with an agreeable taste

Properties: Specific gravity 0.866-0.874; acid value not greater than 0.5; iodine value 75–85; does not dissolve in water; miscible with vegetable oils, mineral oil, alcohol, or most organic solvents

Incompatibilities: May cause some types of rubber to dissolve or swell

Uses: As a vehicle for certain parenteral preparations

OFFICIAL NAME: FRUCTOSE

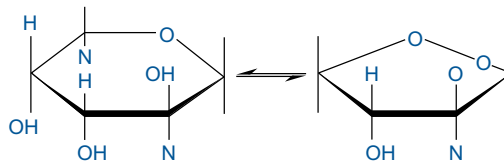
INN: Fructose; Fructosum

Synonyms: Fruit sugar, levulose, fructopyranose

Chemical Name & CAS Number: D-Fructose (58-48-7)

Molecular Formula: $C_6H_{12}O_6$

Structure:



Description: Odorless crystals or powder with a very sweet taste

Properties: pH 5.35 of a 9% solution; density 1.58g/cc; solubility: 1 part in .3 parts water, 15 parts of ethanol

Incompatibilities: Forms brown color in the presence of strong acids or alkalis, aldehyde can react with amines

Uses: As a sweetening agent in tablets and liquids

OFFICIAL NAME: FUMARIC ACID

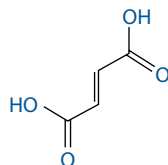
INN: Fumaric Acid

Synonyms: Allomaleic acid; boletic acid

Chemical Name & CAS Number: (E)-2-Butenedioic acid (110-17-8)

Molecular Formula: $C_4H_4O_4$

Structure:



Description: White odorless granules or crystals

Properties: Solubility 1 part in 200 of water, in 33 parts propylene glycol; saturated aqueous solution pH 2.45; density 0.93g/cc

Uses: As an acidifier in liquid products; flavor enhancer in some formulations; and a sialogogue

OFFICIAL NAME: GELATIN

INN: Gelatin

Synonyms: Gelatin

Chemical Name & CAS Number: Gelatin (9000-70-8)

Molecular Formula: Purified proteins obtained by acid hydrolysis (Type A) or alkaline hydrolysis (Type B) of animal collagen

Description: Slightly yellow to amber brittle solid mass that is tasteless. It can be sold in powdered or granular form.

Properties: 1% solutions have a pH of 3.8-6.9 for type A and 5.0-7.4 for Type B; iso-electric points are 7-9 and 4.7-5.3 for Types A and B, respectively; moisture content for both Types is 9-11%. Both types are soluble in hot water, cooling into a solid gel mass below 35°C; viscosities vary, depending on the concentration, water hardness, and temperature.

Incompatibilities: Gelatin will cross-link in the presence of aldehydes, resulting in poor product dissolution; will react with plasticizer resulting in altered performance (i.e., soft gel capsules); can degrade when exposed to acids and bases or prolonged heating of aqueous solutions; hydrolysis can occur by proteolytic enzymes.

Uses: As a base for the manufacture of hard and soft gel capsules, as a coating on tablets (Gel-tabs)

OFFICIAL NAME: GLYCERIN

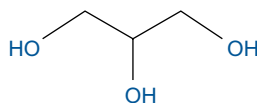
INN: Glycerol, Glycerin

Synonyms: Glycerine, trihydroxypropane glycerol

Chemical Name & CAS Number: 1,2,3-Propanetriol (56-81-5)

Molecular formula: $C_3H_8O_3$

Structure:



Description: Clear, colorless, syrupy liquid with a sweet taste and not more than a slight, characteristic odor, which is neither harsh nor disagreeable; when exposed to moist air, it absorbs water and such gases as H_2S and SO_2 ; solutions are neutral; specific gravity not less than 1.249 (not less than 95% $C_3H_5(OH)_3$); boils at about 290°C under 1 atm, with decomposition, but can be distilled intact in a vacuum

Properties: Miscible with water, alcohol; density 1.26 g/cc at room temperature; very hygroscopic

Incompatibilities: An explosion may occur, if it is triturated with strong oxidizing agents, such as chromium trioxide, potassium chlorate, or potassium permanganate. In dilute solutions, the reactions proceed at a slower rate, forming several oxidation products. Iron is an occasional contaminant and may be the cause of a darkening in mixtures containing phenols, salicylates, tannin, etc. With boric acid or sodium borate, it forms a complex, spoken of as glyceroboric acid, which is a much stronger acid than boric acid.

Uses: For its humectant and emollient activity in topical formulations; as a solvent in parenteral formulations and a plasticizer for gelatin and other polymeric coatings

OFFICIAL NAME: GLYCERYL MONOSTEARATE

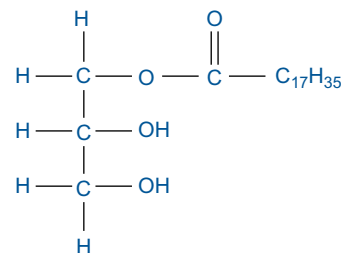
INN: Glycerol monostearate

Synonyms: GMS; stearic monoglyceride; monostearin; glycerol stearate

Chemical Name & CAS Number: Octadecanoic acid, monoester with 1,2,3-propanetriol (31566-31-1)

Molecular Formula: $C_3H_5(OH)_2C_{18}H_{35}O_2$

Structure:



Description: White, wax-like solid or occurs in the form of white, wax-like beads, or flakes; slight, agreeable, fatty odor and taste; affected by light

Properties: Insoluble in water, but may be dispersed in hot water; does not melt below 55°C; dissolves in hot organic solvents, such as alcohol, mineral, fixed oils, etc.; HLB value of 3.8

Incompatibilities: Most acidic substances

Note: There are available self-emulsifying grades of glycerol monostearate that contain an additional surfactant/emulsifier.

Uses: As a thickening and emulsifying/emollient agent for ointments; tablet and capsule lubricant; plasticizer/matrix forming agent in some sustained release products

OFFICIAL NAME: GUAR GUM

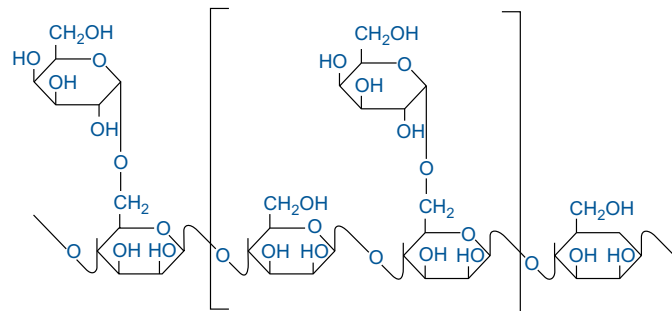
INN: Guar Gum, Guar galactomannan

Synonyms: Guar flour, jaguar gum

Chemical Name & CAS Number: Galactomannan polysaccharide (9000-30-0)

Molecular Formula: $(C_6H_{12}O_6)_n$

Structure:



Description: Odorless, white to off-white, bland tasting powder

Properties: 1% dispersion pH 5.0-7.0; density 1.492 g/cc; insoluble in hot or cold water, but will swell to form a thixotropic gel

Incompatibilities: Alcohol, acetone, tannins, strong acid, or alkalis will cause the gel to break up; guar interferes with the absorption of penicillins

Uses: As a binder or disintegrant in tablets and capsules; suspending and thickening agent for liquids

OFFICIAL NAME: HYDROCHLORIC ACID

INN: Hydrochloric acid

Synonyms: Chlorhydric Acid; Muriatic Acid

Chemical Name & CAS Number: Hydrochloric acid (7647-01-0)

Molecular Formula: HCl

Description: Colorless, fuming liquid; pungent odor; fumes

Properties: pH 0.1 (10% v/v solution), miscible with water soluble in alcohol

Incompatibilities: reacts violently with alkalis resulting in release of large amounts of heat

Uses: As an acidifying agent

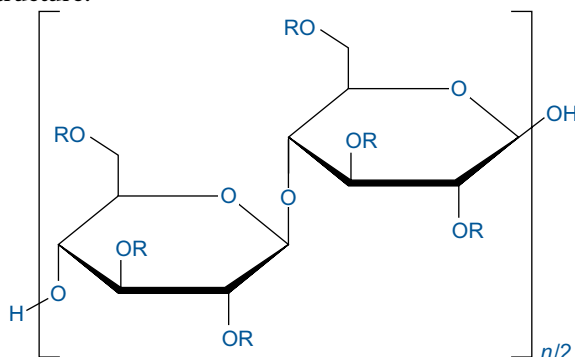
OFFICIAL NAME: HYDROXYETHYL CELLULOSE

INN: Hydroxyethylecellulose; Hydroxyethyl cellulose

Synonyms: HEC; Cellulose, hydroxyethyl ether

Chemical Name & CAS Number: Cellulose, 2-hydroxyethyl ether (9004-62-0)

Structure:



R is H or $[-CH_2CH_2O-]_m H$ where m is a common integral number of cellulose derivatives.

Description: White to off-white, odorless, tasteless, hygroscopic powder

Properties: 1% solution pH 5.5–8.5 (depending upon grade); density approximately 0.6g/cc; moisture content of NMT 5%; particles size can be grade specific; soluble in both hot and cold water, creating a clear solution; viscosity varies, depending upon grade

Incompatibilities: Zein, sodium pentachlorophenate

Uses: Pharmaceutically, as a thickener, protective colloid, binder, stabilizer, and suspending agent in emulsions, ointments, lotions, ophthalmic solutions, suppositories, and tablets; has been used as a controlled release agent in tablets

OFFICIAL NAME: HYDROXYPROPYL CELLULOSE

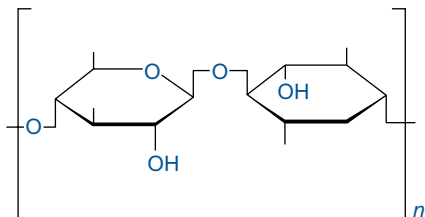
INN: Hydroxypropylcellulose; hydroxypropyl cellulose

Synonyms: Cellulose, hydroxypropyl ether; HPC; hyprollose; hydroxypropylated cellulose

Chemical Name & CAS Number: Cellulose, 2-hydroxypropyl ether (9004-64-2)

Description: Off-white, odorless, tasteless powder; softens at 130°C; burns completely at about 475°C in N_2 or O_2 ; refractive index (2% solution) about 1.337; pH (aqueous solution) 5–8.5; solutions are nonionic

Structure:



R is H or $[CH_2CH(CH_3)O]_m H$ where m is a common integral number of cellulose derivatives.

Properties: Soluble in water below 38°C (insoluble above 45°C); soluble in many polar organic solvents; pH 5.0–8.5 for 1% aqueous solution; density 0.5g/cc.

Uses: As a binder, granulating agent, and film-coating agent in the manufacture of tablets; an alcohol-soluble thickener and suspending agent for elixirs and lotions; and a stabilizer for emulsions

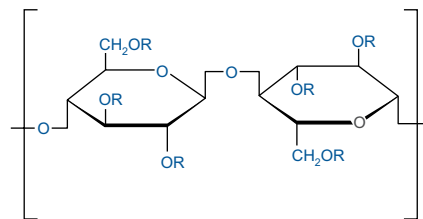
OFFICIAL NAME: HYDROXYPROPYL METHYLCELLULOSE (HYPROMELLOSE)

INN: Hypromellose; Hydroxypropyl methylcellulose; Hydroxypropylmethylcellulose

Synonyms: HPMC; MHPC

Chemical Name & CAS Number: Cellulose, 2-hydroxypropyl methyl ether (9004-65-3)

Structure:



where R is H, CH_3 , or $CH_3CH(OH)CH_2$

Description: White to slightly off-white, fibrous or granular, free-flowing powder

Properties: Swells in water and produces a clear to opalescent, viscous, colloidal mixture; undergoes reversible transformation from sol to gel on heating and cooling, respectively. 1% aqueous solution pH 5.5–8.0; density 1.326g/cc; particle size and viscosity will vary, depending upon the grade

Uses: As a protective colloid that is useful as a dispersing and thickening agent, also used in the preparation of sustained release matrix tablets and as a film coating material

OFFICIAL NAME: ISOPROPYL MYRISTATE

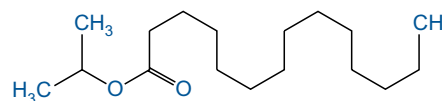
INN: Isopropyl Myristate

Synonyms: Isopropyl ester of myristic acid; myristic acid isopropyl ester; tetradecanoic acid, 1-methylethyl ester

Chemical Name & CAS Number: 1-Methylethyl tetradecanoate (110-27-0)

Molecular formula: $CH_3(CH_2)_{12}COOCH(CH_3)_2$

Structure:



Description: Liquid of low viscosity; practically colorless and odorless; congeals at about 5°C and decomposes at 208°C; withstands oxidation; and does not readily become rancid

Properties: Soluble in alcohol, acetone, chloroform, ethyl acetate, toluene, mineral oil, castor oil, or cottonseed oil; practically insoluble in water, glycerin, or propylene glycol; dissolves many waxes, cholesterol, or lanolin; viscosity 5–7 mPa s at 25°C

Use: As a pharmaceutical aid used in cosmetics and topical medicinal preparations, as an emollient, as a lubricant, and to enhance absorption through the skin

OFFICIAL NAME: KAOLIN

INN: Heavy Kaolin, Kaolin

Synonyms: Bolus alba; china clay; kaolinite; porcelain clay; weisser ton; white bole

Chemical Name & CAS Number: Hydrated aluminum silicate (1332-58-7)

Molecular Formula: $Al_2H_4O_9Si_2$

Structure: $Al_2O_3 \cdot 2SiO_2 \cdot 2H_2O$

Description: White to off-white powder that develops an odor of clay when moistened

Properties: pH 4.0–7.5 for 20% slurry; particle size 0.6–0.8 μ

Incompatibilities: Kaolin has been known to influence to absorption of certain drugs: amoxicillin, cimetidine, digoxin, and phenytoin to name a few

Uses: As a diluent in tablets and capsules; as a suspension vehicle in liquids

OFFICIAL NAME: LACTIC ACID

INN: Lactic acid; Acidum lacticum

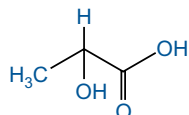
Synonyms: 2-hydroxypropanoic acid; α -hydroxypropionic acid; 2-hydroxypropionic acid; racemic lactic acid

Chemical Name & CAS Number: 2-Hydroxypropionic acid (50-21-5)

Synonyms: Propanoic acid, 2-hydroxy-, 2-Hydroxypropionic Acid; Propanoic Acid; Milk Acid

Molecular formula: $\text{CH}_3\text{CH}(\text{OH})\text{COOH}$

Structure:



Description: Colorless or yellowish, nearly odorless, syrupy liquid

Properties: Miscible with water, alcohol, or ether; insoluble in chloroform; boiling point 122°C; specific gravity 1.21; viscosity of 85% aqueous solution 28.5 mPa s

Incompatibilities: Oxidizing agents, iodides, and albumin. Reacts violently with hydrofluoric acid and nitric acid

Uses: As an acidifier; substrate for biodegradable microspheres (poly-lactic/poly glycolic acid co-block polymers)

OFFICIAL NAME: LACTOSE

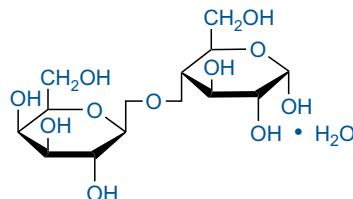
INN: Lactose; Lactose monohydrate; Lactosum

Synonyms: Milk sugar

Chemical Name & CAS Number:

O- β -D-Galactopyranosyl-(1-4)- α -D-glucopyranose anhydrous (63-42-3)O- β -D-Galactopyranosyl-(1-4)- α -D-glucopyranose monohydrate (64044-51-5)Molecular Formula: $\text{C}_{12}\text{H}_{22}\text{O}_{11} \cdot \text{H}_2\text{O}$

Structure:



Description: White or creamy white, hard, crystalline masses or powder; odorless; faintly sweet taste

Properties: 1 g in 5 mL water or 2.6 mL boiling water; very slightly soluble in alcohol; pH 4–6.5 of a 10% aqueous solution; density 1.552g/cc; moisture of 1% for anhydrous and NMT 5% for monohydrate

Incompatibilities: Maillard reaction with amine containing compounds; amino acids, aminophylline and amphetamines

Uses: As a diluent in tablet and capsule formulations. The amorphous and mono-hydrate forms are used in wet granulation processing of materials, whereas the spray-dried anhydrous type is used in direct compression formulations. It is an ingredient of the medium used in penicillin production. It is used extensively as an addition to milk for infant feeding.

OFFICIAL NAME: LANOLIN

INN: Wool Fat; Purified lanolin; Adeps lanae; Lanolin

Synonyms: Lanolin anhydrous; purified lanolin; refined wool fat

Chemical Name & CAS Number: Anhydrous Lanolin (8006-54-0)

Description: Yellowish white, ointment-like mass, with a slight, characteristic odor

Properties: Insoluble in water; soluble in chloroform or ether with separation of its water of hydration; density 0.932-0.945 g/cc at 15°C

Uses: As an emulsifying agent, ointment base

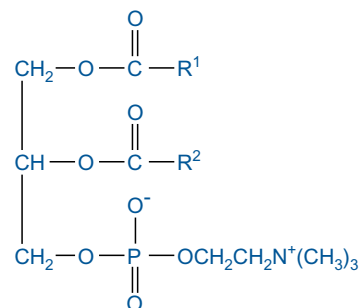
OFFICIAL NAME: LECITHIN

INN: Lecithin

Synonyms: Egg lecithin; mixed soybean phosphatides; ovolécithin; soybean phospholipids

Chemical Name & CAS Number: Lecithin (8002-43-5)

Structure:



Description: Varies greatly, depending on the free fatty acid content; could be a free flowing powder or a viscous dark brown to black syrup-like liquid; changes color (oxidation) quickly when exposed to air

Properties: Powder: density of 0.5g/cc, insoluble in water and polar solvents; liquid: density 0.97 g/cc

Incompatibilities: Will oxidize when exposed to air; hydrolyzes when exposed to esterases

Uses: As an emollient, emulsifying agent, and solubilizing agent

OFFICIAL NAME: MAGNESIUM STEARATE

INN: Magnesium stearate; magnesia stearas

Synonyms: Magnesium octadecanoate, stearic acid magnesium salt; octadecanoic acid magnesium salt

Chemical Name & CAS Number: Octadecanoic acid magnesium salt (557-04-0)

Molecular Formula: $\text{C}_{36}\text{H}_{70}\text{MgO}_4$ Structure: $[\text{CH}_3(\text{CH}_2)_{16}\text{COO}]_2\text{Mg}$

Description: Fine, white, bulky powder; faint, characteristic odor; unctuous, adheres readily to the skin and free from grittiness

Properties: Insoluble in water and alcohol; density of 1.092g/cc; very poorly flowing; surface area 1.6–14.8 m²/g.

Uses: As tablet and capsule lubricant

OFFICIAL NAME: MALTODEXTRIN

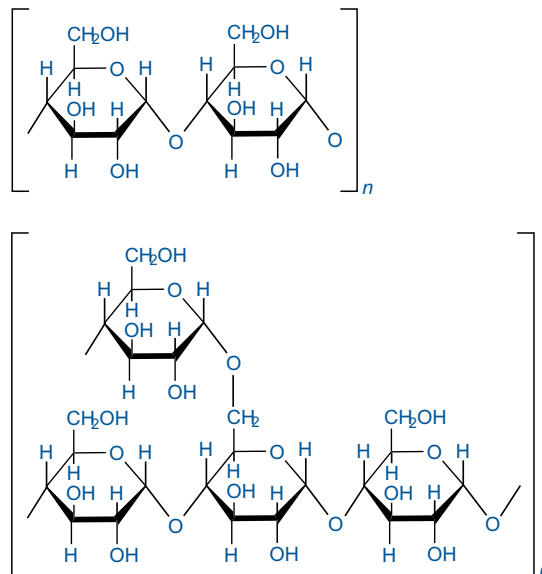
INN: Maltodextrin

Synonyms: Numerous trade names

Chemical Name & CAS Number: Maltodextrin (9050-36-6)

Molecular Formula: $(\text{C}_6\text{H}_{10}\text{O}_5)_n \cdot \text{H}_2\text{O}$

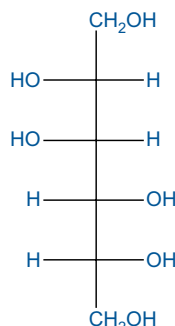
Structure:



Description: Non-sweet, odorless, white powder or granules
Properties: Density range 0.13-0.58 g/cc, depending on manufacturer; freely soluble in water; slightly soluble in ethanol; viscosity varies, depending on manufacturer
Incompatibilities: Certain conditions might result in a discoloration with amino acids and amines
Uses: As a coating agent for tablets; tablet and capsule diluent; tablet binder and viscosity modifier

OFFICIAL NAME: MANNITOL

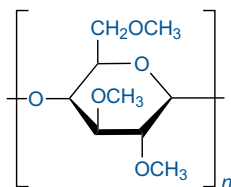
INN: Mannitol; d-Mannite; Mannitolum
Synonyms: Cordycepic acid; d-mannitol; manna sugar; mannite
Chemical Name & CAS Number: D-Mannitol (69-65-8)
Molecular Formula: $C_6H_{14}O_6$
Structure:



Description: White, odorless crystalline powder or free flowing granules with a sweet taste and a cooling effect
Properties: Density 1.514 g/cc; heat of solution -120.9J/g; particle size varies, depending on manufacturer; solubility in water is 1 in 5.5, in ethanol 1 in 83, and glycerin is 1 in 18.
Incompatibilities: None reported in dry applications; at concentrations above 20% in solution, salting effects might be seen with potassium or sodium
Uses: As a sweetening agent for chewable tablets and oral liquids; tablet/capsule diluent; tonicity agent; vehicle for lyophilized products

OFFICIAL NAME: METHYLCELLULOSE

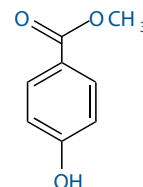
INN: Methylcellulose; Methylcellulosum
Synonyms: Benecel; MC; Methocel
Chemical Name & CAS Number: Cellulose methyl ether (9004-67-5)
Molecular Formula: Long chain substituted cellulose, where 27-32% of the hydroxyl groups are in the form of methyl ether
Structure:



Description: White to off-white, odorless, tasteless granule or powder
Properties: pH of 1% solution 5.5-8.0; density 1.341 g/cc; insoluble in hot water (disperses), in cold water disperses to form a colloidal dispersion that is clear
Incompatibilities: High concentrations of electrolytes increase viscosity; incompatible with aminacrine hydrochloride, chlorocresol, mercuric chloride, phenol, tannic acid, and paraben preservatives
Uses: As a protective colloid for many types of dispersed substances; an effective stabilizer for oil-in-water emulsions; a granulating solution for tablets and capsules. It has been formulated into many coating systems.

OFFICIAL NAME: METHYL PARABEN

INN: Methyl hydroxybenzoate; methyl parahydroxybenzoate; Methylis parahydroxybenzoas
Synonyms: 4-hydroxybenzoic acid methyl ester; methyl-*p*-hydroxybenzoate
Chemical Name & CAS Number: Methyl 4-hydroxybenzoate (99-76-3)
Molecular Formula: $C_8H_8O_3$
Structure:



Description: White to off-white crystalline powder that has a slight burning taste
Properties: Antimicrobial activity against most organisms, most effective in pH range of 4-8; density of 1.352 g/cc; pK_a 8.4 at 22°C; solubility in water 1 in 400, in propylene glycol 1 in 5, in glycerin 1 in 60, in ethanol 1 in 2
Incompatibilities: Nonionic surfactants (e.g., polysorbate), talc, bentonite, tragacanth gum, sodium alginate, sorbitol, and methylcellulose
Uses: As an antimicrobial preservative in all known delivery systems

OFFICIAL NAME: MINERAL OIL

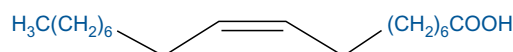
INN: Liquid paraffin; Paraffin liquidum; Mineral oil
Synonyms: Mineral hydrocarbons; heavy mineral oil; liquid petrolatum; paraffin oil; white mineral oil
Chemical name & CAS Number: Mineral oil (8012-95-1)
Molecular Formula: A refined mixture of liquid hydrocarbons (C_{14} to C_{18}) obtained from petroleum
Description: Colorless, transparent, oily liquid, odorless and tasteless when cold, and develops not more than a faint odor of petroleum when heated
Properties: Insoluble in water, glycerin, or alcohol; miscible with most fixed oils, but not with castor oil; soluble in volatile oils; specific gravity 0.818-0.880; kinematic viscosity not more than 33.5 centistokes at 40°C.
Incompatibilities: Strong oxidizing agents
Uses: As an emulsifying agent and emollient in topical formulations mostly ointments

OFFICIAL NAME: NITROGEN

INN: Nitrogen; Nitrogenium
Synonyms: Azote
Chemical Name & CAS Number: Nitrogen [7727-37-9]
Molecular Formula: N_2
Description: Naturally occurring in the atmosphere at about 78%; Nonreactive noncombustible, colorless, tasteless, and odorless gas
Properties: Insoluble in water, unless under pressure; non-flammable; density 0.967 g/cc
Uses: To replace air in the containers of substances that would be affected adversely by air oxidation (nitrogen layer packaging); as a propellant for aerosol drug delivery

OFFICIAL NAME: OLEIC ACID

INN: Oleic acid; Acidum oleicum
Synonyms: Elaic acid; *cis*-9-octadecenoic acid; 9,10-octadecenoic acid; olenic acid
Chemical Name & CAS Number: (Z)-9-Octadecenoic acid (112-80-1)
Molecular Formula: $C_{18}H_{34}O_2$
Structure:



Description: Colorless to pale yellow, oily liquid; lard-like odor and taste

Properties: Specific gravity 0.889-0.895; congeals at a temperature not above 10°C; pure acid solidifies at 4°C; at atmospheric pressure, it decomposes when heated at 80–100°C; on exposure to air, it gradually absorbs oxygen, darkens, and develops a rancid odor; practically insoluble in water; miscible with alcohol, chloroform, ether, benzene, or fixed and volatile oils

Incompatibilities: Reacts with alkali to form soaps. Heavy metals and calcium salts form insoluble oleates. Iodine solutions are decolorized by formation of the iodine addition compound of the acid. It is oxidized to various derivatives by nitric acid, potassium permanganate, and other agents.

Uses: As an emulsifying agent for topical; a penetration enhancer for poorly soluble actives (both topical and oral); a vehicle in soft gelatin capsules

OFFICIAL NAME: OLEYL ALCOHOL

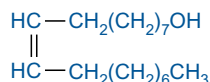
INN: Alcohol oleicus

Synonyms: *cis*-9-octadecen-1-ol; oleic alcohol; oleo alcohol; oleol

Chemical name & CAS Number: (*Z*)-9-Octadecen-1-ol (143-28-2)

Molecular Formula: C₁₈H₃₆O

Structure:



Description: Clear, colorless to light yellow, oily liquid; faint characteristic odor and bland taste; iodine value between 85 and 90; hydroxyl value between 205 and 215.

Properties: Soluble in alcohol, ether, isopropyl alcohol, or light mineral oil; insoluble in water; boiling point 182–184°C.

Use: As an emollient, emulsifying agent

OFFICIAL NAME: PARAFFIN

INN: Hard Paraffin; Paraffin; Paraffin solidum

Synonyms: Mineral hydrocarbons; hard wax; paraffin durum; paraffin wax

Chemical Name & CAS Number: Paraffin (8002-74-2)

Molecular Formula: Purified mixture of saturated hydrocarbons having the formula C_nH_{2n+2}

Description: Colorless or white, more or less translucent mass with a crystalline structure; slightly greasy to the touch; odorless and tasteless

Properties: Freely soluble in chloroform, ether, volatile oils, or most warm fixed oils; slightly soluble in dehydrated alcohol; insoluble in water or alcohol; congeals 47–65°C; density 0.84-0.89 g/cc

Incompatibilities: No known incompatibilities

Uses: As an ointment base

OFFICIAL NAME: PEANUT OIL

INN: Arachis Oil; Peanut Oil; Arachis oleum

Synonyms: Arachis Oil; Groundnut Oil; Nut Oil; Earth Nut Oil; katchung oil

Chemical Name & CAS Number: Peanut Oil (8002-03-7)

Molecular Formula: A mixture of mono-, di- and tri-glyceride esters of fatty acids comprising a combination of arachidic, behenic, palmitic, stearic, lignoceric, linoleic, and oleic acids

Description: Colorless or pale yellow, oily liquid, with a characteristic nutty odor and a bland taste

Properties: Insoluble in water; very slightly soluble in alcohol; miscible with ether, chloroform, or carbon disulfide; specific gravity 0.912-0.920; viscosity approximately 35 mPa s at 37°C

Incompatibilities: Aalkali hydroxides will saponify the oil

Uses: As a solvent for sustained release IM injections

OFFICIAL NAME: PECTIN

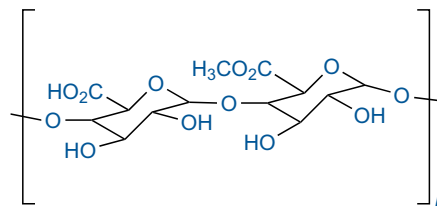
INN: Pectin

Synonyms: Citrus pectin; methopectin; methyl pectin; methyl pectinate; pectinic acid

Chemical Name & CAS Number: Pectin (9000-65-5)

Molecular Formula: Repeating linear units of (1-4)-linked α-D-galactopyranosyluronic acid units with some carboxyl groups esterified with methanol

Structure:



Description: Off-white, odorless powder that has a mucilaginous feel when moistened

Properties: Soluble in water; insoluble in most organic solvents; slightly acidic in solution; solubilizes more easily when mixed with glycerin or 1:3 parts of sucrose to prevent clumping

Incompatibilities: High concentration of alcohol cause precipitation; hydrolysis in the presence of alkalis; high temperatures promotes hydrolysis

Uses: As an emulsifying agent, thickening agent for suspensions

OFFICIAL NAME: PETROLATUM

INN: Yellow soft paraffin; Yellow petrolatum; Petrolatum

Synonyms: Yellow Soft Paraffin; Amber Petrolatum; Yellow Petrolatum; Petroleum Jelly; Paraffin Jelly; Mineral Jelly; vaselinum flavum

Chemical Name & CAS Number: Petrolatum (8009-03-8)

Molecular Formula: Purified mixture of saturated hydrocarbons with the general formula C_nH_{2n+2}

Description: Unctuous mass of yellowish to light amber color; not more than a slight fluorescence after being melted; transparent in thin layers; free or nearly free from odor and taste

Properties: Insoluble in water; almost insoluble in cold or hot alcohol or in cold dehydrated alcohol; specific gravity 0.815-0.880 at 60°C; melts between 38 and 60°C

Uses: As an ointment base, emollient

OFFICIAL NAME: PHENOL

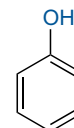
INN: Phenol; Phenolum

Synonyms: Carboic Acid; hydroxybenzene; oxybenzene; phenic acid; phenyl hydrate; phenyl hydroxide; phenylic acid; phenylic alcohol

Chemical Name & CAS Number: Phenol (108-95-2)

Molecular Formula: C₆H₆O

Structure:



Description: Colorless to light pink, interlaced, or separate, needle-shaped crystals, or a white or light pink, crystalline mass; characteristic odor; when undiluted, it whitens and cauterizes the skin and mucous membranes; when gently heated, phenol melts, forming a highly refractive liquid; liquefied by the addition of 10% of water

Properties: Solubility 1 g in 15 mL water; very soluble in alcohol, glycerin, chloroform, ether, or fixed and volatile oils; sparingly soluble in mineral oil; vapor is flammable; gradually darkens on exposure to light and air; specific gravity 1.07; boils at 182°C; congeals at not less than 39°C

Incompatibilities: Produces a liquid or soft mass when triturated with camphor, menthol, acetanilide, acetophenetidin,

aminopyrine, antipyrine, ethyl aminobenzoate, methenamine, phenyl salicylate, resorcinol, terpin hydrate, thymol, and several other substances, including some alkaloids. It also softens cocoa butter in suppository mixtures. Traces of iron in various chemicals (such as alum, borax, etc. may produce a pink to violet color.

Uses: As a caustic, disinfectant, topical anesthetic; a preservative for some injections; used in several proprietary antiseptic mouthwashes, hemorrhoidal preparations, and burn remedies

OFFICIAL NAME: PHENYLETHYL ALCOHOL

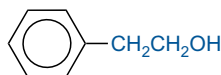
INN: Phenylethyl alcohol

Synonyms: Benzeneethanol; phenethanol; benzyl carbinol; benzylmethanol; β -hydroxyethyl benzene; 2-phenylethyl alcohol; phenylethanol; PEA

Chemical Name & CAS Number: Phenethyl alcohol [60-12-8]

Molecular Formula: $C_8H_{10}O$

Structure:



Description: Colorless liquid with a rose-like odor and a sharp, burning taste

Properties: Solubility is 1 g in 60 mL water or 1 mL alcohol, chloroform, or ether; very soluble in fixed oils, glycerin, or propylene glycol; slightly soluble in mineral oil; solidifies at 27°C; specific gravity 1.017-1.020

Uses: As an antimicrobial preservative

OFFICIAL NAME: PHOSPHORIC ACID

INN: Phosphoric acid; Acidum phosphoricum concentratum

Synonyms: Orthophosphoric Acid; Syrupy Phosphoric Acid; Concentrated Phosphoric Acid; hydrogen phosphate

Chemical Name & CAS Number: Orthophosphoric acid (7664-38-2)

Molecular Formula: H_3PO_4

Description: Colorless, odorless liquid of a syrupy consistency

Properties: Miscible with water or alcohol, with the evolution of heat; specific gravity about 1.71; boiling point 118°C

Uses: As an acidulant

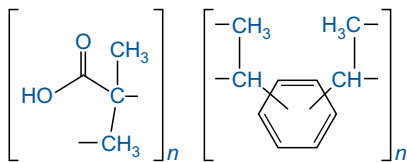
OFFICIAL NAME: POLYACRILIN POTASSIUM

INN: Polacrilin potassium

Synonyms: Methacrylic acid copolymer with divinylbenzene, potassium salt; polacrilinum kali

Chemical Name & CAS Number: 2-Methyl-2-propenoic acid polymer with divinylbenzene, potassium salt (39394-76-5)

Structure:



Description: White to off-white, odorless, tasteless, free flowing powder; bitter taste in an aqueous solution

Properties: Particle size range from 50 to 250 μ ; practically insoluble in water and most other liquids; swells rapidly in the presence of water

Incompatibilities: Strong oxidizing agents; amines (particularly tertiary amines)

Uses: As a tablet disintegrant; taste masker; complexing agent; sustained release ion exchange resin

OFFICIAL NAME: POLOXAMER

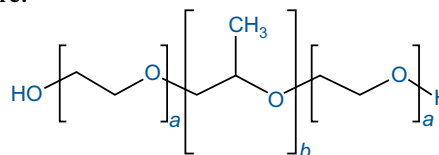
INN: Poloxamer

Synonyms: Poloxalkol; polyethylene-propylene glycol copolymer; polyoxyethylene-polyoxypropylene copolymer

Chemical Name & CAS Number: α -Hydroxyhydroxypoly(oxyethylene)poly(oxypropylene) poly(oxyethylene) block copolymer (9003-11-6)

Molecular Formula: $HO(C_2H_4O)_a(C_3H_6O)_b(C_2H_4O)_aH$

Structure:



Description: Waxy appearing, white colored, freely flowing prilled granules

Properties: 2.4% aqueous solution pH of 5.0–7.4; density 1.06 g/cc; HLB value 0.5–30, depending on grade; moisture content approximately 0.5%; all grades are freely soluble in water, solubility in other solvents is grade dependent

Incompatibilities: Phenols and parabens

Uses: As nonionic emulsifiers and solubilizing agents; tablet binders and coating agents; thermally inverse gelling agents (grade dependent)

OFFICIAL NAME: POLYETHYLENE GLYCOLS

INN: Macrogol; Marcogolum; Polyethylene glycol

Synonyms: PEG; polyoxyethylene glycol

Chemical Name & CAS Number: α -Hydroxyhydroxypoly(oxy-1,2-ethanediyl) (25322-68-3)

Molecular Formula: $HOCH_2(CH_2OCH_2)_mCH_2OH$

Description: Polyethylene glycols 200, 300, 400, and 600 are clear, viscous liquids at room temperature. Polyethylene glycols 900, 1000, 1450, 3350, 4500, 6000, 8000 and 20000 are white, waxy solids.

Properties: All members of this class dissolve in water to form clear solutions and are soluble in many organic solvents. As their molecular weight increases, water solubility, vapor pressure, hygroscopicity, and solubility in organic solvents decrease, whereas freezing or melting range, specific gravity, flash point, and viscosity increase.

Incompatibilities: Some colors may react with these materials, due to the terminal hydroxyl groups; reduced penicillin activity occurs with these materials as is preservative effectiveness of parabens

Uses: As a viscosifier for liquid products; a solubilizer for topical products; a vehicle for some parenteral products; an emollients and suppository bases in semi-solid products; a tablet/capsule lubricant; a solubilizing agent for oral drug delivery.

OFFICIAL NAME: POLYOXYETHYLENE SORBITAN FATTY ACID ESTERS

INN: Polysorbate; Polysotbatum

Synonyms: Crillet; Alkamuls; Atlas; Tween

Chemical Name & CAS Number:

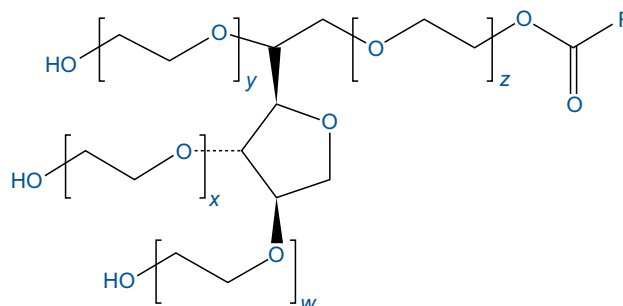
Polyoxyethylene 20 Sorbitan monolaurate (9005-64-5)

Polyoxyethylene 20 Sorbitan monopalmitate (9005-66-7)

Polyoxyethylene 20 Sorbitan monostearate (9005-67-8)

Polyoxyethylene 20 Sorbitan monooleate (9005-65-6)

Structure: Sorbitan esters, poly(oxy-1,2-ethanediyl) derivatives



Description: All polysorbates (20,40,60 & 80) are slightly yellow, oily liquids exhibiting a faint, characteristic odor; and somewhat bitter taste

Properties: Very soluble in water, producing an odorless and nearly colorless solution; soluble in alcohol, cottonseed oil, corn oil, ethyl acetate, methanol, or toluene; insoluble in mineral oil; specific gravity 1.07-1.09; pH (1:20 aqueous solution) 6-8; viscosity range 400-600 mPa s

Uses: As emulsifying agents, forming O/W emulsions; wetting agents in coating applications; solubilizing agents for oil-soluble APIs

OFFICIAL NAME: POLYVINYL ALCOHOL

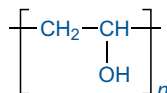
INN: Polyvinyl alcohol

Synonyms: PVA; vinyl alcohol polymer

Chemical Name & CAS Number: Ethenol, homopolymer (9002-89-5)

Molecular Formula: (C₂H₄O)_n

Structure:



Description: White to cream-colored powder or granules; odorless

Properties: Freely soluble in water; insoluble in organic solvents; solution effected more rapidly at somewhat elevated temperatures; viscosity is grade dependant

Use: As a suspending agent; a lubricant and protectant in ophthalmic and nasal preparations, such as artificial tears, contact-lens cleaners, and nasal sprays

OFFICIAL NAME: POTASSIUM SORBATE

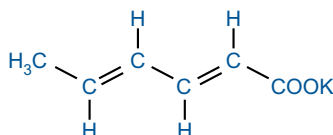
INN: Potassium sorbate; Kalii sorbas

Synonyms: 2,4-hexadienoic acid (E,E)-potassium salt; potassium (E,E) sorbate

Chemical Name & CAS Number: Potassium (E,E)-hexa-2,4-dienoate (24634-61-5)

Molecular Formula: C₆H₇KO₂

Structure:



Description: White crystals or powder with a characteristic odor

Properties: Solubility 1 g in 1.7 mL water, 1.8 mL of propylene glycol or 35 mL alcohol; density 1.36 g/cc; melting point ca. 270°C with decomposition

Uses: As a preservative

OFFICIAL NAME: POVIDONE

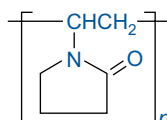
INN: Povidone; Polyvidonum

Synonyms: poly[1-(2-oxo-1-pyrrolidinyl)ethylene]; polyvidone; polyvinylpyrrolidone; PVP; 1-vinyl-2-pyrrolidinone polymer

Chemical Name & CAS Number: 1-ethenyl-2-Pyrrolidinone homopolymer (9003-39-8)

Molecular Formula: (C₆H₉NO)_n

Structure:



Description: White to creamy white, odorless powder, hygroscopic powder

Properties: Soluble in water, alcohol, or chloroform; insoluble in ether; pH (1 in 20 solution) 3-7; density 1.18 g/cc; melting point 150°C; particle size is grade dependant

Uses: As a suspending agent in liquid products and a binder in wet granulation processes

OFFICIAL NAME: PROPYLENE GLYCOL

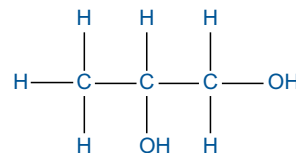
INN: Propylene glycol; Propyleneglycolum

Synonyms: 1,2-Dihydroxypropane; 2-hydroxypropanol; methyl ethylene glycol; methyl glycol; propane-1,2-diol

Chemical Name & CAS Number: 1,2-Propanediol (57-55-6)

Molecular Formula: CH₃CH(OH)CH₂OH

Structure:



Description: Clear, colorless, viscous, and practically odorless hygroscopic liquid; with a slightly acid taste

Properties: Miscible with water, alcohol, acetone, or chloroform; soluble in ether; dissolves many volatile oils; immiscible with fixed oils; specific gravity 1.035-1.037; viscosity approximately 58 mPa s

Uses: As a solvent; preservative; humectant

OFFICIAL NAME: PROPYL PARABEN

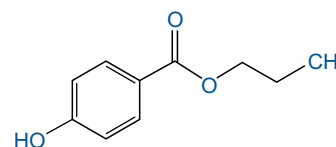
INN: Propyl hydroxybenzoate; Propyl parahydroxybenzoate; Propylis parahydroxybenzoas; Propylparaben

Synonyms: 4-hydroxybenzoic acid propyl ester; propagin; propyl p-hydroxybenzoate

Chemical Name & CAS Number: Propyl 4-hydroxybenzoate (94-13-3)

Molecular Formula: C₁₀H₁₂O₃

Structure:



Description: White crystalline odorless and tasteless powder

Properties: Very poorly soluble in water, 1 in 250 of glycerin, 1 in 4 of propylene glycol; density 1.288 g/cc

Incompatibilities: Nonionic surfactants (incorporation into micelles); may be absorbed by some plastics and discolors in the presence of iron

Use: As an antimicrobial agent

OFFICIAL NAME: SACCHARIN

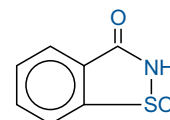
INN: Saccharin; Saccharinum

Synonyms: 1,2-Benzisothiazolin-3-one 1,1-dioxide; benzoic sulfimide; benzosulfimide; saccharin insoluble; o-Benzosulfimide

Chemical Name & CAS Number: 1,2-Benzisothiazol-3(2H)-one, 1,1-dioxide (81-07-2)

Molecular Formula: C₇H₅NO₃S

Structure:



Description: White crystalline, odorless powder; Calcium and Sodium salts have a similar description and use

Properties: Solubility 1 g in 290 mL water, 31 mL alcohol, or 25 mL boiling water; 0.35% aqueous solution pH 2.0; moisture content 0.1%

Incompatibilities: Large molecules resulting in precipitates

Uses: As a sweetening agent, approximately 500 times more sweet than sucrose

OFFICIAL NAME: SODIUM LAURYL SULFATE

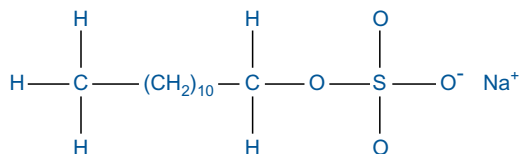
INN: Sodium lauryl sulfate; Natrii laurilsulfas

Synonyms: Dodecyl sodium sulfate; sodium dodecyl sulfate

Chemical Name & CAS Number: Sulfuric acid monododecyl ester sodium salt (151-21-3)

Molecular Formula: $C_{12}H_{25}NaO_4S$

Structure:



Description: Small, white or light yellow crystals, having a slight, characteristic odor of soap

Properties: Solubility 1 g in 10 mL water; 1% solution pH 7.0–9.5; density 1.07 g/cc; moisture content <5%

Incompatibilities: Reacts with cationic surface-active agents with loss of activity, even in concentrations too low to cause precipitation; mildly corrosive to steel, copper, brass, and aluminum

Uses: As an anionic surfactant

OFFICIAL NAME: SODIUM METABISULFITE

INN: Sodium metabisulfite; Natrii metabisulfis

Synonyms: Disodium disulfite; disodium pyrosulfite; sodium pyrosulfite

Chemical Name & CAS Number: Sodium pyrosulfite (7681-57-4)

Molecular Formula: $Na_2S_2O_5$

Description: White crystals or white to yellowish crystalline powder with an odor of sulfur dioxide

Properties: Solubility 1 g in 2 mL water; slightly soluble in alcohol; freely soluble in glycerin; 5% solution pH 3.5–5.0; melting point 150°C

Incompatibilities: Phenylmercuric acetate, possible rubber caps of multi-dose vials

Uses: As an antioxidant

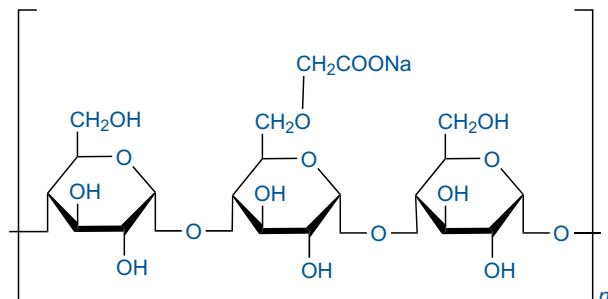
OFFICIAL NAME: SODIUM STARCH GLYCOLATE

INN: Sodium starch glycolate; Carboxymethylamylum natrium

Synonyms: Carboxymethyl starch, sodium salt; SSG, Explotab, Primogel

Chemical name & CAS Number: Sodium carboxymethyl starch (9036-38-1)

Structure:



Description: White to off-white, free flowing, rounded granules that are odorless and tasteless

Properties: Particle size range of 20–100µm; moisture content of <10%; insoluble in water, but does swell to 300 times its original volume

Incompatibilities: Ascorbic acid

Use: As a super disintegrant in tablet and capsule formulations

OFFICIAL NAME: SODIUM STEARYL FUMARATE

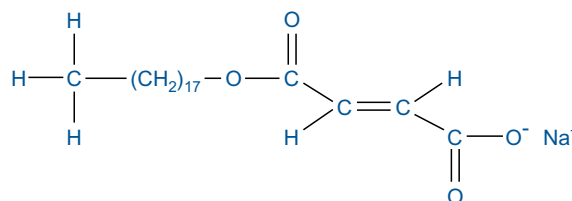
INN: Sodium Stearyl Fumarate

Synonyms: Fumaric acid, octadecyl ester, sodium salt; sodium monostearyl fumarate

Chemical Name & CAS Number: (E)-2-Butenedioic monooctadecyl ester, sodium salt (4070-80-8)

Molecular Formula: $C_{22}H_{39}NaO_4$

Structure:



Description: An odorless, tasteless, fine white powder with agglomerates of flat circular particles

Properties: 5% solution pH 8.3; density 0.3–0.5g/cc; melting point 224–245°C; insoluble in water at room temperature

Incompatibilities: Chlorhexidine acetate

Use: As a tablet and capsule lubricant

OFFICIAL NAME: SORBIC ACID

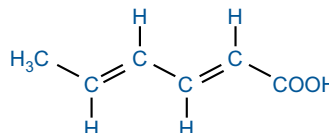
INN: Sorbic acid; Acidum sorbicum

Synonyms: 2,4-hexadienoic acid; 2-propenylacrylic acid

Chemical Name & CAS Number: (E,E)-Hexa-2,4-dienoic acid (22500-92-1)

Molecular Formula: $C_6H_8O_2$

Structure:



Description: Free-flowing, white, crystalline powder, with a characteristic odor

Properties: Solubility 1 g in 400 mL water, 10 mL alcohol or 19 mL propylene glycol; melting point 135°C; density 1.2 g/cc

Incompatibilities: Bases, oxidizing agents, and reducing agents. Some functionality is lost in the presence of nonionic surfactants

Uses: As a preservative

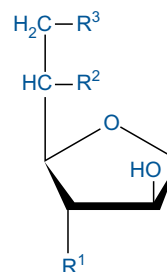
OFFICIAL NAME: SORBITAN ESTERS

INN: Sorbitan; Sobritani

Synonyms: Spans

Chemical Name & CAS Number: Sorbitan esters (monolaurate [1338-39-2]; monooleate [1338-43-8]; monopalmitate [26266-57-9]; monostearate [1338-41-6]; trioleate [26266-58-0]; tristearate [26658-19-5])

Structure:



Description: Monolaurate: Amber, oily liquid; Monooleate: Amber liquid; Monopalmitate: Tan, granular waxy solid; Monostearate: Cream to tan beads; Trioleate: Amber, oily liquid; viscosity about 200 cps; HLB number 1.8; acid number 15 max; saponification number 170–190; hydroxyl number 55–70. Tristearate: Tan, waxy beads; HLB number 2.1; acid number 12–15; saponification number 176–188; hydroxyl number 66–80

Properties: Monolaurate: Soluble in methanol or alcohol; dispersible in distilled water and hard water (200 ppm); insoluble in hard water (20,000 ppm); viscosity 4250 cps; HLB number 8.6; acid number 7.0. Monooleate: Soluble in most

mineral or vegetable oils; slightly soluble in ether; dispersible in water; insoluble in acetone; viscosity approximately 1000 cps; HLB number 4.3 acid number 8.0. Monopalmitate: Dispersible in distilled water or hard water (200 ppm); soluble in ethyl acetate; insoluble in cold distilled water or hard water (20,000 ppm); HLB number 6.7; acid number 7.5. Monostearate: Soluble (above melting point) in vegetable oils or mineral oil; insoluble in water, alcohol, or propylene glycol; HLB number 4.7; acid number 5–10. Trioleate: Soluble in mineral oil, vegetable oils, alcohol, or methanol; insoluble in water; HLB number 1.8; acid number 15; viscosity approximately 200 cps. Tristearate: Soluble in isopropyl alcohol; insoluble in water; HLB number 2.1; acid number 12–15.

Uses: As a nonionic surfactant used as emulsifying agents.

OFFICIAL NAME: SORBITOL

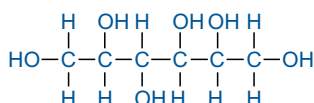
INN: Sorbitol; D-Sorbitol; Sorbitolum

Synonyms: 1,2,3,4,5,6-hexanehexol; sorbite; D-sorbitol

Chemical Name & CAS Number: D-Glucitol (50-70-4)

Molecular Formula: C₆H₁₄O₆

Structure:



Description: White, hygroscopic powder, granules, or flakes with a sweet taste

Properties: Solubility 1 g in about 0.45 mL of water; slightly soluble in alcohol, methanol, or acetic acid; 10% solution pH 4.5–7.0; density 1.507 g/cc; melting point of 96°C

Incompatibilities: di- and tri-valent ionic compounds (chelate formation), penicillins, and iron oxide

Uses: As a sweetener (chewable tablets); humectant liquid products, plasticizer for gelatin and other polymers; and, in 70% (w/w) solution, a vehicle

OFFICIAL NAME: STARCH

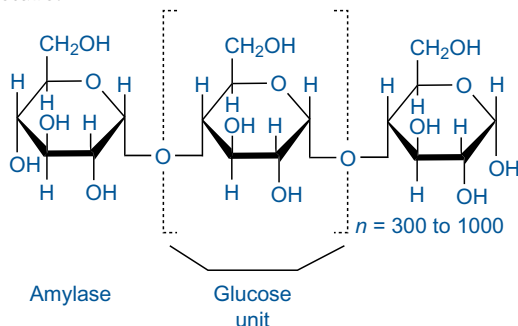
INN: Maize-, Potato-, Rice-, Tapioca-, Wheat- starch; Maydis-, Oryzae-, Solani-, Tritic- amylum

Synonyms: Corn Starch; Wheat Starch; Potato Starch

Chemical Name & CAS Number: Starch (9005-25-8)

Molecular Formula: (C₆H₁₀O₅)_n

Structure:



Description: Irregular, angular, white masses or fine powder; odorless; slight, characteristic taste

Properties: Insoluble in cold water or alcohol; when it is boiled with about 20 times its weight of hot water for a few minutes and then cooled, a translucent, whitish jelly results; aqueous suspension neutral to litmus

Incompatibilities: No known incompatibilities

Uses: As a filler, binder, and disintegrant for tablets and capsules

Note: Under the title **Pregelatinized Starch**, the NF recognizes starch that has been processed chemically or mechanically to rupture all or part of the granules in the presence of water and subsequently dried. This alteration results in a material that is more soluble in cold water and more easily used in manufacturing as a granulating agent/binder. Some

types may be modified to render them compressible and flowable for use in direct compression applications.

OFFICIAL NAME: STEARIC ACID

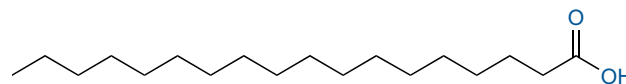
INN: Stearic acid; Acidum stearicum

Synonyms: Stearic acid; cetylacetic acid; stereophanic acid

Chemical Name & CAS Number: Octadecanoic acid (57-11-4)

Molecular Formula: C₁₈H₃₆O₂

Structure:



Description: Hard, white, or faintly yellowish, somewhat glossy and crystalline solid, or a white or yellowish white powder; an odor and taste suggestive of tallow

Properties: Practically insoluble in water; 1 g in about 20 mL alcohol; melting point 55°C; density 0.98g/cc; moisture content 0%

Incompatibilities: Insoluble stearates are formed with di- and tri-valent metal ions. Ointment bases made with stearic acid may show evidence of drying out or lumpiness, due to such a reaction when zinc or calcium salts are present.

Uses: As a tablet lubricant; ointment base

OFFICIAL NAME: STEARYL ALCOHOL

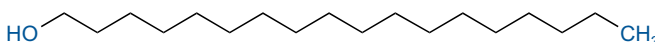
INN: Stearyl alcohol; Alcohol stearylicus

Synonyms: n-octadecanol; octadecyl alcohol; stenol

Chemical Name & CAS Number: 1-Octadecanol (112-92-5)

Molecular Formula: C₁₈H₃₈O

Structure:



Description: White, unctuous flakes or granules having a faint, characteristic odor and a bland taste

Properties: Insoluble in water; soluble in alcohol or vegetable oils; melting point range 55–60°C; density 0.884–0.906g/cc

Incompatibilities: Strong oxidizing agents

Uses: As a stiffening agent for semi-solids

OFFICIAL NAME: SUCRALOSE

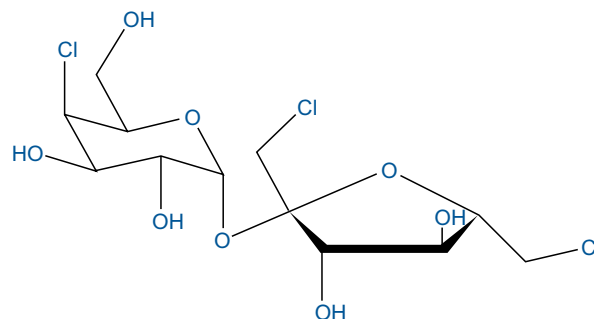
INN: Sucralose

Synonyms: Splenda; TGS; 1',4',6'-trichlorogalactosucrose

Chemical Name & CAS Number: 1,6-Dichloro-1,6-dideoxy-β-D-fructofuranosyl-4-chloro-4-deoxy-α-D-galactopyranoside (56038-13-2)

Molecular Formula: C₁₂H₁₉Cl₃O₈

Structure:



Description: White to off-white free flowing crystalline powder

Properties: Freely soluble in both water and ethanol; density 1.63 g/cc; particle size 90% < 12μ; 10% solution pH 5–6

Incompatibilities: No known incompatibilities

Uses: As a sweetener

OFFICIAL NAME: SUCROSE

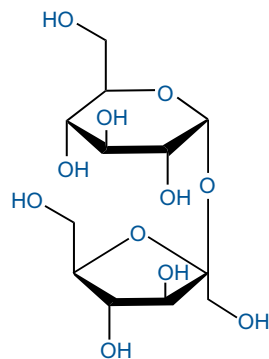
INN: Sucrose; Saccharum

Synonyms: α-D-glucopyranosyl-β-D-fructofuranoside; Sugar; Cane Sugar; Beet Sugar

Chemical Name & CAS Number: β -D-fructofuranosyl- α -D-glucopyranoside (57-50-1)

Molecular Formula: $C_{12}H_{22}O_{11}$

Structure:



Description: Occurs as a colorless mass or block, or as a white odorless and sweet tasting crystalline powder

Properties: Solubility 1 g in 0.5 mL water, 170 mL alcohol; density 1.6g/cc; hygroscopic capable of absorbing 1% moisture

Incompatibilities: Sucrose has no known incompatibilities; however, the naturally occurring trace materials in sucrose could result in issues. These trace materials include heavy metals, which could interfere with some APIs (e.g., ascorbic acid), and sulfite, which could result in discoloration of tablet coatings.

Uses: As a sweetener in oral liquids and solids; coating material for tablets; diluent for parenteral products

OFFICIAL NAME: TALC

INN: Purified talc; Talc; Talcum

Synonyms: Powdered talc; purified French chalk; soapstone; steatite

Chemical Name & CAS Number: Talc (14807-96-6)

Molecular Formula: $Mg_3(Si_2O_5)_4(OH)_4$

Description: Very fine, white, or grayish white, odorless, non-gritty crystalline powder

Properties: Insoluble in water, dilute acids and alkali; particle size (grade dependant) 99% < 74 μ or 99% < 44 μ ; 20% aqueous dispersion pH 6.5–10

Incompatibilities: Quaternary ammonium compounds

Uses: As a lubricant or glidant

OFFICIAL NAME: TARTARIC ACID

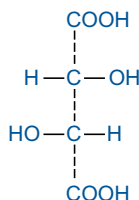
INN: Tartaric acid; Acidum tartaricum

Synonyms: L-(+)-2,3-dihydroxybutanedioic acid; d-tartaric acid; L-(+)-tartaric acid

Chemical name & CAS Number: [R-(R*,R*)]-2,3-dihydroxybutanedioic acid (87-69-4)

Molecular Formula: $C_4H_6O_6$

Structure:



Description: Large, colorless, or translucent crystals, or a white crystalline powder; odorless; extremely tart taste; stable in the air; solutions acid to litmus; dextrorotatory

Properties: Solubility 1 g in 0.8 mL water or 3 mL alcohol; 1.5% solution pH 2.2; density 1.76g/cc

Incompatibilities: Silver salts and reacts with metal carbonates and bicarbonates

Uses: As an acidulant; sialogogue and activator for carbonation

OFFICIAL NAME: TITANIUM DIOXIDE

INN: Titanium dioxide; Titanii dioxidum

Synonyms: Anatase titanium dioxide; brookite titanium dioxide

Chemical Name & CAS Number: Titanium oxide (13463-67-7)

Molecular Formula: TiO_2

Description: White, amorphous, odorless, tasteless powder

Properties: Particle size < 1 μ ; density 3.8-4.2g/cc; moisture 0.44%; insoluble in water

Incompatibilities: Possesses a catalytic effect in the presence of some actives

Use: As an opacifier for coatings and semisolids

OFFICIAL NAME: TRAGACANTH

INN: Tragacanth; Tragacantha

Synonyms: Gum Tragacanth; Gum Dragon; Goat's Thorn; Persian tragacanth

Chemical Name & CAS Number: Tragacanth gum (9000-65-1)

Molecular Formula: A mixture of water soluble and insoluble polysaccharides derived from dried gum of *Astragalus gummifer*. It's comprised of 60–70% bassorin and 30–40% soluble gum (*tragacanthin*).

Description: Flattened, lamellated, frequently curved fragments or straight or spirally twisted linear pieces 0.5-2.5 mm in thickness; white to weak-yellow in color; translucent; horny in texture; odorless; insipid, mucilaginous taste. When powdered, it is white to yellowish white.

Properties: 1% dispersion pH 5–6; practically insoluble in water and alcohol; when mixed with water, it does swell to 10 times its original volume creating a viscous dispersion.

Incompatibilities: Above pH of 7 tragacanth, reduces activity of benzalkonium chloride

Use: As a suspending agent; granulating agent

OFFICIAL NAME: TRIACETIN

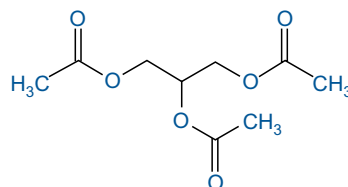
INN: Triacetin; Glycerolum triacetat

Synonyms: Glycerol triacetate; glyceryl triacetate; triacetyl glycerin

Chemical Name & CAS Number: 1,2,3-Propanetriol triacetate (102-76-1)

Molecular Formula: $C_9H_{14}O_6$

Structure:



Description: Colorless viscous liquid with a slight fatty odor

Properties: Density 1.16 g/cc; solubility 1 in 14 of water, miscible with ethanol; viscosity 17 mPa.s

Incompatibilities: Metal ions, oxidizing agents

Use: As a humectant, solvent or plasticizer for polymers

OFFICIAL NAME: TRIETHANOLAMINE

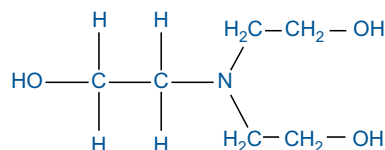
INN: Triethanolamine; Trolaminum; Trolamine

Synonyms: Triethylolamine; trihydroxytriethylamine; tris(hydroxyethyl)amine; TEA

Chemical Name & CAS Number: 2,2',2''-Nitrilotriethanol (102-71-6)

Molecular Formula: $C_6H_{15}NO_3$

Structure:



Description: Colorless to pale yellow, clear viscous liquid with a slight odor of ammonia

Properties: Miscible with water; 0.1N solution pH 10.5; freezing point 21.6°C

Incompatibilities: Higher fatty acids; mineral acids; heavy metal salts; if hydroxyl groups are replaced with halogen, the resultant product is toxic similar to mustard gas

Use: As a polymer plasticizer; humectants; emollient

OFFICIAL NAME: VANILLIN

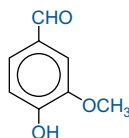
INN: Vanillin; Vanillinum

Synonyms: 4-Hydroxy-*m*-anisaldehyde; *p*-hydroxy-*m*-methoxybenzaldehyde; vanillic aldehyde

Chemical Name & CAS Number: 4-Hydroxy-3-methoxybenzaldehyde (121-33-5)

Molecular Formula: C₈H₈O₃

Structure:



Description: Fine, white to slightly yellow crystals, usually needle-like, with an odor and taste suggestive of vanilla

Properties: Solubility 1 g in about 100 mL water, about 20 mL glycerin, or 20 mL water at 80°C and freely soluble in alcohol; affected by light and air; solutions are acid to litmus; melting point 81–83°C; density is 0.6g/cc

Incompatibilities: Combines with glycerin, forming a compound almost insoluble in alcohol; decomposed by alkali; oxidized slowly by the air

Uses: As a flavor

OFFICIAL NAME: WATER

INN: Purified Water; Aqua purificata

Synonyms: Aqua; hydrogen oxide

Chemical Name & CAS Number: Purified Water (7732-18-5)

Molecular Formula: H₂O

Description: Clear, colorless, odorless, and tasteless liquid

Properties: Boiling point 100°C, freezing point 0°C; miscible with most polar solvents

Incompatibilities: With drug subject to hydrolysis; reacts violently with alkali metals

Use: Most widely used excipient in pharmaceutical manufacturing; solvent for granulating agents; solvent for syrups, suspensions; vehicle for aqueous solutions; in oral solid and liquid processing; in parenteral processing as water for injection or sterile water for injection; in topical, ocular, otic delivery systems; to clean all processing equipment

Special Classes of Water:

Purified Water: Obtained by distillation, ion-exchange treatment, reverse osmosis, or any other suitable process; contains no added substance

Water for Injection: Purified by distillation or by reverse osmosis under aseptic conditions; not sterilized by filtration or in the final container

Bacteriostatic Water for Injection: Sterile water for injection containing one or more suitable antimicrobial agents

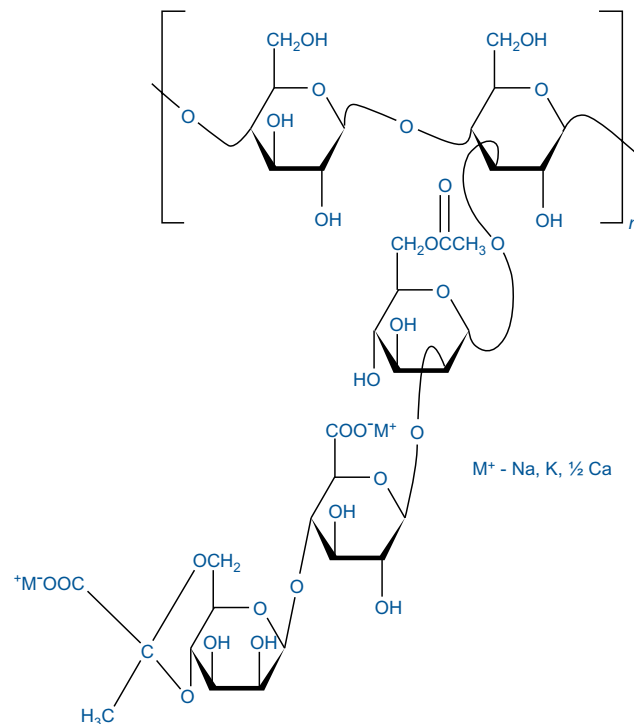
OFFICIAL NAME: XANTHAN GUM

INN: Xanthan gum; Xanthani gummi

Synonyms: Corn sugar gum; Keltrol

Chemical Name & CAS Number: Xanthan Gum (11138-66-2)

Structure:



Description: White or cream-colored, tasteless, free-flowing powder

Properties: Soluble in hot or cold water; 1% solution pH 6–8; 1200–1600 mPa.s for 1% solution at 25°C

Incompatibilities: Cationic surfactants, polymers, and preservatives will cause precipitation; exhibits thixotropic gel characteristics at concentrations greater than 1%

Uses: As a stabilizing agent; viscosity increasing agent

OFFICIAL NAME: XYLITOL

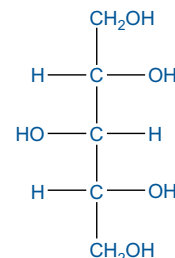
INN: Xylitol; Xylitolum

Synonyms: *meso*-xylitol

Chemical Name & CAS Number: xylo-Pentane-1,2,3,4,5-pentol (87-99-0)

Molecular Formula: C₅H₁₂O₅

Structure:



Description: White, odorless, yet sweet tasting granular solid, consisting of equi-dimensional particles

Properties: 5% solution pH 5.0–7.0; Density 1.52 g/cc; moisture content of up to 3% depending upon environmental conditions; particle size is grade dependant; solubility is 1 in 1.6 of water, 1 in 15 of propylene glycol and 1 in 80 of ethanol

Incompatibilities: Oxidizing agents

Use: As a non-cariogenic sweetening agent for tablets, syrups, and coatings

OFFICIAL NAME: ZEIN**INN:** Zein**Chemical Name & CAS Number:** Zein (9010-66-6)**Description:** Straw to pale yellow amorphous powder or flake with a characteristic odor and bland taste**Properties:** Density 1.23 g/cc; practically insoluble in water and ethanol; soluble in glycols**Incompatibilities:** Oxidizing agents**Use:** As a wet binder or tablet coating

Chapter 37

Basic Pharmacokinetics and Pharmacodynamics

Raymond E. Galinsky, PharmD and Craig K. Svensson, PharmD, PhD

ANALYTICAL CONSIDERATIONS 705

THE CRITICAL NATURE OF THE CONCENTRATION VERSUS EFFECT RELATIONSHIP 705

DRUG CONCENTRATION VERSUS TIME PROFILE 705

PHARMACOKINETIC MODELS 706

RATES AND ORDERS OF REACTIONS 706

INSTANTANEOUS INPUT WITH INSTANTANEOUS DISTRIBUTION 708

INSTANTANEOUS INPUT WITH NONINSTANTANEOUS DISTRIBUTION 710

CONTINUOUS INPUT 712

MULTIPLE-DOSE ADMINISTRATION 713

NONCOMPARTMENTAL ANALYSIS FOLLOWING INSTANTANEOUS INPUT 713

ABSORPTION 715

ORGAN-SPECIFIC CLEARANCE 717

DOSE-AND TIME-DEPENDENT PHARMACOKINETICS 720

STEREOCHEMICAL CONSIDERATIONS 721

KINETICS OF PHARMACOLOGIC EFFECT 722

The goal of pharmacotherapy is to provide optimal drug therapy in the treatment or prevention of disease. A major barrier to the achievement of this goal is the large variability in the pharmacological effect that is observed following drug administration (Fig 37-1).¹ The ability to implement drug therapy in a safe and rational manner necessitates an understanding of the factors that cause this variability. One of the most important factors is the concentration of drug that is achieved at the site of action.

ANALYTICAL CONSIDERATIONS

Any discussion of pharmacokinetics presumes that the drug concentrations can be determined with a high degree of accuracy and precision. One of the most frequent causes of high variability in pharmacokinetic parameters is poor data resulting from imprecise analytical procedures. Evaluation of pharmacokinetic data in the literature must begin with an assessment of the validity of the assay used under the conditions in which the study was conducted. An assay must be tested for specificity, sensitivity, reproducibility, stability, and accuracy. Because drug metabolites are frequently present in the fluid to be measured and are similar in structure to the parent compound, differentiation of drug from any putative metabolites must be ensured.

THE CRITICAL NATURE OF THE CONCENTRATION VERSUS EFFECT RELATIONSHIP

The quantitative response to a drug depends highly on the concentration of drug at the site of action. In most situations one cannot quantify drug concentration at the actual site of action. Rather, drug concentrations are measured in an easily accessible site that is believed to be in equilibrium with the site of action (e.g., blood or one of its components). Figure 37-2² provides a good illustration of a drug whose pharmacological effect is particularly sensitive to changes in blood concentration. Numerous studies have been published that substantiate the critical nature of the concentration-effect relationship for a wide variety of drugs.

It is recognized now that drug therapy may be optimized by designing regimens that account for the concentration of a drug necessary to achieve a desired pharmacological response.

However, there is often significant difficulty in achieving such target concentrations. In particular, it often is observed that if a fixed dose of a drug is administered to a group of individuals, the drug concentration measured in plasma can vary widely. For example, the peak concentration of 6-mercaptopurine achieved in a group of 20 patients who received a standard 1 mg/m² dose is shown in Figure 37-3.³ The concentrations ranged from 0 to 660 ng/mL. Taken together, this suggests that variability in drug concentration is a major source of variability in drug effect, and there may be a significant degree of variability among individuals in the drug concentrations produced by a given dose of drug.

A basic understanding of the factors that control drug concentration at the site of action is important for the optimal use of drugs. This is the area of study referred to as *pharmacokinetics*, which is the study of the time course of drug absorption, distribution, metabolism, and elimination.

DRUG CONCENTRATION VERSUS TIME PROFILE

Blood (or its components, plasma or serum) represents the most frequently sampled fluid used to characterize the pharmacokinetics of drugs. Drug concentration in the blood is the sum of several processes (Fig 37-4).⁴ Initially visual characterization of the processes controlling the concentration of drug in the blood can be made by constructing a drug concentration versus time profile (i.e., a plot of drug concentration in the blood versus time). As can be seen from Figure 37-5, several useful pieces of information can be derived from such a profile. For example, the time at which the peak concentration occurs can be approximated and the peak concentration quantified. If the minimum concentration needed to maintain a desired effect is known, the onset and duration of effect also can be approximated. While useful information can be drawn casually from a simple graph as depicted in Figure 37-5, a more rigorous description of the pharmacokinetics of a drug is necessary to achieve the accuracy in dosage regimen design required for the safe and effective use of drugs. This higher degree of accuracy necessitates the development of mathematical models for describing the time course of absorption, distribution, metabolism, and elimination.

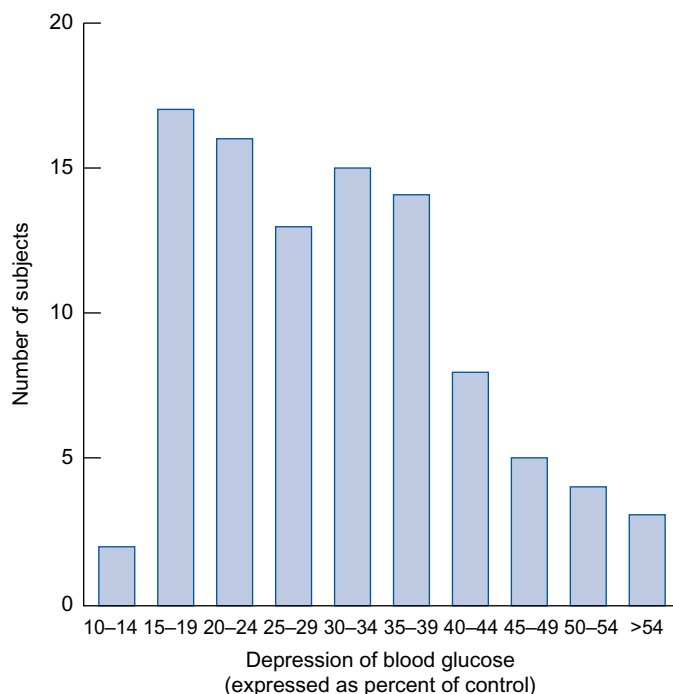


Figure 37-1. Decrease in blood glucose in 97 subjects 30 min after an intravenous dose of 1 g of tolbutamide. Note the large variability observed after the equivalent dose was administered in this group. (Data from Swerdloff RS *et al.* Influence of age on the intravenous tolbutamide response test. *Diabetes* 1967; 16: 161–170.)

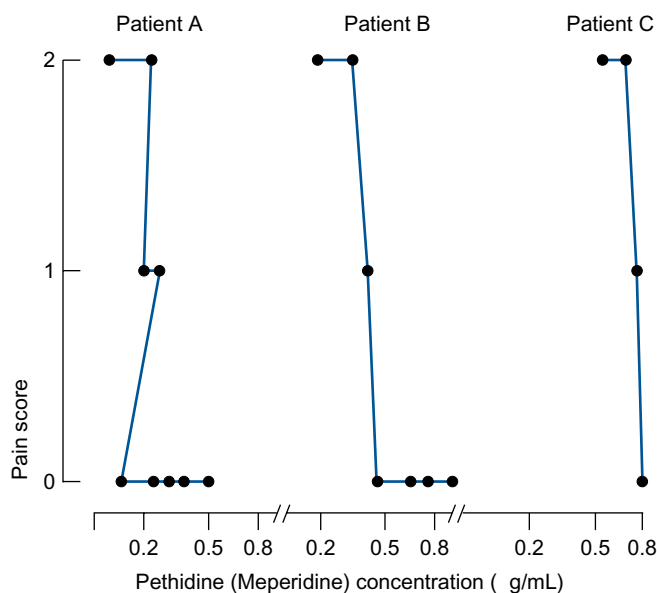


Figure 37-2. Blood pethidine concentration-response curves for three individual patients, illustrating a typical range in interpatient responses. (From Austin KL, Stapleton JV, Mather LE. Relationship Between Blood Meperidine Concentrations and Analgesic Response: A Preliminary Report. *Anesthesiology* 1980; 53:460–66.)

PHARMACOKINETIC MODELS

One of the primary objectives of pharmacokinetic models is to develop a quantitative method to describe the relationship of drug concentration, or amount in the body, as a function of time. The complexity of the pharmacokinetic model will vary with the route of administration, the extent and duration of

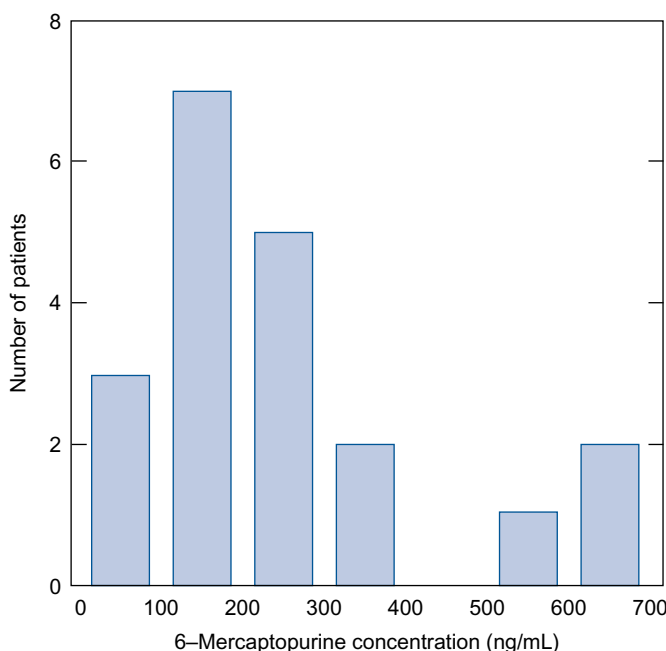


Figure 37-3. Distribution of peak 6-mercaptopurine concentrations achieved in a group of 20 patients receiving an oral dose of 1 mg/m². (Data from Sulh H *et al.* Pharmacokinetic determinants of 6-mercaptopurine myelotoxicity and therapeutic failure in children with acute lymphoblastic leukemia. *Clin Pharmacol Ther* 1986; 40: 604–609.)

distribution into various body fluids and tissues, the processes of elimination, and the intended application of the pharmacokinetic model. Often, numerous potential mathematical models exist for a particular drug. In such cases, the simplest model that will adequately and accurately describe the pharmacokinetics of the drug is the model that should be chosen.

There are a wide variety of potential uses for pharmacokinetic models, which include:

1. Prediction of drug concentration in blood/plasma or tissue.
2. Calculation of a dosage regimen.
3. Quantitative assessment of the effect of disease on drug disposition.
4. Elucidation of the mechanism of disease-induced alterations in drug disposition.
5. Determination of the mechanism for drug-drug interactions.
6. Prediction of drug concentration versus effect relationships.

There are three primary types of pharmacokinetic models: compartmental, noncompartmental, and physiological.

Compartmental models describe the pharmacokinetics of drug disposition by grouping body tissues that are kinetically indistinguishable and describe the transfer of drug between body tissues in terms of rate constants.

Noncompartmental models describe the pharmacokinetics of drug disposition using time- and concentration-averaged parameters.

Physiological models attempt to describe drug disposition in terms of realistic physiological parameters, such as blood flow and tissue-partition coefficients.

RATES AND ORDERS OF REACTIONS

Many pharmacokinetic models use parameters that are analogous to rate constants in chemical kinetics. For example,

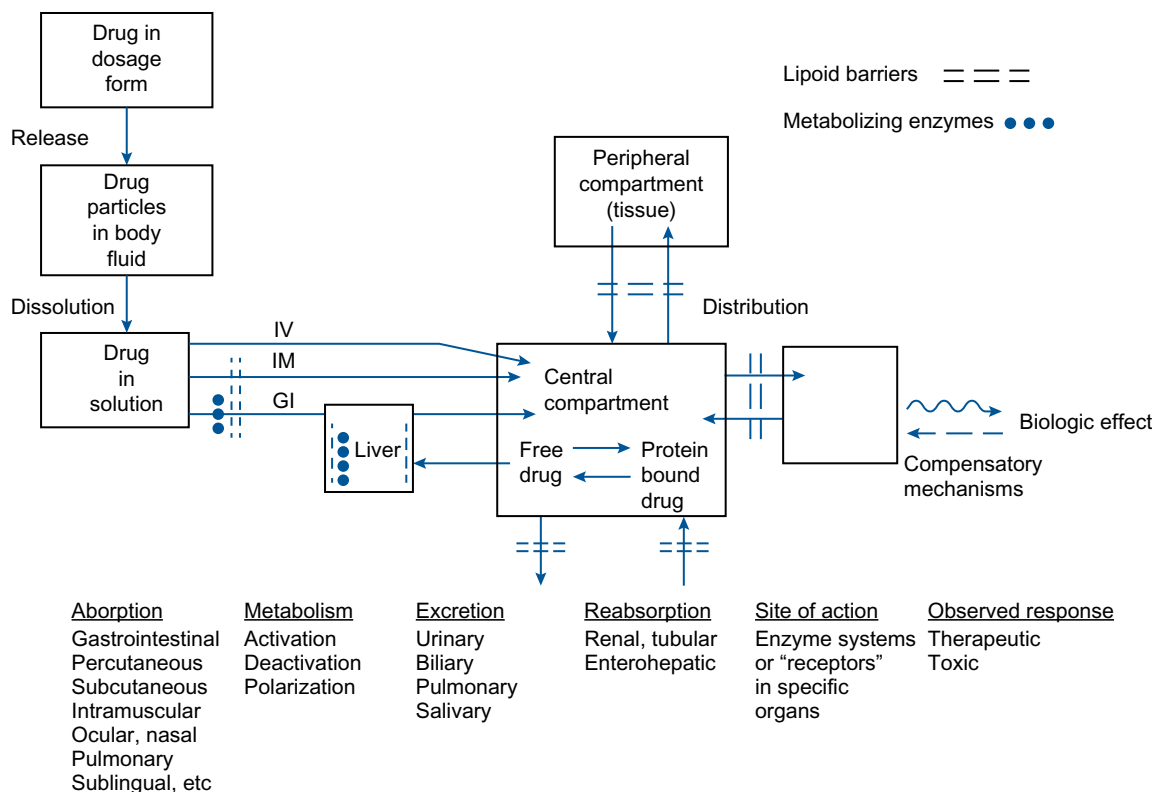


Figure 37-4. Diagram illustrating the factors that influence onset, duration, and intensity of drug effects. Note that the drug must dissolve before being absorbed and that it passes across many lipid barriers and some metabolizing systems before reaching the site of action. (From Barr WH. Principles of biopharmaceutics. *Am J Pharm Educ* 1968; 32: 958-959.)

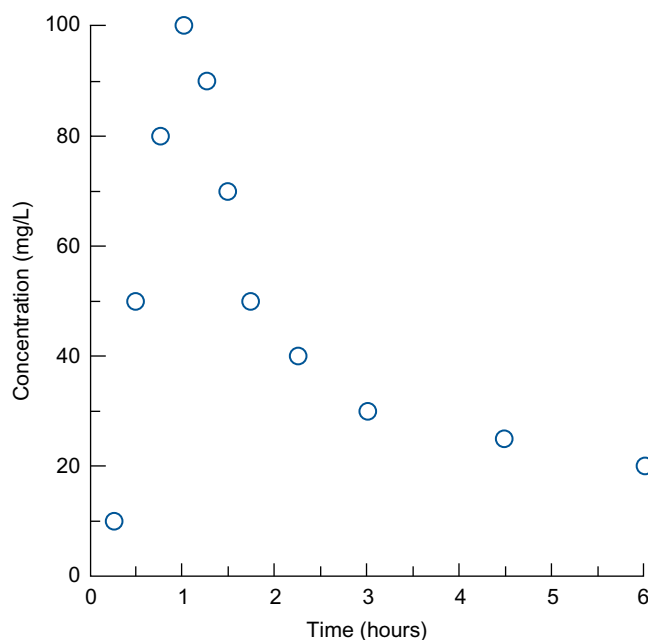


Figure 37-5. Hypothetical plot of drug-concentration data after oral administration of a drug.

consider the case of a drug (D) that is metabolized to a metabolite (M).



This reaction may be described as a function of either the disappearance of the drug or as a function of the appearance of

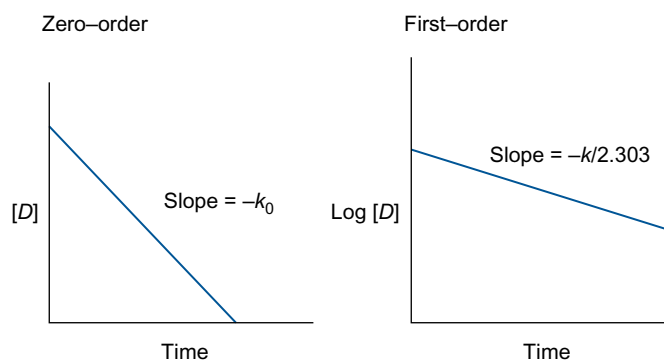


Figure 37-6. Plots illustrating a zero-order and a first-order reaction.

the metabolite. If the *amount* of the drug that is converted to a metabolite is a constant with respect to time, the reaction is said to be *zero-order* and is expressed as

$$-\frac{dD}{dt} = k_0 \quad (1)$$

where k_0 is the zero-order rate constant with units of mass per time (e.g., mg/min). A plot of the time-course of the amount of the drug in the body that is converted to a metabolite by zero-order kinetics is shown in Figure 37-6. Integration of Equation 1 yields an equation for a straight line, which describes the amount of the drug in the body at any time (t):

$$\text{Amount}_t = -k_0 t + \text{Amount}_{t=0} = -k_0 t + \text{Dose} \quad (2)$$

Zero-order rate processes typically are found when an enzyme or transport system becomes saturated and the rate process becomes constant and cannot be increased by increases in

the concentration of substrate. Zero-order rate processes are typical of constant-rate intravenous infusions and prolonged-release dosage forms.

If the amount of the drug in the body is converted to a metabolite at a rate that is a constant *fraction* of the amount of the drug in the body, the conversion of D to M is said to be a first-order reaction described by:

$$\frac{dD}{dt} = -kD \quad (3)$$

where k is the first-order rate constant expressed in units of reciprocal time (e.g., min^{-1}). Rearrangement of Equation 3 leads to:

$$\frac{dD}{D} = -kdt \quad (4)$$

and integration of this expression yields:

$$\int_0^t \frac{dD}{D} \Rightarrow \ln D = -kt + \ln D_0 \quad (5)$$

where $\ln$ is the natural logarithm. This equation also can be expressed in the exponential form:

$$D_t = D_0 e^{-kt} \quad (6)$$

Graphically, the integrated form usually is expressed in terms of $\log_{10}$ rather than in natural logarithms (see Fig 37-6):

$$\log D = \frac{-kt}{2.303} + \log D_0 \quad (7)$$

INSTANTANEOUS INPUT WITH INSTANTANEOUS DISTRIBUTION

The disposition of a drug from its site of administration and its distribution and elimination from the body occurs via the vascular system. Most drugs are low-molecular-weight compounds of sufficient lipophilicity that they are able to distribute readily into the intra- and extracellular fluid compartments in the body. The transfer of drug from the circulation to these fluid compartments and then into tissues is called distribution. The pharmacokinetic parameter *volume of distribution* is a proportionality constant that relates drug concentration in a reference fluid, typically plasma, to the amount of drug distributed throughout the body.

$$\text{Volume of distribution}(V_D) = \frac{\text{Amount of drug in body}(D_B)}{\text{Drug concentration}(CP)} \quad (8)$$

Drugs that distribute widely to tissues will have large volumes of distribution and low plasma concentrations relative to the dose administered, whereas drugs that are highly bound to plasma proteins (e.g., warfarin, phenylbutazone) or do not readily enter cells (e.g., amikacin) will have low volumes of distribution and high plasma concentrations relative to the administered dose.

Oie and Tozer⁵ have developed a physiological model for expression of the apparent volume of distribution, which takes into account the extracellular water, including plasma- and protein-binding of the drug in both plasma and tissue. For an average 70-kg male, total body water is about 42 L, of which 3 L is plasma and 12 L is extracellular fluid space. Moreover, 55–60% of the albumin in the extracellular space is found outside of plasma. Thus, for drugs that are largely bound to albumin, the apparent volume of distribution can be expressed as:

$$V_D = 7 + 8f_u + V_T \left[\frac{f_u}{f_{uT}} \right] \quad (9)$$

where f_u is the fraction of the drug in plasma that is unbound (often referred to as the *free fraction*), f_{uT} is the *free fraction* of

drug in tissue, and V_T is the volume of intracellular tissue water. Equation 9 can be simplified further to:

$$V_D = 7 + 8f_u + 27 \left[\frac{f_u}{f_{uT}} \right] \quad (10)$$

This model has been extremely useful in predicting the magnitude of changes in the apparent volume of distribution due to alterations in (1) plasma protein binding, (2) tissue protein binding, and (3) the volume of extracellular and intracellular fluid. For example, if a drug distributes into extracellular fluid but not intracellular fluid, the apparent volume of distribution can be expressed as:

$$V_D = 7 + 8f_u \quad (11)$$

and will vary between 7 and 15 L, depending upon the extent of plasma protein binding to albumin. For such a compound, with a relatively small volume of distribution, alterations in the plasma protein binding will not produce proportional changes in the apparent volume of distribution. Indeed, as reported by Williams *et al.*,⁶ the free fraction of tolbutamide in plasma increases in patients with cirrhosis by 28%, from 0.068 to 0.087, yet the apparent volume of distribution increases less than 10%, from 0.15 to 0.164 L/kg. Conversely, drugs with a volume of distribution greater than total body water indicate drug distribution and binding to tissue proteins and other cellular components. Such compounds also may be bound highly to plasma proteins. With drugs having a large volume of distribution (≥ 50 –100 L), the contribution of plasma and extracellular water space can be ignored, and Equation 3 simplifies to:

$$V_D = 27 \left[\frac{f_u}{f_{uT}} \right] \quad (12)$$

Changes in the plasma or plasma-free fraction will produce proportional changes in the apparent volume of distribution. For example, a two-fold increase in the free fraction of drug in tissue, f_{uT} , will decrease the apparent volume of distribution by two-fold. Less drug will be distributed to tissue, reflected by an increase in plasma concentrations. The volume necessary to account for the total amount of drug in the body will appear to have been decreased.

Once a drug is in the vascular system, it is transported by the blood to tissues where it can be eliminated from the circulation by distribution into tissue, metabolism by the tissue, or excretion from the tissue (see Fig 37-4). All of these processes lower the plasma concentration of drug. Each separate process may be described by a first-order rate constant, and the overall change in the plasma concentration is the net effect of all of these parallel, competing, first-order processes.

Intravenous injection of a drug that has nearly instantaneous distribution and first-order elimination can be described by an open one-compartment model (Fig 37-7). The body behaves as if it were a homogeneous compartment. In the one-compartment model, distribution is very rapid and can be considered instantaneous and is, therefore, ignored. After intravenous administration the plasma concentration declines exponentially according to:

$$C = C_0 e^{-\lambda t} \quad (13)$$

where C_0 is the initial concentration and λ is the overall elimination rate constant. Such an exponential elimination of theophylline, given intravenously, is shown in Figure 37-8.⁷ According to Equation 13, if the data of Figure 37-8 are plotted on semilog paper, a straight line should result, and such a plot is shown in Figure 37-9. Several derived data can be obtained from a plot of log concentration versus time. Extrapolation to zero time (i.e., the y-intercept) gives an estimated theoretical concentration in plasma at time zero, from which the apparent volume of distribution (V_D) can be estimated by simply dividing the dose by C_0 .

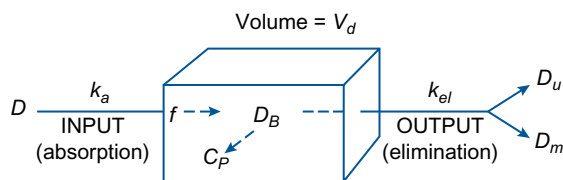


Figure 37-7. The open one-compartment pharmacokinetic model. The amount of the drug dose (D) that enters the body is D_B ; in the case of intravenous injection, this is the entire dose, whereas in the case of extravascular administration, some fraction of the dose (F) is absorbed with a rate constant of k_a . The compartment has an apparent volume of distribution (V_d), into which drug distributes instantaneously to achieve a concentration of C_p . Drug is eliminated from the compartment with a rate constant k_{el} . D_u is the amount excreted into urine, feces, bile, expired air, sweat, milk, etc; D_m is the amount of drug metabolized.

It is a theoretical concentration because neither the injection nor the distribution is actually instantaneous.

The *half-life* of a drug is the time required to reduce the amount of drug in the body or the plasma concentration by 50%. For a first-order process, the half-life is constant and is independent of the starting value of the amount of drug in the body (or plasma concentration). The plasma half-life, $t_{1/2}$, can be determined directly from the graph or from the elimination rate constant, λ or k_{el} , by means of the relationship:

$$t_{1/2} = \frac{0.693}{\lambda} = \frac{0.693}{k_{el}} \quad (14)$$

As shown, the half-life is related inversely to the elimination rate constant. When the elimination rate constant, k_{el} , is determined from the slope of the concentration versus time plot, one must keep in mind that the data need to be plotted on a semilog scale. From Figure 37-9, the k_{el} is determined to be 0.22 hr^{-1} . This is the instantaneous rate constant and indicates that 22% of the theophylline in the body is lost per hour. The rate constant

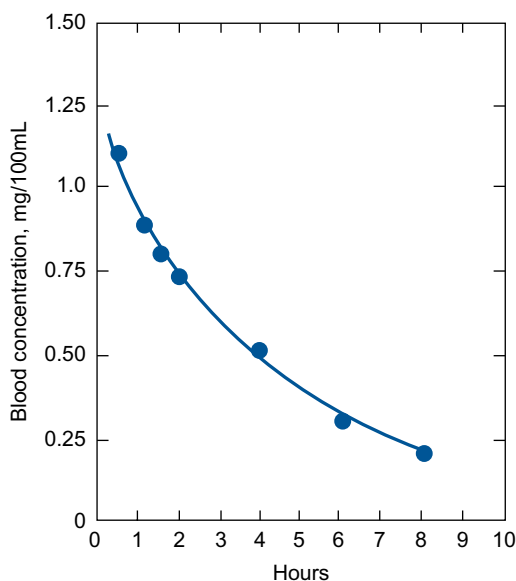


Figure 37-8. Elimination curve of average blood levels of theophylline in 11 human subjects after intravenous administration of 0.5 g aminophylline per 70 kg to each. (Data from Truitt Jnr EB *et al.* Theophylline blood levels after oral, rectal, and intravenous administration, and correlation with diuretic action. *J Pharmacol Exp Ther* 1950; 100: 309–315.)

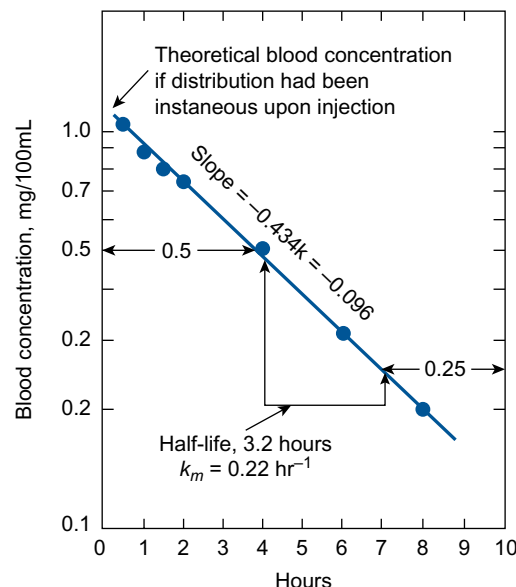


Figure 37-9. Semilog plot of the elimination curve in Figure 37-8. Note the log scale of the ordinate.

for elimination (see Fig 37-7) is shown without reference to the route of elimination. It must be recognized that k_{el} represents the overall elimination by all competing, parallel pathways and is equal to the sum of the rate constants that define the various simultaneous (i.e., parallel) contributory processes (e.g., metabolism, renal excretion, or biliary secretion). Thus, the overall rate constant, $k_{el} = k_1 + k_2 + k_3 + \dots + k_N$, where $k_1 + k_2 + k_3 + \dots + k_N$ are the rate constants of the separate contributory processes.

Half-life is a clinically useful pharmacokinetic parameter in that it indicates when the next dose of a drug needs to be administered and is therefore helpful in designing an optimal dosing regimen. The half-life also is useful in determining:

1. The fluctuation of plasma concentrations between doses;
2. The time required to reach steady-state equilibrium after beginning continuous drug administration; and
3. The persistence of drug in the system once drug administration has ceased.

Under some conditions, it is not possible to obtain plasma concentration data over sufficient time to obtain accurate estimates of the half-life for designing dosage regimens. The elimination rate constant, and hence the half-life, may be estimated from the excretion rate of unchanged drug. Because the first-order elimination rate constant is independent of the amount of drug in the body, the instantaneous excretion rate, dD_u/dt , is directly proportional to the total amount of drug in the body.

$$\frac{dD_u}{dt} = k_u D_B \quad (15)$$

where D_B and D_u are the amount of drug in the body at time zero and the amount of drug excreted in the urine, respectively, and k_u is the urinary excretion rate constant. A plot of $\log dD_u/dt$ versus time yields a straight line with slope of $-k_e/2.303$. One also may estimate the half-life from urinary excretion data using the cumulative amount of drug excreted (sigma-minus) method. Using this approach:

$$D_u \infty = D_B \frac{k_u}{k_{el}} (1 - e^{-k_{el} t}) \quad (16)$$

k_u/k_{el} represents the fraction of the drug in the body that eventually is excreted in urine as unchanged drug and D_u^∞ represents the total amount of unchanged drug excreted in urine. A plot of the log of the amount of drug remaining to be excreted ($D_u^\infty - D_u$) versus time yields a slope equal to $-k_e/2.303$. This method requires collecting urine for at least six to eight half-lives to achieve an accurate measure of D_u^∞ .

For a drug eliminated by first-order kinetics, the elimination rate constant, k_{el} , can be expressed as the fraction of the volume of distribution that is presented to an eliminating organ and cleared of drug per unit time (clearance) relative to the total volume of distribution, V_D . Thus, k_{el} represents the fractional removal rate of drug from the system, and the elimination rate constant can be expressed in terms of clearance and volume of distribution:

$$\frac{\text{Clearance}}{V_D} = \text{Elimination-rate constant}(k_{el}) \quad (17)$$

As written in Equation 17, the elimination rate constant (and hence, plasma half-life) is a dependent parameter that, by itself, is not always the most reliable indicator of drug removal from the body. Disease or altered physiology (e.g., aging, pregnancy) can alter protein binding, thereby affecting the apparent volume of distribution, or alter organ function, thereby affecting clearance, but these changes may not be reflected by changes in the half-life. For example, the volume of distribution may be altered due to changes in tissue or plasma protein binding, independent of specific organ function (clearance). In this instance, the half-life of a drug may be changed, but clearance could remain constant. Although a useful parameter, one must always bear in mind that half-life is a dependent or derived parameter that does not reliably reflect irreversible removal of drug from the body. A more accurate way to express half-life (Equation 14), therefore, is:

$$t_{1/2} = 0.693 \frac{V_D}{CL} \quad (18)$$

Clearance is the most useful pharmacokinetic indicator of irreversible loss of drug from the body and refers to a volume of fluid from which the drug appears to be removed in a given amount of time. Clearance also can be expressed as the quotient of overall rate of elimination of a drug relative to the drug concentration at a particular organ of elimination,

$$\text{Clearance} = \frac{\text{Rate of elimination}}{\text{Concentration}} \quad (19)$$

and, if time-averaged over the time course of plasma concentrations, drug clearance can be expressed as:

$$\text{Clearance} = \frac{\text{Amount of drug removed}}{AUC} \quad (20)$$

where AUC is the area under the concentration-versus-time curve. Total body clearance, CL_T , also can be estimated as the quotient of dose and area under the concentration-versus-time curve from zero to infinity.

$$CL_T = \frac{\text{Dose}_{IV}}{AUC_0^\infty} \quad (21)$$

Total body clearance is the sum of all the separate clearances that contribute to drug elimination:

$$CL_T = CL_{\text{RENAL}} + CL_{\text{HEPATIC}} + CL_{\text{OTHER}} \quad (22)$$

INSTANTANEOUS INPUT WITH NONINSTANTANEOUS DISTRIBUTION

The one-compartment model adequately describes the pharmacokinetics of drugs with instantaneous distribution.

However, for some compounds, distribution requires some finite time to reach equilibrium. During this time, the drug undergoes distribution and elimination, and drug concentrations decrease rapidly. When distribution equilibrium is established, the loss of drug from the body is due to elimination, and plasma concentrations decline more slowly. This biexponential time-course of plasma concentrations can be described by a two-compartment model. In this model, the body appears to behave as if it is comprised of two compartments, a central compartment and a peripheral compartment. By convention, drug absorption (or injection) and drug elimination occur from the central compartment, and the peripheral compartment is closed and communicates with the environment only through the central compartment (Fig 37-10). The movement of drug between compartments, following rapid intravenous injection with elimination from the central compartment, can be described by:

$$\frac{dD_1}{dt} = k_{21}D_2 - k_{12}D_1 - k_{10}D_1 \quad (23)$$

and

$$\frac{dD_2}{dt} = k_{12}D_1 - k_{21}D_2$$

where D_2 is the amount of drug in the peripheral or tissue compartment, D_1 is the amount of drug in the central compartment, k_{21} and k_{12} are the apparent first-order intercompartmental distribution rate constants, and k_{10} or k_{el} is the apparent first-order elimination rate constant from Compartment 1.

After intravenous injection of a drug that obeys two-compartment pharmacokinetics, the plasma concentration declines in a complex biexponential fashion. When plotted on semilog graph paper, the separate processes of distribution and elimination can be identified by the method of residuals (Fig 37-11).⁸

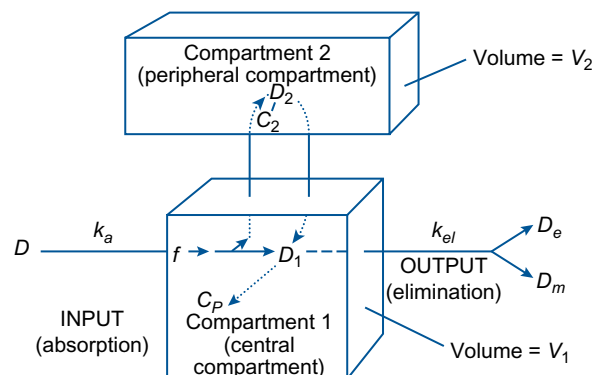


Figure 37-10. Diagram of open two-compartment pharmacokinetic model. The amount of dose that enters the body for an intravenous injection is the entire dose, and administration is instantaneous. The amount of dose absorbed from an extravascular dose is fd , where f is the fraction of dose absorbed with a rate constant, k_a . Some of the absorbed drug enters Compartment 2 with a first-order rate constant of k_{12} and is returned to Compartment 1 with a first-order rate constant of k_{21} . D_1 is the amount of drug in Compartment 1, and D_2 in Compartment 2; C_1 and C_2 are the respective concentrations in Compartments 1 and 2 ($C_1 = C_p$). Drug is eliminated from Compartment 1 with a first-order rate constant, k_{el} , which, however, is obscured by the lag in transfer of drug from Compartment 2 to Compartment 1. D_u is the amount excreted into urine, feces, expired air, sweat, milk, etc; D_m is the amount of drug metabolized. The relative volumes V_1 and V_2 may vary greatly, V_1 sometimes being the larger and other times the smaller.

Figure 37-11 shows such a resolution of the biexponential decay into the two components of distribution and elimination. From the slopes and intercepts of the residuals, the plasma concentration, C , at any time, t , can be described as the sum of two exponentials, namely:

$$C = C_1 e^{-\lambda_1 t} + C_2 e^{-\lambda_2 t} \quad (24)$$

where

$$C_1 = \frac{\text{Dose}(\lambda_1 - k_{21})}{V_1(\lambda_1 - \lambda_2)} \quad (25)$$

and

$$C_2 = \frac{\text{Dose}(k_{21} - \lambda_2)}{V_1(\lambda_1 - \lambda_2)} \quad (26)$$

Distribution is more rapid than elimination, such that at some point, the first term in Equation 24, $C_1 e^{-\lambda_1 t}$ approaches zero, and the biological half-life can be determined from the slope of the terminal phase:

$$t_{1/2} = \frac{0.693}{\lambda_2} \quad (27)$$

C_1 , C_2 , λ_1 , and λ_2 are hybrid constants, representing the intercepts, C_1 , C_2 , and slopes, λ_1 , λ_2 , of the two exponential phases, which can be obtained by computer-fitting the biexponential data. The zero-time intercept and the volume of the central compartment, V_1 , and the actual pharmacokinetic parameters k_{12} , k_{21} , and k_{el} can be derived from the hybrid rate constants, using the following relationships. At time $t = 0$

$$C_0 = C_1 + C_2; V_1 = \frac{\text{Dose}}{C_1 + C_2} \quad (28)$$

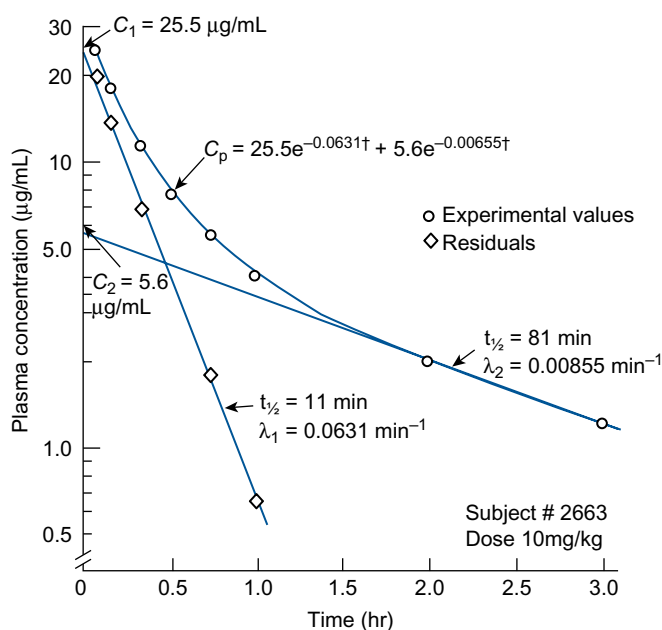


Figure 37-11. Resolution of the plasma concentration curve for pralidoxime into its distribution and elimination components after intravenous administration. Note that plasma concentration is plotted on a logarithmic scale. The time constant for the elimination phase is determined from the slope, $-0.434 \lambda_2$; it is a hybrid constant and λ_2 is not the same as k_{el} (see text). Likewise, the time constant for distribution, λ_1 , is obtained from the slope, $-0.434 \lambda_1$, of the distribution line; λ_1 is also a hybrid constant. (From Gibaldi M, Perrier D. *Pharmacokinetics*, 2nd edn. New York: Dekker, 1982.)

The hybrid rate constants, λ_1 and λ_2 , can be defined using the following two equations:

$$\lambda_1 \lambda_2 = k_{21} k_{el} \quad (29)$$

$$\lambda_1 + \lambda_2 = k_{12} + k_{21} + k_{el} \quad (30)$$

Thus,

$$k_{el} = \frac{\lambda_1 \lambda_2}{k_{21}} \quad (31)$$

$$k_{21} = \frac{C_1 \lambda_2 + C_2 \lambda_1}{\lambda_1 - \lambda_2} \quad (32)$$

$$k_{12} = \lambda_1 + \lambda_2 - k_{21} - k_{el} \quad (33)$$

The reader is referred to Gibaldi and Perrier (see *Bibliography*) for a more in-depth derivation of these expressions. The slope of the final phase of biexponential disposition, λ_2 , can be related to the elimination-rate constant, k_{el} , by:

$$\lambda_2 = f_C k_{el} \quad (34)$$

where f_C is the fraction of the drug that is in the central compartment after distribution equilibrium has been achieved. After distribution, the fraction of the drug in the central compartment is a constant.

$$f_C = \frac{k_{21} - \lambda_2}{k_{21} + k_{12} - \lambda_2} \quad (35)$$

The terminal disposition constant, λ_2 , reflects disposition from the entire body and is a function of distribution and elimination. The rate constant, k_{el} , represents only elimination from the theoretical central compartment.

The volume of distribution, V_D , can be determined in a two-compartment system; however, the estimation is complicated by the noninstantaneous nature of the distribution phase between the two compartments and results in the apparent volume of distribution being time-dependent. From Equation 18, it can be seen that the volume of distribution of the central compartment, V_1 , can be obtained following administration of an intravenous dose, D_{IV} , of drug from:

$$V_1 = \frac{D_{IV}}{C_1 + C_2} \quad (36)$$

Clearance can be calculated from the product of k_{el} and V_1 , and the volume of the central compartment can be expressed as:

$$V_1 = \frac{D_{IV}}{k_{10} [\text{AUC}]_{0 \rightarrow \infty}} \quad (37)$$

The most accurate method of estimating the volume of distribution is to estimate the steady-state volume of distribution. The volume of distribution at the steady state, V_{SS} , represents the steady state with respect to distribution of the drug from the central compartment to the tissue compartments and is not altered by changes in drug elimination or clearance. The total amount of the drug in the body at the steady state is the sum of the amounts in all compartments, thus:

$$V_{SS} = V_1 + \frac{k_{12}}{k_{21}} V_1 \quad (38)$$

Notice that V_{SS} is independent of the elimination rate constant, k_{el} , and λ_2 .

The volume of distribution by area, V , is an alternate method of estimating the apparent volume of distribution and relies on the assumption that the plasma and the amount of drug in the body decline in parallel during the postdistributive phase.

$$V_\beta = \frac{V_1 k_{el}}{\lambda_2} \quad (39)$$

The least accurate method of estimating the volume of distribution for a drug that follows biexponential elimination kinetics is by extrapolation, V_{EXTRAP} , because changes in distribution can alter the estimation of the hybrid intercept, C_2 .

$$V_{EXTRAP} = \frac{D_{IV}}{C_2} \quad (40)$$

Distribution to various tissues depends upon both blood flow to that tissue and the rate of uptake (effective partition coefficient) into a particular tissue and its cells. The overall pattern of drug distribution is governed by both tissue perfusion and diffusion of drug within tissues. Tissues with the highest blood flow, such as liver, kidney, and brain, equilibrate more rapidly than tissues that are perfused less well, such as skin and fat. Once in the tissue vasculature, drug distribution into tissue is controlled largely by diffusional barriers of cell membranes. Rowland and Tozer (see *Bibliography*) present a useful expression for the first-order rate constant for distribution of drug into tissue.

$$k_{TISSUE} = \frac{(Q/V_{TISSUE})}{k_{PARTITION}} \quad (41)$$

where $k_{PARTITION}$ is the equilibrium distribution ratio of tissue and venous drug concentration, Q is tissue blood flow, V_{TISSUE} is the tissue volume, and the quotient of Q and V_{TISSUE} is the tissue perfusion rate. The time to reach tissue equilibrium is the reciprocal of Equation 41. For a poorly perfused tissue such as fat, the $k_{PARTITION}$ may be quite high and the Q/V_{TISSUE} low, resulting in a long time to reach tissue equilibrium. Even for highly perfused tissues, such as the brain, the distribution of some drugs may be quite variable and will depend upon diffusion across cell membranes. In this case, distribution is said to be diffusion-rate-limited and will depend upon both the oil-to-water partition coefficient and the degree of ionization at physiological pH.

For most drugs, distribution usually occurs more rapidly than elimination, resulting in complete distribution before most of the drug has been eliminated. For some drugs, once injected, distribution is so rapid that the overall plasma-concentration time-course represents elimination (see Fig 37-7). Thus, both administration and distribution appear to be instantaneous, and the pharmacokinetics can be modeled by the simplest one-compartment model (see Fig 37-7). For such a drug, the volume of distribution can be calculated as the quotient of the intravenous dose and the extrapolated plasma concentration at time zero, C_0 .

CONTINUOUS INPUT

It is sometimes desirable to administer a drug continuously to maintain constant plasma concentration. This is often the case for drugs with very rapid elimination or for those that have a low therapeutic index. Continuous input commonly is thought of in terms of intravenous infusion; however, sustained-release oral dosage forms and delivery of drugs through the skin from patches also are examples of continuous input, and the pharmacokinetics of drug administration is similar for all systems with continuous input.

With constant intravenous infusion, the plasma concentration rises in a logarithmic fashion and eventually reaches a plateau (Fig 37-12).⁸ The time to reach the plateau or steady-state concentration is determined by the elimination rate constant.

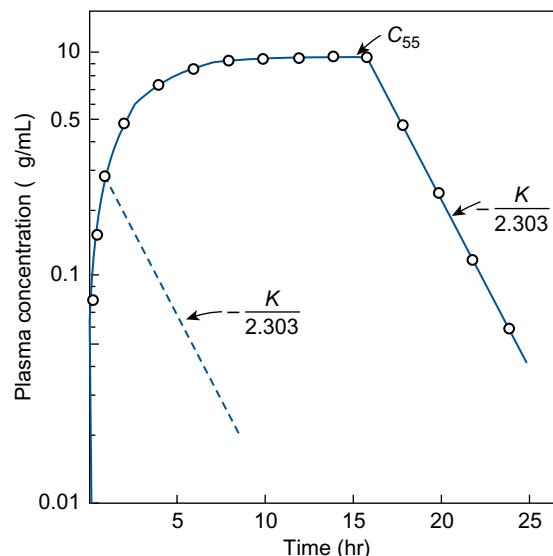


Figure 37-12. Semilogarithmic plot of plasma concentration during and after cessation of a constant intravenous infusion of a drug in a one-compartment system. Whether infusion is stopped prior to the attainment of a plateau or after, the plasma concentration will fall logarithmically with a slope of $-0.434k_{el}$. In the figure, K is k_{el} and $1/2.303 = 0.434$. C_{ss} is the steady-state concentration, C_p^{ss} . (From Gibaldi M, Perrier D. *Pharmacokinetics*, 2nd edn. New York: Dekker, 1982.)

The rate of change of drug in the body (D_B) during a constant rate infusion (R_0) is the difference between the zero-order infusion rate and the first-order elimination rate.

$$\frac{dD_B}{dt} = R_0 - (k_{el} \cdot D_B) \quad (42)$$

The plasma concentration (C) at any time during the constant infusion is:

$$C = \frac{R_0}{CL_T} (1 - e^{-k_{el} t_{inf}}) \quad (43)$$

where t_{inf} equals the time of the infusion. As the time of the infusion increases, the exponential expression approaches zero, and the concentration approaches steady state. At steady state the rate of infusion is equal to the rate of elimination, and the simplified expression can be expressed as:

$$C_{ss} = \frac{R_0}{CL_T} \quad (44)$$

The fraction of the steady state that is achieved in some time, t_{inf} , after the start of the infusion can be calculated as:

$$\frac{C}{C_{ss}} = (1 - e^{-k_{el} t_{inf}}) \quad (45)$$

and can be expressed in terms of half-lives as:

$$\frac{C}{C_{ss}} = (1 - 2^{-n}) \quad (46)$$

where n is the ratio of infusion time and half-life. For example, when the infusion time equals three half-lives ($n = 3$), the concentration is at 87.5% of the steady state, and when the infusion has lasted for four half-lives ($n = 4$), the concentration has achieved 93.75% of the steady state. Theoretically, one never reaches steady-state conditions because this is an exponential process; however, for clinical purposes one can assume, with little error, that steady-state concentrations are achieved within four to five half-lives.

If a drug has a relatively long half-life and the therapeutic situation demands rapid attainment of therapeutic plasma concentrations, it is sometimes desirable to give a loading dose at the beginning of the constant-rate infusion. The loading dose should approximate the amount of drug in the body at steady state. If the apparent volume of distribution and target concentration is known, the loading dose can be calculated simply as:

$$\text{Dose}_{\text{LOADING}} = (\text{Target concentration})(V_D) \quad (47)$$

The plasma concentration is the sum of the contributions from the loading dose and the infusion and can be estimated at any time after the loading dose has been given and infusion started from:

$$C = \frac{\text{Dose}_{\text{LOADING}}}{V_D} (e^{-k_{el}t}) + \frac{R_0}{CL_T} (1 - e^{-k_{el}t_{inf}}) \quad (48)$$

and if the half-life of the drug is known, the loading dose can be estimated from the quotient of infusion rate, R_0 , and elimination rate constant, k_{el} . For some drugs, such as lidocaine, the entire loading dose cannot be given in a single bolus injection because there is a significant distribution phase. In such a case, fractional loading-dose schemes can be used in which the loading dose is divided into several smaller bolus doses and given during the beginning of the infusion.

Finally, it should be noted that whether or not a loading dose is given, the attainment of steady state is determined by the elimination half-life and not by the rate of the infusion or the use of bolus loading doses to achieve concentrations rapidly.

MULTIPLE-DOSE ADMINISTRATION

Continuous administration of a drug is often impractical, and multiple-dose regimens are used to maintain the concentration of a drug within an acceptable range that minimizes the development of toxicity and avoids loss of efficacy. Usually, the dose of a drug is administered with a constant dose interval, referred to as τ . Some features of a multiple dosage scheme are shown in Figure 37-13. The drug is administered at a fixed dose and a

fixed interval. Each successive dose is administered before the previous dose has been eliminated entirely, and thus drug accumulation occurs. As with the constant intravenous infusion, the time to reach a steady-state fluctuation depends upon the elimination half-life and not on the size of the dose or the dosing interval. In Figure 37-13, the dose, D , is administered at a dosing interval equal to the half-life. After the first dose is given, the amount of drug in the body is equal to that dose. When the next dose is given, the amount of drug in the body is equal to $D + 0.5D$. At the end of each dose interval, the total amount of drug in the body is half of the postinjection peak and is the sum of the amount remaining from all of the previous doses. The maximum, $C_{\text{MAX,SS}}$, and minimum, $C_{\text{MIN,SS}}$, concentrations at steady state are described by:

$$C_{\text{MAX,SS}} = \frac{\text{Dose} / V_D}{(1 - e^{-k_{el}\tau})} \quad (49)$$

and

$$C_{\text{MIN,SS}} = C_{\text{MAX,SS}} e^{-k_{el}\tau} \quad (50)$$

The concentration at the midpoint of a dosing interval at the steady state is a time-averaged concentration, C_{AV} , over the entire dosing interval and is described by:

$$C_{\text{AV}} = \frac{D_{\text{IV}}}{V_D k_{el} \tau} = \frac{1.44 t_{1/2} D_{\text{IV}}}{V_D \tau} \quad (51)$$

NONCOMPARTMENTAL ANALYSIS FOLLOWING INSTANTANEOUS INPUT

There are numerous disadvantages associated with viewing drug disposition from a compartmental perspective, not the least of which is the lack of physiological relevance of such models. Noncompartmental models have been developed and are generally the *preferred* method for assessing overall drug disposition, in part because parameters such as volume of distribution and clearance can be calculated directly from the

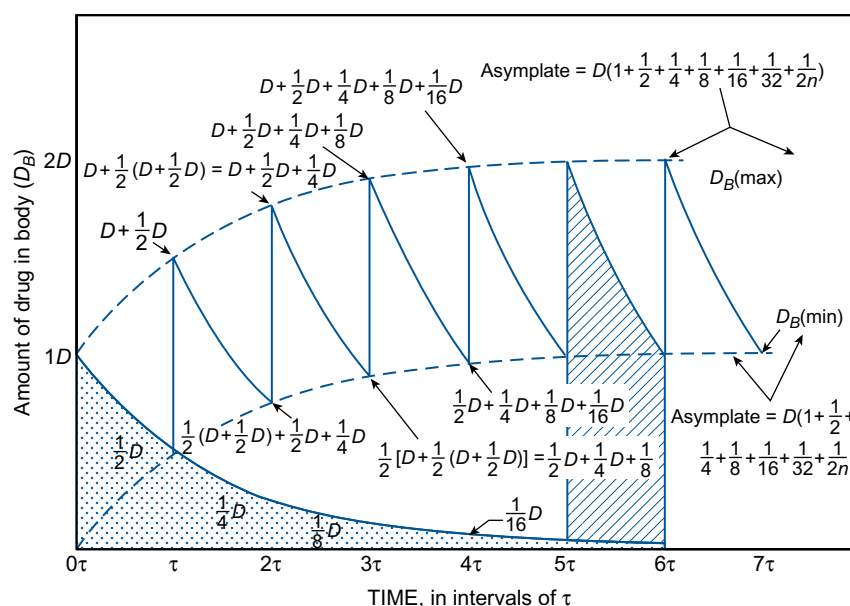


Figure 37-13. The accumulation of drug in the body during a regimen of multiple dosing. Dose, D , is administered intravenously at intervals, τ , equal to the half-life, $t_{1/2}$. Thus, after each dose, the amount in the body, D_B , has decreased to half the previous peak amount at the time each dose is administered. When the cumulated amount in the body after injection reaches $2D$, the body content will fluctuate from $2D$ to $1D$ during each dose interval thereafter. Approximately five half-lives are required before this leveling off (plateau) of the body content occurs. The stippled area is the area under the elimination curve of a single injection if no second dose had been given. The cross-hatched area is the area under the curve during a single-dose interval at steady state. The two areas are equal.

data, without computer fitting. Moreover, parameters estimated by these methods generally are less sensitive to variability in the data. Noncompartmental methods characterize drug disposition using time- and concentration-averaged parameters and have been described by Jusko.⁹ This analysis also is known as nonparametric analysis and assumes that all processes are first-order and that the parameters of the model reflect steady-state behavior. A primary tool used in this methodology is that of statistical moments.¹⁰

STATISTICAL MOMENTS

The use of statistical moments in the analysis of the time-course of drug concentrations is especially useful because it frees the investigator from the use of such models as the compartmental, which often are derived empirically and do not represent physiological events.

The time-course of drug concentration in blood generally can be viewed as a statistical distribution curve and described in a similar manner as any other array of data. A moment is simply a mathematical description of a discrete distribution of data. In the field of statistics, for example, the sample size (n), mean, and variance are the zero (M_0), first (M_1) and second (M_2) moments, respectively, for an array of data. In physics, for example, weight, center of mass, and moment of inertia represent M_0 , M_1 , and M_2 , respectively.

In statistics, the mean of a population is estimated by the sample mean. Similarly, in pharmacokinetics one may calculate estimates of the true function that describes the drug concentration versus time using statistical moments. Assume the existence of a theoretical relationship for $C(t)$ as a function of time. The non normalized moments, S_r , where $r = 0, 1, 2, \dots m^{\text{th}}$ moment, about the origin are calculated as:

$$S_r = \int_0^{\infty} t^r C(t) dt \quad (r = 0, 1, 2, \dots m) \quad (52)$$

Hence,

$$S_0 = \int_0^{\infty} C(t) dt = AUC \quad (53)$$

$$S_1 = \int_0^{\infty} tC(t) dt = AUMC \quad (54)$$

where $AUMC$ is the area under the C - t versus time curve, whereas S_0 and S_1 are the zero and first nonnormalized moments, respectively. These two parameters, AUC and $AUMC$, are derived from the drug concentration versus time data and are used to calculate the pharmacokinetic parameters of interest.

The use of noncompartmental methods requires a means of determining the AUC . While several methods are available for such determinations, the simplest is the use of the trapezoidal rule.¹¹ This permits calculation of the AUC or the $AUMC$ from zero to the time of the last sample (t_n). However, it generally is necessary to determine the area from zero to infinity. If one assumes that there is log-linear decline from t_n to infinity, then

$$AUC_{t^n}^{\infty} = \int_{t^n}^{\infty} C dt = \frac{C^n}{\lambda_z} \quad (55)$$

where λ_z is the slope of the terminal exponential. Thus, the AUC from zero to infinity can be calculated as

$$AUC_0^{\infty} = AUC_0^{t^n} + \frac{C^n}{\lambda_z} \quad (56)$$

For the $AUMC$,

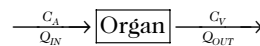
$$AUMC_{t^n}^{\infty} = \frac{C^n}{(\lambda_z)^2} + \frac{t^n C^n}{\lambda_z} \quad (57)$$

and

$$AUMC_0^{\infty} = AUMC_0^{t^n} + \frac{C^n}{(\lambda_z)^2} + \frac{t^n C^n}{\lambda_z} \quad (58)$$

PHARMACOKINETIC PARAMETERS DERIVED FROM STATISTICAL MOMENTS

There are four parameters of primary interest that are derived using statistical moments, the most important of which is clearance. A conceptual consideration of clearance is provided by considering an organ through which blood containing a drug flows.



where Q_{IN} , Q_{OUT} , C_A , and C_V are the blood flow in and out and the concentration of the drug in arterial and venous blood, respectively. If $C_V < C_A$, the organ is capable of elimination and is referred to as a clearing organ, or an organ of elimination. The elimination of a drug can be described using mass balance considerations:

$$\text{Rate of the drug entering the organ} = QC_A \quad (59)$$

$$\text{Rate of the drug leaving the organ} = QC_V \quad (60)$$

Rate of elimination of the drug:

$$\text{Rate of elimination of the drug} = QC_A - QC_V = Q(C_A - C_V) \quad (61)$$

The ratio of the rate of drug elimination and the rate at which the drug enters the organ is defined as the extraction ratio, E .

$$E = \frac{\text{Rate of elimination}}{\text{Rate of entry}} = \frac{Q(C_A - C_V)}{QC_A} = \frac{(C_A - C_V)}{C_A} \quad (62)$$

The extraction ratio is a measure of the efficiency with which an organ eliminates a drug. From this parameter, the organ clearance of a drug can be described as:

$$CL_T = QE = \frac{Q(C_A - C_V)}{C_A} \quad (63)$$

Recall that clearance is defined as the volume of blood from which all of the drug would appear to be removed per unit time. By analogy to the definition of organ clearance, one can define the total or systemic clearance as the ratio of overall elimination rate, dX/dt , to drug concentration in blood, C :

$$CL_T = \frac{dX/dt}{C} \quad (64)$$

Integrating from zero to infinity yields:

$$CL_T = \frac{\int_0^{\infty} \frac{dX}{dt} dt}{\int_0^{\infty} C dt} \quad (65)$$

where the numerator is the total amount of the drug ultimately eliminated (the IV dose) and the denominator is the AUC from zero to infinity. Thus, the total clearance is the quotient of intravenous dose and the AUC from zero to infinity.

$$CL_T = \frac{\text{Dose}_{IV}}{AUC_0^{\infty}} \quad (66)$$

The volume of distribution at the steady-state, V_{ss} , most reliably measured during a steady-state infusion, now can be determined using data from single-dose experiments and employing statistical moment analysis.¹²

$$V_{ss} = \frac{(\text{Dose}_{IV})(AUMC)}{(AUC)^2} \quad (67)$$

Equation 67 assumes that:

1. All processes involved in drug disposition (e.g., distribution, elimination) are linear.
2. The drug is administered to and eliminated via the sampling site.
3. There is instantaneous input.

If the drug is administered via a short infusion, the volume of distribution at the steady state can be estimated from:

$$V_{ss} = \frac{(R_0 T)(AUMC)}{(AUC)^2} - \frac{R_0 T^2}{2(AUC)} \quad (68)$$

Where R_0 is the rate of infusion and T is the duration of the infusion.

Another important pharmacokinetic parameter that can be determined using statistical moment analysis is the *systemic availability*, F , which is a measure of the fraction of the administered dose that reaches the systemic circulation following oral administration. This parameter can be calculated as:

$$F = \frac{AUC_{PO} Dose_{IV}}{AUC_{IV} Dose_{PO}} \quad (69)$$

where AUC_{PO} and $Dose_{PO}$ are the oral area under the concentration-versus-time curve and oral dose, respectively.

When administering a drug, the amount administered in terms of gross weight (e.g., mg, g, or μ g) often is considered. It is, however, probably more appropriate to focus on *molecules* when considering pharmacokinetic and pharmacodynamic events. Even the administration of a relatively small dose of drug may represent a large number of molecules. Consider the administration of 1 mg of a drug with a molecular weight of 300 daltons. The number of molecules in this dose is approximately 2×10^{18} . Instantaneous administration of the entire dose will result in drug molecules spending various amounts of time in the body. After the intravenous injection of a drug, one can imagine that some of the drug molecules are eliminated immediately, whereas some of the molecules require a longer time to be eliminated, and some molecules even require a very long time to be eliminated. The time spent in the body, for a given molecule, is its residence time. The *mean residence time*, MRT , is the sum of all the residence times divided by the number of molecules. A conceptual understanding of this can be gained from the following example.

Assume a child receives 20 dimes for his birthday and immediately places them in his piggy bank. Over the next month, he periodically removes one or more dimes from the piggy bank to purchase candy. Specifically, 3 days after placing the coins in his bank, he removes 5 dimes, on day 10 he removes 4 dimes, on day 21 he removes 6 dimes, and on day 30 he removes 5 dimes. At the 30th day after placing the coins in his bank, all of the coins have been removed. Hence, the *elimination* of dimes from the bank is complete. The MRT of dimes in the piggy bank is simply the sum of the times that coins spend in the bank divided by the number of dimes placed in the bank:

$$MRT = \frac{3 + 3 + 3 + 3 + 10 + 10 + 10 + 10 + 21 + 21 + 21 + 21 + 30 + 30 + 30 + 30}{20}$$

$$MRT = \frac{3 \times 5 + 10 \times 4 + 21 \times 6 + 30 \times 5}{20}$$

$$MRT = 16.55 \text{ days}$$

This provides a relationship with which one can determine the MRT for any given number of drug molecules, A_i , which spend a given amount of time in the body of, t_i , thus:

$$MRT = \frac{\sum_{i=1}^n A_i t_i}{A_{TOTAL}} \quad (70)$$

where n equals the total number of residence times. The mean rate of drug leaving the body relative to the total amount eliminated also can be expressed in terms of concentration.

$$MRT = \frac{\int_0^{\infty} tC(t)dt}{\int_0^{\infty} C(t)dt} = \frac{AUMC}{AUC} \quad (71)$$

Equation 71 is not a definition of MRT , rather, it is the derived expression from which one can calculate MRT when clearance is constant. The mean residence time assumes instantaneous administration, and therefore, it is technically incorrect to calculate the mean residence time following an oral dose using the quotient of $AUMC_{PO}$ and AUC_{PO} . When calculated in this manner, it often is stated that the MRT is a function of the route of administration. Actually, MRT is *independent* of the route of administration because the mean time that molecules reside in the body is not influenced by the route of administration.¹³ However, the interpretation of the ratio of $AUMC$ and AUC does change as a function of administration because this ratio only yields the MRT when the input is instantaneous.

A better way to express the route dependence of the $AUMC/AUC$ is to refer to this ratio as the *mean transit time*, MTT . The MTT is the average time required for drug molecules to leave a kinetic system after administration. Thus, because an IV bolus assumes instant input,

$$AUMC_{IV} / AUC_{IV} = MRT = MTT_{IV} \quad (72)$$

whereas,

$$\begin{aligned} AUMC_{PO} / AUC_{PO} &= MTT_{PO} = MTT_{IV} + MAT \\ &= MRT + MAT \end{aligned} \quad (73)$$

where MAT is the mean absorption time. Thus, for oral absorption, the $AUMC/AUC$ provides an MTT in the kinetic system that is composed of the gastrointestinal (GI) tract and the body. The MRT also can be calculated as the quotient of the V_{ss} and clearance. Finally, one can relate the MRT to the elimination half-life by considering the situation in which a drug displays monoexponential decline. The MRT can be written as:

$$MRT = \frac{AUMC}{AUC} = \frac{C_0 / \lambda^2}{C_0 / \lambda} = \frac{1}{\lambda} \quad (74)$$

and represents the time required for 63.2% of an intravenous dose to be eliminated from the body.

ABSORPTION

If a drug is administered intravenously in a single, rapid injection, the process of absorption is bypassed. The time for this injection is typically so short compared with other pharmacokinetic processes that it is ignored. As previously described for a one-compartment model, peak plasma concentration and distribution equilibrium are achieved instantaneously. This is depicted in Figure 37-14A.¹⁴ In the model for the figure, there is no elimination, and the concentration remains constant following administration. With a constant intravenous infusion (B), the concentration rises rectilinearly so long as the infusion is maintained at a constant, zero-order rate. With other routes of administration, there are delays in the appearance of drug in the vascular system because the drug must be absorbed from the site of administration (oral, intramuscular, subcutaneous, and rectal). Drug absorption depends upon both the physicochemical properties of the drug (pK_a , dosage form, partition coefficient) and the physiology of the site of absorption (surface area, blood flow). Most drugs are absorbed by simple diffusion, and the kinetics are first-order. Zero-order absorption occurs for some processes that are saturable and for sustained-release dosage forms. Absorption and elimination of a drug are a sequential process, and the rate of change of drug in the body is the difference between the rate of uptake (absorption) and rate of efflux (elimination). For a drug that is absorbed by a first-order process and eliminated by a first-order process, with

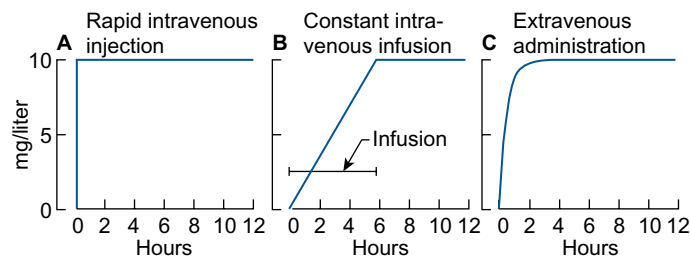


Figure 37-14. Time-concentration curves for injection (A), infusion (B), and extravenous (C) administration of drug in the one-compartment model. The volume of the compartment is 100 L (V_d 100 L); the amount of drug administered in each instance is 1000 mg. Drug elimination has been set to zero so that the time-concentration curve for each model of administration can be examined without the complication of simultaneous elimination. (Adapted from Bigger JT. Pharmacologic and clinical control of antiarrhythmic drugs. *Am J Med* 1975; 58: 479–488.)

instantaneous distribution, the rate of change of the amount of drug in the body can be expressed as:

$$\frac{dD_B}{dt} = \text{RATE}_{IN} - \text{RATE}_{OUT} \quad (75)$$

For a drug that is absorbed from the GI tract, the rate of change of the amount of drug in the body is

$$\frac{dD_B}{dt} = Fk_a D_{GI} - kD_{BODY} \quad (76)$$

The time-course of absorption and elimination is shown in Figure 37-15.⁷ The plasma concentration at any time t is equal to:

$$C = \frac{k_a D_0 F}{V_d (k_a - k_{el})} (e^{-k_{el}t} - e^{-k_a t}) \quad (77)$$

where F is the fraction of the dose, D_0 , that is absorbed from the GI tract and k_{el} and k_a are the first-order rate constants for elimination and absorption, respectively. The time to reach the maximum concentration, t_{MAX} , can be determined from:

$$t_{MAX} = \frac{2.3 \log(k_a / k)}{k_a - k} \quad (78)$$

and this time substituted into Equation 77 will determine the maximum concentration, C_{MAX} . The rising phase of the plot (see Fig 37-15) is not log-linear because absorption and elimination are occurring simultaneously. At C_{MAX} , the absorption rate is equal to the elimination rate, and after absorption is complete, the plot declines in a log-linear manner. This log-linear line described by the elimination phase, when extrapolated to zero time, yields a theoretical zero-time concentration. The absorption rate constant, k_a , can be obtained from the difference between the empirical curve and the extrapolated line using the *method of residuals*. This is a commonly used technique in pharmacokinetics to separate a curve into its component parts and is often referred to as *feathering*, *stripping*, or *peeling* the curve. The reader is referred to Gibaldi and Perrier (see *Bibliography*) for a more comprehensive discussion with examples of the application of this technique.

That the peak concentration should vary with the dose is self-evident from Equations 77 and 78 and from Figure 37-16. The time of the peak concentration is the same for all doses. The time to peak concentration can be affected by both the absorption rate and the elimination rate. In Figure 37-17, the effect of altering the absorption rate on the time to peak concentration is shown. With faster absorption, the time to peak concentration occurs earlier and is higher than with slower absorption. Figure 37-18 shows the effect of altering the elimination rate constant on the t_{MAX} . With a rapid elimination rate (shorter

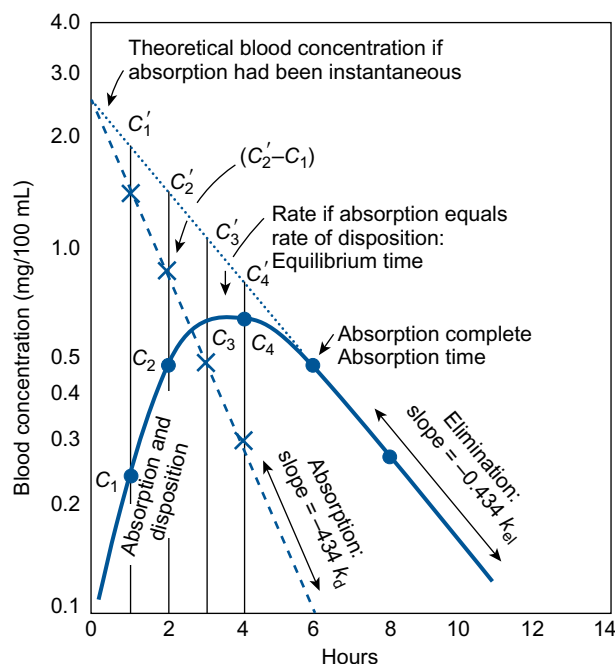


Figure 37-15. Kinetics of absorption and disposition of theophylline in a human subject after oral administration of 0.5 g of aminophylline per 70 kg. Blood concentration is plotted on a log scale. (Data from Truitt, Jnr EB *et al.* Theophylline blood levels after oral, rectal, and intravenous administration, and correlation with diuretic action. *J Pharmacol Exp Ther* 1950; 100: 309–315.)

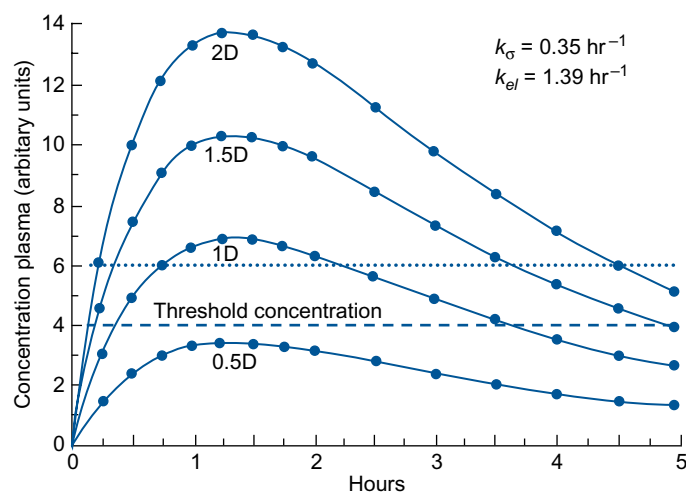


Figure 37-16. The effect of the size of the dose of a drug on the peak concentration, time of peak concentration, and duration of action. The data were calculated from a one-compartment model.

half-life), the peak concentration occurs sooner and is lower than with slower elimination (longer half-life).

The maximum concentration at the steady state for an oral regimen is given by:

$$C_{MAX,SS} = \frac{FDose}{V_d} \left(\frac{1}{1 - e^{-k\tau}} \right) e^{-k t_{PEAK}} \quad (79)$$

The minimum concentration at the steady state is:

$$C_{MIN,SS} = \frac{k_a FD}{V(k_a - k)} \left(\frac{1}{1 - e^{-k\tau}} \right) e^{-k\tau} \quad (80)$$

To design multiple oral dosing regimens, the equations that describe the maximum and minimum concentrations at the

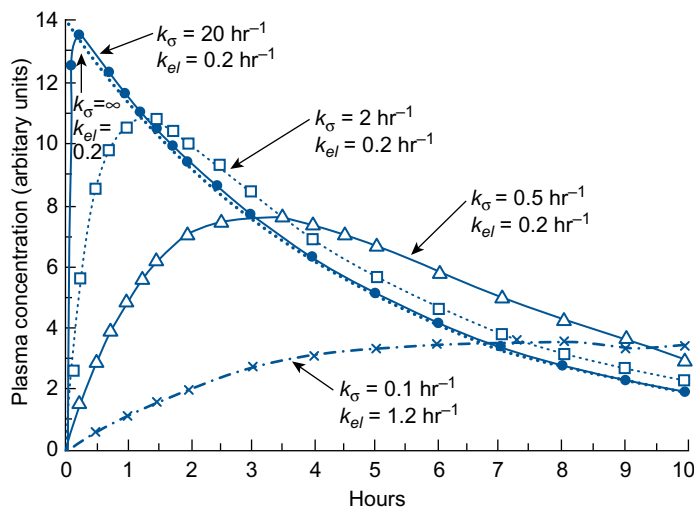


Figure 37-17. The effect of differences in the rate of absorption of drugs on the peak concentration, time of peak concentration, and sojourn in the body. The rate of elimination is the same for all curves. The dotted line ($k_a = \infty$) is approximately what the concentration curve would be, had the drug been given intravenously. The data were calculated from a one-compartment model.

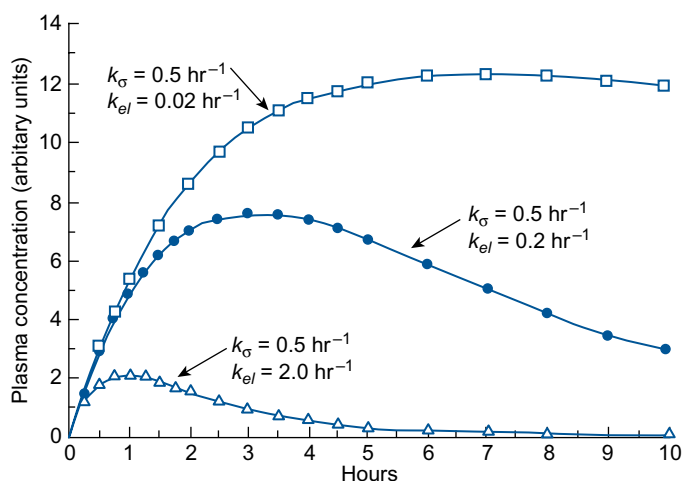


Figure 37-18. The effect of differences in the rate of elimination of drugs on the peak concentration, time of peak concentration, and sojourn in the body. The rate of absorption is the same for all curves. The data were calculated from a one-compartment model.

steady state are somewhat unwieldy. In clinical practice, values for k_a are not always readily available, and in such instances the equations for the maximum and minimum concentrations at the steady state following intravenous administration will generally suffice as long as one recalls that because absorption is not instantaneous, the peak concentration and time to peak concentration will not occur immediately.

ORGAN-SPECIFIC CLEARANCE

The total clearance of a drug from the body almost always involves more than one organ of elimination. The anatomy of the human body is such that the clearance from the composite clearing organs occurs in parallel and is, therefore, additive. For example, if a drug is eliminated solely by hepatic and renal elimination, the total clearance, CL_T , of the drug is given as

$$CL_T = CL_H + CL_R \quad (81)$$

where CL_H and CL_R are the hepatic and renal clearance, respectively. Measurement of the total amount of drug excreted unchanged in urine, D_U , after an intravenous dose, D_{IV} , allows the calculation of the fraction of the drug eliminated renally, F_r , where

$$F_r = \frac{D_U}{D_{IV}} \quad (82)$$

The renal clearance, CL_R , may be determined as the product of total clearance and the fraction of the drug eliminated by the kidney. If the liver is the only other eliminating organ, the hepatic clearance is given by

$$CL_H = CL_T - CL_R \rightarrow (1 - F_r)CL_T \quad (83)$$

One exception to the principle of the additivity of organ clearances is pulmonary drug elimination. This exception is because the lung is in circulatory series with the rest of the body organs such that 100% of cardiac output passes through the lungs. Few drugs exhibit significant elimination by the lungs so that this exception is rarely of concern in the overall assessment of drug elimination.

HEPATIC CLEARANCE

It was shown previously that the ratio of the rate of drug elimination and the rate at which drug enters the organ of elimination is defined as the extraction ratio, E , and is a measure of the efficiency with which an organ eliminates a given drug. One can define the organ clearance of a drug as the product of blood flow to the organ, Q , and the extraction ratio, and for hepatic clearance the equation becomes:

$$CL_H = Q_H E \quad (84)$$

where Q_H is the hepatic blood flow. While an initial examination of this simplistic model for hepatic clearance would suggest that CL_H is directly proportional to Q_H , this conclusion is not correct because E varies inversely with Q_H . Specifically, as Q_H increases, E decreases. This observation indicates that a more complex model of hepatic clearance is necessary if quantitative and qualitative predictions of hepatic drug clearance are to be made. In particular, this parameter must be described in terms that are physiologically independent.

Numerous models have been proposed and tested to describe the hepatic clearance of drugs. While a discussion of the advantages and disadvantages of the various models proposed is beyond the scope of this chapter, the *venous equilibrium* model of hepatic clearance has shown substantial utility in the prediction of both pathophysiological and drug-induced changes in hepatic clearance. For a good discussion of the various models of hepatic clearance, see the review by Morgan and Smallwood.¹⁵ In the venous equilibrium model, the hepatic extraction is described by:

$$E = \frac{f_{ub} CL_{u,int}}{Q_H + f_{ub} CL_{u,int}} \quad (85)$$

where f_{ub} and $CL_{u,int}$ are the unbound fraction in blood and the unbound intrinsic hepatic clearance, respectively. The unbound intrinsic clearance reflects the ability of the liver to remove drug from blood in the absence of other confounding factors, such as Q_H and f_{ub} . Since it has already been shown that hepatic clearance is the product of Q_H and E ,

$$CL_H = \frac{(Q_H)(f_{ub} CL_{u,int})}{Q_H + f_{ub} CL_{u,int}} \quad (86)$$

This model for hepatic clearance provides a powerful tool for predicting changes in drug clearance and, subsequently, steady-state drug concentrations, when certain limiting conditions are met. In particular,

when $Q_H \gg f_{ub} CL_{u,int}$, CL_H can be approximated by $f_{ub} CL_{u,int}$ (87)

When $Q_H \ll f_{ub} CL_{u,int}$, CL_H can be approximated by Q_H (88)

Compounds with a high $f_{ub} CL_{u,int}$ are said to exhibit perfusion rate-limited elimination; that is, their elimination rate will be rate-limited by hepatic blood flow. Compounds with a low $f_{ub} CL_{u,int}$ are said to be perfusion rate-independent. These limiting conditions allow us to place many drugs into classifications that exhibit similar pharmacokinetics. For example, agents with an $f_{ub} CL_{u,int} < 0.2$ L/min can be classified as low intrinsic clearance drugs, whereas those with an $f_{ub} CL_{u,int} > 5$ L/min are defined as exhibiting a high intrinsic clearance (Table 37-1).

The venous equilibrium model also serves as a useful tool in the assessment of the impact of changes in protein binding on hepatic clearance. Recall in Equation 87 that for a drug exhibiting a low intrinsic clearance, changes in protein binding will result in proportional changes in hepatic clearance. This type of drug is said to exhibit *restrictive clearance*; that is, only the free (or unbound) drug is available for clearance by the liver. High intrinsic clearance drugs, on the other hand, are said to exhibit *nonrestrictive clearance*.

These relationships provide important insight into the effect of changes in protein-binding on the steady-state concentration of drugs. Consider the case of a drug being administered as a constant-rate intravenous infusion. As described previously:

$$C_{ss} = R_0 / CL_T \rightarrow R_0 / CL_H \quad (89)$$

for a drug solely eliminated by the liver. In the case of a drug with a low intrinsic clearance, Equation 89 can be simplified to:

$$C_{ss} = R_0 / (f_{ub} CL_{u,int}) \quad (90)$$

If f_{ub} were to be increased, for example, by displacement from protein-binding sites, the steady-state concentration would decrease. This may lead one to the conclusion that the infusion rate needs to be increased to maintain the original steady-state concentration. However, one needs to examine the effects of altered physiology on the free or unbound drug concentration, $C_{u,ss}$:

$$C_{u,ss} = f_{ub} C_{ss} \rightarrow C_{ss} = C_{u,ss} / f_{ub} \quad (91)$$

substituting for C_{ss} in Equation 90 and solving for $C_{u,ss}$ yields

$$C_{u,ss} = R_0 / CL_{u,int} \quad (92)$$

It can be seen that the steady-state concentration of unbound (active) drug is independent of changes in the free fraction, and no dosage adjustment would be necessary. This conclusion also is valid following the oral administration of a drug with a low intrinsic clearance.

In contrast, for a drug with a high intrinsic clearance, a change in f_{ub} will result in a proportional change in the steady-state unbound drug concentration during a constant-rate infusion.

Table 37-1. Examples of Drugs with Low and High Intrinsic Clearances that are Eliminated Largely by Hepatic Metabolism

Low intrinsic clearance	High intrinsic clearance
Antipyrine	Chlorpromazine
Barbiturates	Encainide
Diazepam	Meperidine
Digitoxin	Metoprolol
Isoniazid	Organonitrates
Phenytoin	Propafenone
Theophylline	Propranolol
Tolbutamide	Tricyclic antidepressants
Warfarin	Verapamil

An additional consideration, which must be accounted for with high intrinsic clearance drugs, is the impact of the first-pass effect. When a drug is absorbed from the stomach and small intestine, the venous blood from the sites of absorption enters into the portal venous flow. This results in all the absorbed drug passing through the liver prior to entry into the systemic circulation. For drugs that exhibit a high intrinsic clearance, the consequence of presystemic hepatic metabolism is a substantial reduction in the systemic availability of the drug when administered orally. This phenomenon explains the marked discrepancy between an oral and an intravenous dose of a given drug required to achieve identical plasma concentrations. For example, the therapeutic dose of propranolol ranges between 1 and 6 mg intravenously, whereas the oral doses necessary to achieve therapeutic effect range from 40 to 200 mg.

The systemic availability, F , of a drug that is absorbed completely from the GI tract after oral administration is the fraction of the absorbed dose that escapes extraction and is given as:

$$F = 1 - E \quad (93)$$

Rearranging Equation 85 and substituting for E in Equation 93 yields

$$F = \frac{Q_H}{Q_H + f_{ub} CL_{u,int}} \quad (94)$$

Similar to the limiting conditions described for CL_H , one can define two limiting conditions for systemic availability, F . Specifically, when $Q_H \ll f_{ub} CL_{u,int}$, F approaches 1.0, whereas when $Q_H \gg f_{ub} CL_{u,int}$, F approaches zero. These limiting conditions indicate that a drug with a high $f_{ub} CL_{u,int}$ will exhibit low systemic availability after oral administration because of extensive first-pass metabolism. On the other hand, drugs with a low $f_{ub} CL_{u,int}$ will not be subject to significant first-pass metabolism.

Generally, the parameter most commonly determined to assess overall drug availability after oral administration is the area under the drug concentration-versus-time curve, AUC . Recall from Equation 67 that the total clearance is equal to the quotient of intravenous dose and AUC from zero to infinity. If the drug is eliminated entirely by metabolism, the hepatic clearance is defined as:

$$CL_H = \frac{Dose_{IV}}{AUC_{IV}} = \frac{FDose_{PO}}{AUC_{PO}} \quad (95)$$

Substituting Equation 86 for CL_H and Equation 94 for F yields:

$$\frac{(Q_H)(f_{ub} CL_{u,int})}{Q_H + f_{ub} CL_{u,int}} = \left(\frac{Dose_{PO}}{AUC_{PO}} \right) \left(\frac{Q_H}{Q_H + f_{ub} CL_{u,int}} \right) \quad (96)$$

Simplifying,

$$f_{ub} CL_{u,int} = Dose_{PO} / AUC_{PO} \quad (97)$$

Thus, for a high intrinsic hepatic clearance drug, the AUC_{PO} is independent of Q_H . Additionally, the steady-state AUC for unbound drug is independent of the free fraction. For a more in-depth discussion of these concepts on hepatic clearance, see the paper by Wilkinson and Shand.¹⁶

RENAL CLEARANCE

Physiologists studied the renal clearance of endogenous and exogenous substances long before the use of clearance concepts became popular in pharmacokinetics. Indeed, the basis for the understanding of drug clearance in pharmacokinetics has its roots in the decades of work by renal physiologists. Moreover, there are significant differences in the complexity of processes involved in hepatic and renal handling of drugs. In the kidney there are three primary processes (and one minor process) responsible for the renal elimination of drugs, namely filtration, secretion, and reabsorption (and metabolism), respectively. Each of the major processes is affected by common, yet unique, determinants.

The rate of filtration in the kidney for a drug is given as:

$$\text{Rate of filtration} = (GFR)(C_u) \quad (98)$$

where GFR is the glomerular filtration rate and C_u is the previously defined free-drug concentration. The clearance by filtration is the quotient of the rate of filtration at a given concentration; therefore, the renal clearance, CL_R , of a drug due to filtration is:

$$CL_R = (GFR)(f_u) \quad (99)$$

The renal clearance of a drug that is eliminated only by filtration can be estimated if the glomerular filtration rate and the free fraction of drug are known. There are two substances commonly used to estimate the GFR , namely creatinine, an endogenous by-product of muscle metabolism, and inulin, a polysaccharide. Both are essentially 100% eliminated in the urine by filtration and, thus, the CL_T of these two substances can be used as reasonable estimates for the GFR .

Another primary process involved in the renal elimination of drugs is active tubular secretion, ATS . There are several active-transport systems in the proximal renal tubule that are capable of excreting drugs from the blood into the urine. There appears to be a multiplicity of systems for the ATS of cations and anions. Whenever the renal clearance is greater than the product of the GFR and the free fraction, there must be net tubular secretion in addition to clearance by filtration. The renal clearance due to ATS , which is CL_{ATS} , is given as:

$$CT_{ATS} = \frac{(Q_{RP})(f_u CL_{u,s,int})}{Q_{RP} + f_u CL_{u,s,int}} \quad (100)$$

where Q_{RP} and $CL_{u,s,int}$ represent effective renal plasma flow and unbound intrinsic secretory clearance, respectively. Similar to the situation described for hepatic clearance, drugs may exhibit a high or low intrinsic secretory clearance. The impact of changes in plasma protein-binding on renal clearance would substantially differ between these two situations.

For drugs that undergo both filtration and ATS , the renal clearance is simply the sum of the clearance due to filtration and the clearance due to secretion,

$$CL_R = (f_u)(GFR) + CL_{ATS} \quad (101)$$

In addition to these two processes, some drugs undergo tubular reabsorption, whereby some fraction of the drug excreted into the urine by filtration and secretion is reabsorbed into the body. Therefore, the full expression for renal clearance, taking into account all three processes, is given by:

$$CL_R = (f_u)(GFR) + CL_{ATS} - F_{TR}((f_u)(GFR) + CL_{ATS}) \quad (102)$$

or $CL_R = (1 - F_{TR})((f_u)(GFR) + CL_{ATS})$

where F_{TR} is the fraction undergoing tubular reabsorption. These relationships provide the basis for determining the primary mechanisms involved in the renal handling of a given drug. Specifically, determination of the ratio of renal clearance and that of inulin provides a clinically useful means to determine the processes that are primarily responsible for renal excretion. If the ratio is equal to 1, the drug would appear to be filtered exclusively by the kidney. Both ATS and TR also could be occurring, but at equal rates (an unlikely occurrence). If the ratio of renal clearance to inulin clearance is greater than 1, it is clear that the drug is undergoing net tubular secretion as well as filtration. Similarly, if the ratio is less than 1, the drug must be undergoing net tubular reabsorption.

Assessment of the mechanisms for the renal excretion of specific drugs is important because different factors will alter CL_R . For example, if a drug undergoes net tubular secretion, other drugs secreted by the same transport processes may compete for secretory sites, resulting in an overall decrease in the renal excretion of the drug. Alternatively, if a drug undergoes tubular re-absorption and is a weak acid (e.g., salicylate) or a weak

base (e.g., amphetamine), the renal clearance may be altered by manipulation of the urine pH or urine flow. See the review by Tucker¹⁷ for a description of the methods for calculation of CL_R .

PROTEIN-BINDING

Drugs circulating in the blood may bind reversibly to a number of components, including plasma proteins. This reversible binding may be described by simple mass-law relationships,

$$[D_{unbound}] + [P] \leftrightarrow [D - P] \quad (103)$$

where $[D_{unbound}]$, $[P]$, and $[D - P]$ are the molar concentrations of the free drug, the protein to which the drug binds, and the drug-protein complex, respectively. It is obvious from this relationship that the amount of drug-protein complex formed is a function both of the concentration of the drug and protein and the affinity between the protein and the drug. Thus, changes in the protein concentration may alter binding, as may changes in the total drug concentration. For most drugs and their respective binding proteins, the concentration of protein far exceeds the concentration of drug, such that the fraction of drug that is bound to protein is independent of drug concentration in plasma or blood.

Because plasma proteins often have a molecular size that restricts their passage across cell membranes and capillary walls, drugs that are bound to plasma proteins are restricted similarly. Thus, plasma protein-binding can have a marked effect on the distribution and elimination of drugs. It is a basic tenet of pharmacology that only the unbound (free) drug is pharmacologically active, because it is assumed that the unbound drug is able to traverse biological membranes and reach the site of drug action. While there have been few direct tests of this hypothesis, those investigations that have been conducted support the assumption that the free drug is the principal pharmacologically active species.

There are several methods by which drug protein-binding may be described quantitatively, though the most frequent and useful is the free fraction, f_u . The f_u can be determined as:

$$f_u = \frac{C_u}{C_u + C_B} = \frac{C_u}{C_T} \quad (104)$$

where C_u is the concentration of unbound or free drug, C_B is the concentration of drug bound to protein, and C_T is the concentration of total drug (bound plus free). Obviously, f_u can range from 0 to 1. This relationship provides a means by which free drug *in vivo* can be calculated if the total concentration and the free fraction are known.

$$C_u = (C_T)(f_u) \quad (105)$$

Recognizing changes in protein-binding is important because it may substantially alter the pharmacokinetics of a drug. The previous section described the impact of protein-binding on clearance, referenced to total drug concentration. Protein-binding changes also may result in alterations in other pharmacokinetic parameters. The volume of distribution at the steady state can be expressed as:

$$V_{SS} = V_{blood} + V_{TW} \frac{f_u}{f_{uT}} \quad (106)$$

where V_{TW} is the volume of tissue water and f_{uT} is the free fraction of drug in tissue. From the relationship in Equation 106 (which is analogous to Equation 12, for drugs with a volume of distribution >50 L), it is clear that a decrease in protein-binding (i.e., an increase in f_u) will result in an increase in the V_{SS} . The impact of changes in f_u on the half-life of a drug with a large volume of distribution can be assessed from Equation 18 and either Equations 87 or 88. The impact resulting from a change in protein-binding of a drug depends upon the magnitude of such binding alterations on both V_{SS} and CL_T .

The two major drug-binding proteins in plasma are albumin and α_1 -acid glycoprotein (AAG). Albumin is the major protein

both in plasma and in the extracellular space outside of plasma and is present in concentrations ranging from 3.5 to 5.5 g/dL in normal, healthy individuals. Albumin is the primary binding protein for acidic drugs, such as salicylate, tolbutamide, or warfarin. Numerous diseases can result in marked reductions in the concentration of albumin, including nephrotic syndrome, severe burns, liver disease, malnutrition, and some chronic inflammatory conditions.¹⁸ Thus, disease is most likely to produce an increase in f_u for those drugs highly bound to albumin.

The substance AAG belongs to the family of *acute-phase reactants*, endogenous substances that are markedly increased in concentration secondary to some type of stress. While normal AAG concentrations range from 80 to 120 mg/dL, concentrations may increase above 300 mg/dL in patients experiencing major stress, such as surgery, trauma, or burns. More moderate elevations of AAG have been observed in patients following a myocardial infarction or in inflammatory diseases, such as Crohn disease. The increase in AAG concentration results in a decrease in f_u , and AAG is the major binding protein for many basic lipophilic drugs.

A major source of drug interactions is competition for protein-binding sites. Each albumin molecule contains at least four different drug-binding sites, two of which are the sites where most drugs that bind to albumin interact with the other molecule. If two drugs bind to the same site on a protein, they may compete with each other for binding. Thus, the addition of a drug to the existing therapeutic regimen of a patient may result in displacement of existing bound drug molecules from their protein-binding sites. However, as described in the *Hepatic Clearance* section of this chapter, these types of interactions rarely are clinically significant (i.e., these interactions do not significantly alter the free-drug concentration). Hence, while protein-binding interactions are probably the most widely reported drug interaction, they rarely necessitate alterations in drug therapy.¹⁹

ROLE OF TRANSPORTERS IN ALTERING PHARMACOKINETICS

As previously discussed, tissue (cellular) distribution of drugs is a function of tissue perfusion, drug permeability and plasma and tissue protein binding. Influx and efflux drug transporters are also now acknowledged to be major determinants controlling cellular drug accumulation. For a number of chemically diverse drugs (atorvastatin, methotrexate, enalapril, and valsartan) hepatic uptake by a specific transporter (i.e., OATP1B1 or organic anion transporter protein) is a prerequisite to hepatic biotransformation. In the case of atorvastatin (and other statin drugs), the liver is both a distribution tissue and an organ of elimination. A decrease in hepatic OATP activity, either by concomitant administration of a transporter inhibitor, such as rifampin or cyclosporine, or by the presence of transporter gene polymorphism, can produce a decrease in both volume of distribution and drug clearance.²⁰

DOSE-AND TIME-DEPENDENT PHARMACOKINETICS

Up to this point, the processes for drug absorption, distribution, metabolism, and elimination have been assumed to be characterized by first-order rate constants, and the general concepts and equations presented are applicable to a wide variety of drugs, with modification. Moreover, with any of the pharmacokinetic models (compartmental, noncompartmental, or physiological), a number of basic assumptions apply, in particular, the principle of superposition holds. In other words, measurements of the concentration of drug plasma, urinary excretion of unchanged drug, or amount of metabolite recovered in bile increase proportionally with increases in dose. When these measurements or other observations are corrected for dose, the values are identical or superimposable. Thus, the

pharmacokinetic parameters V_D , CL_T , and F remain constant with respect to time and with dose or concentration.

But the processes controlling the disposition of drugs are biological and therefore involve processes that are mediated by specialized carriers or enzymes. Under some conditions these processes can become saturated, and changes in dose may produce nonproportional changes (e.g., in concentration, amount of metabolite(s) produced, etc.) Table 37-2 delineates some of the various causes of nonlinear pharmacokinetic behavior.

Nonlinearity is a term applied to all situations in which a semilogarithmic plot of plasma concentration versus time data cannot be resolved completely into log-linear components (i.e., first-order processes). There are a wide variety of causes for nonlinearity, such as capacity-limited metabolism, capacity-limited absorption, saturable first-pass metabolism, changes in blood supply to the site of absorption and/or the organ of elimination, low or erratic dissolution or release rates from dosage forms, low solubility of the drug, or drug-induced changes in organ function or body temperature. Nonlinear drug disposition primarily has been determined by measuring the pharmacokinetics at several dosage levels. When a capacity-limited enzyme metabolism is the source of the nonlinearity, the Henri Michaelis-Menten equation,

$$\text{Velocity} = \frac{V_{\text{MAX}}C_{\text{SS}}}{K_M + C_{\text{SS}}} \quad (107)$$

can be applied to assess the velocity versus substrate (drug) concentration relationship. There are several techniques for the determination of the direct cause of nonlinearity in drug kinetics, including direct calculation of CL_T , CL_{ORAL} , F , V_{SS} , and V_1 . Most commonly, lack of superposition (disproportionate increase in AUC with increasing dose) is an indication of nonlinearity in the system.

Table 37-2. Examples of Mechanisms for Dose- and Time-Dependent Pharmacokinetics (Nonlinear Drug Disposition)

Kinetic Process and Mechanism	Examples
Gastrointestinal absorption	
Saturable transport	Riboflavin, penicillins
Intestinal metabolism	Salicylamide
Biotransformation	
Saturable metabolism	Phenytoin, salicylate
Product inhibition	Phenytoin (rat)
Cosubstrate depletion	Acetaminophen
Plasma protein-binding	Prednisolone, disopyramide
Renal excretion	
Glomerular filtration/protein-binding	Naproxen
Tubular secretion	<i>p</i> -Aminohippuric acid, mezlocillin
Tubular reabsorption	Riboflavin, cephapirin
Biliary excretion	
Biliary secretion	Iodipamide, BSP
Enterohepatic cycling	Cimetidine, isotretinoin
Tissue distribution	
Plasma protein-binding	Prednisolone, ceftriaxone
Hepatic uptake	Indocyanine green, warfarin (rat)
CSF transport	Benzylpenicillins
Cellular uptake	Methicillin (rabbit)
Tissue-binding	Cyclosporine, dideoxyinosine (rat)

PROTEIN-BINDING

When the amount of drug bound to plasma proteins approaches saturation, the percentage of the drug that is unbound may vary considerably with increasing dose. Under the conditions of saturation, for example, after a salicylate overdose, the f_u may vary considerably with the total amount of drug in the body, and hence, certain pharmacokinetic parameters, such as apparent volume of distribution, will be influenced.

TIME-DEPENDENT KINETICS

Carbamazepine is a drug with low-to-intermediate intrinsic clearance, which also induces an increase in the activity of the biotransforming enzyme system by which it is metabolized. This increase will increase total clearance and decrease half-life. Because such autoinduction of metabolism does not occur until several dose-intervals of repetitive dosing, the pharmacokinetics vary with time and are called time-dependent. *Allosteric* (feedback) inhibition by accumulated metabolites of a drug, or an effect of a drug to impair its route of elimination, also will cause dose-and time-dependent changes in pharmacokinetics. Drugs that cause the depletion of some slowly replaceable intermediary factor, such as the depletion of norepinephrine by reserpine or the depletion of inorganic sulfate by acetaminophen, will manifest time-dependent effects.

DOSE DEPENDENT KINETICS

When the elimination route is saturated by either capacity-limited metabolism or capacity-limited renal excretion, it is evident that the total clearance of a drug will decrease and the half-life will increase with increases in doses. Examples of important drugs that demonstrate dose-dependent kinetics are salicylic acid, phenylbutazone, probenecid, levodopa, phenytoin, heparin, and dicumarol. Ethanol obeys essentially zero-order elimination kinetics at blood concentrations above 0.02 to 0.04%, which is a fact of considerable social and legal importance.

Salicylic acid is one of the most interesting examples of dose-dependent kinetics from multiple sources. Salicylic acid is eliminated from the body by at least five parallel, competing processes.²¹ Two of these are saturable processes for the formation of salicyluric acid (the glycine conjugate of salicylic acid) and salicylphenol glucuronide. The other three processes of elimination, excretion of unchanged salicylic acid in urine and formation of gentisic acid and salicyl glucuronide, are first-order processes. The half-life of salicylic acid increases from about 3 hours to over 20 hours as the dose is increased upward from 300 mg to 10 g. At low doses, the half-life is about 3 hours, the apparent volume of distribution is approximately 9L, and a total clearance can be estimated to be 2 L/hr. The binding of salicylic acid to albumin also is capacity-limited (saturable), and saturation occurs even at therapeutic (low) doses of the drug. Therefore, as the amount of salicylic acid in the body increases, the $CL_{u,int}$ decreases, whereas the f_u increases. These two effects tend to oppose each other, such that the total clearance of salicylic acid remains relatively unchanged within the anti-inflammatory range of unbound concentrations (between 10 and 60 mg/L; Fig 37-19).²² Finally, the renal excretion of salicylic acid ($pK_a = 3.5$) can be increased by increasing urinary pH, resulting in a decrease in renal tubular reabsorption (data not shown). The toxicological consequences of a salicylic acid overdose are well known, but it is not always appreciated that as the concentration of salicylic acid increases, the total amount of drug in the body increases out of proportion to the total plasma concentration.

STEREOCHEMICAL CONSIDERATIONS

Chiral drugs constitute approximately 60% of the drugs that are currently commercially available. Most of these are marketed as racemic mixtures. These facts obviously indicate the importance of understanding the impact of stereochemistry on both

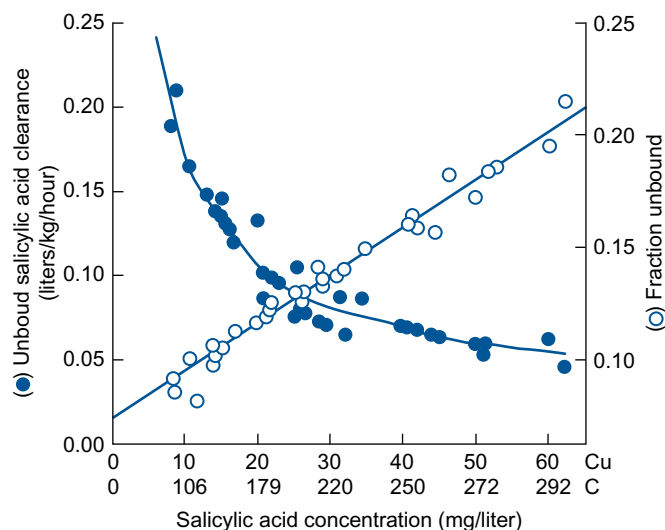


Figure 37-19. The clearance of unbound drug (●), determined under steady-state conditions, and the fraction unbound in plasma (○) vary inversely with each other as the salicylic acid concentration is increased. The corresponding total plasma salicylic acid concentrations are superimposed on the linear scale of the concentration of unbound drug; 1 mg/L = 7.2 micromolar. (Redrawn from Furst DE *et al.* Salicylate clearance, the resultant of protein binding and metabolism. *Clin Pharmacol Ther* 1979; 26: 380–389).

pharmacokinetics and pharmacodynamics. While it has long been appreciated that optical isomers often differ in the potency of pharmacological or toxic effect, it is only recently that significant attention has been paid to the influence of chirality on pharmacokinetic processes involved in absorption, distribution, and elimination. This is primarily due to the previous lack of analytical methodology required to separate drug enantiomers. The recent development of reasonably inexpensive methods for the separation of stereoisomers has led to a more comprehensive assessment of the pharmacokinetics of drugs that are administered as racemic mixtures.

Enantiomers possess identical physical and chemical properties, despite significant differences in spatial configuration. Thus, biological processes that are passive in nature (and thereby depend only upon physical and chemical characteristics of the molecule) do not display selectivity for one isomer over another. In contrast, biological processes that require the interaction of a drug molecule with a macromolecule (such as protein-binding or metabolism) may exhibit stereoselectivity. This knowledge permits some generalizations about when pharmacokinetic processes may differ between enantiomers.

ABSORPTION

Since most drugs are absorbed by passive diffusion, most will not exhibit stereoselective alterations in absorption. Drugs that are absorbed by a carrier-mediated or active process may display such stereoselectivity. Indeed, demonstration of stereoselective absorption would be strong evidence that a drug is absorbed via a carrier-mediated process.

PROTEIN-BINDING

Drug association with plasma proteins requires interaction of a small molecule with a macromolecule, which depends upon the spatial configuration of both components. It should not be surprising, therefore, that plasma protein-binding has been found to exhibit stereoselectivity for some drugs, including disopyramide, ibuprofen, mexilitene, propranolol, and verapamil.

METABOLISM

Biotransformation requires the interaction of a drug with an enzyme, an interaction in which spatial arrangement is critical.

Many drugs that undergo metabolism exhibit stereoselective hepatic clearance. For example, the oral clearance of verapamil (a high intrinsic clearance drug) displays profound stereoselectivity such that the oral clearance ratio of the *R* to *S* isomers is approximately.⁴

RENAL EXCRETION

Filtration in the kidney is a passive process; however, if a drug exhibits stereoselective protein-binding, one might anticipate that the drug enantiomers would exhibit differential filtration rates. Active tubular secretion, being an active process, also may demonstrate stereoselectivity for some drugs. Indeed, numerous drugs, including chloroquine, disopyramide, and terbutaline, have been found to be secreted stereoselectively by the kidney. While passive tubular reabsorption would not be expected to show stereoselective effects, active reabsorption may demonstrate these effects, as has been shown for certain endogenous substances such as glucose and amino acids.

KINETICS OF PHARMACOLOGIC EFFECT

In addition to considering the relationship between drug concentration and time, proper design of therapeutic regimens often necessitates an understanding of the relationship between concentration and response. These two relationships can be combined to produce a time course of pharmacologic response, and this is referred to as pharmacodynamics. This area of investigation provides important quantitative information regarding the onset, duration, and intensity of pharmacologic effect, often in relation to drug concentration. Full characterization of these elements can involve the development of pharmacokinetic/pharmacodynamic models. Modeling the kinetics of effect requires an understanding of the mechanism of pharmacologic action, of which there are many. These include receptor stimulation (e.g., β_2 -adrenoceptor agonists), receptor antagonism (e.g., H_1 -histamine receptor antagonists), transporter inhibition (probenecid, diuretics), enzyme inhibitors (e.g., angiotensin converting enzyme), substrate replacement (e.g., thyroxine, testosterone), non-receptor-mediated drug action (e.g., chelation) and chemotherapy (e.g., antibiotics).

Clinically useful information can often be derived from relatively simplistic models. The relationship between drug response and concentration is usually graded, that is, the rise in concentration results in a progressively increasing magnitude of effect (Fig 37-20). There is essentially always a ceiling, or maximum, to the intensity of effect, such that further increases in drug concentration will not result in additional increases in the intensity of pharmacologic response. In practice, plasma concentrations rarely, if ever, reach the maximum effect because toxicities often develop. Thus, only the linear portion of the concentration-effect curve may be observed. The placement of the curve along the *x*-axis may vary among compounds, resulting in differences in the potency the drugs. Moreover, the steepness of the curve (the rate at which the intensity of the effect changes as a function of concentration) will vary among drug responses. For some drugs, such as narcotic agonists, a small change in the drug concentration can result in marked changes in drug effect (see Fig 37-20).

There are certain agents that appear to exhibit an all-or-none response. Rather than the typical graded increase in effect as concentration increases, these compounds exhibit a threshold of response that, once reached, result in the detection in the response. While a given individual's response to the drug is identified by the presence or absence of the effect, the fraction of subjects in a sample population responding will increase as the concentration of drug is increased. A good example of this is seen in the conversion to normal sinus rhythm in patients with atrial fibrillation treated with an antiarrhythmic agent (Fig 37-21). In this instance, there is variation within the patient population in the concentration that must be achieved to convert a specific patient from atrial fibrillation to normal sinus rhythm.

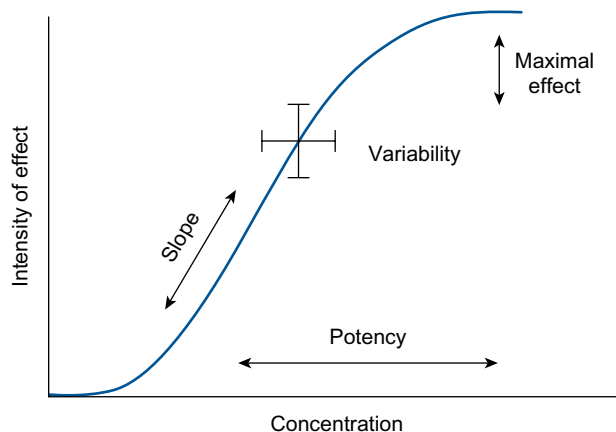


Figure 37-20. A graded pharmacologic response to a drug, showing a progressive increase in the intensity of effect as concentration increases until the maximal effect is achieved, when effect is plotted as a function of the log concentration. (Redrawn from Nies AS, Spielberg SP, Principles of Therapeutics, In: Hardman JG et al. eds. *Goodman and Gilman's The Pharmacologic Basis of Therapeutics*, 9th edn San Francisco: McGraw Hill, 1996.)

Thus, the higher the concentration achieved in a patient, the more likely he or she is to exhibit the pharmacologic effect.

Although visual inspection of concentration-response graphs can yield much useful information, rigorous comparisons between agents require a more quantitative analysis of concentration response data. The models that are used for such analyses are highly dependent upon the mechanism of action of the drug. In general, the mechanisms of pharmacologic agents can be divided into three categories: direct acting reversible agents, indirect acting reversible agents, and irreversibly acting agents.

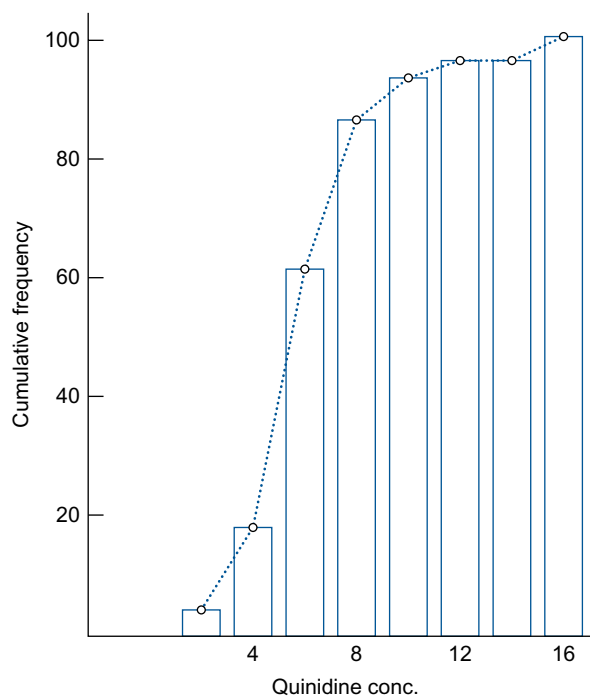


Figure 37-21. Cumulative frequency of conversion to normal sinus rhythm (expressed as percent of patients) as a function of quinidine concentration in a group of subjects with atrial fibrillation treated with the antiarrhythmic quinidine. (Redrawn from Gibaldi M. *Biopharmaceutics and Clinical Pharmacokinetics*, 4th edn, Philadelphia: Lippincott Williams & Wilkins, 1991.)

INTENSITY OF EFFECT—THE INTENSITY (OR MAGNITUDE) OF PHARMACOLOGIC EFFECT IS GIVEN AS

$$\text{Intensity of Effect} = \frac{E_{\max} C^{\gamma}}{EC_{50}^{\gamma} + C^{\gamma}} \quad (108)$$

where $E_{\max}$ = maximum effect, EC_{50} = concentration necessary to achieve 50% $E_{\max}$, C = concentration of drug, and γ = Hill coefficient

This equation is similar to that which describes the binding of oxygen to hemoglobin. One significant difference is that when used to quantify pharmacologic effect, the Hill coefficient has no physical or mechanistic meaning. However, the larger the value of the Hill coefficient, the steeper will be the concentration versus response relationship. For some measured pharmacologic responses, there is a baseline physiological value that must be taken into consideration when modeling the pharmacologic effect. For example, there are a number of therapeutic agents that induce methemoglobinemia after ingestion. There is, however, a low level of methemoglobin (1-2%) in drug-naïve subjects that needs to be accounted for in modeling the intensity of effect of agents that induce methemoglobinemia. Equation 108 can be adjusted to account for baseline effect:

$$\text{Intensity of Effect} = E_0 \frac{E_{\max} \cdot C^{\gamma}}{EC_{50}^{\gamma} + C^{\gamma}} \quad (109)$$

where E_0 = baseline effect.

Such a quantitative approach permits comparison of the intensity of drug effect between various agents. For example, while the antimicrobial agent dapsone is widely reported to cause methemoglobinemia in patients treated with the drug, the antimicrobial agent sulfamethoxazole is rarely associated with methemoglobinemia, despite the fact that both form a reactive hydroxylamine metabolite and that sulfamethoxazole is administered at much higher doses than dapsone. It was unclear whether this difference was due to differences in the potency of their respective hydroxylamine metabolite or due to differences in the pharmacokinetics of the drugs and/or metabolites. This question was addressed by comparing the ability of the two hydroxylamine metabolites for forming methemoglobin when incubated with human erythrocytes *in vitro* and modeling the effects using Equation 109.²³ As shown in Table 37-3, the maximal effect of the two metabolites studied were similar, but the potency (EC_{50}) of dapsone hydroxylamine was 20-fold greater than that of sulfamethoxazole hydroxylamine.

DURATION OF EFFECT

A measurable pharmacologic effect will be observable as soon as drug concentration reaches the minimal effect concentration (MEC). The duration of effect is determined by how long the concentration remains above the MEC, which is influenced by the dose administered and the rate of elimination of the drug. Consider the linear range of the log concentration-effect curve, such that $0.2E_{\max} < E < 0.8E_{\max}$. The effect within this boundary may be expressed as:

Table 37-3. Pharmacodynamics of Dapsone Hydroxylamine (DNOH) and Sulfamethoxazole Hydroxylamine (SNOH) Induced Methemoglobin Formation in Human Erythrocytes

	DNOH	SNOH
$E_{\max}$ (%)	26 (5)	463 (105)
EC_{50} (0M)	89 (4)	84 (5)

Data from Reilly TP *et al.* Methemoglobin formation by hydroxylamine metabolism of sulfamethoxazole and dapsone: implications for differences in adverse drug reactions. *J Pharmacol Exp Ther* 199; 288: 951–959.

$$E = m \log C + r \quad (110)$$

where m = slope of E versus $\log C$ plot and r = constant. This equation can be rearranged, such that:

$$\log C = \frac{E - r}{m} \quad (111)$$

Recall that for a drug exhibiting instantaneous distribution, following instantaneous administration:

$$\log C = \log C_0 - \frac{\lambda}{2.303} t \quad (112)$$

The maximal effect elicited by this dose, E_m , will occur when $C = C_0$. Thus,

$$\log C_0 = \frac{E_m - r}{m} \quad (113)$$

Substituting the equivalents in Eq. 111 and 113 into Eq. 112 yields

$$E = E_m - \frac{m\lambda}{2.303} t \quad (114)$$

Thus, the intensity of effect should decline at a constant rate that is a function of the elimination rate constant and the slope of the response versus log concentration curve. It should be noted that though the decrease in concentration is first-order, the decrease in effect is zero-order. But how can we quantify the actual duration of action following a given dose of a drug? Recall from Equation 13 that

$$C = \frac{D_{iv}}{V_d} e^{-\lambda t} \quad (115)$$

The duration time (t_d) is that time at which the plasma concentration drops just below the MEC, such that:

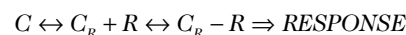
$$MEC = \frac{D_{iv}}{V_d} e^{-\lambda t} \quad (116)$$

Solving to duration time:

$$t_d = \frac{1}{\lambda} [\log D_{iv} - \log (MEC)V_d] \quad (117)$$

DIRECT ACTING, REVERSIBLE AGENTS

It could be argued that few, if any, drugs are truly “direct acting.” Most drugs interact with a receptor that produces the effect. Sometimes this interaction results in a cascade of events that ultimately produce the measured pharmacologic response. Thus, it might be appropriate to designate a category of drugs as *rapid acting, reversible agents*. The drug-receptor interaction is easily reversible, the pharmacological effect is easily reversible, and the time course of the effect is not delayed with respect to the time course of the drug-receptor interaction. For example, the β_1 -adrenoceptor antagonists (metoprolol) and the suppression of exercise-induced heart rate. There is a reversible (i.e., non-covalent) interaction with a cellular macromolecule, which, as a consequence of the binding of the drug, stimulates a cellular response. This can be expressed as:



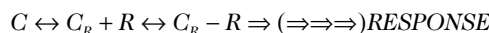
where C = concentration in plasma, C_R = concentration at receptor site, R = receptor, $C_R - R$ = drug receptor complex.

The intensity and duration of response can be readily quantified by measuring the “response” at several different concentrations.

INDIRECT ACTING, REVERSIBLE AGENTS

For many drugs, the response measured clinically is several steps removed from the initial biochemical effect of the drug. The time course of the effect, therefore, lags behind the time

course of concentrations. In these circumstances there may appear to be no direct association or relationship between the concentration of the drug in blood or plasma and the pharmacologic response.



and “ $\Rightarrow$ ” represents a “transduction” of the response that may depend on several factors, including the rate of turnover and trafficking of endogenous substrates or other mediators of drug effect. In this instance the drug-receptor interaction may not easily be reversible, and the pharmacologic effect is prolonged, such as the acetylation by aspirin of the serine moiety at the active site of cyclooxygenase. For some drugs with an indirect pharmacologic response, the drug-receptor interaction is easily reversible, but the pharmacologic effect is prolonged. An example might be corticosteroids, which bind with nuclear receptor proteins, resulting in RNA transcription and protein synthesis.

A classic example of an indirect pharmacologic response is the oral anticoagulant warfarin (Fig. 37-22). In this instance, the dissociation between the pharmacologic effect and the concentration is readily understood by a consideration of the mechanism action.

Warfarin inhibits the synthesis of the vitamin K-dependent factors that determine the prothrombin time. However, there is a time lag for this effect to be observed since circulating clotting factors must be degraded by the normal metabolic processes before an anticoagulant effect is observed. Thus, the turnover rate of the clotting factors, and not the time course of warfarin concentrations, becomes key determinant for the onset, duration, and offset of effect. This was accounted for by models that incorporated the endogenous degradation and synthesis rates of the clotting factors, thereby providing a good relationship between the concentration and “true” pharmacologic effect.²⁴

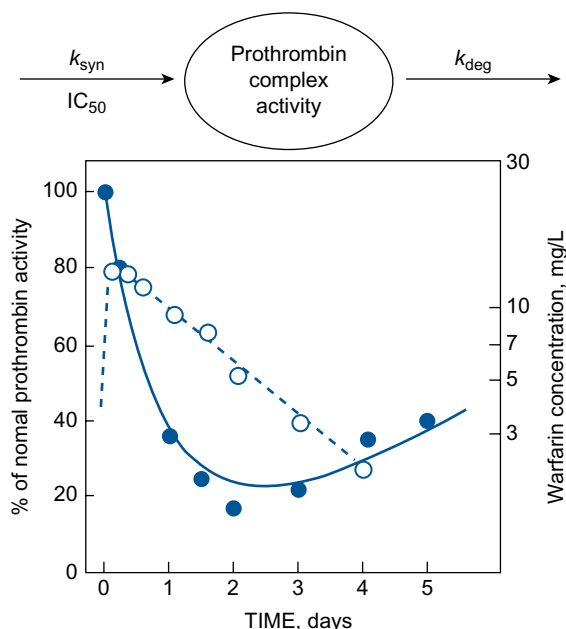


Figure 37-22. Time course of plasma warfarin concentration (top panel) and pharmacologic response, measured as percent decrease in normal prothrombin activity (bottom panel). Open circles represent warfarin concentration, while closed circles represent prothrombin activity. Note that the pharmacologic effect does not become evident until the concentration has reduced substantially from its maximal value. (Redrawn from Nagashima R *et al.* Kinetics of pharmacologic effects in man: the anticoagulant effect of warfarin. *Clin Pharmacol Ther* 1969; 10: 22–35).

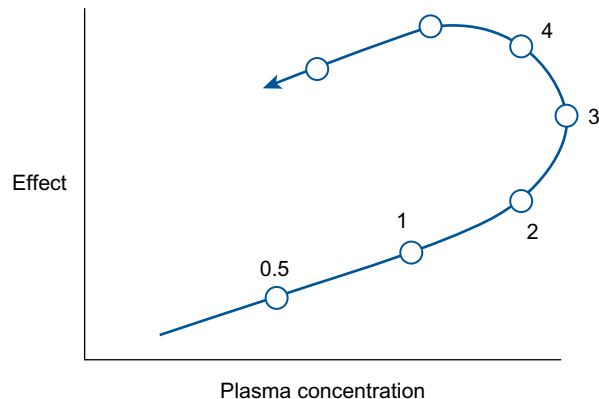


Figure 37-23. A counterclockwise hysteresis is observed following administration of a drug that exhibits noninstantaneous distribution and for which the site of action is in the slowly equilibrating tissues. Open circles represent a single effect measurement at a given concentration, and numbers indicate time (in hours) that measurements were made.

If distribution of a drug to its site of action is delayed, a counterclockwise hysteresis curve can be observed when plotting the effect versus plasma concentration after a dose of the drug, particularly when administered via a slow infusion (Fig 37-23). Thus, the intensity of effect observed for a given concentration shortly after administration (while concentrations are rising) is less than the intensity observed at that same concentration during the decline in concentration. This is due to the fact that during the distribution phase most of the drug is in the rapidly equilibrating tissues (i.e., the central compartment), while in the latter time periods the larger fraction of the drug is in the slowly equilibrating tissues (i.e., the peripheral compartment). Hence, the relationship between effect and plasma concentration for such a drug will be time dependent. This type of hysteresis can also be observed if the response is due to an active metabolite rather than the parent drug. A clockwise hysteresis is observed when tolerance occurs.

Sheiner *et al*²⁵ and Sheiner and Holford²⁶ originally modeled the time delay in pharmacological response with a hypothetical “effect compartment.” Using regression analysis to estimate the rate constants for the “effect compartment,” the hysteresis loop was collapsed, thereby making it possible to model indirect effects. Jusko and colleagues^{27,28} have developed a series of comprehensive physiologic indirect response models that incorporate an understanding of the mechanism(s) producing a particular pharmacologic effect. These indirect response models envision a given pharmacologic response that is based on the inhibition or stimulation of the input or the output factors controlling the pharmacologic effect. These complex models have increased our understanding of a variety of pharmacologic effects, such as induction of protein synthesis, cell trafficking, and altered hormone secretion.

IRREVERSIBLE ACTING AGENTS

Modeling the pharmacodynamics of irreversible acting agents is considerably more complex but has been successfully accomplished. The most intensely studied agents are those used in the treatment of cancer or infections. Modeling the ability of such agents to kill tumor or bacterial cells necessitates incorporation of cell-cycle kinetics. Such models have been useful in determining whether agents are best administered infrequently at high doses or in continuous exposure regimens.

REFERENCES

1. Swerdloff RS *et al.* Influence of age on the intravenous tolbutamide response test. *Diabetes* 1967; 16: 161–170.
2. Edwards DJ *et al.* Clinical pharmacokinetics of pethidine: 1982. *Clin Pharmacokinet* 1982; 7: 421–433.

3. Sulh H *et al.* Pharmacokinetic determinants of 6-mercapto-purine myelotoxicity and therapeutic failure in children with acute lymphoblastic leukemia. *Clin Pharmacol Ther* 1986; 40: 604–609.
4. Barr WH. Principles of biopharmaceutics. *Am J Pharm Educ* 1968; 52: 958–959.
5. Øie S, Tozer T. Effect of altered plasma protein binding on apparent volume of distribution. *J Pharm Sci* 1979; 68: 1203–1205.
6. Williams RL *et al.* Influence of acute viral hepatitis on disposition and plasma binding of tolbutamide. *Clin Pharmacol Ther* 1977; 21: 301–9.
7. Truitt Jnr, EB *et al.* Theophylline blood levels after oral, rectal, and intravenous administration, and correlation with diuretic action. *J Pharmacol Exp Ther* 1950; 100: 309–315.
8. Gibaldi M, Perrier D. *Pharmacokinetics*, 2nd edn. New York: Dekker, 1982: 30, 276, 441.
9. Jusko WJ. In: Burton ME *et al.* eds. *Applied Pharmacokinetics and Pharmacodynamics: Principles of Therapeutic Drug Monitoring*, 4th edn. Philadelphia, PA: Lippincott Williams & Williams, 2006: 8–29.
10. Nüesch EA. Noncompartmental approach in pharmacokinetics using moments. *Drug Metab Rev* 1984; 15: 103–131.
11. Yeh KC, Kwan KC. A comparison of numerical integrating algorithms by trapezoidal, Lagrange, and spline approximation. *J Pharmacokinet Biopharm* 1978; 6: 79–98.
12. Benet LZ, Galeazzi RL. Noncompartmental determination of the steady-state volume of distribution. *J Pharm Sci* 1979; 68: 1071–1074.
13. Karol MD. Mean residence time and the meaning of AUMC/AUC. *Biopharm Drug Dispos* 1990; 11: 179–181.
14. Bigger JT. Pharmacologic and clinical control of antiarrhythmic drugs. *Am J Med* 1975; 58: 479–488.
15. Morgan DJ, Smallwood RA. Clinical significance of pharmacokinetic models of hepatic elimination. *Clin Pharmacokinet* 1990; 18: 61–76.
16. Wilkinson GR, Shand DG. Commentary: a physiological approach to hepatic drug clearance. *Clin Pharmacol Ther* 1975; 18: 377–390.
17. Tucker GT. Measurement of the renal clearance of drugs. *Br J Clin Pharmacol* 1981; 12: 761–770.
18. Svensson CK *et al.* Free drug concentration monitoring in clinical practice. Rationale and current status. *Clin Pharmacokinet* 1986; 111: 450–469.
19. Benet LZ, Hoener BA. Changes in plasma protein binding have little clinical relevance. *Clin. Pharmacol Ther* 2002; 71: 115–121.
20. Grover A, Benet LZ. Effects of drug transporters on volume of distribution. *AAPS Journal* 2009; 11: 250–261.
21. Levy G, Tsuehuya T. Salicylate accumulation kinetics in man. *N Engl J Med* 1972; 287: 430–432.
22. Furst DE *et al.* Salicylate clearance, the resultant of protein binding and metabolism. *Clin Pharmacol Ther* 1979; 26: 380–389.
23. Reilly TP *et al.* Methemoglobin formation by hydroxylamine metabolites of sulfamethoxazole and dapsone: implications for differences in adverse drug reactions. *J Pharmacol Exp Ther* 1999; 288: 951–959.
24. Nagashima R *et al.* Kinetics of pharmacologic effects in man: the anticoagulant action of warfarin. *Clin Pharmacol Ther* 1969; 10: 22–35.
25. Sheiner LB *et al.* Simultaneous modeling of pharmacokinetics and pharmacodynamics: application to d-tubocurarine. *Clin Pharmacol Ther* 1979; 25: 359–371.
26. Sheiner LB, Holford NH. Understanding the dose-effect relationship: clinical application of pharmacokinetic-pharmacodynamic models. *Clin Pharmacokinetics* 1981; 6: 429–453.
27. Sharma A, Jusko WJ. Characteristics of indirect pharmacodynamic models and applications to clinical drug responses. *Br J Clin Pharmacol* 1998; 45: 229–239.
28. Jusko WJ, Ko HC. Physiologic indirect response models characterize diverse types of pharmacodynamic effects. *Clin Pharmacol Ther* 1994; 56: 406–419.

BIBLIOGRAPHY

- Burton ME *et al.* *Applied Pharmacokinetics and Pharmacodynamics: Principles of Therapeutic Drug Monitoring: Principles of Therapeutic Drug Monitoring*, 4th edn. Philadelphia: Lippincott, Williams & Wilkins, 2006.
- Gibaldi M, Perrier D. *Pharmacokinetics*, 2nd edn. New York: Dekker, 1982.
- Ho R, Gibaldi M. *Biotechnology and Biopharmaceuticals: Transforming Proteins and Genes into Drugs*. Hoboken, NJ: Wiley-Liss, 2003.
- Pecile A, Rescigno A. *Pharmacokinetics. Mathematical and Statistical Approaches to Metabolism and Distribution of Chemicals and Drugs*. New York: Plenum Press, 1988.
- Pratt WB, Taylor P. *Principles of Drug Action. The Basis of Pharmacology*, 3rd edn. New York: Churchill Livingstone, 1990.
- Reidenberg MM, Erill S, eds. *Drug-Protein Binding, Esteve Found Symp I*. New York: Praeger, 1986.
- Rowland M, Tozer TN. *Clinical Pharmacokinetics and Pharmacodynamics. Concepts and Applications*, 4th edn. Philadelphia: Lippincott, Williams & Wilkins, 2011.
- Shargel L, Wu-Pong S, Yu A. *Applied Biopharmaceutics and Pharmacokinetics*, 5th edn. New York: McGraw-Hill, 2004.
- Winter ME. *Basic Clinical Pharmacokinetics*, 5th edn. Spokane, WA: Applied Therapeutics, 2009.

Chapter 38

Drug Action and Effect

Donald N. Franz, PhD

DRUG ACTION AND EFFECT 727

DEFINITIONS AND CONCEPTS 727

DOSE-EFFECT RELATIONSHIPS 728

DRUG RECEPTORS AND RECEPTOR THEORY 732

OCCUPATION AND OTHER THEORIES 732

MECHANISMS OF DRUG ACTION 734

TYPES OF TARGETS FOR DRUG ACTION 734

RECEPTOR BINDING 735

RECEPTOR STRUCTURE AND FUNCTION 735

A drug is an agent intended for use in the diagnosis, mitigation, treatment, cure, or prevention of disease in humans or other animals. One of the most remarkable qualities of drugs is the great diversity of their actions and effects on the body, which enables their selective use in the treatment of a wide range of conditions involving virtually every body organ, tissue, and cell. Some drugs selectively stimulate or depress the cardiac muscle, the central nervous system (CNS), or the gastrointestinal tract. Mydriatic drugs dilate the pupil of the eye, whereas miotic drugs constrict or decrease pupillary size. Some drugs can induce vomiting, whereas others prevent vomiting. Diuretic drugs increase the flow of urine, whereas other drugs decrease it. Laxatives evacuate the bowel; other drugs cause constipation. The diverse effects of drugs in the body include the treatment of infections, cancer, cardiovascular disease, asthma, glaucoma, Alzheimer disease, and many others. Drugs can aid in diagnosis, treatment, replenishment, prevention, as well as many other applications.

Drug discovery and development is a complex phenomenon that involves the collective contributions of many scientific specialists. After a new potential drug substance has undergone definitive chemical and physical characterization, the basic pharmacology, or the nature and mechanism of action of the drug on the biologic system, must be determined. The drug's effects, or multiple effects, as well as its selectivity, dose, potency and efficacy, dose-effect relationships, dose-intensity of effect relationships, dose-frequency of response relationships, and variations in response and responsiveness must be ascertained. Additional studies on how the new drug works investigate drug receptors and receptor theory, occupation and other theories, mechanisms of drug action, types of targets for drug action, receptor binding, receptor structure and function, voltage-sensitive channels, ligand-activated ion channels, G protein-coupled receptors, tyrosine kinase-linked receptors, intracellular receptors that control DNA transcription, enzyme inhibition, and receptor regulation.

DRUG ACTION AND EFFECT

The word *drug* imposes an action-effect context within which the properties of a substance are described. The description must include the pertinent properties of the recipient of the drug. Thus, when a drug is defined as an analgesic, it is implied that the recipient reacts to a noxious stimulus in a certain way, called *pain*. (Studies indicate that pain is not simply the *perception* of a certain kind of stimulus but rather, a *reaction* to the perception of a variety of kinds of stimuli or stimulus patterns.) Both because the pertinent properties are locked into the complex and somewhat imprecise biological context and because the types of possible response are many, descriptions of the properties of drugs tend to emphasize the qualitative features of the effects they elicit. Thus, a drug may be described as having analgesic,

vasodepressor, convulsant, antibacterial, or other properties. The specific effect (or use) categories into which many drugs may be placed are covered in the section Pharmaceutical and Medicinal Agents and are not elaborated upon in this chapter. However, the description of a drug does not end with the enumeration of the responses it may elicit. Certain intrinsic properties of the drug-recipient system can be described in quantitative terms and are essential to the full description of the drug and to the validation of the drug for specific uses. In the following section, Definitions and Concepts, certain general terms are defined in qualitative language; the subsequent section, Dose-Effect Relationships, lays the foundation for an appreciation of some of the quantitative aspects of pharmacodynamics.

DEFINITIONS AND CONCEPTS

The vocabulary that is unique to the field of pharmacology is relatively small; the general vocabulary is that of the biological sciences and chemistry. Nevertheless, there are a few definitions that are important to the proper understanding of pharmacology. It is necessary to differentiate among *action*, *effect*, *selectivity*, *dose*, *potency*, and *efficacy*.

ACTION VERSUS EFFECT

The *effect* of a drug is an alteration of function of the structure or process upon which the drug acts. It is common to use the term *action* as a synonym for *effect*. However, action precedes effect. *Action* is the alteration of condition that brings about the effect.

The final effect of a drug may be far removed from its site of action. For example, the diuresis subsequent to the ingestion of ethanol does not result from an action on the kidney but, rather, from a depression of activity in the region of the hypothalamus, which regulates the release of antidiuretic hormone from the posterior pituitary gland. The alteration of hypothalamic function is, of course, also an effect of the drug, as is each subsequent change in the chain of events leading to diuresis. The action of ethanol was exerted only at the initial step, with each subsequent effect becoming then the action to a following step.

MULTIPLE EFFECTS

No known drug is capable of exerting a single effect, although a number are known that appear to have a single mechanism of action. Multiple effects may derive from a single mechanism of action. For example, the inhibition of acetylcholinesterase by physostigmine will elicit an effect at every site where acetylcholine is produced, is potentially active, and is hydrolyzed by cholinesterase. Thus, physostigmine elicits a constellation of effects.

A drug also can cause multiple effects at several different sites by a single action at only one site, providing that the

function initially altered at the site of action ramifies to control other functions at distant sites. Thus, a drug that suppresses steroid synthesis in the liver not only may lower serum cholesterol, impair nerve myelination and function, and alter the condition of the skin (as a consequence of cholesterol deficiency), but also may affect digestive functions (because of a deficiency in bile acids) and alter adrenocortical and sexual hormonal balance.

Although a single action can give rise to multiple effects, most drugs exert multiple actions. The various actions may be related, as for example, the sympathomimetic effects of phenylephrine that accrue to its structural similarity to norepinephrine and its ability to exert sympathetic responses, or the actions may be unrelated, as with the actions of morphine to interfere with the release of acetylcholine from certain autonomic nerves, block some actions of 5-hydroxytryptamine (serotonin), and release histamine. Many drugs bring about immunological (i.e., allergic or hypersensitivity) responses that bear no relation to the other pharmacodynamic actions of the drug.

SELECTIVITY

Although most drugs have the potential to elicit multiple effects, one effect is generally more readily elicitable than another. This differential responsiveness is called *selectivity*. It usually is considered to be a property of the drug, but it is also a property of the constitution and biodynamics of the recipient subject or patient.

Selectivity may come about in any of several ways. The subcellular structure (receptor) with which a drug combines to initiate one response may have a higher affinity for the drug than that for some other action. Atropine, for example, has a much higher affinity for muscarinic receptors that subserve the function of sweating than it does for the nicotinic receptors that subserve voluntary neuromuscular transmission, so that suppression of sweating can be achieved with only a tiny fraction of the dose necessary to cause paralysis of the skeletal muscles. A drug may be distributed unevenly, so that it reaches a higher concentration at one site than throughout the tissues generally; for example, chloroquine is much more effective against hepatic than intestinal (colonic) amebiasis because it reaches a much higher concentration in the liver than in the wall of the colon. An affected function may be much more critical to, or have less reserve in, one organ than in another, so that a drug will be predisposed to elicit an effect at the more critical site. Some inhibitors of dopa decarboxylase (also known as 5-hydroxytryptophan decarboxylase) depress the synthesis of histamine more than that of either norepinephrine or 5-hydroxytryptamine (serotonin), even though histidine decarboxylase is less sensitive to the drug, simply because histidine decarboxylase is the only step and, hence, is rate-limiting in the biosynthesis of histamine. Dopa decarboxylase does not become rate limiting in the synthesis of either norepinephrine or 5-hydroxytryptamine until the enzyme is nearly completely inhibited. Another example of the determination of selectivity by the critical balance of the affected function is that of the mercurial diuretic drugs. An inhibition of only 1% in the tubular resorption of glomerular filtrate usually will double urine flow, because 99% of the glomerular filtrate normally is resorbed. Aside from the question of the possible concentration of diuretics in the urine, a drug-induced reduction of 1% in sulfhydryl enzyme activity in tissues other than the kidney usually is not accompanied by an observable change in function. Selectivity also can be determined by the pattern of distribution of inactivating or activating enzymes among the tissues and by other factors.

DOSE

Even the uninitiated person knows that the dose of a drug is the amount administered. However, the appropriate dose of a drug is not some unvarying quantity, a fact sometimes overlooked by pharmacists, official committees, and physicians. The practice

of pharmacy is entrapped in a system of fixed-dose formulations, so that fine adjustments in dosage often are difficult to achieve. Fortunately, a rather wide latitude in dosages usually is allowable. It is obvious that the size of the recipient individual should have a bearing upon the dose, and the physician may elect to administer the drug on a body-weight or surface-area basis rather than as a fixed dose. Usually, however, a fixed dose is given to all adults, unless the adult is exceptionally large or small. The dose for infants and children often is determined by one of several formulas that take into account age or weight, depending on the age group of the child and the type of action exerted by the drug. Infants are relatively more sensitive to many drugs, often because systems involved in the inactivation and elimination of the drugs may not be developed fully in the infant.

The patient's nutritional status, his or her mental outlook, the presence of pain or discomfort, the severity of the condition being treated, the presence of secondary disease or pathology, and genetic factors as well as many other factors affect the dose of a drug necessary to achieve a given therapeutic response or to cause an untoward effect. Even two apparently well-matched normal persons may require widely different doses for the same intensity of effect. Furthermore, a drug is not always employed for the same effect and, therefore, is not always given in the same dose. For example, the dose of a progestin necessary for an oral contraceptive effect is considerably different from that necessary to prevent spontaneous abortion, and the dose of an estrogen used for the treatment of menopausal symptoms is much lower than that used for the treatment of prostatic carcinoma.

The wise physician knows that the dose of a drug is not a rigid quantity, but, rather, that which is necessary and can be tolerated, and he or she individualizes the regimen accordingly. The wise pharmacist also recognizes that official or manufacturer's recommended doses sometimes are quite narrowly defined and should serve only as a useful guide rather than as an absolute.

POTENCY AND EFFICACY

The *potency* of a drug is the reciprocal of dose. Thus, it will have the units of persons/unit weight of drug or body weight/unit weight of drug, for example. Potency generally has little utility other than to provide a means of comparing the relative activities of drugs in a series; *relative potency*, relative to some prototypic member of the series, is a parameter commonly used among pharmacologists and in the pharmaceutical industry.

Whether a given drug is more potent than another has little bearing on its clinical usefulness, as long as the potency is not so low that the size of the dose is physically unmanageable or the cost of treatment is higher than with an equivalent drug. A drug that is less potent but more selective is the one to be preferred. Promotional arguments in favor of a more potent drug, therefore, are irrelevant to the important considerations that should govern the choice of a drug. However, drugs of the same class sometimes differ in the maximum intensity of effect; that is, some drugs of the class may be less efficacious than others, regardless of how large a dose is used.

Efficacy connotes the property of a drug to achieve the desired response, and *maximum efficacy* denotes the maximum achievable effect. Even huge doses of codeine often cannot achieve the relief from severe pain that relatively small doses of morphine can; thus, codeine is said to have a lower maximum efficacy than morphine. Efficacy is one of the primary determinants of which drug is chosen.

DOSE-EFFECT RELATIONSHIPS

It is crucial to know how changes in the intensity of response to a drug vary with the dose. Both the physician, who prescribes or administers a drug, and the manufacturer, who must package the drug in appropriate dose sizes, must translate such

knowledge into everyday practice. Theoretical or molecular pharmacologists also study such relationships in inquiries into mechanism of action and receptor theory. It is necessary to define two types of relationships: (1) dose-intensity relationship, i.e., the manner in which the intensity of effect in the individual recipient relates to dose; and (2) dose-frequency relationship, i.e., the manner in which the number of responders among a population of recipients relates to dose.

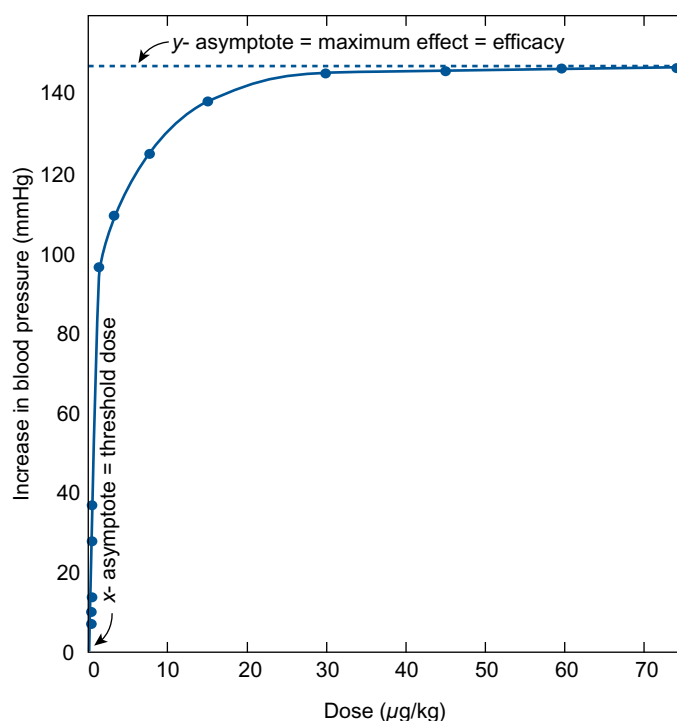
DOSE-INTENSITY OF EFFECT RELATIONSHIPS

Whether the intensity of effect is determined *in vivo* (e.g., the blood-pressure response to epinephrine in the human patient) or *in vitro* (e.g., the response of the isolated guinea pig ileum to histamine), the dose-intensity of effect (often called dose-effect) curve usually has a characteristic shape, one that closely resembles one quadrant of a rectangular hyperbola.

In the dose-intensity curve depicted in Figure 38-1, the curve appears to intercept the x axis at 0 only because the lower doses are quite small on the scale of the abscissa, the smallest dose being 1.5×10^{-3} micrograms. Actually, the x intercept has a positive value, because a finite dose of drug is required to bring about a response, this lowest effective dose being known as the *threshold dose*. Statistics and chemical kinetics predict that the curve should approach the y axis asymptotically. However, if the intensity of the measured variable does not start from zero, the curve may have a positive y intercept (or negative x intercept), especially if the ongoing basal activity before the drug is given is closely related to that induced by the drug.

In practice, instead of an asymptote to the y axis, dose-intensity curves nearly always show an upward concave foot at the origin of the curve, so that the curve has a lopsided sigmoid shape. At high doses, the curve approaches an asymptote that is parallel to the x axis, and the value of the asymptote establishes the maximum possible response to the drug, or *maximum efficacy*. However, experimental data in the regions of the asymptotes generally are too erratic to permit an exact definition of the curve at very low and very high doses. The example shown in Figure 38-1 represents an unusually good set of data.

Because the dose range may be 100- or 1000-fold from the lowest to the highest dose, it has become the practice to plot



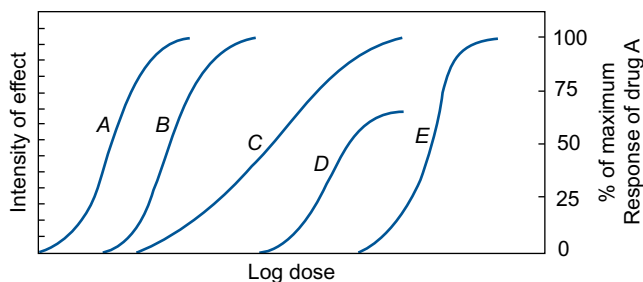


Figure 38-3. Log dose–intensity of effect curves of five different hypothetical drugs (see text for explanation).

compared. Drugs A, B, C, and E can achieve the same maximum effect, which suggests that the same effector system may be common to all. D may be working through the same effector system, but there are no *a priori* reasons to think this is so. Only A and B have parallel curves and common slopes. Common slopes are consistent with, but in no way prove, the idea that A and B not only act through the same effector system but also by the same mechanism. Although drug-receptor theory (see Drug Receptors and Receptor Theory, later in this chapter) requires that the curves of identical mechanism have equal slopes, exceptions do exist. Furthermore, mass-law statistics require that all simple drug-receptor interactions generate the same slope; only when slopes depart from this universal slope in accordance with distinctive characteristics of the response system do they provide evidence of specific mechanisms.

The relative potency of any drug may be obtained by dividing the ED₅₀ of the standard, or prototypic, drug by that of the drug in question. Any level of effect other than 50% may be used, but it should be recognized that when the slopes are not parallel, the relative potency depends on the intensity of effect chosen. Thus, the potency of A relative to C (see Figure 38-3) calculated from the ED₅₀ will be smaller than that calculated from the ED₂₅.

The low maximum intensity inducible by D poses even more complications in the determination of relative potency than do the unequal slopes of the other drugs. If its dose–intensity curve is plotted in terms of percentage of its own maximum effect, its relative inefficacy is obscured, and the limitations of relative potency at the ED₅₀ level will not be evident. This dilemma underscores the fact that drugs can be compared better from their entire dose–intensity curves than from a single derived number such as ED₅₀ or relative potency.

Drugs that elicit multiple effects will generate a dose–intensity curve for each effect. Even though the various effects may be qualitatively different, the several curves may be plotted together on a common scale of abscissa, and the intensity may be expressed in terms of percentage of maximum effect; thus, all curves can share a common scale of ordinates in addition to a common abscissa. Separate scales of ordinates could be employed, but this would make it harder to compare data.

The selectivity of a drug can be determined by noting what percentage of maximum of one effect can be achieved before a second effect occurs. As with relative potency, selectivity may be expressed in terms of the ratio between the ED₅₀ for one effect and that for another effect, or a ratio at some other intensity of effect. As with relative potency, difficulties follow from nonparallelism. In such instances, selectivity expressed in dose ratios varies from one intensity level to another.

When the dose–intensity curves for a number of subjects are compared, they are found to vary considerably from individual to individual in many respects, e.g., threshold dose, mid-point, maximum intensity, and sometimes even slope. An average dose–intensity curve can be constructed by averaging the intensities of the effect at each dose.

Average dose–intensity curves have limited application in comparing drugs. A single line expressing an average response

has little value in predicting individual responses unless it is accompanied by some expression of the range of the effect at the various doses. This may be done by indicating the standard error of the response at each dose. Occasionally, a simple scatter diagram is plotted in lieu of an average curve and statistical parameters. An average dose–intensity curve also may be constructed from a population in which different individuals receive different doses; if sufficiently large populations are employed, the average curves determined by the two methods will approximate each other.

The determination of such average curves from a population large enough to be statistically meaningful requires a great deal of work. Retrospective clinical data occasionally are treated in this way, but prospective studies rarely are designed in advance to yield average curves. The usual practice in comparing drugs is to employ a quantal (i.e., all-or-none) endpoint and plot the frequency or cumulative frequency of response over the dose range, as discussed in the following section.

DOSE-FREQUENCY OF RESPONSE RELATIONSHIPS

When an endpoint is truly all-or-none, such as death, it is easy to plot the number of responding individuals (e.g., dead subjects) at each dose of drug or intoxicant. Many other responses that vary in intensity can be treated as all-or-none if simply the presence or absence of a response (e.g., cough or no cough, convulsion or no convulsion) is recorded, without regard to the intensity of the response when it occurs. When the response changes from the basal or control state in a less abrupt manner (e.g., tachycardia, miosis, rate of gastric secretion), it may be necessary to designate arbitrarily some particular intensity of effect as the end point. If the end point is taken as an increase in heart rate of 20 beats/min, all individuals whose tachycardia is less than 20 beats/min would be recorded as nonresponders, whereas all those with 20 or above would be recorded as responders. When the percentage of responders in the population is plotted against the dose, a characteristic dose–response curve, more properly called a *dose–cumulative frequency* or *dose–percentage* curve, is generated. Such a curve is, in fact, a cumulative frequency distribution curve, with the percentage of responders at a given dose being the frequency of response.

Dose–cumulative frequency curves usually are of the same geometric shape as dose–intensity curves (i.e., sigmoid) when frequency is plotted against log dose (Figure 38-4). The tendency of the cumulated frequency of response (i.e., percentage) to be linearly proportional to the log of the dose in the middle of the dose range is called the *Weber-Fechner law*, although it is not invariable, as a true natural law should be. In many instances, the cumulative frequency is simply proportional to dose rather than log dose. The Weber-Fechner law applies to either dose–intensity or dose–cumulative frequency data. The similarity between dose–frequency and dose–intensity curves may be more than fortuitous—the intensity of response usually has an approximately linear relationship to the percentage of responding *units* (e.g., smooth muscle cells, nerve fibers) and, hence, also is a type of cumulative frequency of response. These are the same kind of statistics that govern the law of mass action.

If only the increase in the number of responders with each new dose is plotted, rather than the cumulative percentage of responders, a bell-shaped curve is obtained. This curve is the first derivative of the dose–cumulative frequency curve and is a *frequency distribution* curve. The distribution will be symmetrical—i.e., *normal* or Gaussian (see Chapter 10)—only if the dose–cumulative frequency curve is symmetrically hyperbolic. Because most dose–cumulative frequency curves are more nearly symmetrical when plotted semilogarithmically (i.e., as log dose), dose–cumulative frequency curves usually are *log-normal*.

Because the dose–intensity and dose–cumulative frequency curves are similar in shape, it follows that the curves have similar defining characteristics, such as ED₅₀, maximum effect

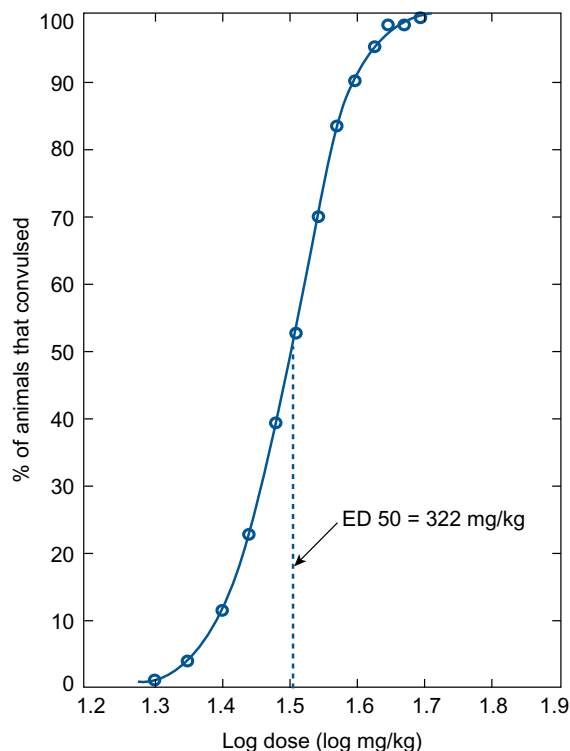


Figure 38-4. The relationship of the number of responders in a population of mice to the dose of pentylenetetrazole.

(maximum efficacy), and slope. In dose-cumulative frequency data, the ED₅₀ (*median effective dose*) is the dose to which 50% of the population responds (see Figure 38-4). If the frequency distribution is normal, the ED₅₀ is both the arithmetic mean and the median dose and is represented by the midpoint on the curve; if the distribution is log-normal, the ED₅₀ is the median dose but not the arithmetic mean dose. The efficacy is the cumulative frequency summed over all doses; it is usually, but not always, 100 percent. The slope is characteristic of both the drug and the test population. Even two drugs of identical mechanism may give rise to different slopes in dose-percentage curves, whereas in dose-intensity curves the slopes are the same.

Dose-cumulative frequency relationships are characterized by statistical parameters such as standard deviation, in addition to ED₅₀, maximum cumulative frequency (efficacy) and slope (see Chapter 10).

Several formulations are available for dose-cumulative frequency curves, some of which are employed only to define the linear segment of a curve and to determine the statistical parameters of this segment. For the statistical treatment of dose-frequency data, see Chapter 10. One simple mathematical expression of the entire log-symmetrical sigmoid curve is:

$$\log \text{dose} = K + f \log \left(\frac{\% \text{ response}}{100\% - \text{response}} \right) \quad (1)$$

where percentage response may be either the percentage of maximum intensity or the percentage of a population responding. The equation is thus basically the same for both log normal dose-intensity and log normal dose-percentage relationships. K is a constant that is characteristic of the midpoint of the curve, or ED₅₀, and $1/f$ is characteristically related to the slope of the linear segment, which, in turn, is closely related to the standard deviation of the derivative log-normal frequency-distribution curve.

The comparison of dose-percentage relationships among drugs is subject to the pitfalls indicated for dose-intensity comparisons, namely, that when the slopes of the curves are not the same (i.e., the dose-percentage curves are not parallel), it is

necessary to state at which level of response a potency ratio is calculated. As with dose-intensity data, potencies generally are calculated from the ED₅₀, but potency ratios may be calculated for any arbitrary percentage response. The expression of selectivity is, likewise, subject to similar qualifications, inasmuch as the dose-percentage curves for the several effects usually are nonparallel.

The term *therapeutic index* is used to designate a quantitative statement of the selectivity of a drug when a therapeutic effect is being compared with an undesired effect. If the untoward effect is designated T (for toxic) and the therapeutic effect is designated E , the therapeutic index may be defined as TD₅₀/ED₅₀ or a similar ratio at some other arbitrary levels of response. The TD and the ED are not required to express the same percentage of response; some clinicians use the ratio TD₁/ED₉₉ or TD₅/ED₉₅, based on the rationale that if the untoward effect is serious, it is important to use a most-severe therapeutic index in passing judgment upon the drug. Unfortunately, therapeutic indices are known in humans for only a few drugs.

Each untoward effect that a drug may elicit will have a different therapeutic index, and, if there is more than one therapeutic effect, it will have a family of therapeutic indices for each therapeutic effect. However, in clinical practice, it is customary to distinguish among the various toxicities by indicating the percentage incidence of a given side-effect.

VARIATIONS IN RESPONSE AND RESPONSIVENESS

The preceding discussion of dose-frequency relationships and information presented in Chapter 10 show that in a normal population, there may be quite a large difference between the dose required to elicit a given response in the least-responsive member of the population and that needed to elicit the response in the most-responsive member. The difference ordinarily will be a function of the slope of the dose-percentage curve or, in statistical terms, of the standard deviation. If the standard deviation is large, the extremes of responsiveness of responders are likewise large.

In a normal population, 95.46% of the population responds to doses within two standard deviations from the ED₅₀ and 99.73% within three standard deviations. In log-normal populations, the same distribution applies when standard deviation is expressed as log dose.

In the population represented in Figure 38-4, 2.25 percent of the population (two standard deviations from the median) would require a dose more than 1.4 times the ED₅₀; an equally small percentage would respond to 0.7 of the ED₅₀. The physician who is unfamiliar with statistics is apt to consider the 2.25 percent of patients at either extreme to be abnormal reactors. The statistician will argue that these 4.5 percent are within the normal population and that only those who respond well outside the normal population, at least three standard deviations from the median, deserve to be called abnormal.

Regardless of whether physician's or statistician's criteria of abnormality obtain, the term *hyporeactive* applies to those individuals who require abnormally high doses, and *hyperreactive* applies to those who require abnormally low doses. The terms *hyporesponsive* and *hyperresponsive* also may be used. It is incorrect to use the terms *hyposensitive* and *hypersensitive* in this context; *hypersensitivity* denotes an allergic response to a drug and should not be used to refer to hyperreactivity. The term *supersensitivity* correctly applies to hyperreactivity that results from denervation of the effector organ; it often is more definitively called *denervation supersensitivity*. Sometimes hyporeactivity is the result of an immunochemical deactivation of the drug, or *immunity*. Hyporeactivity should be distinguished from an increased dose requirement that results from a severe pathological condition. Severe pain requires large doses of analgesics, but the patient is not a hyporeactor; what has changed is the baseline from which the endpoint quantum is measured. The responsiveness of a patient to certain drugs

sometimes may be determined by his or her history of previous exposure to appropriate drugs.

Tolerance is a diminution in responsiveness as use of the drug continues, which results in an increase in the dose requirement. It may be due to an increase in the rate of elimination of drug (as discussed elsewhere in this chapter), to reflex or other compensatory homeostatic adjustments, to a decrease in the number of receptors or in the number of enzyme molecules or other coupling proteins in the effector sequence, to exhaustion of the effector system or depletion of mediators, to the development of immunity, or to other mechanisms. Tolerance may be gradual, requiring many doses and days to months to develop, or acute, requiring only the first or a few doses and only minutes to hours to develop. Acute tolerance is called *tachyphylaxis*.

Drug resistance is the decrease in responsiveness of microorganisms, neoplasms, or pests to chemotherapeutic agents, anti-neoplastics, or pesticides, respectively. It is not tolerance in the sense that the sensitivity of the individual microorganism or cancer cell decreases; rather, it is the survival of normally unresponsive cells, which then pass the genetic factors of resistance on to their progeny.

Patients who fail to respond to a drug are called *refractory*. Refractoriness may result from tolerance or resistance, but it also may result from the progression of pathological states that negate the response or render the response incapable of surmounting an overwhelming pathology. Rarely, it may result from a poorly developed receptor or response system.

Sometimes a drug evokes an unusual response that is *qualitatively* different from the expected response. Such an unexpected response is called a *meta-reaction*. A not uncommon meta-reaction is seen when phenobarbital acts as a CNS stimulant rather than depressant, an effect seen especially in women. Pain and certain pathological states sometimes favor meta-reactivity. Responses that are different in infants or the aged from those in young and middle-aged people are not meta-reactions if the response is usual in the age group. The term *idiosyncrasy* also denotes meta-reactivity, but the word has been so abused that it is recommended that it be dropped. Although hypersensitivity may cause unusual effects, it is not included in the category "meta-reactivity."

DRUG RECEPTORS AND RECEPTOR THEORY

Most drugs act by combining with some key substance in the biological milieu that has an important regulatory function in the target organ or tissue. This biological partner of the drug goes by the name *receptive substance* or *drug receptor*. The receptive substance is considered to be mainly a cellular constituent, although in a few instances it may be extracellular, as the cholinesterases are, in part. The receptive substance is thought of as having a special chemical affinity and structural requirements for the drug. Drugs such as emollients, which have a physical rather than chemical basis for their action, obviously do not act upon receptors. Drugs such as demulcents and astringents, which act in a nonselective or nonspecific chemical way, also are not considered to act upon receptors, since the candidate receptors have neither sharp chemical nor biological definition. Even antacids, which react with the extremely well-defined hydronium ion, cannot be said to have a receptor, because the reactive proton has no permanent biological residence.

Because of early preoccupation with physical theories of action and the classical and illogical dichotomy of chemical and physical molecular interaction, there is a reluctance to admit receptors for drugs such as general anesthetics or certain electrolytes, for example, which usually are not considered to combine selectively with distinct cellular or organelle membrane constituents. The word *receptor* often is used inconsistently and intuitively. However, the term is a legitimate symbol for that biological structure with which a drug interacts to initiate a response. Ignorance of the identities of many receptors does

not detract from, but rather increases, the importance of the term and general concept.

Once a receptor is identified, it commonly is no longer thought of as a receptor, although such identification may afford the basis of profound advances in receptor theory. Because the effects of anticholinesterases are derived only indirectly from inhibition of cholinesterase, and no drugs are known that stimulate the enzyme, it may be argued that it is not a receptor. Nevertheless, a number of drugs ultimately act indirectly through the inhibition of such modulator enzymes, and it is important for the theoretician to develop models based on such indirect interrelations.

Enzymes, of course, readily suggest themselves as candidates for receptors. However, there is more to cellular function than enzymes. Receptors may be membrane or intracellular constituents that govern the spatial orientation of enzymes, gene expression, compartmentalization of the cytoplasm, contractile or compliant properties of subcellular structures, or permeability and electrical properties of membranes. For nearly every cellular constituent there can be imagined a possible way for a drug to affect its function; therefore, few cellular constituents can be dismissed *a priori* as possible receptors. All the receptors for neurotransmitters and autonomic agonists are membrane proteins with agonist-binding groups projecting into the extracellular space. The transducing apparatus, whereby an occupied receptor elicits a response, is called a *coupling system*. Excitatory neurotransmitters in the CNS, and acetylcholine receptors elsewhere, are coupled to ion channels that, when opened, permit rapid ingress, especially of sodium ions. Each ion channel is composed of five subunits, and each subunit has four transmembrane, spanning regions. γ -Aminobutyric acid (GABA) and glycine are coupled to inhibitory chloride channels. Each of these receptors is composed of pentameric proteins, each of which has two to four different types of subunits. Benzodiazepine receptors are coupled to the GABA receptor. Beta-adrenergic receptors, histamine (H_2) receptors, and a number of receptors for polypeptide hormones interact with a stimulatory guanosine diphosphate (GDP)/guanosine triphosphate (GTP) binding protein (G-protein) that can activate the enzyme adenylate cyclase. The cyclase then produces cAMP, which, in turn, activates protein kinases. Other receptors interact with inhibitory G-proteins. Some receptors couple to guanylate cyclase.

Alpha-adrenergic α_1 -receptors, some muscarinic (M_1 and M_3) receptors, and various other receptors couple to the membrane enzyme phospholipase C, which cleaves inositol phosphates from phosphoinositides. The cleavage product, 1,4,5-inositol triphosphate (IP_3), then causes an increase in intracellular calcium, whereas the product diacylglycerol (DAG) activates kinase C. There are a number of other less ubiquitous coupling systems as well. Substances such as cAMP, cyclic guanosine-3',5'-monophosphate (cGMP), IP_3 , and DAG are called *second messengers*.

There may be several different receptors for a given agonist. Differences may be shown not only in the types of coupling systems and effects but also by differential binding of agonists and antagonists, desensitization kinetics, physical and chemical properties, genes, and amino acid sequences. The differentiation among receptor subtypes is called *receptor classification*. Receptor subtypes are designated by Greek or Arabic alphabetical prefixes and/or numerical subscripts. There are at least two each of beta-adrenergic, histaminergic, serotonergic, GABAergic, and benzodiazepine receptors; three each of muscarinic and alpha-adrenergic receptors; and five opioid receptor subtypes.

OCCUPATION AND OTHER THEORIES

Drug-receptor interactions are governed by the law of mass action. However, most chemical applications of mass law are concerned with the rate at which reagents disappear or products

are formed, whereas receptor theory usually concerns itself with the fraction of the receptors combined with a drug. The usual concept is that only when the receptor actually is occupied by the drug is its function transformed in such a way as to elicit a response. This concept has become known as the *occupation theory*. The earliest clear statement of its assumptions and formulations is often credited to Clark in 1926¹, but both Langley and Hill also made important contributions to the theory in the first two decades of the 20th century.

In all receptor theories, the terms agonist, partial agonist, and antagonist are employed. An *agonist* is a drug that combines with a receptor to initiate a response.

In the classical occupation theory, two attributes of the drug are required: (1) *affinity*, a measure of the equilibrium constant of the drug-receptor interaction; and (2) *intrinsic activity*, or *intrinsic efficacy* (not to be confused with efficacy as intensity of effect), a measure of the ability of the drug to induce a positive change in the function of the receptor.

A *partial agonist* is a drug that can elicit some but not a maximal effect and that antagonizes an agonist. In the occupation theory it would be a drug with a favorable affinity but a low intrinsic activity.

A *competitive antagonist* is a drug that occupies a significant proportion of the receptors and thereby preempts them from reacting maximally with an agonist. In the occupation theory, the prerequisite property is affinity without intrinsic activity.

A *noncompetitive antagonist* may react with the receptor in such a way as not to prevent agonist-receptor combination but to prevent the combination from initiating a response, or it may act to inhibit some subsequent event in the chain of action-effect-action-effect that leads to the final overt response.

The mathematical formulation of the receptor theories derives directly from the law of mass action and chemical kinetics. Certain assumptions are required to simplify calculations. The key assumption is that the intensity of effect is a direct linear function of the proportion of receptors occupied. The correctness of this assumption is most improbable on the basis of theoretical considerations, but empirically it appears to be a close enough approximation to be useful. A second assumption upon which formulations are based is that the drug-receptor interaction is at equilibrium. Another common assumption is that the number of molecules of receptor is negligibly small compared with that of the drug. This assumption is undoubtedly true in most instances, and departures from this situation greatly complicate the mathematical expression of drug receptor interactions.

The first clearly stated mathematical formulation of drug-receptor kinetics was that of Clark,¹ in his equation:

$$Kx^n = \frac{y}{100 - y} \quad (2)$$

where K is the affinity constant, x is the concentration of drug, n is the molecularity of the reaction, and y is the percentage of maximum response. Clark assumed that y was a linear function of the percentage of receptors occupied by the drug, so y could also symbolize the percentage of receptors occupied. When the equation is rearranged to solve for y , we get:

$$y = \frac{100Kx^n}{1 + Kx^n} \quad (3)$$

A Cartesian plot of this equation is identical in form to that shown in Figure 38-1. When y is plotted against $\log x$ instead of x , the usual sigmoid curve is obtained. Thus, it can be seen that the dose-intensity curve derives from mass action equilibrium kinetics, which, in turn, derive from the statistical nature of molecular interaction. The fact that dose-intensity and dose-percentage curves have the same shape shows that they involve similar statistics.

If Equation 2 is put into log form,

$$\log K + n \log x = \log \frac{y}{100 - y} \quad (4)$$

a plot of $\log y/100 - y$ against $\log x$ then will yield a straight line with a slope of n ; n is theoretically the number of molecules of drug that react with each molecule of receptor. Currently, there are no known examples in which more than one molecule of agonist combines with a single receptor, hence, n should equal 1, universally. Nevertheless, n often deviates from 1. Deviations occur because of cooperative interactions among receptors (*co-operativity*), spare receptors (see below), amplifications in the response system (i.e., *cascades*), receptor coupling to more than one sequence (e.g., to both adenylate cyclase and calcium channels), and other reasons. In these departures from $n = 1$, the slope becomes a characteristic of the mechanism of action and response system.

The probability that a molecule of drug will react with a receptor is a function of the concentration of both drug and receptor. Unlike the concentration of a drug, the concentration of receptor molecules cannot be manipulated. However, as each molecule of drug combines with a receptor, the population of free receptors is diminished accordingly. If the drug is a competitive antagonist, it will diminish the probability of an agonist-receptor combination in direct proportion to the percentage of receptor molecules preempted by the antagonist. Consequently, the intensity of effect will be diminished. However, the probability of agonist-receptor interaction can be increased by increasing the concentration of agonist, and the intensity of effect can be restored by appropriately larger doses of agonist. Addition of more antagonist will, again, diminish the response, which can, again, be overcome or surmounted by more agonist.

Clark¹ showed empirically and by theory that as long as the ratio of antagonist to agonist was constant, the concentration of the competitive drugs could be varied over an enormous range without changing the magnitude of the response (Figure 38-5). Because the presence of competitive antagonist only diminishes the probability of agonist-receptor combination at a given concentration of agonist and does not alter the molecularity of the reaction, it also follows that the effect of the competitive antagonist is to shift the dose-intensity curve to the right in proportion to the amount of antagonist present; neither the shape nor the slope of the curve is changed (Figure 38-6).

Many refinements of the Clark formula have been made, but they are not discussed in this chapter; details and citations of relevant literature can be found in various works on receptors cited in the Bibliography. Several refinements have been introduced to facilitate studies of competitive inhibition. The introduction of the concepts of intrinsic activity² and efficacy³ required appropriate changes in mathematical treatment.

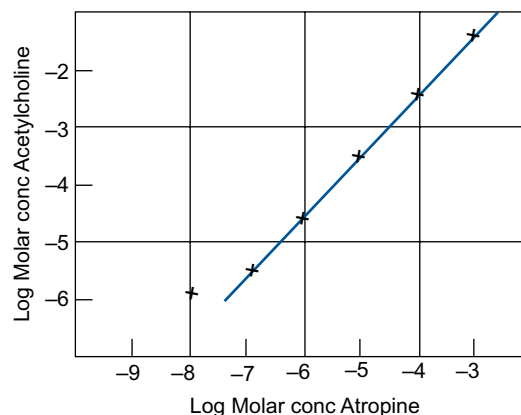


Figure 38-5. Direct proportionality of the dose of agonist (acetylcholine) to the dose of antagonist (atropine) necessary to cause a constant degree of inhibition (50%) of the response of the frog heart. (Adapted from Clark AJ. The antagonism of acetylcholine by atropine. *J Physiol [London]* 1926; 61: 547-556.)

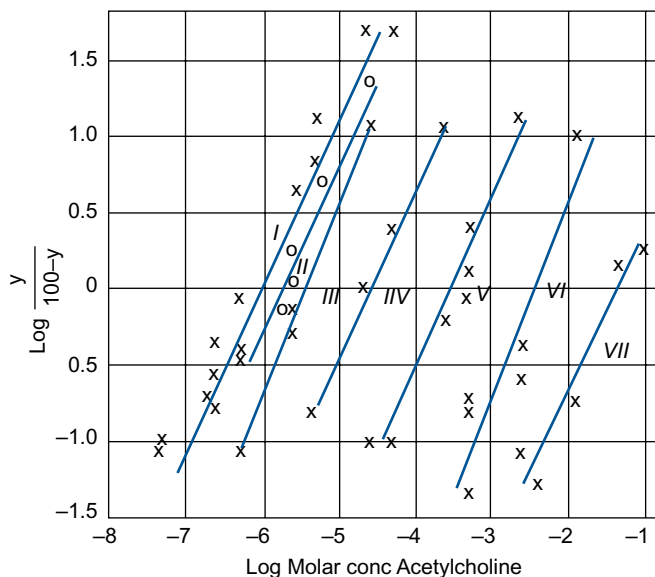


Figure 38-6. Effect of an antagonist to shift the log dose-intensity curve to the right without altering the slope. The effector is the isolated heart. I, no atropine; II, atropine, 10^{-8} M; III, 10^{-7} M; IV, 10^{-6} M; V, 10^{-5} M; VI, 10^{-4} M; VII, 10^{-3} M; Y, % of maximum intensity of response. The function $\log y/(100 - y)$ converts the log dose-intensity relationship to a straight line. (Adapted from Clark AJ. The antagonism of acetylcholine by atropine. *J Physiol [London]* 1926; 61: 547–556.)

Another important concept has been added to the occupation theory, that of *spare receptors*. Clark assumed that the maximal response occurred only when the receptors were completely occupied, which does not account for the possibility that the maximum response might be limited by some step in the action-effect sequence subsequent to receptor occupation. Work with isotopically labeled agonists and antagonists and with dose-effect kinetics has shown that the maximal effect sometimes is achieved when only a small fraction of the receptors are occupied. The mathematical treatment of this phenomenon has enabled theorists to explain several puzzling observations that previously appeared to contradict occupation theory.

The classical occupation theory fails to explain several phenomena satisfactorily, and it is unable to generate a realistic model of intrinsic activity and partial agonism. A rate theory, in which the intensity of response is proportional to the rate of drug-receptor interaction instead of occupation, was proposed to explain some of the phenomena that occupation theory could not, but the rate theory was unable to provide a realistic mechanistic model of response generation, and it had other serious limitations as well.

The phenomena that neither the classical occupation nor the rate theory could explain can be explained by various theories in which the receptor can exist in at least two conformational states, one of which is the active one; the drug can react with one or more conformers. In a *two-state model*,⁴



where R is the inactive and R^* is the active conformer. The agonist combines mainly with R^* , the partial agonist can combine with both R and R^* , and the antagonist can combine with R, with the equilibrium being shifted according to the extent of occupation of R and R^* . Other variations of occupation theory treat the receptor as an aggregate of subunits that interact cooperatively.⁵

MECHANISMS OF DRUG ACTION

Drugs are distributed to many or all parts of the body by the circulation. However, they do not act everywhere; they would

have extremely limited usefulness if they did. Clinically useful drugs act only on certain existing biological systems. Although drugs cannot create new systems, some drugs can temporarily or permanently damage existing functional systems that are susceptible to them, thereby producing toxic effects. Almost all drugs act more or less selectively on large specific proteins, glycoproteins, or lipoproteins located on the cell membrane or in the cell cytoplasm, nuclei, or other intracellular organelles. These specific proteins are referred to as *receptors*. Although they often are regarded as drug receptors, they are in reality receptors for endogenous substances that mediate normal biological and physiological regulatory processes.

Virtually all cells of the body have multiple receptors, since they are regulated by a variety of endogenous substances that act continuously, intermittently, or only occasionally. Similarly, cells theoretically can be influenced by a variety of drugs that act on the different receptors that the cells contain. The chemical nature of many of the endogenous substances that activate receptors is known, but new ones continue to be identified and sought. For example, the former mystery of why animals have receptors for morphine, which is produced by some species of poppy plants, was solved when endogenous opioid peptides were identified in the brain and some peripheral tissues in the mid-1970s.

Drugs that selectively activate receptors and produce the same effects normally produced by a respective endogenous substance are called *agonists*. Drugs that selectively block receptors are called *antagonists* because they antagonize, or block, the normal effects of the respective endogenous substance. Pure antagonists do not activate their receptors. Some experimental drugs stimulate or activate certain enzymes, but none are useful therapeutic agents, because their effects are too widespread. Forskolin is one such example; it directly stimulates the enzyme adenyl cyclase to synthesize cAMP, which is a second messenger in many cellular systems throughout the body.

On the other hand, many very useful therapeutic drugs are *enzyme inhibitors*, which selectively inhibit the normal activity of only one type of enzyme, thereby reducing the ability of the enzyme to act on its normal biochemical substrate. In this context, the enzymes are the drug receptors. Although the chemical nature of receptors and enzymes and their interactions with drugs often was vaguely understood in the past, the application of new techniques in molecular biology, biochemistry, and pharmacology since the mid-1980s has made unprecedented progress in defining the structures of receptors and enzymes and the consequences of drug-receptor interactions.

TYPES OF TARGETS FOR DRUG ACTION

Drug *effects* are the result of drug *actions*. Drug action may be defined as the drug-receptor interaction, whereas drug effects are the consequences of that action. For example, the interaction of epinephrine with β -receptors in the heart sets into motion a cascade of intracellular events (actions) that lead to increases in heart rate and strength of contraction (effects). The interaction of epinephrine with α -receptors in the vasculature sets into motion a cascade of intracellular events (actions) that lead to vasoconstriction and increased blood pressure (effects).

Typical responses that involve drug-receptor interactions are those that involve agonist or antagonist interactions at a receptor. Agonists also can act through various transduction mechanisms to produce a variety of intracellular changes that alter cellular activity. (Transduction mechanisms are considered in more detail near the end of this section.) Agonist actions may be direct, as with acetylcholine acting on the nicotinic receptors at the neuromuscular junction to briefly open sodium channels. This produces rapid depolarization of skeletal muscle, leading to muscle contraction. Drugs also can act directly on ion channels to block their activity. For example, lidocaine and other local anesthetics block sodium channels in nerve fibers (i.e.,

axons) so that the conduction of action potentials is blocked, and the area served by those nerve fibers is anesthetized. Drugs also can act directly on ion channels to modulate their activity. The benzodiazepines, characterized by diazepam, produce multiple effects (e.g., sedation, hypnosis, anticonvulsant and antianxiety activity, and muscle relaxation) by modifying the actions of GABA on its receptors in the CNS. GABA is the predominant inhibitory neurotransmitter in the CNS, and it acts on GABA_A receptor complexes by opening chloride channels on neurons to hyperpolarize them and render them less excitable. The benzodiazepines act on a different receptor on the GABA_A receptor complex to enhance the actions of GABA on its receptors, thereby rendering target neurons even less excitable.

Many drugs act by inhibiting enzymes so that they cannot perform their normal functions as efficiently. One such drug, omeprazole, reduces the ability of parietal cells in the stomach to produce hydrochloric acid by inhibiting the enzyme, or proton pump, H⁺, K⁺-ATPase, which is found only in these parietal cells. It is used to facilitate healing of peptic ulcers and control esophageal reflux (heartburn). The body's normal enzymes also can convert false substrates into active drugs. For example, α -methyl dopa is converted into α -methylnorepinephrine by the enzymes that normally synthesize dopamine and norepinephrine from dopa. α -Methylnorepinephrine acts on brain receptors to reduce sympathetic activity to blood vessels, thereby reducing blood pressure in hypertensive patients. Antimetabolites used to treat cancer also are false substrates, which are similar in structure to endogenous metabolites involved in cell-cycle reactions but function abnormally to interfere with synthesis of essential metabolites. Some drugs are, or have been, designed to be inactive until they are converted, usually by liver drug-metabolizing enzymes such as cytochrome P450, to active drug; the inactive drug is called a *prodrug*.

Various carriers are used by cells to take up neurotransmitters that have been released. The actions of dopamine released from dopamine nerve terminals in the brain are terminated by reuptake into the nerve terminals by a dopamine carrier. The dopamine then is reused for neurotransmission. If the carrier is blocked by a reuptake blocker such as cocaine, dopamine concentrations between the nerve terminals and the dopamine receptors build up for a time and produce greater effects.

Finally, antibiotics and antiviral, antifungal, and antiparasitic drugs owe their selectivities to selective actions on certain biochemical processes that are essential to the offending organism but are not shared by the mammalian host. The penicillins and related antibiotics interfere with the synthesis of rigid cell walls by growing bacteria, but mammalian cells are contained only by plasma membranes and, therefore, are not affected by penicillins. Antiparasitic drugs target enzymes found only in parasites, enzymes that are indispensable only in parasites, or biochemical functions with different pharmacological properties in the parasite and the host.

RECEPTOR BINDING

Drugs that bind to certain receptors selectively at pharmacological concentrations are known as *receptor ligands*; they can be agonists or antagonists. Many drugs also bind nonselectively to nonreceptor proteins throughout the body where they exert no pharmacological actions or effects. Many drugs bind to plasma proteins, especially albumin. Albumin-bound drug can act as a reservoir for free drug, with which it is in equilibrium, and competition among drugs for plasma protein binding can lead to increased free drug levels and drug interactions as they displace one another.

Drugs and endogenous ligands or substrates bind selectively to certain receptors because of both a chemical attraction and a proper fit to the protein. The lock-and-key analogy provides a useful concept of proper fit. Carried a step further, an agonist fits the lock and turns it, but an antagonist only fits the lock but cannot turn it; yet, it does block entry of the agonist

key. Generally, a number of drugs with both characteristics can combine with the same receptor. The study of structure-activity relationships among similar drugs and their receptors always has been an important and fruitful approach of both pharmacology and medicinal chemistry. Highly selective drugs tend to bind to only one or several closely related receptors. However, some drugs can combine with and activate or inactivate a number of different receptors that have similar structures, thereby diminishing selectivity and magnifying side effects.

The types of chemical bonds by which drugs bind to their receptors are as follows, in decreasing order of bond strength: covalent, ionic, hydrogen, hydrophobic, and van der Waals bonds. Relatively few drugs form covalent bonds with their receptors. Covalent bonds are irreversible and very long-lasting; new receptors or enzymes must be synthesized to restore function, and this process takes a week or two. Most drugs rely on combinations of the other, weaker bonds to bind tightly but reversibly to receptors. For example, the binding of acetylcholine, a relatively simple molecule, to nicotinic receptors at the neuromuscular junction involves ionic, hydrogen, and van der Waals bonds, with the ionic and hydrogen bonds being the most important. It is no accident that receptor-binding drugs are partially ionized at body pH, because receptor proteins also are partially ionized. Drugs and proteins contain positively charged nitrogen groups and negatively charged carboxyl groups that strongly attract one another and usually provide the initial drug-receptor bonds. Hydrogen bonds, formed between bound hydrogen atoms and oxygen, nitrogen, fluoride, or sulfur atoms, further orient the drug molecule to its receptor to enhance the proper fit. One or several hydrogen bonds can be involved. Hydrophobic bonds form among nonpolar ring structures (e.g., benzene) or chains of methylene groups to stabilize orientation further. Finally, the very weak van der Waals forces provide some additional, electrostatic bonding over very short distances.

Drug molecules that contain asymmetrical carbon atoms can exist as stereoisomers, only one of which is oriented to bond well with its receptors. For example, the side chain of epinephrine contains an asymmetrical carbon atom in the α position of the side chain, with a hydroxyl group attached, permitting epinephrine to exist in D- and L- forms (mirror images). The endogenous L-form is about 1000 times more potent than the synthesized D-form, because the L-form has a much greater binding affinity for its receptors because of its preferred configuration. In the past, drugs synthesized as mixtures of stereoisomers were formulated as racemic mixtures, but improved chemical separation techniques now often allow isolation of the more active isomer for formulation.

RECEPTOR STRUCTURE AND FUNCTION

The number of receptors and their subtypes continues to grow at a rapid pace as a result of the identification of new endogenous ligands and application of newly developed techniques to study them. Despite this large number, most receptors can be classified structurally and functionally into only a few basic types, which are described in this section. No attempt is made to provide detailed descriptions of individual receptors within each category. Rather, one or two examples will suffice for each, with brief reference to some prominent types that are therapeutically relevant.

VOLTAGE-SENSITIVE CHANNELS

While not generally classified as receptors, voltage-sensitive channels contain receptors that are acted upon by drugs or toxins to block or modify their normal function. The voltage-sensitive sodium channels in axons allow initiation and conduction of action potentials in response to a voltage change in the plasma membrane. When sodium channels open, sodium ions rush into the cytoplasm, thereby causing depolarization and propagation of the action potential. The crucial component

of the sodium channel is a single protein composed of a chain of about 2000 amino acids and called the α subunit. Several β subunits with minor roles also are associated with the α subunit. The α subunit has four repeating domains composed of about 250 amino acids each, and each domain contains six, α -helical, 22- to 25-amino acid, transmembrane, spanning segments. Each domain forms one of four clusters of the six membrane-spanning regions to encircle the sodium channel so formed. On end, the channel resembles 24 cylinders neatly arranged around the sodium channel that, at rest, is charged positively due to positive charges on the four transmembrane helices that surround the channel. Upon activation, these particular helices are thought to rotate upward, thereby moving the positive charges away from the channel and allowing the positive sodium ions to rush through. The channel remains open for only about 1 msec because the voltage changes attract a protein loop of the channel in the cytoplasm to shut the channel like a tetherball. Local anesthetics block the sodium channel from the cytoplasmic side by binding to receptors inside the channel. Several neurotoxins block from the outside.

Axons are repolarized by brief (~ 1 msec) opening of voltage-activated potassium channels that are constructed similarly to sodium channels but are composed of four identical subunits of peptide that associate in the membrane to form the potassium channel. Each subunit spans the membrane six times. It probably functions much like the sodium channel, including inactivation by a tetherball segment of cytoplasmic peptide. Quinidine, an antiarrhythmic drug, will block this potassium channel in the heart.

Voltage-activated calcium channels of the L-type are composed of five similar protein subunits that assemble across heart muscle and vascular smooth muscle membranes to form the calcium channel. Its arrangement in the membrane is similar to that of the sodium and potassium channels. Calcium channel blockers such as verapamil and nifedipine are used to treat several cardiovascular conditions by virtue of their ability to block calcium channels in the heart and blood vessels.

LIGAND-ACTIVATED ION CHANNELS

The best-characterized ligand-activated ion channel is the nicotinic receptor complex at the neuromuscular junction. As the name implies, these channels are activated by receptor ligands—in this case, acetylcholine. The nicotinic receptor complex is composed of five subunit proteins with similar structures that associate across the plasma membrane to form a sodium channel. The receptor complex is formed from two α and one each of β , γ , and δ subunits (Fig 38-7). In contrast to the voltage-activated ion channels, each of the five proteins

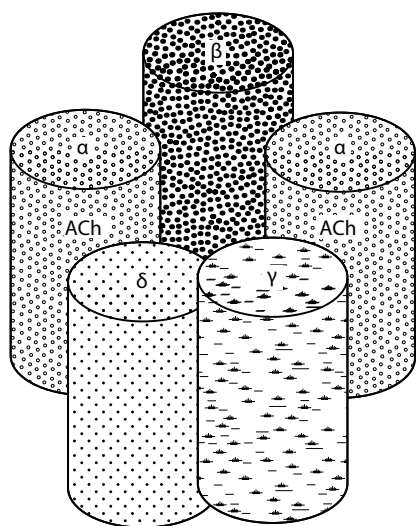


Figure 38-7. Nicotinic receptor complex. ACh, acetylcholine.

crosses the membrane only four times. The two α subunits contain the nicotinic receptors, which acetylcholine activates, and both must be activated to open the sodium channel to 6.5 Å for about 4 msec. The receptors can be blocked by neuromuscular blocking agents such as curare. The nicotinic receptors on autonomic ganglia are similar in structure but are composed of a different set of subunits, which accounts for the long-known differences in selective antagonists at the two sites.

Other ligand-activated ion channels – GABA_A, glycine, and glutamate – have structures similar to that of the nicotinic receptor complex. GABA and glycine channels are chloride channels, which permit chloride influx into neurons to produce hyperpolarization and decreased neuronal excitability. Glutamate channels are primarily sodium channels, and they also contain modifying receptors for glycine and polyamines. The GABA_A receptor complex contains not only receptors for GABA but also separate receptors for benzodiazepines (e.g., diazepam), barbiturates, and steroids, which modify the actions of GABA on the chloride channel. The convulsant activity of strychnine is due solely to its ability to block glycine receptors, primarily in the brainstem and spinal cord.

G PROTEIN-COUPLED RECEPTORS

The G protein-coupled receptors make up a very large family of receptors that are activated by monoamines (e.g., epinephrine, norepinephrine, dopamine, serotonin), acetylcholine (muscarinic receptors), opioids, and a host of active peptides, including a number of hormones. Structurally, these receptors are single proteins, most of which are composed of chains of 350 to 550 amino acids and cross the plasma membrane seven times in a serpentine arrangement (Fig 38-8). Each of the seven transmembrane domains is composed of 22 to 30 amino acids configured into an α -helix. The third of three intracellular (cytoplasmic) loops is much longer than the other two and is responsible for coupling with the G proteins. Rather than residing at the extracellular surface of the receptor, the actual receptor-binding sites often lie *within* the membrane between the seven transmembrane domains. For example, the β -adrenergic receptor lies 11 Å below the extracellular surface, or about one-third of the distance through the membrane. The positively charged nitrogen on the side chain of the epinephrine molecule forms an ionic bond with the negatively charged carboxyl group on an

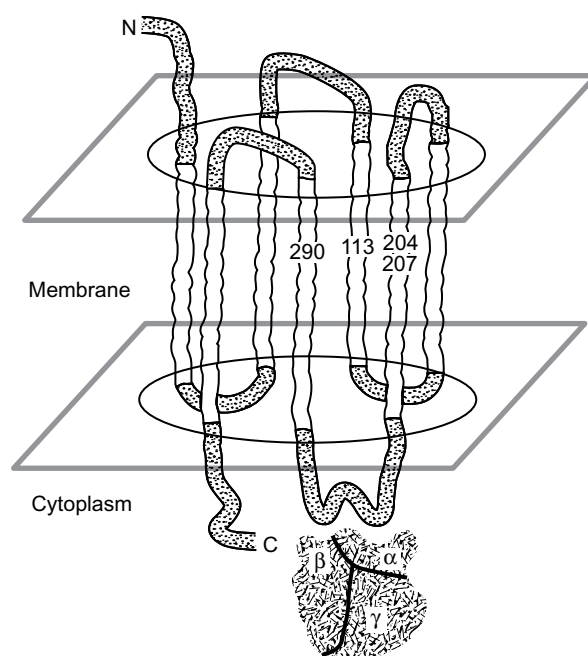


Figure 38-8. G-protein-coupled receptor complex.

aspartate amino acid (residue 113) in the third transmembrane domain (TM3). The two catechol hydroxyl groups of epinephrine form hydrogen bonds with the free hydroxyl groups of two serine amino acids at residues 204 and 207 in TM5, and the aromatic ring of epinephrine forms a hydrophobic bond with that of a phenylalanine at residue 290 in TM6. The location of G protein-coupled receptors within the membrane underscores the importance of size and configuration in the molecular structure of both agonists and antagonists for these receptors. Some negatively charged and peptide ligands do bind to an extracellular domain, however.

Among some families of G protein-coupled receptors there is considerable structural homology; i.e., the same amino acids and the same sequences make up large portions of a number of different receptors. Consequently, a number of antagonist receptor ligands bind to these similar arrangements of amino acids in the transmembrane domains. For example, many of the antipsychotic drugs (i.e., neuroleptics) are antagonists not only at dopamine receptors, where they are thought to exert their therapeutic effects, but also at α_1 -adrenergic, serotonin, histamine, and muscarinic receptors, thereby producing hypotension, sedation, blurred vision, dry mouth, and constipation as side-effects.

The G proteins closely associated with the third cytoplasmic loop of the receptors are heterotrimers composed of three different subunits: α , β , and γ . Upon receptor activation, the α subunit exchanges a bound GDP for a GTP and dissociates from the $\beta\gamma$ subunits to activate a membrane enzyme such as adenylyl cyclase or to influence an ion channel. In some cases, the $\beta\gamma$ subunits may interact with the same or a different intracellular effector. The duration of action of the active GTP- α subunit is determined by the hydrolysis of GTP to GDP by a GTPase, which is intrinsic to the α subunit, and its reassociation with the $\beta\gamma$ subunits. This process is of longer duration than the association of the ligand with the ligand-G protein-coupled receptor, resulting in amplification of the original signal.

In the case of adenylyl cyclase activation, this enzyme synthesizes cAMP from ATP. As a *second messenger*, cAMP then goes on to activate one or several protein kinase As that phosphorylate one or several other proteins to produce the appropriate cellular effects. The targeted protein may be an enzyme, a transport protein, a contractile protein, or an ion channel. The specificity of these regulatory effects depends on the distinct protein substrates that are expressed in different cells (e.g., liver vs. smooth muscle). The actions of cAMP are terminated by several types of intracellular phosphodiesterases that convert cAMP to 5'-AMP. Competitive inhibition of phosphodiesterases to prolong the actions of cAMP is one of the mechanisms by which caffeine produces its effects.

As if the foregoing is not sufficiently complicated, the activity of adenylyl cyclase also can be inhibited by activation of different G protein-coupled receptors. The G proteins coupled to inhibitory receptors are designated G_i proteins, as opposed to those coupled to stimulatory receptors and designated G_s proteins. G_i proteins also are heterotrimers, and receptor activation of G_i also leads to GTP binding to the α subunit and its dissociation from the $\beta\gamma$, but G_i proteins differ structurally from G_s proteins. Examples of G_s -coupled receptors are β -adrenergic, dopamine-1, histamine-2, glucagon, and adrenocorticotrophic hormone. Examples of G_i -coupled receptors are α_2 -adrenergic, dopamine-2, muscarinic, and opioid. A number of different G_s and G_i protein-coupled receptors can exist on the same cell, so that the activity of adenylyl cyclase can be fine-tuned between zero and maximum.

Another important group of G protein-coupled receptors activate the enzyme phospholipase C (PLC) to hydrolyze a minor component of the plasma membrane, phosphatidylinositol-4,5-bisphosphate, into two second messengers, diacylglycerol (DAG) and IP_3 . In contrast to the cAMP systems, receptors coupled to PLC are only excitatory. Examples are α_1 -adrenergic, muscarinic, Substance P, and thyrotropin-releasing hormone

receptors. The second messenger DAG is confined to the membrane, where it activates a protein kinase C, of which nine distinct types have been identified. The other second messenger, IP_3 , diffuses through the cytosol to release calcium from intracellular stores. Calcium is involved in many cellular regulatory activities, including activation of calcium-calmodulin, which regulates the activities of other enzymes including other kinases. The kinases, in turn, phosphorylate enzymes, ion channels, or other proteins to produce cellular effects. When the phosphoinositide and cAMP signaling systems coexist, they can oppose or complement one another in complex ways.

A third second-messenger system uses cyclic guanosine-3',5'-monophosphate (cGMP) in intestinal mucosa and vascular smooth muscle. It is synthesized from GTP by activation of guanylyl cyclase and activates protein kinase G, which then dephosphorylates myosin light chains in vascular smooth muscle, thereby producing muscle relaxation. Agonists, e.g., acetylcholine and histamine, cause the release of nitric oxide from vascular endothelial cells, which then diffuses into the smooth muscle cells to activate guanylyl cyclase. A direct receptor-mediated activation is produced by atrial natriuretic factor, a blood-borne peptide hormone. In this case, the receptor domain is outside the membrane and is connected through a single transmembrane domain to the intracellular guanylyl cyclase enzyme, which is activated by receptor binding.

TYROSINE KINASE-LINKED RECEPTORS

The tyrosine kinase-linked receptors are composed of an extracellular receptor domain, a single transmembrane domain, and an intracellular catalytic domain that catalyzes phosphorylation of tyrosine residues on target proteins. Some receptors are composed of single proteins, whereas others are assembled from two subunits (e.g., insulin receptors). Activation of insulin receptors triggers increased uptake of glucose and amino acids and regulates metabolism of glycogen and lipids in the cell. The catalytic actions persist for a number of minutes after insulin leaves the binding site. Several growth factors also exert their complex cellular effects by activating tyrosine kinase or similar receptors. Growth factors trigger changes in membrane transport and other metabolic events, including regulation of DNA synthesis.

INTRACELLULAR RECEPTORS THAT CONTROL DNA TRANSCRIPTION

Activation of intracellular receptors for steroids (e.g., glucocorticoids, mineralocorticoids, sex steroids, vitamin D) and thyroid hormones stimulates the transcription of certain genes by binding to specific DNA sequences in the nucleus. The receptors generally are composed of a single protein with a ligand-binding domain, a DNA-binding domain, and a transcription-activating domain. In the inactivated state, the receptor protein is bound to another protein, a heat shock protein (hsp 90), which dissociates upon activation by a hormone, permitting DNA binding and transcription of mRNA, which then is translated into new protein. This process typically takes several hours, and the effects can last for days or weeks if there is a slow turnover of the newly synthesized proteins. A similar process accounts for the induction of drug-metabolizing enzymes in the liver by certain drugs and other chemicals. In this process, formation of a heterodimeric complex between a second protein and the ligand-bound receptor is required for DNA binding.

ENZYME INHIBITION

Enzymes are very large, complex proteins or associated proteins that evolved to catalyze specific biochemical reactions that are essential to normal cellular function. A number of very selective drugs exert their effects by inhibiting particular enzymes, so that their abilities to process their normal substrates are blocked or impaired. Enzyme inhibitors can produce competitive blockade at a substrate or cofactor binding site on the enzyme. For example, the stimulant effect of digitalis glycosides

on cardiac muscle contraction is mediated by competitive inhibition of a sodium pump, Na^+, K^+ -ATPase, which leads indirectly to an increase in intracellular calcium to interact with contractile proteins. Other enzyme inhibitors act noncompetitively at allosteric sites (sites remote from the substrate binding site), which prevent the enzyme from performing its catalytic function. For example, aspirin binds irreversibly to a site on cyclooxygenase that is remote from the binding site for arachidonic acid, which is normally converted to prostaglandins by the enzyme. The binding of related drugs such as ibuprofen is reversible. Irreversible inhibition by the formation of covalent bonds between a drug and an enzyme typically is long lasting because new enzyme must be synthesized to restore function.

RECEPTOR REGULATION

The regulation of receptor numbers or density normally is constant, because synthesis keeps pace with degradation of the proteins. However, continuous stimulation of receptors with agonists can lead to desensitization or down-regulation of receptor sensitivity or number. Desensitization can occur rapidly without a change in receptor number, whereas down-regulation usually implies a decline in receptor number. For example, excess use of β -adrenergic agonists for treating bronchial asthma can lead to loss of receptor sensitivity to the agonist, caused by changes in coupling mechanisms to the G proteins. Chronic blockade of receptors can lead to up-regulation, which, in some cases, is due to synthesis of new receptors. An example is chronic blockade of β -adrenergic receptors in the heart, in which new β -receptors are synthesized, leading to supersensitivity upon abrupt withdrawal of the blocker. Another form of supersensitivity is demonstrated by denervation of skeletal muscle, which is followed by a proliferation of nicotinic receptors within and adjacent to the neuromuscular junction.

REFERENCES

- Clark AJ The antagonism of acetylcholine by atropine. *J Physiol (London)* 1926 61: 547–556.
- Ariens EJ, ed. *Molecular Pharmacology*, vol 1. New York: Academic Press, 1964:176.
- Stephenson RP A modification of receptor theory. *Br J Pharmacol* 1956; 11: 379–393.
- Rang HP Receptor mechanisms. *Br J Pharmacol* 1973; 48: 475–495.
- Colquhoun D. In: Rang HP, ed. *Drug Receptors*. Baltimore: University Park, 1973.

BIBLIOGRAPHY

- Bertolino M, Llinas R. The central role of voltage activated and receptor operated calcium channels in neuronal cells. *Annu Rev Pharmacol Toxicol* 1992; 32: 399.
- Boelsterli U. *Mechanistic Toxicology: The Molecular Basis of How Chemicals Disrupt Biological Targets*. New York: Taylor and Francis, 2003.
- Bowery NG. *Allosteric Receptor Modulation in Drug Targeting*. New York: Informa Healthcare, 2006.
- Brunton L et al. *Goodman and Gilman's The Pharmacological Basis of Therapeutics*, 12th edn. Columbus, OH: McGraw-Hill Professional 2010.
- Conn PM, Kordon C. *Insights into Receptor Function and New Drug Development Targets (Research and Perspectives in Endocrine Interactions)*. New York: Springer, 2010.

- Denison MS, Nagy SR. Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals. *Annu Rev Pharmacol Toxicol* 2003; 43: 309.
- Ding X, Kaminsky LS. Human extrahepatic cytochromes P450: function in xenobiotic metabolism and tissue selective chemical toxicity in the respiratory and gastrointestinal tract. *Annu Rev Pharmacol Toxicol* 2003; 43: 149.
- Foreman JC et al. *Textbook of Receptor Pharmacology*, 3rd edn. Boca Raton, FL: CRC Press, 2010.
- Gibson GG, Skett PL. *Introduction to Drug Metabolism*, 2nd edn. London: Chapman & Hall, 1994.
- Gomtsyan A, Faltynek CR. *Vanilloid Receptor TRPV1 in Drug Discovery: Targeting Pain and Other Pathological Disorders*. Hoboken, NJ: Wiley, 2010.
- Hulme EC et al. Muscarinic receptor subtypes. *Annu Rev Pharmacol Toxicol* 1990; 30: 633.
- Ioannides C, ed. *Cytochromes P450, Metabolic and Toxicological Aspects*. Boca Raton, FL: CRC Press, 1996.
- Kalow W. Pharmacogenetics in biological perspective. *Pharmacol Rev* 1997; 49: 369.
- Leifert WR. *G Protein-Coupled Receptors in Drug Discovery (Methods in Molecular Biology)*. New York: Humana Press, 2009.
- Levine RR. *Pharmacology: Drug Actions and Reactions*, 4th edn. Boston: Little, Brown, 1990.
- Lewis DFW. *Guides to Cytochromes P450*. New York: Taylor & Francis, 2002.
- Loh HH et al. Molecular characterization of opioid receptors G proteins in signal transduction. *Annu Rev Pharmacol Toxicol* 1990; 30: 123.
- Meyer UA. Drugs in special patient groups: clinical importance of genetics in drug effects. In: Melman KL et al. eds. *Clinical Pharmacology*, 3rd edn. New York: McGraw-Hill, 1992.
- Mulder GJ, ed. *Conjugation Reactions in Drug Metabolism*. New York: Taylor & Francis, 1990.
- Mulder GJ. Glucuronidation and its role in regulation of biological activity of drugs. *Annu Rev Pharmacol Toxicol* 1992; 32: 25.
- Neve KA. *Functional Selectivity of G Protein-Coupled Receptor Ligands: New Opportunities for Drug Discovery (The Receptors)*. New York: Humana Press, 2009.
- Nguyen T et al. Regulatory mechanisms controlling gene expression mediated by the antioxidant response element. *Annu Rev Pharmacol Toxicol* 2003; 43: 233.
- Ortiz de Montellano PR, ed. *Cytochrome P450. Structure, Mechanism, and Biochemistry*, 3rd edn. New York: Kluwer Academic Plenum, 2003.
- Parkinson A. Biotransformation of xenobiotics. In: Claassen CD, ed. *Casarett and Doull's Toxicology: The Basic Science of Poisons*, 6th edn. New York: McGraw Hill, 2007.
- Pratt WB, Taylor P. *Principles of Drug Action*. New York: Churchill Livingstone, 1990.
- Prull CR et al. *A Short History of the Drug Receptor Concept (Science, Technology & Medicine in Modern History)*. Basingstoke Hampshire UK: Palgrave Macmillan, 2009.
- Smit MJ et al. *Chemokine Receptors as Drug Targets (Methods and Principles in Medicinal Chemistry)*. Hoboken, NJ: Wiley-VCH, 2010.
- Sueyoshi T, Negishi M. Phenobarbital response elements of cytochrome P450 genes and nuclear receptors. *Annu Rev Pharmacol Toxicol* 2001; 41: 123.
- Thummel KE, Wilkinson GR. In vitro and in vivo drug interactions involving human CYP3A. *Annu Rev Pharmacol Toxicol* 1998; 38: 389.
- Tukey RH, Strassburg CP. Human UDP-glucuronosyltransferases: metabolism, expression and disease. *Annu Rev Pharmacol Toxicol* 2000; 40: 581.
- Yamazaki M et al. Recent advances in carrier-mediated hepatic uptake and biliary excretion of xenobiotics. *Pharm Res* 1996; 13: 497.
- Zaki Y et al. Opioid receptor types and subtypes: the receptor as a model. *Annu Rev Pharmacol Toxicol* 1996; 36: 379.

Chapter 39

Drug Absorption, Distribution, Metabolism, and Excretion

Michael R. Franklin, PhD

INTRODUCTION 739

STRUCTURE AND PROPERTIES OF MEMBRANES 739

DIFFUSION AND TRANSPORT 740

PHYSICOCHEMICAL FACTORS IN PENETRATION 741

ABSORPTION OF DRUGS 742

ROUTES OF ADMINISTRATION 742

FACTORS THAT AFFECT ABSORPTION 743

DISTRIBUTION 746

BIOTRANSFORMATION 748

EXCRETION 753

INTRODUCTION

Drugs differ widely in their pharmacodynamic effects and clinical applications, as well as in penetration, absorption, and usual route of administration. They also differ in their distribution among the body tissues, and in disposition and mode of termination of action. Certain general principles that help explain the differences have both pharmaceutical and therapeutic implications. These principles facilitate an understanding of both the features that are common to a class of drugs and the differences among the members of that class.

To have the desired action, a drug must achieve absorption and transport to the appropriate tissue or organ, penetrate to the responding cell surface or subcellular structure, and elicit a response or alter ongoing processes. The drug may be distributed simultaneously or sequentially to a number of tissues, be bound or stored, be metabolized to inactive or active products, or be excreted. The basic entry, movement, and disposition of drugs and metabolites within the body are summarized in Figure 39-1. Each of the processes or events depicted relates importantly to therapeutic and toxic effects of a drug and to the mode of administration, and drug design must take each into account. The extent to which all the components of absorption, distribution, metabolism (biotransformation), and elimination apply varies enormously with the drug or xenobiotic (the latter being a term widely used to refer to not only drugs but any chemical not part of the normal biochemistry and physiology of the body) and the dose (level of exposure), and to some extent is subject to inter-individual variation, the latter often arising from genetic and disease state influences.

Pharmacokinetics is the science that treats the rate and extent of absorption, rates of distribution among body compartments, rate of elimination, and related phenomena. Because of its importance, other chapters in this book are devoted to the subject. This chapter will consider the physiological bases of the processes.

STRUCTURE AND PROPERTIES OF MEMBRANES

In almost all stages of absorption, distribution, metabolism (biotransformation), and elimination, a drug must pass through several to many biological membranes during the processes. Since membranes are traversed in all of these events, a brief description of biological membranes and membrane processes is in order, as well as the relationship of the physicochemical properties of a drug molecule to penetration and transport.

Numerous sophisticated techniques have established the nature of the plasma, mitochondrial, nuclear, and other cell membranes. The description of the plasma membrane that follows is much oversimplified, but it will suffice to provide a

background for an understanding of drug penetration into and through membranes. The cell membrane has been characterized as a bimolecular layer of lipid material entrained between two parallel monomolecular layers of protein. However, rather than forming a continuous layer, the protein layer comprises “islands” sporadically scattered over the surfaces. For many proteins, much of the protein is below the surface and within the fatty bilayer. The lipid bilayer can be envisaged as a somewhat orderly, lamellar array of phospholipid molecules associated tail-to-tail, each tail being an alkyl chain or steroid group, and the heads being polar groups. The disorder that does exist is the result of the different degrees of saturation of the fatty acids and the interspersed cholesterol molecules that break up the close packing of the fatty acid tails. Cholesterol maintains the mechanical stability of cell membranes, is a determinant of membrane fluidity, and—with relevance to drug passage across membranes—decreases permeability to water-soluble molecules. Moreover, the lamellar portion is penetrated by large globular proteins with a highly hydrophobic interior (like the lipid layers), and by some fibrous proteins as well.

The plasma membrane is asymmetrical. The lipid composition varies from cell type to cell type and perhaps from site to site on the same membrane. There are, for example, differences between the membrane of the endoplasmic reticulum and the plasma membrane, even though the membranes are co-extensive. The membrane surface facing the cytoplasm is rich in phosphatidylethanolamine and phosphatidylserine, while the surface facing the outside is rich in phosphatidylcholine and sphingomyelin. Oligosaccharide chains linked to lipids (glycolipids), and oligo- and polysaccharide chains attached to proteins (glycoproteins) are confined to non-cytosolic facing surfaces. Sugar moieties attached to the outer proteins are most often attached to the asparagine residue. These sugar moieties are important to both cellular and immunological recognition and adhesion, and they have other functions as well. Where membranes are double, the inner and outer layers differ considerably; the inner and outer membranes of mitochondria have strikingly different compositions and properties.

The cell membrane appears to be perforated by water-filled pores of various sizes, varying from about 4 to 10 Å, most of which are about 7 Å. Probably all major ion (water-filled) channels penetrate the large proteins assemblies that traverse the membrane. Through these pores pass inorganic ions and small organic molecules. Among the common inorganic ions, because sodium ions are more hydrated than potassium and chloride ions, they are larger and do not pass as freely through the pores as do potassium and chloride. Ion (ion plus water) movement can be by diffusion down a chemical concentration gradient. However, movement of ions through the pores can be controlled by counterion transport, or expenditure of intracellular

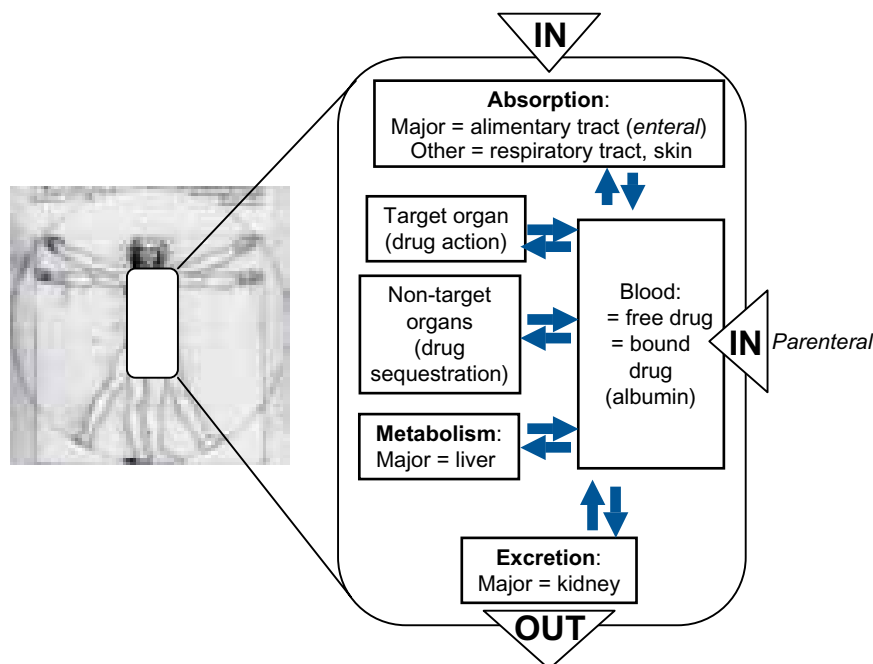


Figure 39-1. Interrelationship of the major components of drug entry, distribution, metabolism, and elimination within the body.

energy—adenosine triphosphate (ATP) hydrolysis—by so-called ABC transporters (see Excretion in this chapter). Vascular endothelium appears to have pores at least as large as 40 Å, but these seem to be interstitial passages rather than transmembrane pores. Lipid molecules small enough to pass through the cell membrane pores may do so, but they have a higher probability of entering into the lipid layer (since pores constitute less than 1 percent of the cell surface upon which the drug molecule impinges); from there these molecules will equilibrate chemically with the interior of the cell. Other proteins may be confined to one or the other surface and not traverse the membrane. Often proteins on the inner surface protein are linked to intracellular structural proteins that contribute to cell shape.

DIFFUSION AND TRANSPORT

Transport is the movement of a drug from one place to another within the body. The drug may diffuse freely in uncombined form with a kinetic energy appropriate to its thermal environment, or it may move in combination with extracellular or cellular constituents, sometimes in connection with energy-yielding processes that allow the molecule or complex to overcome barriers to simple diffusion.

SIMPLE NONIONIC DIFFUSION AND PASSIVE TRANSPORT

Molecules in solution move in a purely random fashion, provided they are not charged and moving in an electrical gradient. Such random movement is called diffusion; if the molecule is uncharged, it is called nonionic diffusion. In a population of drug molecules, the probability that during unit time any drug molecule will move across a boundary is directly proportional to the number of molecules adjoining that boundary and, therefore, to the drug concentration. Except at dilutions so extreme that only a few molecules are present, the actual rate of movement (molecules per unit time) is directly proportional to the probability of movement and, therefore, to the concentration. Once molecules have passed through the boundary to the opposite side, their random motion may cause some to return and others to continue to move farther away from the boundary. The rate of return is likewise proportional to the concentration on the opposite side

of the boundary. It follows that, although molecules are moving in both directions, there will be a net movement from the region of higher to that of lower concentration, and the net transfer will be proportional to the concentration differential. If the boundary is a membrane, which has both substance and dimension, the rate of movement is also directly proportional to the permeability and inversely proportional to the thickness. These factors combine into Fick's law of diffusion:

$$dQ/dt = \bar{D}A(C_1 - C_2)/x$$

where Q is the net quantity of drug transferred across the membrane, t is time, C_1 is the concentration on one side and C_2 that on the other, x is the thickness of the membrane, A is the area, and $\bar{D}$ is the diffusion coefficient, related to permeability. Since a biological membrane is heterogeneous, with pores of different sizes and probably with varying thickness and composition, both $\bar{D}$ and x probably vary from place to place. Nevertheless, some mean values can be assumed. It is customary to combine the membrane factors into a single constant, called a permeability constant or coefficient, P , so that $P = \bar{D}/x$, and A in the equation above has unit value. The rate of net transport (diffusion) across the membrane then becomes:

$$dQ/dt = P(C_1 - C_2)$$

As diffusion continues, C_1 approaches C_2 , and the net rate, dQ/dt , approaches zero in exponential fashion, characteristic of a first-order process. Equilibrium is defined as that state in which $C_1 = C_2$. The equilibrium is, of course, dynamic, with equal numbers of molecules being transported in each direction during unit time. If water also is moving through the membrane, it may either facilitate the movement of drug or impede it, according to the relative directions of movement of water and drug; this effect of water movement is called solvent drag.

IONIC OR ELECTROCHEMICAL DIFFUSION

If a drug is ionized, the transport properties are modified. The probability of penetrating the membrane is still a function of concentration, but it is also a function of the potential difference or electrical gradient across the membrane. A cationic

drug molecule will be repelled from the positive charge on the outside of the membrane, and only those molecules with a high kinetic energy will pass through the ion barrier. If the cation is polyvalent, it may not penetrate at all.

Once inside the membrane, a cation will be simultaneously attracted to the negative charge on the intracellular surface of the membrane and repelled by the outer surface; it is said to be moving along the electrical gradient. If it is also moving from a higher toward a lower concentration, it is said to be moving along its electrochemical gradient, which is the sum of the influences of the electrical field and the concentration differential across the membrane.

Once inside the cell, cations will tend to be kept inside by the attractive negative charge on the interior of the cell, and the intracellular concentration of drug will increase until—by sheer numbers of accumulated drug particles—the rate of outward diffusion or mass escape equals that of inward transport. At this point electrochemical equilibrium is said to have occurred. At electrochemical equilibrium at body temperature (37°C), ionized drug molecules will be distributed according to the Nernst equation:

$$\pm \log C_o/C_i = ZE/61$$

where C_o is the molar extracellular and C_i the intracellular concentration, Z is the number of charges per molecule, and E is the membrane potential in millivolts. $\log C_o/C_i$ is positive when the molecule is negatively charged and negative when the molecule is positively charged.

FACILITATED DIFFUSION

Sometimes a substance moves more rapidly through a biological membrane than can be accounted for by the process of simple diffusion. This accelerated movement is termed facilitated diffusion. It is due to the presence of a special molecule within the membrane, called a carrier, with which the transported substance combines. There is considered to be greater permeability to the carrier-drug complex than to the drug alone, so that the transport rate is enhanced. After the complex has traversed the membrane, it dissociates. For the carrier process to be continuous, either the carrier must return to the original side of the membrane to be used again, or it must constantly be produced on one side and eliminated on the other. Many characteristics of facilitated diffusion, formerly attributed to ion carriers, can be explained by ion exchange. Facilitated diffusion only transports a molecule along its electrochemical gradient.

ACTIVE TRANSPORT

Active transport can be defined as energy-dependent movement of a substance through a biological membrane against an electrochemical gradient. It is characterized as follows:

1. The substance is transported from a region of lower to one of higher electrochemical activity.
2. Metabolic poisons (that most often reduce ATP concentrations) interfere with transport.
3. The transport system shows a requirement for specific chemical structures.
4. Closely related chemicals are competitive for the transport system.
5. The transport rate approaches an asymptote (i.e., saturates) as concentration increases.

Characteristics 3, 4, and 5 are in common with those of carrier-mediated facilitated diffusion.

Many drugs are secreted by active transport from the renal tubules into urine, from liver cells into bile or blood, from intestinal cells into the lumen of the gastrointestinal (GI) tract, or from the cerebrospinal fluid into blood, but the role of active transport of drugs in the distribution into most body compartments and tissues has been less extensively documented, although it is now an active area of research. Active transport is

often required for the movement of drug metabolites, entities that generally have less lipid solubility than the parent drug, across cell membranes.

PINOCYTOSIS AND EXOCYTOSIS

Many (perhaps all) cells are capable of a type of phagocytosis called pinocytosis. The cell membrane has been observed to invaginate into a sac-like structure containing extracellular materials and then pinch off the sac at the membrane, so that the sac remains as a vesicle or vacuole within the interior of the cell. Because metabolic activity is required and because an extracellular substance can be transported against an electrochemical gradient, pinocytosis shows some of the same characteristics as active transport. However, pinocytosis is relatively slow and inefficient compared with most active transport, except in GI absorption, where for some xenobiotics pinocytosis may be of some importance.

It is not known to what extent pinocytosis contributes to the transport of most drugs, but many macromolecules and even larger particles can be absorbed by the gut. Exocytosis is the reverse of pinocytosis. Granules, vacuoles, or other organelles within the cell move to the cell membrane, fuse with it, and extrude their contents into the interstitial space.

PHYSICOCHEMICAL FACTORS IN PENETRATION

Drugs and other substances may traverse the membrane primarily either through the pores, or by movement into the membrane lipids and subsequent diffusion from the membrane into the cytosol or other fluid on the far side of the membrane. The physicochemical prerequisites differ according to which route is taken. To pass through the pores, the *diameter* of the molecule must be smaller than the pore, but the molecule can be longer than the pore diameter. The probability that a long, thin molecule will be suitably oriented, however, is low unless there is also bulk flow, and therefore transmembrane passage of such molecules is slow.

Water-soluble molecules with low lipid solubility are usually thought to pass through the membrane mainly via the pores. If there is a membrane carrier or active-transport system, a low solubility of the drug in membrane lipids is no impediment to penetration, because the drug-carrier complex is assumed to have an appropriate solubility, and energy from an active-transport system enables the drug to penetrate the energy barrier imposed by the lipids. Actually, the lipids are not an important energy barrier; rather, the barrier is the force of attraction of the solvent water for its dipolar-to-polar solute, so that it is difficult for the solute to leave the water and enter the lipid.

Drugs with a high solubility in the membrane lipids pass easily through the membrane. Even when their dimensions are small enough to permit passage through pores, lipid-soluble drugs primarily pass through the membrane lipids, not only because chemical partition favors the lipid phase but also because, as mentioned previously, the surface area occupied by pores is only a small fraction of the total membrane area.

LIPID SOLUBILITY AND PARTITION COEFFICIENTS

Over a century ago, the importance of lipid solubility in the penetration and absorption of drugs was being investigated. Eventually it was recognized that more important than lipid solubility was the lipid-to-water partition (or distribution) coefficient; in other words, a high lipid solubility does not favor penetration unless the water solubility is low enough so that the drug is not entrained in the aqueous phase. When the water solubility of a substance is so low that a significant concentration in water or extracellular fluid cannot be achieved, absorption may be negligible despite a favorable partition coefficient. Hence, such substances as mineral oil or petrolatum are virtually unabsorbed. The optimal partition coefficient for permeation of the skin appears to be lower than that for the permeation of the cell membrane, perhaps being as low as unity.

DIPOLARITY, POLARITY, AND NONIONIC DIFFUSION

The partition coefficient of a drug depends upon the polarity and the size of the molecule. Drugs with a high dipole moment, even though nonionized, have low lipid solubility and hence penetrate poorly. An example of a highly dipolar substance with a low partition coefficient, which does not penetrate into cells, is sulfoxazole. Sulfadiazine is somewhat less dipolar, has a chloroform-to-water partition coefficient 10 times that of sulfoxazole, and readily penetrates cells. Ionization not only greatly diminishes lipid solubility but also may impede passage through charged membranes.

It is often stated that ionized molecules do not penetrate membranes, except for ions of small diameter. This is not necessarily true because of the presence of membrane carriers for some ions that effectively shield or neutralize the charge (formation of ion pairs). The renal tubular transport systems, which transport such obligate ions as tetraethylammonium, probably form ion pairs. Furthermore, if an ionized molecule has a large non-polar moiety such that appreciable lipid solubility is imparted to the molecule despite the charge, the drug may penetrate, though usually at a slow rate. Nevertheless, when a drug is a weak acid or base, the nonionized form, with a favorable partition coefficient, passes through a biological membrane so much more readily than the ionized form that for all practical purposes, only the nonionized form is said to pass through the membrane. This has become known as the principle of nonionic diffusion.

ABSORPTION OF DRUGS

Absorption is the process of movement of a drug from the site of application into the extracellular compartment of the body. Inasmuch as there is a great similarity among the various membranes through which a drug may pass to gain access to the extracellular fluid, it might be expected that the particular site of application (or route) would make little difference to the successful absorption of the drug. Actually it makes a great deal of difference; many factors, other than the structure and composition of the membrane, determine the ease with which a drug is absorbed. These factors are discussed in the following sections, along with an account of the ways that drug formulations can be manipulated to alter the ability of a drug to be absorbed readily.

ROUTES OF ADMINISTRATION

Drugs can be administered by many different routes, including oral, rectal, sublingual or buccal, parenteral, inhalation, and topical. The choice of a route depends upon both convenience and necessity.

ORAL ROUTE

This is obviously the most convenient route for access to the systemic circulation, providing that various factors do not militate against it. Oral administration does not always give rise to plasma concentrations sufficiently high to be effective; some drugs are absorbed unpredictably or erratically; patients occasionally have an absorption malfunction. Drugs cannot be given by mouth to patients who have GI intolerance, are being prepared for anesthesia, or have had GI surgery. Oral administration is also precluded in comatose patients.

In the drug development setting, Lipinski's "Rule of Five" predicts that, in general, an orally active drug has no more than one violation of the following criteria:

- not more than five hydrogen bond donors (oxygen or nitrogen atoms with one or more hydrogen atoms)
- not more than ten (2×5) hydrogen bond acceptors (nitrogen or oxygen atoms)
- a molecular mass not greater than (100×5) 500 daltons
- an octanol-to-water partition coefficient $\log P$ not greater than 5.

RECTAL ROUTE

Drugs that are ordinarily administered by the oral route can usually be administered by injection or by the alternative lower enteral route, through the anal portal into the rectum or lower intestine. With regard to the latter, rectal suppositories or retention enemas were formerly used quite frequently, but their popularity has abated somewhat as a result of improvements in parenteral preparations. Nevertheless, they continue to be valid—and sometimes very important—ways of administering a drug, especially in pediatric and geriatric patients, and retention enema may offer a useful substitute for the oral route. However, rectal suppositories may be inadequate when rapid absorption and high plasma levels are required.

SUBLINGUAL OR BUCCAL ROUTE

Even though an adequate plasma concentration may eventually be achievable by the oral route, it may rise much too slowly for use in some situations when a rapid response is desired. In such situations parenteral therapy is usually indicated. However, patients with angina pectoris may get quite prompt relief from an acute attack by the sublingual or buccal administration of nitroglycerin, so that parenteral administration can be avoided. When only small amounts of drugs are required to gain access to the blood, the buccal route may be very satisfactory, providing the physicochemical prerequisites for absorption by this route are present in the drug and dosage form.

PARENTERAL ROUTES

These routes, by definition, include any route other than the oral-GI (enteral) tract, but in common medical usage the term excludes topical administration and includes only various hypodermic routes. Parenteral administration includes the intravenous, intramuscular, and subcutaneous routes. Parenteral routes are an option whenever enteral routes are contraindicated or are inadequate.

The intravenous route may be preferred on occasion, even when a drug may be well absorbed by the oral route. There is no delay imposed by absorption before the administered drug reaches the circulation, and blood levels rise virtually as rapidly as the time necessary to empty the syringe or infusion bottle. Consequently, the intravenous route is the preferred route when an emergency calls for an immediate response.

In addition to the rapid rise in plasma concentration of drug, another advantage of intravenous administration is the greater predictability of the peak plasma concentration, which with some drugs can be calculated with a fair degree of precision. Smaller doses are generally required by the intravenous than by other routes, but this usually affords no advantage, inasmuch as the sterile injectable dosage form costs more than enteric preparations, and the requirements for medical or paramedical supervision of administration also may add to the cost and inconvenience.

Because of the rapidity with which drug enters the circulation, dangerous side effects to the drug may occur that often are not extant by other routes. The principal untoward effect is a depression of cardiovascular function. Consequently, some drugs must be given quite slowly to avoid vasculotoxic concentrations of drug in the plasma. Acute, serious allergic responses are also more likely to occur by the intravenous route than by other routes.

Many drugs are too irritating to be given by the oral, intramuscular, or subcutaneous route, and must of necessity be given intravenously. However, such drugs also may cause damage to the veins (phlebitis) or, if extravasated, cause necrosis around the injection site. Consequently, such irritant drugs may be diluted in isotonic solutions of saline, dextrose, or other media, and given by slow infusion, providing that the slower rate of delivery does not negate the purpose of the administration in emergency situations.

Absorption by the intramuscular route is relatively fast, and this parenteral route may be used when an immediate effect

is not required but a prompt effect is desirable. Intramuscular deposition can also be made of certain repository preparations where rapid absorption is not desired. Absorption from an intramuscular depot is more predictable and uniform than from a subcutaneous site. Irritation around the injection site is a frequent accompaniment of intramuscular injection, depending upon the drug and other ingredients. Because of the dangers of accidental intravenous injection, medical supervision is generally required. Sterilization is necessary.

In subcutaneous administration, the drug is injected into the connective tissue just below the skin. Absorption is slower than by the intramuscular route but nevertheless can be prompt with some drugs. Often, however, absorption by this route may be no faster than by the oral route. Therefore, when a fairly prompt response is desired with some drugs, the subcutaneous route may not offer much advantage over the oral route, unless for some reason the drug cannot be given orally.

The slower rate of absorption by the subcutaneous route is usually the reason for choosing the route, and the drugs given by this route are usually those for which it is desirable to distribute the drug's action over several hours, to avoid either too intense a response, too short a response, or frequent injections. Examples of drugs given by this route are insulin and sodium heparin, neither of which is absorbed orally, and both of which should be absorbed slowly over many hours. In the treatment of asthma, epinephrine is usually given subcutaneously to avoid the dangers of rapid absorption and consequent dangerous cardiovascular effects. Many repository preparations, including tablets or pellets, are given subcutaneously. As with other parenteral routes, irritation may occur. Sterile preparations are also required. However, medical supervision is not always required, and self-administration by this route is customary with certain drugs, such as insulin.

Intradermal injection, in which the drug is injected into the dermis instead of below it, is rarely used, except in certain diagnostic and test procedures, such as screening for allergic or local irritant responses.

Occasionally, even by the intravenous route, it is not possible, practical, or safe to achieve plasma concentrations high enough so that an adequate amount of drug penetrates into special compartments (e.g., the cerebrospinal fluid) or various cavities (e.g., the pleural cavity). The brain is especially difficult to penetrate with water-soluble drugs. The name blood-brain barrier is applied to the impediment to penetration. When drugs do penetrate, the choroid plexus often secretes them back into the blood very rapidly, so that adequate levels of drugs in the cerebrospinal fluid can be difficult to achieve. Consequently, intrathecal or intraventricular administration may be indicated.

Body cavities such as the pleural cavity are normally wetted by a small amount of effusate that is in diffusion equilibrium with the blood and hence is accessible to drugs. However, infections and inflammations may cause the cavity to fill with serofibrinous exudate that is too dense to be in rapid diffusion equilibrium with the blood. Intracavitary administration thus may be required. It is extremely important that sterile, non-irritating preparations be used for intrathecal or intracavitary administration.

INHALATION ROUTE

Inhalation may be employed for delivering gaseous or volatile substances into the systemic circulation, as with most general anesthetics. Absorption is virtually as rapid as the drug can be delivered into the alveoli of the lungs, since the alveolar and vascular epithelial membranes are quite permeable, blood flow is abundant, and there is a very large surface for absorption.

Aerosols of nonvolatile substances can also be administered by inhalation, but the route is used infrequently for delivery into the systemic circulation because of various factors that contribute to blood concentrations that are erratic or difficult to achieve. Whether or not an aerosol reaches and is retained in pulmonary alveoli depends critically upon particle size.

Particles larger than 1 micrometer in diameter tend to settle in the bronchioles and bronchi, whereas particles smaller than 0.5 micrometer fail to settle and mainly are exhaled. Aerosols are employed mostly when the purpose of administration is an action of the drug upon the respiratory tract itself. An example of a drug commonly given as an aerosol is isoproterenol, which is employed to relax the bronchioles during an asthma attack.

TOPICAL ROUTE

Although the stratum corneum is not a membrane in the same sense as a cell membrane, it offers a barrier to diffusion, which is of significance in the topical application of drugs. The stratum corneum consists of several layers of dead, keratinized, cutaneous epithelial cells enmeshed in a matrix of keratin fibers and bound together with cementing desmosomes and penetrating tonofibrils of keratin. Varying amounts of lipids and fatty acids from dying cells, sebum, and sweat are contained among the dead squamous cells. Immediately beneath the layer of dead cells and above the viable epidermal epithelial cells is a layer of keratohyaline granules and various water-soluble substances, such as amino acids, purines, monosaccharides, and urea.

Both the upper and lower layers of the stratum corneum are involved in the cutaneous barrier to penetration. The barrier to penetration from the surface is in the upper layers for water-soluble substances and the lower layers for lipid-soluble substances, and the barrier to the outward movement of water is in the lowest layer.

Topical administration is employed to deliver a drug at, or immediately beneath, the point of application. Although occasionally enough drug is absorbed into the systemic circulation to cause systemic effects, absorption is too erratic for the topical route to be used routinely for systemic therapy. However, various transdermal preparations are employed quite successfully for systemic use. A large number of topical medicaments are applied to the skin, although topical drugs are also applied to the eye, nose, throat, ear, vagina, etc.

In humans, percutaneous absorption probably occurs mainly from the surface. Absorption through the hair follicles occurs, but the follicles in humans occupy too small a portion of the total integument to be of primary importance. Absorption through sweat and sebaceous glands generally appears to be minor. When the medicament is rubbed on vigorously, the amount of the preparation that is forced into the hair follicles and glands is increased. Rubbing also forces some material through the stratum corneum without molecular dispersion and diffusion through the barrier. When the skin is diseased or abraded, the cutaneous barrier may be disrupted or defective, so that percutaneous absorption may be increased. Since much of a drug that is absorbed through the epidermis diffuses into the circulation without reaching a high concentration in some portions of the dermis, systemic administration may be preferred in lieu of, or in addition to, topical administration.

FACTORS THAT AFFECT ABSORPTION

In addition to the physicochemical properties of drug molecules and biological membranes, various factors affect the rate of absorption and determine, in part, the choice of route of administration.

CONCENTRATION

It is self-evident that the concentration, or more exactly, the thermodynamic activity, of a drug in a drug preparation will have an important bearing upon the rate of absorption, since the rate of diffusion of a drug away from the site of administration is directly proportional to the concentration. Thus, a 2 percent solution of lidocaine will induce local anesthesia more rapidly than a 0.2 percent solution. However, drugs administered in solid form are not necessarily absorbed at the maximal rate (see, Physical State of Formulation and Dissolution Rate below).

After oral administration the concentration of drugs in the gut is a function of the dose, but the relationship is not necessarily linear. Drugs with a low aqueous solubility quickly saturate the GI fluids, so that the rate of absorption tends to reach a limit as the dose is increased. The peptizing and solubilizing effects of bile and other constituents of the GI contents assist in increasing the rate of absorption but are in themselves somewhat erratic. Furthermore, many drugs affect the rates of gastric, biliary, and small intestinal secretion, which causes further deviations from a linear relationship between concentration and dose.

Drugs that are administered subcutaneously or intramuscularly also may not always show a direct linear relationship between the rate of absorption and the concentration of drug in the applied solution, because osmotic effects may cause dilution or concentration of the drug if the movement of water or electrolytes is different from that of the drug. Whenever possible, drugs for hypodermic injection are prepared as isotonic solutions. Some drugs affect the local blood flow and capillary permeability, so that at the site of injection there may be a complex relationship between concentration achieved and the concentration administered.

PHYSICAL STATE OF FORMULATION AND DISSOLUTION RATE

The rate of absorption of a drug may be affected greatly by the rate at which the drug is made available to the biological fluid at the site of administration. The intrinsic physicochemical properties, such as solubility and the thermodynamics of dissolution, are only some of the factors that affect the rate of dissolution of a drug from a solid form. Other factors include not only the unavoidable interactions among the various ingredients in a given formulation but also deliberate interventions to facilitate dispersion or retard it. There also are factors that affect the rate of delivery from liquid forms. For example, a drug in a highly viscous vehicle is absorbed more slowly from the vehicle than a drug in a vehicle of low viscosity; in oil-in-water emulsions the rate depends upon the partition coefficient. These manipulations are the subject of biopharmaceutics.

AREA OF ABSORBING SURFACE

The area of absorbing surface is an important determinant of the rate of absorption. To the extent that the therapist must work with the absorbing surfaces available in the body, the absorbing surface is not subject to manipulation. However, the extent to which the existing surfaces may be used is subject to variation. In those rare instances in which percutaneous absorption is intended for systemic administration, the entire skin surface is available.

Subsequent to subcutaneous or intramuscular injections, the site of application can be massaged to spread the injected fluid from a compact mass to a well-dispersed deposit. Alternatively, the dose can be divided into multiple small injections, although this recourse is generally undesirable.

The different areas for absorption afforded by the various routes partly account for differences in the rates of absorption by those routes. The large alveolar surface of the lungs allows extremely rapid absorption of gases, vapors, and properly aerosolized solutions; with some drugs the rate of absorption can be nearly as fast as with intravenous injection. In the gut the small intestine is the site of the fastest, and hence the greatest, absorption because of the small lumen and highly developed villi and microvilli; the stomach has a relatively small surface area, so that even most weak acids are absorbed predominantly in the small intestine despite a pH partition factor that should favor absorption from the stomach (see The pH Partition Principle below).

VASCULARITY AND BLOOD FLOW

Although the thermal velocity of a freely diffusible, average drug molecule is on the order of meters per second, in solution

the rate at which it will diffuse away from a reference point will be much slower. Collisions with water and/or other molecules that cause a random motion, and the forces of attraction between the drug and water or other molecules, slow the net mean velocity.

The time taken to traverse a given distance is a function of the square of the distance; on average it would take about 0.01 second for a net outward movement of 1 micrometer, 1 second for 10 micrometers, 100 seconds for 100 micrometers, etc. In a highly vascular tissue such as skeletal muscle, in which there can be more than 1000 capillaries per square millimeter of cross section, a drug molecule would not have to travel more than a few micrometers, hence less than a second on average, to reach a capillary from a point of extravascular injection.

Once the drug reaches the blood, diffusion is not important to transport and the rate of blood flow determines the movement. The velocity of blood flow in a capillary is about 1 mm/second, which is 100 times faster than the mean net velocity of drug molecules 1 mm away from their injection site. The velocity of blood flow is even faster in the larger vessels. Overall, less than a minute is required to distribute drug molecules from the capillaries at the injection site to the rest of the body.

From the above discussion it follows that absorption is most rapid in the vascular tissues. Drugs are absorbed more rapidly from intramuscular sites than from less vascular subcutaneous sites, etc. Despite the small absorbing surface for buccal or sublingual absorption, the high vascularity of the buccal, gingival, and sublingual surfaces favors an unexpectedly high rate of absorption. Because of hyperemia, absorption will be faster from inflamed than from normal areas, unless the presence of edema lengthens the mean distance between capillaries and thus negates the effects of hyperemia on absorption.

Vasoconstriction can have a profound effect upon the rate of absorption. When a local effect of a drug is desired, as in local anesthesia, absorption away from the infiltrated site can be impeded greatly by vasoconstrictors included in the preparation. Unwanted vasoconstriction (e.g., cold-induced) can sometimes cause serious problems. During shock, blood flow is also diminished.

Extravascularly injected molecules too large to pass through the capillary endothelium will of necessity enter the systemic circulation through the lymph. Thus, the lymph flow can be important to the absorption of a few drugs.

MOVEMENT

Several factors combine so that movement at the site of injection increases the rate of absorption. In the intestine, segmental movements and peristalsis aid in dividing and dispersing the drug mass. The continual mixing of the chyme helps keep the concentration maximal at the mucosal surface. The pressures developed during segmentation and peristalsis may also favor a small amount of filtration. Movement at the site of hypodermic injection also favors absorption, as it tends to force the injected material through the tissue, increasing the surface area of drug mass and decreasing the mean distance to the capillaries. Movement also increases the flow of blood and lymph. The selection of a site for intramuscular injection may be determined by the amount of expected movement, according to whether the preparation is intended as a fast-acting or a repository preparation.

GASTRIC MOTILITY AND EMPTYING

The motility of the stomach is more important to the rate at which an orally administered drug is passed on to the small intestine than it is to the rate of absorption from the stomach itself, since for various reasons noted above, absorption from the stomach may be of minor importance.

The average emptying time of the unloaded stomach is about 40 minutes, and the half-time is about 10 minutes, although it varies according to stomach contents, to reflex and psychological factors, and to the action of certain autonomic drugs or some types of disease. The effect of food to delay absorption is

partly due to its action to prolong emptying time. The emptying time causes a delay in the absorption of drug, which may be unfavorable or favorable according to what is desired. In the case of therapy with antacids, gastric emptying is a nuisance, since it removes the antacid from the stomach where it is needed.

SOLUBILITY AND BINDING

The dissolution of drugs of low solubility is generally a slow process. Indeed, low solubility is the result of a low rate of departure of drug molecules from the undispersed phase. Furthermore, since the concentration around the drug mass is low, the concentration gradient from the site of deposition to the plasma is small, and the rate of diffusion is low, accordingly.

When it is desired that a drug have a prolonged action but not a high plasma concentration, a derivative of low solubility is often sought. The insoluble estolates and other esters of several steroids have durations of action of weeks because of the slow rates of absorption from the sites of injection. Insoluble salts or complexes of acidic or basic drugs also are employed as repository preparations; for example, the procaine salt of penicillin G has a low solubility and is used in a slow-release form of the antibiotic.

The solubility of certain macromolecules depends critically on the ionization of substituent groups. When they are amphiprotic, they are least soluble at their isoelectric pH. Insulin is normally soluble at the pH of the extracellular fluid, but by combining insulin with the right proportion of a basic protein, such as protamine, the isoelectric pH can be made to be approximately 7.4 from 5.1, and the complex can be used as a low-solubility, prolonged-action drug.

Some drugs may bind with natural substances at or near the site of application. The strongly ionized mucopolysaccharides in connective tissue, ground substance, and mucous secretions of the gut retard the absorption of a number of drugs, especially large cationic or polycationic molecules. In the gut, the binding is the least at low pH, which should favor absorption of large cations from the stomach; however, absorption from the stomach is slow (see above), so that the absorption of large cations occurs mainly in the upper duodenum where the pH is still relatively low. Pharmacologically inactive quaternary ammonium compounds are sometimes included in an oral preparation of a quaternary ammonium drug for the purpose of saturating the binding sites of mucin and other mucopolysaccharides, and thereby enhancing the absorption of drug.

In addition to mucopolysaccharides in mucous secretions, food in the GI tract binds many drugs and slows absorption. Antacids, especially aluminum hydroxide plus other basic aluminum compounds and magnesium trisilicate, bind amine and ammonium drugs, interfering with absorption.

DONNAN EFFECT

The presence of a charged macromolecule on one side of a semi-permeable membrane (impermeable to the macromolecule) will alter the concentration of permeant ionized particles according to the Donnan equilibrium. Accordingly, drug molecules of the same charge as the macromolecule will be constrained to the opposite side of the membrane. The pull of large anionic blood plasma proteins will concentrate small membrane-diffusible cations around them, but not small diffusible anions, which will tend to concentrate on the other side. The presence of appropriately charged macromolecules will not only influence the distribution of drug ions in accordance with the Donnan equation but also increase the rate of transfer of the drug across the membrane, because of mutual ionic repulsion. This effect is sometimes used to facilitate the absorption of ionizable drugs from the GI tract. The Donnan effect also operates to retard the absorption of drug ions of opposite charge; however, the mutual electrostatic attraction of a macromolecule and a drug ion generally results in actual binding, which is more important than the Donnan effect.

VEHICLES AND ABSORPTION ADJUVANTS

Drugs that are to be applied topically to the skin and mucous membranes are often dissolved in vehicles that are thought to enhance penetration. For a long time it was thought that oleaginous vehicles promoted the absorption of lipid-soluble drugs. However, the role and effect of the vehicle has proven to be quite complex. In the skin at least five factors are involved:

1. The effect of the vehicle to alter the hydration of the keratin in the barrier layer
2. The effect of the vehicle to promote or prevent the collection of sweat at the surface of the skin
3. The partition coefficient of the drug in a vehicle–water system
4. The permeability of the skin to the undissolved drug
5. The permeability of the skin to the vehicle.

The effect of the vehicle to aid in the access of the drug to the hair follicles and sebaceous glands may also be involved, although in humans the follicles and glands are probably ordinarily of minor importance to absorption.

A layer of oleaginous material over the skin prevents the evaporation of water, so that the stratum corneum may become macerated and more permeable to drugs. In dermatology it is sometimes the practice to wrap the site of application with plastic wrap or some other waterproof material for the purpose of increasing the maceration of the stratum corneum. However, the layer of perspiration that forms under an occlusive vehicle may become a barrier to the movement of lipid-soluble drugs from the vehicle to the skin, but it may facilitate the movement of water-soluble drugs. Conversely, polyethylene glycol vehicles remove the perspiration and dehydrate the barrier, which decreases the permeability to drugs; such vehicles remove the aqueous medium through which water-soluble drugs may pass down into the stratum corneum but at the same time facilitate the transfer of lipid-soluble drugs from the vehicle to the skin.

Even in the absence of a vehicle, it is not clear what physicochemical properties of a drug favor cutaneous penetration—high lipid solubility being a prerequisite according to some authorities, and an ether-to-water partition coefficient of approximately unity according to others. Yet, the penetration of ethanol and dibromomethane are nearly equal, and other such enigmas exist. It is not surprising, then, that the effects of vehicles are not altogether predictable.

A general statement might be made that if a drug is quite soluble in a poorly absorbed vehicle, the vehicle will retard the movement of the drug into the skin. For example, salicylic acid is 100 times as permeant when absorbed from water as from polyethylene glycol, and pentanol is five times as permeant from water as from olive oil. Yet, ethanol penetrates five times faster from olive oil than from either water or ethanol, all of which denies the trustworthiness of generalizations about vehicles.

For several decades there has been some interest in certain highly dielectric, aprotic solvents, especially dimethyl sulfoxide. Such substances generally prove to be excellent solvents for both water- and lipid-soluble compounds, and for some compounds not soluble in either water or lipid solvents. The extraordinary solvent properties probably are due to a high polarizability and van der Waals bonding capacity, a high degree of polarization (dipole moment), and a lack of association through hydrogen bonding. As a vehicle, dimethyl sulfoxide greatly facilitates the permeation of the skin and other biological membranes by numerous drugs, including such large molecules as insulin.

OTHER FACTORS THAT AFFECT DRUG ABSORPTION

Several other, less well-defined factors affect the absorption of drugs, some of which may operate, in part, through factors already cited above. Disease or injury has a considerable effect upon absorption. For example, debridement of the stratum corneum increases the permeability to topical agents, meningitis

increases the permeability of the blood–brain barrier, biliary insufficiency decreases the absorption of lipid-soluble substances from the intestine, and acid–base disturbances can affect the absorption of weak acids or bases. Certain drugs, such as ouabain, that affect active-transport processes may interfere with the absorption of certain other drugs. The condition of the ground substance, or intracellular cement, probably bears on the absorption of certain types of molecules. Hyaluronidase, which depolymerizes the mucopolysaccharide ground substance, can be demonstrated to facilitate the absorption of some, but not all, drugs from subcutaneous sites.

DISTRIBUTION

The term drug disposition is used variously to include all processes that tend to lower the plasma concentration of drug, as opposed to drug absorption, which elevates the plasma level. Consequently, the distribution of drugs to the various tissues is a component of disposition, as are biotransformation and excretion, two components that contribute to drug elimination. The term distribution denotes the partitioning of a drug among the numerous locations where a drug may be contained within the body. Biotransformations are the alterations in the chemical structure of a drug that are imposed upon it by the life processes. Excretion is, in a sense, the converse of absorption—namely, the transportation of the drug or its products out of the body. The term applies whether or not special organs of excretion are involved.

The body can be considered to comprise a number of compartments: enteric (GI), plasma, interstitial, cerebrospinal fluid, bile, glandular secretions, urine, storage vesicles, cytoplasm or intracellular space, etc. Some of these compartments, such as urine and secretions, are open-ended, but since their contents relate to those in the closed compartments, they must also be included.

At first thought, it may seem that if a drug was distributed passively (i.e., by simple diffusion) and the plasma concentration could be maintained at a steady level, the concentration of a drug in the water in all compartments ought to become equal. It is true that some substances, such as ethanol, are distributed nearly equally throughout the body water, but they are more the exception than the rule. Such substances are mainly small, uncharged, non-dissociable, highly water-soluble molecules.

The condition of small size and high water solubility allows passage through the pores without the necessity of carrier or active transport. Small size also places a limit on van der Waals binding energy and configurational complementarity, so that binding to proteins in plasma, or cells, is slight. The presence of a charge on a drug molecule makes for unequal distribution across charged membranes, in accordance with the Donnan distribution. Dissociability causes unequal distribution when there is a pH differential between compartments, as discussed under The pH Partition Principle (see next section). Thus, even if a drug is distributed passively, its distribution may be uneven throughout the body. When active transport occurs into, or rapid biotransformation occurs within some compartments, uneven distribution is also inevitable.

THE pH PARTITION PRINCIPLE

An important consequence of nonionic diffusion is that a difference in pH between two compartments will have an important influence upon the partitioning of a weakly acidic or basic drug between those compartments. The partition is such that the nonionized form of the drug has the same concentration in both compartments, since it is the form that is freely diffusible; the ionized form in each compartment will have the concentration that is determined by the pH in that compartment, the pK_a , and the concentration of the nonionized form. The governing effect of pH and pK_a on the partition is known as the pH partition principle.

To illustrate the principle, consider the partition of salicylic acid between the gastric juice and the interior of a gastric mucosal cell (Figure 39-2). Assume the pH of the gastric juice to be 1, which it can occasionally become, and the pK_a of salicylic acid is 4. With the Henderson–Hasselbalch equation, $pH = pK_a + \log [\text{ionized}]/[\text{nonionized}]$. It may be calculated that the drug is only 0.1 percent ionized at pH 1. The intracellular pH of most cells is about 7. Assuming the pH of the mucosal cell to be the same, it may be calculated that salicylic acid will be 99.9 percent ionized within the cells. Since the concentration of the nonionized form is theoretically the same in both gastric juice and mucosal cells, it follows that the total concentration of the drug (ionized + nonionized) within the mucosal cell will be 1000 times greater than that in gastric juice. Such a relatively high intracellular concentration can have important osmotic and toxicological consequences. If stomach antacid or proton pump inhibitor drug, or even ingested food materials raised the stomach content pH to 5, the lumen-to-cell concentration would only be 100 times greater, an order of magnitude difference in the “concentration differential.”

Had the drug been a weak base instead of an acid, the high concentration would have been in the gastric juice. In the small intestine, where the pH can range from 7.5 to 8.1, the partition of a weak acid or base will be the reverse of that in the stomach, but the concentration differential will be lower, because the pH differential from intestinal lumen to mucosal cells, etc., will be lower. The reversal of partition as the drug moves from the stomach to the small intestine accounts for the phenomenon that some drugs can be absorbed from one GI segment and returned to another. The weak base atropine is absorbed from the small intestine, but because of pH partition, it is secreted into the gastric juice.

The pH partition of drugs has never been demonstrated to be as marked as that illustrated in Figure 39-2 and in the text. Not only do many drug ions probably pass through the pores of the membrane to a significant extent, but some may also pass through the lipid phase. Furthermore, formation of ion pairs in carrier transport also bypasses nonionic diffusion. All processes that tend toward an equal distribution of drugs across membranes and among compartments will cause further deviations from theoretical predictions of pH partition.

ELECTROCHEMICAL AND DONNAN DISTRIBUTION

A drug ion may be distributed passively across a membrane in accordance with the membrane potential, the charge on the

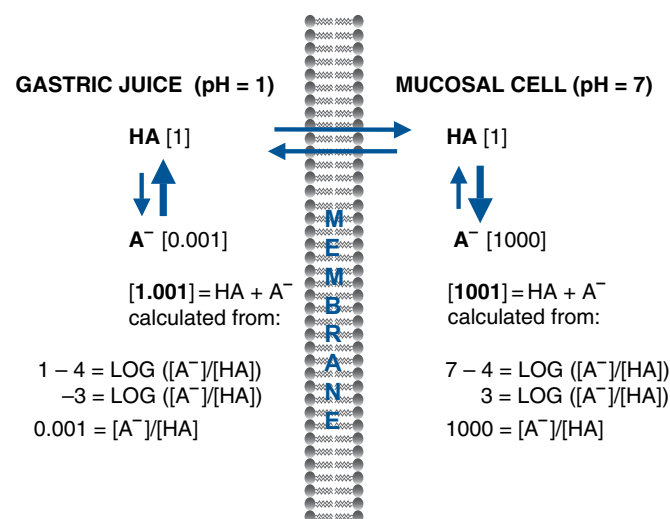


Figure 39-2. A scheme illustrating and calculating (using the Henderson–Hasselbalch equation) the extent of ion trapping between two body compartments, and for the compartments shown, how it facilitates the absorption of weak acids in the stomach. HA, weak acid; A⁻, ionic form of weak acid.

drug ion, and the Donnan effect. The relationship of the membrane potential to the passive distribution of ions is expressed quantitatively by the Nernst equation and has already been discussed. Barring active transport, pH partition, and binding, the drug will be said to be distributed according to the electrical gradient or to its equilibrium potential. If the membrane potential is 90 mV, the concentration of a univalent cation will be 30 times as high within the cell as without; if the drug cation is divalent, the ratio will be 890. The distribution of anions would be just the reverse. If the membrane potential is but 9 mV, the ratio for a univalent cation will be only 1.4 and for a divalent cation only 2.0. It thus can be seen how important membrane potential may be to the distribution of ionized drugs.

Large potentials derive from active transport of ions but small potentials may result from Donnan distribution. According to the Donnan membrane theory, the ratio of intracellular to extracellular concentration of a permeant univalent anion is equal to the ratio of extracellular to intracellular concentration of a permeant univalent cation. A more general mathematical expression that includes ions of any valence is

$$\left(\frac{A_i}{A_e}\right)^{1/Z_a} = \left(\frac{C_e}{C_i}\right)^{1/Z_c} = r$$

where A_i is the intracellular and A_e the extracellular concentration of anion, Z_c is the valence of cation, Z_a is the valence of anion, C_i is the intracellular and C_e the extracellular concentration of cation, and r is the Donnan factor. The value of r depends upon the average molecular weight and valence of the macromolecules (mostly protein) within the cell, and the intracellular and extracellular volumes. Since the macromolecules within the cell are charged negatively, the cation concentration will be higher within the cell; that is, $C_i > C_e$. Since a Donnan distribution results in a membrane potential, the distribution of drug ion also will be in keeping with the membrane potential.

The Donnan distribution also applies to the distribution of a charged drug between the plasma and interstitial compartment, because of the presence of anionic proteins in the plasma. The same equation applies, but requires changing the subscript i to p (for plasma), and e to i , (for interstitial). The Donnan factor, r , for plasma-to-interstitial space partition is about 1.05:1.

BINDING AND STORAGE

Drugs are frequently bound to plasma proteins (especially albumin), interstitial substances, intracellular constituents, and bone and cartilage. If binding is extensive and firm, it will have a considerable impact upon the distribution, excretion, and sojourn of the drug in the body. Obviously, a drug that is bound to a protein or any other macromolecule will not pass through the membrane in the bound form; only the unbound form can negotiate among the various compartments.

The partition among compartments is determined by the binding capacity and binding constant in each compartment. As long as the binding capacity exceeds the quantity of drug in the compartment, the following equation generally applies:

$$\log D_b = \log K + a \log D_f$$

where D_b is the concentration of bound drug, D_f is the concentration of free drug, and a and K are constants characteristic of the drug and binding macromolecule. The equation is that of a Freundlich isotherm. As the binding capacity is approached, the relationship no longer holds. For a non-dissociable drug at equilibrium, D_f will be the same in all communicating compartments, so that it would be possible to calculate the partition if K and a are known for each compartment. Except for plasma, the values of K and a are generally unknown, but the percentage bound is often known.

From the percentage bound, the partition also can be calculated. A drug 50 percent bound in one compartment [free (1) = bound (1); total = 2] and 90 percent bound in a second compartment [free (1), bound (9); total = 10] partitions 2:10, since only free (unbound) drug can pass through the membrane. However, the logarithmic relationships shown serve as a reminder that the percentage bound changes with the concentration, so that the partition will vary with the dose. If the drug is a weak acid or base, the nonionized free form negotiates among the compartments, but the ionized form is often the more firmly bound, and calculations must take into account the dissociation constant and the different K and a values of the ionized and nonionized forms.

It is commonly misbelieved that binding in the plasma interferes with the activity of a drug, and that the intracellular binding in a responsive cell increases activity or toxicity. Both binding in plasma and in the tissues decreases the concentration of free drug, but this is easily remedied by adjusting the dose to give a sufficient concentration for pharmacological activity. The distribution and activity of the free form are not affected by binding. The principal effect of binding is to increase the initial dose requirement for the drug and create a reservoir of drug from which the drug can be withdrawn as the free form is excreted or metabolized. However, if the binding is extremely firm and release is slow, the rate of release may not be sufficient to sustain the free form at a level sufficient for pharmacological activity; in such instances the bound drug cannot be considered a reserve.

The effect of binding upon the sojourn of a drug may be considerable. For example, quinacrine, which may be concentrated in the liver to as much as several thousand times the concentration in plasma, may remain in the body for months. Some iodine-containing, radiopaque diagnostic agents are bound strongly to plasma protein and may remain in the plasma for as long as 2 years. In pathological conditions such as nephrosis, diabetes, or cirrhosis, in which plasma protein levels may be decreased, the plasma protein binding, loading dose, and duration of action all may be decreased.

If a drug is bound to a functional macromolecule, binding may relate to pharmacological activity and toxicity, providing that the binding is at a critical center of the macromolecule. The binding by nucleic acids of certain antimalarials, such as quinacrine, undoubtedly contributes to the parasitocidal actions as well as to toxicity.

Most drugs are bound to proteins by relatively weak forces, such as van der Waals forces (from London, Keesom, and/or Debye forces), or hydrogen or ionic bonds. Consequently, binding constants generally are small, and binding is usually readily reversible. The larger the molecule, the greater is the van der Waals bonding, so that large drug molecules are more likely to be bound strongly than are small ones.

Just as shape and the nature of functional groups are important to drug-receptor combination, so they also are to binding. Drugs of similar shape and/or chemical affinities may bind at the same sites on a binding protein and hence compete with one another. For example, phenylbutazone displaces warfarin from human plasma albumin, which may cause an increase in the anticoagulant effect of warfarin. Some drugs may also displace protein-bound endogenous constituents. For example, sulfisoxazole displaces bilirubin from plasma proteins; in infants with kernicterus the freed bilirubin floods the central nervous system and causes sometimes fatal toxicity.

Depending on the lipid-to-water partition coefficient, a drug may be taken up into fatty tissue. The ratio of concentration in fat to that in the plasma will not be the same as dictated by the partition coefficient, because of the content of water and nonlipids in adipose tissue, and because electrolytes and other solutes alter the dielectric constant and hence solubilities from those of pure water. Lipoproteins and even non-polar substituents on plasma proteins also take up lipid-soluble molecules, so that solubility in plasma can be considerably higher than that

in water. The relatively high solubility of ether in plasma makes plasma a pool for diethyl ether, the filling of which delays the onset of anesthesia. However, ether and other volatile anesthetics are taken up gradually into the adipose tissue, which acts as a store of the anesthetic. The longer the anesthetic is administered, the greater the store, and the longer it takes for anesthesia to terminate when inhalation has been discontinued.

Another notable substance that is taken up readily into fat is thiopental. Even though there is a high solubility of this barbiturate in fat, the low rate of blood flow in fat limits the rate of uptake. Because the blood flow in the brain is very high, thiopental rapidly enters brain tissue. However, it soon equilibrates with the other tissues, and the brain concentration falls as that in the other tissues (e.g., muscle or liver) increases. As the brain concentration falls, anesthesia ceases. Gradually, the fat accumulates the drug at the expense of other compartments.

NONEQUILIBRIUM AND REDISTRIBUTION

Thus far, the distribution of drugs has been discussed mainly as though equilibrium or steady-state conditions exist after a drug is absorbed and distributed. However, since most drugs are administered at intervals and the body content of drug rises and falls with absorption and biotransformation–excretion, neither a true equilibrium among the body compartments nor a steady state exists.

The term equilibrium is used misleadingly to describe the conditions that exist when the plasma concentration and the concentration in a tissue are equal. But equilibrium between plasma and fat occurs much later than equilibrium between plasma and muscle, so that no true equilibrium really exists among all the compartments. Furthermore, a crossover point (equal concentration) for plasma and a tissue is not necessarily an equilibrium point, because the rates of ingress and egress from the tissue are not necessarily equal when the internal and external concentrations are equal. Recall that there are numerous factors that make for unequal distribution (pH partition, Donnan effect, electrochemical distribution, active transport, binding, etc.).

For instance, the distribution of thiopental discussed earlier continually changes during a period (e.g., 3 hours) of observation. At the end of such a period, the content in fat is still increasing, while that in each of the other compartments is decreasing. This time-dependent shift in partition is called redistribution. Eventually the content in fat will reach a peak, which would represent as nearly a true equilibrium point as could be achieved in the dynamic situation where biotransformation and a slight amount of excretion of the drug is taking place. Once the concentration in the fat has reached its peak, its content will decline in parallel with that in the other tissues, and the partition among the compartments would remain essentially constant. Redistribution, then, takes place only until the concentration in the slowest-filling compartment reaches its peak, so long as the kinetics of elimination are constant.

An index of distribution known as the volume of distribution (amount of drug in the body divided by plasma concentration) is of considerable usefulness in pharmacokinetics but is of limited value in defining the way in which a drug is partitioned in the body.

The word “space” is often used synonymously with volume of distribution. It is employed especially when the distributed substance has a volume of distribution that is essentially identical to a physical real space or body compartment. *N*-Acetyl-4-aminoantipyrine is distributed evenly throughout the total body water and is not bound to proteins or other tissue constituents. Thus, the acetylaminantipyrine space, or volume of distribution, coincides with that of total body water. Inulin, sucrose, sulfate, and a number of other substances essentially are confined to extracellular water, so that an inulin space, for example, measures the extracellular fluid volume. Evans blue is confined to the plasma, so that the Evans blue space is the

plasma volume. Such space measurements with standard space indicators are a necessary part of studies on the distribution of drugs, since it is desirable to compare the volume of distribution of a drug with the physiological spaces.

BIOTRANSFORMATION

Most drugs are acted upon by enzymes in the body and converted to metabolic derivatives called metabolites. The process of conversion is called biotransformation. Metabolites are usually more polar and less lipid soluble than the parent drug because of the introduction of oxygen into the molecule, hydrolysis to yield more highly polar groups, or conjugation with a highly polar substance. As a consequence, metabolites often show less penetration into tissues and less renal tubular resorption than the parent drug, in accordance with the principle of the low penetration of polar and high penetration of lipid-soluble substances. For similar reasons, metabolites, particularly conjugates, are usually less active than the parent drug and often inactive. Even if they are appreciably active, they generally are excreted more rapidly. Therefore, the usual net effect of biotransformation can be said to be one of inactivation or detoxification.

Although there are numerous examples in which biotransformation does result in inactivation, there are also examples in which the parent drug has little or no activity of its own but is converted to a pharmacologically active metabolite: codeine to morphine [by cytochrome P450 2D6 (CYP2D6)] or codeine glucuronide [by UDP-glucuronosyltransferase-2B7 (UGT2B7)] for analgesia; parathion to paraoxon for anticholinesterase activity; chloroguanide to cycloguanil (by CYP2C19) for antimalarial activity; lovastatin to lovastatin β -hydroxyacid for HMG CoA reductase inhibition; irinotecan (CPT-11) to SN38 for topoisomerase I inhibition; and enalapril to enalaprilat for angiotensin-converting enzyme inhibition (the last three by hydrolysis reactions). For cytochrome P450 nomenclature, see next heading, Endoplasmic Reticulum and Microsomal System; for UDP-glucuronosyltransferases nomenclature, see Phase II Conjugation Enzymes in this chapter.

When a delayed or prolonged response to a drug is desired or an unpleasant taste or local reaction is to be avoided, it is a common pharmaceutical practice to prepare an inactive or non-offending precursor, such that the active form can be generated in the body. This practice has been termed drug latency. Chloramphenicol palmitate and the estolates of various steroid hormones are examples of deliberately latentiated drugs. Because inactive metabolites do not always result from biotransformation, the term detoxification should not be used as a synonym for biotransformation.

Biotransformations take place principally in the liver, although the kidney, skeletal muscle, intestine, or even plasma may be important sites of the enzymatic attack on some drugs. Biotransformations in plasma are mostly hydrolytic.

ENDOPLASMIC RETICULUM AND MICROSOMAL SYSTEM

Many biotransformations in the liver occur in the endoplasmic reticulum. The endoplasmic reticulum is a tubular system that courses through the interior of the cell but also appears to communicate with the interstitial space, and its membrane is continuous with the cell membrane. Some of the reticulum is lined with ribonucleoprotein particles, called ribosomes, that are engaged in protein synthesis; this is the rough endoplasmic reticulum. The smooth endoplasmic reticulum lacks such a granular appearance. The smooth endoplasmic reticulum is the most heavily invested with numerous enzymes that biotransform many drugs and some endogenous substances.

When a broken-cell homogenate of the liver is prepared, the reticulum becomes fragmented, and the fragments form vesicular structures called microsomes which can be isolated

from other subcellular organelles by differential centrifugation. Microsomes are separated from cytosol by 60 minutes' centrifugation of the 10000g supernatant of a liver homogenate at 100000g. Although the microsomes are artifacts, it is often the practice to refer to drug metabolism as occurring in microsomes rather than in the endoplasmic reticulum.

The microsomal system is peculiar in that both oxidations, and some reductions, usually require the reducing cofactor, reduced nicotinamide adenine dinucleotide phosphate (NADPH). This is because most microsomal oxidations proceed by way of the introduction of oxygen rather than by dehydrogenation, and NADPH is essential to reduce one of the atoms of molecular oxygen. For the major source of microsomal oxidations, the drug first binds to an oxidized (ferric) cytochrome P450. The ferric heme of the drug–cytochrome P450 complex is then reduced to ferrous by NADPH–cytochrome P450 reductase; the reduced complex then combines with oxygen. With an additional electron from NADPH–cytochrome P450 reductase, the oxidized metabolite is released, a molecule of water is formed from the second oxygen atom of molecular oxygen, and ferric cytochrome P450 is regenerated. Cytochrome P450 is a generic term for a superfamily of enzymes.

The general designation of the cytochromes P450 is CYP followed by a number (the family) and a letter (the subfamily) subdivisions. The classification is based on amino acid sequence homology. To belong to the same family, the homology must be greater than 40 percent, and to the same subfamily it must be greater than 59 percent. The subfamily individual is indicated by a number that is based upon the chronological discovery order, independent of the animal species in which it was found. Thus, for example, humans do not have six forms of CYP2D6; CYP2D1 through CYP2D5 are found in other animal species. The major forms involved in drug metabolism in the human liver are CYP1A1 and CYP1A2, CYP2A6, CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP2D6, CYP2E1, CYP3A4, CYP3A5, and CYP3A7. In concentration, CYP3A forms compose 40 percent of the liver P450, CYP2C forms compose 25 percent, and CYP1A2 about 15 percent. CYP3A forms are involved in the metabolism of about 40 percent of the commonly used drug that undergo oxidative metabolism. In contrast to the balance in “abundance” and “demand” of the CYP3A forms, CYP2D6

represents less than 5 percent of the liver P450 but metabolizes about one-fourth of currently used drugs. Further complicating abundance and demand considerations are the presence of genetic polymorphisms. Polymorphisms are indicated by an asterisk (*) followed by a number greater than 1 (1 is wild type; e.g., CYP2D6*4 is the most common activity-decreasing CYP2D6 polymorphism in Caucasians, resulting in 5–10 percent of the population being “poor metabolizers”). The different isozymes present in humans, together with which drugs they metabolize, are of increasing importance in understanding drug interactions and toxicities, and individual responses to standardized doses.

In addition to cytochrome P450, the endoplasmic reticulum contains non-heme-containing flavoprotein monooxygenases (abbreviated FMO), which are also responsible for the oxidative metabolism of drugs. The mechanism of oxidation differs from that of cytochrome P450, involving a hydroperoxyflavin-oxidizing species, and their substrate selectivity (drugs containing a nucleophilic heteroatom) is much less. Of five of these enzymes present in the human genome, FMO3 is the major human liver form.

DRUG-METABOLIZING ENZYME INDUCTION

Some of the enzymes of the microsomal system are quite easily induced; that is, a drug may increase considerably the activity of the enzyme by increasing the biosynthesis of the enzyme. An increase in the amount of endoplasmic reticulum sometimes occurs concomitantly with enzyme induction. Although most extensively studied by monitoring cytochrome P450 changes, induction is not confined to cytochrome P450s, nor is it confined to enzymes in the endoplasmic reticulum.

At least four major mechanisms of induction have been elucidated for cytochrome P450s in the liver, and additional mechanism(s) are present for some other enzymes. For the cytochrome P450s, the same general scheme applies to all (Figure 39-3). The cytosol contains proteins that have a high affinity for certain inducing agents. When the inducing agent binds to its receptor, there is a conformational change and/or displacement of chaperone proteins that allows the receptor–inducer complex to translocate into the nucleus. Within the nucleus the complex links with an additional nuclear factor, and the ensemble interacts with a regulatory element on the

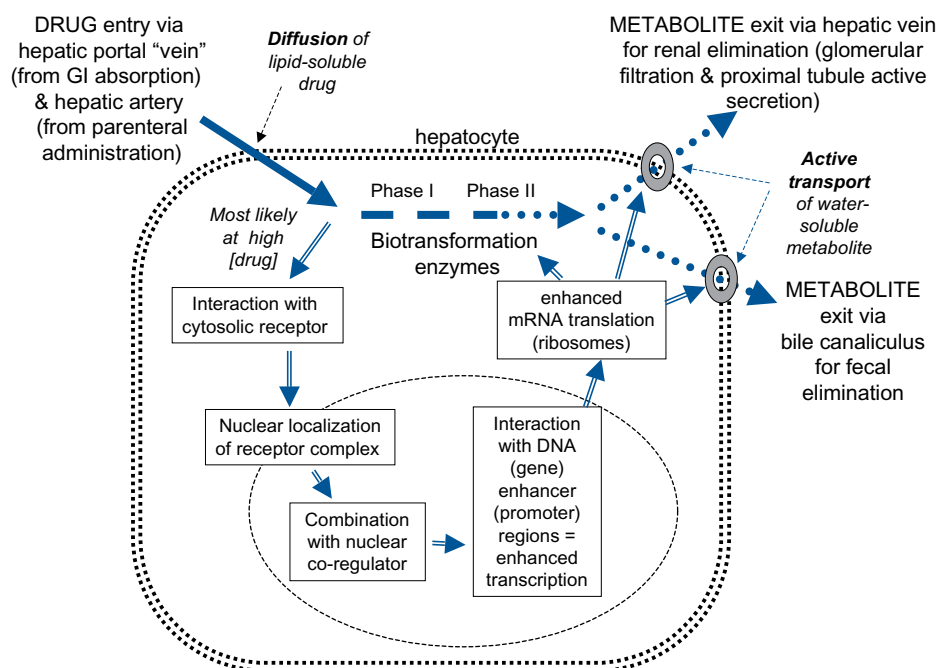


Figure 39-3. The general pathway resulting in hepatic drug-metabolizing enzyme and membrane transporter induction to counteract possible accumulation of drugs and drug metabolites in the hepatocyte.

DNA, causing the assembling of factors and enzymes necessary for RNA transcription at the gene promoter region. The genes of several cytochrome P450s, as well as other enzymes, have identical or closely similar regulatory elements, so the binding of multiple ensembles results in the enhanced production of messenger RNAs from a number of genes and, after translation on the ribosomes attached to the endoplasmic reticulum, an enhanced abundance of drug-metabolizing enzymes. This phenomenon is generally viewed as a self-protective mechanism against the accumulation of excessive amounts of exogenous chemicals (either from high dose or excessively frequent dosing) within the liver. In the event that the liver contains insufficient amounts of enzyme to metabolize exogenous chemicals to promote their excretion, the induction mechanism provides an acute response to remedy the deficiency.

For cytochrome P450s, it has been noted that CYP1A2 is induced preferentially by polycyclic aromatic hydrocarbons and other chemicals contained in cigarette smoke and charcoal-broiled meats, as well as by components in cruciferous vegetables. CYP2A6 is induced by barbiturates, as are CYP2C9 and CYP3A4. CYP2C9, CYP2C19, and CYP3A4 are all induced by rifampicin, but CYP3A4 is additionally induced by many drugs including carbamazepine, phenytoin, glucocorticoids (dexamethasone), clotrimazole, sulfapyrazone, and macrolide antibiotics such as troleandomycin. CYP2E1 can be induced by ethanol and isoniazid. Ironically, given its low percentage abundance in the human liver, there are no known inducers of CYP2D6. The changes elicited by rifampicin and polycyclic aromatic hydrocarbons occur through two different pathways. For rifampicin, the cytosolic receptor is the pregnane X receptor (abbreviated PXR), the associating nuclear factor is the 9-*cis*-retinoic acid-activated receptor (abbreviated RXR), and the regulatory element is termed a pregnane X receptor response element (abbreviated PXRE). For the planar polycyclic aromatic hydrocarbons, the cytosolic receptor is the aryl hydrocarbon receptor (abbreviated AhR), the associating nuclear factor is the aryl hydrocarbon receptor nuclear translocator (abbreviated ARNT), and the regulatory region with its defined base sequence different from that of the PXRE is termed a xenobiotic-response element (abbreviated XRE). Other delineated pathways for cytochrome P450 induction include a pathway responsive to phenobarbital and a pathway responsive to peroxisome proliferators. Enzymes in addition to cytochrome P450s are induced by all these mechanisms, commonly uridine diphospho-glucuronosyltransferases and glutathione *S*-transferases (GSTs). A further and different type of mechanism for induction of non-P450 enzymes exists, a pathway that is responsive to certain chemicals that alter the redox state of the liver cell. Activation of this pathway results in the release of nuclear factor erythroid 2-related factor 2 protein (abbreviated Nrf2) from its Kelch-like ECH-associated protein 1 (abbreviated KEAP1) extranuclear anchor and its interaction with the antioxidant response element (abbreviated ARE) of certain genes, most notably GSTs, epoxide hydrolase, and NADH quinone oxidoreductase.

Regulation of enzyme synthesis can vary with species. Induction via the PXR receptor differs markedly between rodents and human, with rodent PXR almost unresponsive to rifampicin, the prototype PXR agonist in humans.

Treatment of an experimental subject with an inducing agent will often increase the rate of metabolism of the inducing agent. Whether this will necessitate larger and more frequent dosing, or smaller dosing and/or decreased frequency, will depend on whether the metabolism results in pharmacological inactivation (requiring the former) or whether the agent is a prodrug activated by metabolism (requiring the latter).

Induction may also create other therapeutic problems. Given the promiscuity of most drug-metabolizing enzymes, the metabolism of many other drugs may also be affected, producing what are now referred to as drug-drug interactions. For example, the use of phenobarbital or rifampicin during treatment with warfarin increases the dose requirement for warfarin. If the

physician is unaware of this interaction and fails to increase the dose, the patient may suffer a thrombotic episode. If the dose of warfarin has been increased and the phenobarbital or rifampicin is then discontinued, the rate of metabolism of warfarin may drop to its previous level, so that the patient is overdosed, with hemorrhagic consequences. Knowing which enzymes are induced by which drugs, and which drugs are metabolized by a given enzyme or enzymes is now a key component of therapy involving multiple drugs and in medicating patients with multiple conditions, each possibly requiring treatment with a different class of pharmacological agent. The required knowledge base has also been expanded to include nutritional supplements and other plant products that are now known to contain inducing chemicals, in some cases quite potent. Adverse drug-drug interactions have been documented in consumers of St John's wort, which contains the potent PXR agonist, hyperforin.

DRUG-METABOLIZING ENZYME INHIBITION

Some drugs inhibit rather than induce the microsomal enzymes. Whether this results in a need to increase or decrease the doses for adequate therapeutic effect also depends on whether the metabolism inhibited results in drug inactivation or prodrug activation. With multiple drug therapy, as with enzyme induction, enzyme inhibition can result in the altered metabolism of drugs other than that designated as the inhibitor. Where two drugs are in competition for metabolism at the same enzyme, the mechanism of inhibition appears as competitive. In this scenario, the enzyme is not "inhibited" *per se*; it may be performing metabolism at a similar rate as if only a single drug were present, but now the metabolism is split between two compounds. Thus, to the external observer, the metabolism of each appears to be less (i.e., inhibited). Given the promiscuity of drug-metabolizing enzymes, this is a relatively common occurrence; however, whether the effect is of clinical significance will depend on the steepness of the dose-response curve, the relative concentration of the two or more drugs, and the relative inhibitory affinity of the drugs. With *in vitro* kinetic analysis, competitive inhibition will show as an increase in the Michaelis constant, K_m (1/affinity; concentration of drug necessary to elicit half the maximum velocity), with no change in V_{max} . For some drug-enzyme combinations, a drug may not be metabolized by the enzyme it inhibits, but it nevertheless binds and blocks the metabolism of others. An example is quinidine, which potently binds to and inhibits CYP2D6 but is metabolized by CYP3A4. At the extreme, this high-affinity inhibition will manifest itself as non-competitive inhibition, because the inhibitory drug binds with such avidity that other drugs are unable to outcompete or displace it. Non-competitive inhibition is a phenomenon often seen with cytochrome P450, because several chemical entities present in drug molecules have a relatively high affinity for the heme of this enzyme. Perhaps the most notable group of agents in this regard are the antifungal N-substituted imidazoles and triazoles. With *in vitro* kinetic analysis, non-competitive inhibition will show as a decrease in V_{max} , but no change in K_m . Practically, few inhibitory interactions manifest as purely competitive or purely non-competitive. For some interactions, and especially with cytochrome P450s, this lack of purity has a defined mechanistic basis—mechanism-based (suicide) inhibition. During the metabolism of a drug, an intermediate or product is formed that fails to leave the enzyme, often because of an extremely strong affinity for the enzyme (quasi-irreversible), or because of the formation of a covalent bond with protein (irreversible), or for cytochrome P450, the heme moiety. The list of compounds exhibiting suicide inhibition continues to expand. Erythromycin and some related antibiotics were among the earliest drugs identified that demonstrated quasi-irreversible inhibition of cytochrome P450.

Occasionally, an inhibition of the metabolism of drug can arise from depletion of reaction cofactors, or for phase II conjugations (see below Types of Biotransformations), cosubstrates. Sulfate conjugation is particularly prone to this effect,

as cellular content of 3'-phosphoadenosine-5'-phosphosulfate (PAPS) can be limiting, or depleted by high concentrations of drugs undergoing sulfate conjugation. Depletion of hepatocyte glutathione concentration, an effect often seen as a result of exposure to hepatotoxins, affects the ability of the liver to undertake glutathione conjugation, and since this is often the method of sequestration of harmful electrophiles, further exacerbates liver damage. Glucuronidation, on the other hand, rarely suffers from this effect because uridine diphosphoglucuronic acid (UDPGA) is readily generated from plentiful liver glycogen stores.

The activity of the microsomal biotransformation enzymes is affected by many factors other than the presence of drugs. Age, sex, nutritional status, pathological conditions, and genetic factors are among the influences that have been identified. Age in particular has received considerable attention. Neonates and infants have a less well-developed microsomal biotransformation system as compared to adults.

TYPES OF BIOTRANSFORMATIONS

Biotransformations can be degradative, wherein the drug molecule is diminished to a smaller structure, or synthetic, wherein one or more atoms or groups may be added to the molecule. However, it is more useful to categorize biotransformations with respect to metabolic (non-conjugative) biotransformations and conjugative biotransformations. The former is called phase I and the latter phase II. The general map of major pathways of metabolism (by chemical grouping) is shown in Figure 39-4. In phase I, pharmacodynamic activity may be lost; however, active and chemically reactive intermediates may also be generated. The polarity of the molecule may or may not be increased sufficiently to increase excretion markedly. In phase II, metabolites from phase I may be conjugated, and sometimes the original drug may be conjugated, thus bypassing phase I. Phase II frequently generates metabolites of high polarity, which are excreted readily. Rarely do drugs undergo a single phase I and phase II reaction. This has been interpreted to indicate the defensive objective of drug metabolism; however, the body's ability to modify a foreign compound to enhance its elimination will probably be employed. Thus, phase I reactions may occur at more than one site on the drug molecule (not necessarily to

the same extent), and more than one conjugation may occur for the reactive center introduced by phase I metabolism or already present in the parent drug. The former is illustrated with the drug tamoxifen, which undergoes *N*-demethylation at one end of the molecule, and aromatic hydroxylation at the other. This latter is evident for the drug acetaminophen, where the aromatic hydroxyl group can undergo either glucuronidation or sulfation, and both conjugates are found in the urine.

Phase I biotransformations are for convenience often considered in terms of the chemical reaction: oxidation, reduction, and hydrolysis. Phase II biotransformations are usually considered in terms of the endogenous biochemical entity added.

The activity and selectivity of the enzymes of biotransformation vary greatly from species to species, so that care must be exercised in extrapolating experimental findings in laboratory animals to humans.

Phase I Oxidation

Oxidation is a very common type of biotransformation. Oxidations that occur primarily in the liver microsomal system include side-chain (alkyl) hydroxylation; aromatic hydroxylation; deamination (which is oxidative and results in the intermediate formation of RCHO); *N*-, *O*-, and *S*-dealkylation (which involve hydroxylation of the alkyl group followed by oxidation to the aldehyde); and sulfoxide formation. As indicated earlier, these microsomal oxidations are catalyzed by two types of enzymes, most frequently by the hemoprotein cytochrome P450, but also by flavin monooxygenases. The reactions as listed do not divide themselves among various members of the cytochrome P450 superfamily; thus, depending on the drug molecule, a given enzyme (e.g., CYP3A4) might catalyze aromatic hydroxylation of one molecule, *N*-demethylation of another, and *O*-demethylation of a third. It might also catalyze these reactions at different sites on a single drug molecule. A single drug molecule may also be metabolized differently by two different enzymes; the *N*-demethylation of tamoxifen mentioned above is largely catalyzed by CYP3A4; the aromatic hydroxylation of tamoxifen is largely catalyzed by CYP2D6.

The major oxidations that occur elsewhere, other than the microsomes, are generally dehydrogenations with NAD⁺ as the

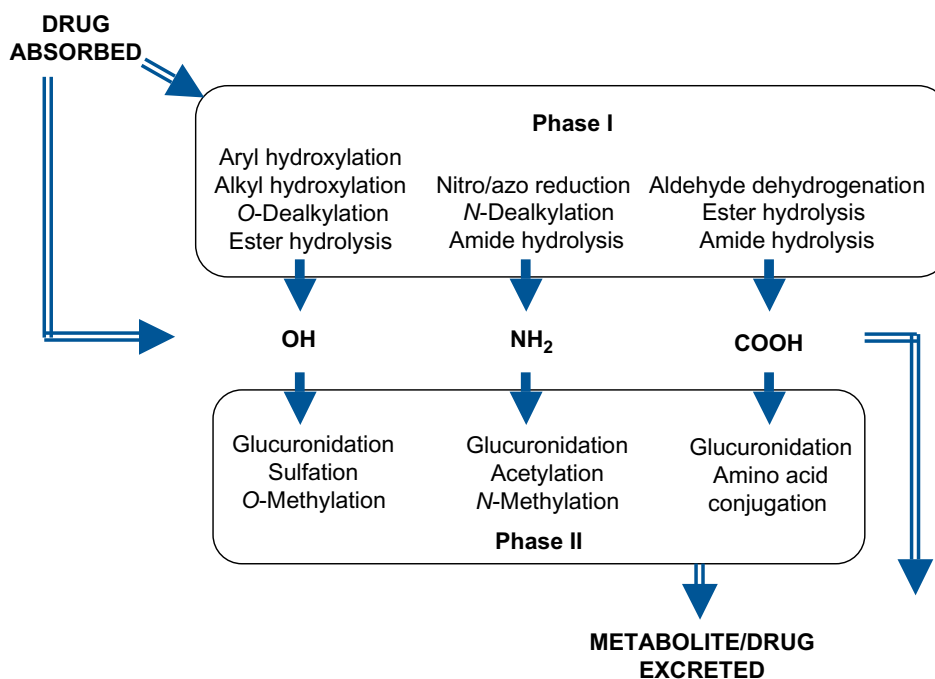


Figure 39-4. The major reactions associated with the formation and conjugation of the major chemical entities during the hepatic metabolism of drugs, and their place within the general disposition of drugs and other xenobiotics.

acceptor. Examples are the oxidation of alcohols by alcohol dehydrogenase, the oxidation of aldehydes by aldehyde dehydrogenase, and the deamination of monoamines by monoamine oxidase and diamines by diamine oxidase. There are numerous alcohol and aldehyde dehydrogenase enzymes, each with some degree of selectivity.

Phase I Reduction

Reduction is a relatively uncommon reaction that occurs mainly in liver microsomes (if at all). However, it occasionally takes place in other tissues. Examples are the reduction of nitro and nitroso groups (as in chloramphenicol, nitroglycerin, and organic nitrites), of the azo group (as in Prontosil, the first commercially available antibiotic), and of certain aldehydes to the corresponding alcohols.

Phase I Hydrolysis

Hydrolysis is a common biotransformation among esters and amides. Esterases are located in many structures besides the microsomes. For example, cholinesterases are found in plasma, erythrocytes, liver, nerve terminals, junctional interstices, and postjunctional structures, and procaine esterases are found in plasma. Many esterases will hydrolyze both esters and amides, albeit the metabolism of amides is generally slowly than that of esters. Various phosphatases and sulfatases are also distributed widely in tissues and plasma, although few drugs are appropriate substrates. The hydrolytic deamidation of meperidine occurs primarily in the hepatic microsomes.

The hydration of epoxides, often generated by cytochrome P450 oxidations to form dihydrodiols, is an important detoxification reaction. For the metabolism of xenobiotic epoxides, two epoxide hydrolases exist, one in the endoplasmic reticulum (microsomal EH) and another in the cytoplasm (soluble EH).

Other biotransformations include desulfuration, in which oxygen may replace sulfur, and dehalogenation. Desulfuration takes place in the liver. Thiopental is converted in part to pentobarbital by desulfuration, and parathion is transformed to paraoxon. Dehalogenation of certain insecticides and various halogenated hydrocarbons may take place, principally in the liver but not in the microsomes.

Phase II Conjugation Chemistry

In considering the outcome of the bulk of phase I reactions, the formation of three chemical entities (reactive centers), hydroxyl ($-\text{OH}$), amine ($-\text{NH}_2$), and carboxyl ($-\text{COOH}$) compose the majority (Figure 39-4.). Aromatic and alkyl oxidation, and *O*-dealkylation and ester hydrolysis, produce metabolites with a hydroxyl group; *N*-demethylation, nitro and azo reduction, and amide hydrolysis produce metabolites with an amine group; aldehyde dehydrogenation, and ester and amide hydrolysis produce metabolites with a carboxyl group. Major conjugations therefore center on the additional metabolism of these functional groups. Conjugation is the biosynthetic process of combining a drug or drug metabolite with a highly polar and water-soluble endogenous biochemical to yield, with a couple of exceptions, an even more water-soluble and easily excreted, usually pharmacologically inactive, product. In almost all conjugations, the endogenous biochemical cosubstrate involved in the reaction is required to be in a "high-energy" form. The exception is conjugation with glutathione (γ -glutamylcysteinyl glycine), where the center of reactivity, the nucleophilic sulfhydryl of cysteine, is sufficiently reactive toward electrophiles that the conjugation can occur directly, and in some cases in the absence of an enzyme. Also unusual with glutathione conjugation is a further metabolism of the conjugate. Through a series of enzymatic reactions, the γ -glutamyl and glycyl residues are removed, the remaining cysteine conjugate is *N*-acetylated, and the product spontaneously dehydrates to form a mercapturic acid, the excreted product.

Glucuronic acid is the most frequent partner added to the drug in conjugation. Actually, the drug reacts with UDPGA rather than with simple glucuronic acid. UDPGA is formed from the NAD^+ -supported oxidation of UDP-glucose, which in turn is formed from glucose-1-phosphate (readily available from the catabolism of liver glycogen stores) and UTP. The drug or drug metabolite combines at the number 1 carbon (aldehyde "end") and not at the carboxyl group of glucuronic acid. The hydroxyl group of an alcohol or a phenol attacks the number 1 carbon of the pyran ring to replace uridine diphosphate. The product is a hemiacetal-like derivative. Since the product is not an ester, the term glucuronide is appropriate. Rarely, thiols and amines may also form analogous glucuronides. Carboxyl compounds form esters, appropriately called glucuronates, in replacing the uridine diphosphate. Some acyl glucuronates are subject to intramolecular acyl migration, and some of these products are noted for their toxicity. The greater instability of acyl glucuronates as compared to ether glucuronides in alkaline conditions requires some care in their detection and quantification from biological matrices.

The alternative conjugation of carboxylic acids with an amino acid, usually glycine but also glutamine, is an amidation conjugative process. An example is benzoic acid, which is conjugated with glycine to form hippuric acid. In amino acid conjugation, the drug or drug metabolite is first adenylated (by reaction with ATP) and then converted to its coenzyme A derivative before conjugation with the amino acid.

Sulfuric acid is also a frequent conjugant, especially with phenols and to a lesser extent with simple alcohols. The cosubstrate for this reaction is PAPS. The sulfated product is called an ethereal sulfate. Occasionally sulfuric acid conjugates with aromatic amines to form sulfamates. Phosphoric acid also conjugates with phenols and aromatic amines.

Aromatic amines and occasionally aliphatic amines or heterocyclic nitrogens frequently are acetylated. Acetyl coenzyme A is the biological reagent rather than acetic acid itself. Unlike most other conjugates, the acetylate (amide) is usually less water soluble than the parent compound. The acetylation of the *p*-amino group of the sulfonamides is a prime example of this type of conjugation. Many amines, especially derivatives of β -phenylethylamine and heterocyclic compounds, are methylated in the body. The products are usually biologically active, sometimes more so than the parent compound. *N*-Methylation may occur in the cytoplasm of the liver and elsewhere, especially in chromaffin tissue in the case of phenylethylamines. Phenolic compounds may be *O*-Methylated. *O*-Methylation is the principal route of biotransformation of catecholamines such as epinephrine and norepinephrine, the methyl group being introduced on the *meta*-hydroxy substituent. Both *N*- and *O*-methylation require *S*-adenosylmethionine as the cosubstrate.

Phase II Conjugation Enzymes

Although many conjugations occur in the liver, some occur in the kidney or in other tissues. All the drug conjugation reactions are catalyzed by specialized enzymes, termed transferases, and all are present in multiple forms. Glucuronidation is catalyzed by UDP-glucuronosyltransferases (UGTs), located in the endoplasmic reticulum. UGTs are classified in two major classes (UGT1As and UGT2Bs) based on amino acid homology, but the two classes also differ in gene structure and substrate selectivity. UGT1As preferentially react with planar drugs, and UGT2Bs preferentially react with three-dimensionally bulkier molecules. Whereas the various UGT2Bs exist as separated six-exon genes in the genome, the UGT1As exist as a nested complex and the individual enzymes are formed from transcription and translation of a combination of one of several first exons to a common set of four exons. In both UGT1As and UGT2Bs, the amino-terminal region of the enzyme is responsible for drug (the aglycone) recognition, while the carboxy-terminal end is involved in binding UDPGA and as such appears to reside on

the cytoplasmic side of the endoplasmic reticulum. As with cytochrome P450s, these enzymes are inducible, and the two classes differ in their response to various drugs and other chemicals.

Sulfation is catalyzed by sulfotransferases (SULTs), located in the cytoplasm. The many isozymes exhibit substrate selectivity, and some differ in thermal stability. Unlike most major drug-metabolizing enzymes, SULTs do not appear to respond to drug-mediated induction.

Glutathione conjugations are catalyzed by GSTs, also located in the cytoplasm. The multiple isozymes involved in xenobiotic or drug metabolite metabolism exist in four major classes: alpha, mu, pi, and theta. The isozymes have relatively low substrate (electrophile) selectivity.

Acetylation is catalyzed by cytoplasmic *N*-acetyltransferases, NAT1, and in the liver, NAT2. NAT2 exhibits a genetic polymorphism, resulting in "fast" and "slow" acetylator phenotypes with differing incidences in various populations (slow acetylators are more common in Middle Eastern, less common in Asian populations). The acetylation reaction proceeds through an acetylated enzyme intermediate. Methylation reactions are catalyzed by cytoplasmic *O*-, *N*-, and *S*-methyltransferases, and each of these exists in multiple forms.

EXCRETION

Although some drugs are not biotransformed in the body, most that are readily absorbed require modification for efficient elimination. Although the kidney is the most important organ of excretion, some substances are excreted in bile, sweat, saliva, or gastric juice, or from the lungs. Urine analysis will often reveal the presence of unchanged drug in addition to metabolites, but for the majority of drugs they constitute only a minor fraction.

RENAL EXCRETION

The excretory unit of the kidney is called the nephron. There are several million nephrons in the human kidney. The nephron is essentially a filter funnel called Bowman's capsule, with a long stem called a renal tubule. The final collecting duct is also functionally a part of the nephron. The blood vessels that invest the capsule and the tubule are also an essential part of the nephron.

Bowman's capsule is packed with a tuft of branching interconnected capillaries that provide a large surface area of capillary endothelium through which fluid and small molecules can filter into the capsule and begin passage down the tubule. The capillary network, together with Bowman's capsule, constitutes the glomerulus. The glomerular capillary endothelium and the supporting layer of Bowman's capsule have channels ranging upward to 40 Å in diameter. Consequently, all unbound solutes in plasma pass or are forced by pressure into the glomerular filtrate.

The postglomerular blood vessels, which lie close to the tubules, are critically important to renal function in that substances resorbed from the filtrate by the tubule are returned to the blood along these vessels. The tubule first makes a number of convolutions (called a proximal convoluted tubule), then courses down and back up a long loop (called the loop of Henle), makes more convolutions (the distal convoluted tubule), and finally joins the collecting duct. The loop of Henle is divided into a proximal (descending) tubule, a thin segment, and a distal (ascending) tubule.

As the glomerular filtrate passes through the proximal tubule, some solute may be resorbed (tubular resorption) through the tubular epithelium and returned to the blood. Resorption occurs in part by passive diffusion and in part by active transport, especially with sodium and glucose. Chloride follows sodium obligatorily.

In the proximal region the tubule is quite permeable to water, so that resorbed solutes are accompanied by enough water to keep the resorbate isotonic. Consequently, although the filtrate

becomes diminished in volume by approximately 80 percent in the proximal tubule, it is not concentrated.

Some acidification occurs in the proximal tubule as the result of carbonic anhydrase activity in the tubule cells and the diffusion of hydronium ions into the lumen. In the lumen the hydronium ion reacts with bicarbonate ion, which is converted to resorbable nonionic CO₂.

There is also active transport of organic cations and anions into the lumen (tubular secretion), each by a separate system. These active-transport systems are extremely important in the excretion of a number of drugs; for example, penicillin G is secreted rapidly by the anion transport system, and tetraethylammonium ion by the cation transport system. Probenecid is an inhibitor of anion secretion and, hence, decreases the rate of loss of penicillin from the body.

Tubular secretion can be considered to occur in two phases: movement from the blood into the tubule cell and from the cell into the urine. There are active-transport systems at both tubule cell surfaces. Transporters moving compounds into the cell are predominantly members of the solute carrier protein family. Different transporters have differing substrate preferences, but for drug elimination they can be broadly divided into organic-anion transporters and organic-cation transporters. It should be noted that glucuronic acid and sulfate conjugates are organic anions existing almost entirely in the ionized form at physiological pH values, so in addition to increasing water solubility, metabolism has made the drug more suitable for transporter elimination. The affinity of the transporters for their substrates is generally higher than the affinity of the drug or metabolite for albumin, so in contrast to glomerular filtration in the Bowman's capsule, in the proximal tubule, drugs and metabolites can be stripped from albumin for elimination. On the lumen (urine) cell surface, these transporters are augmented by transporters of the ATP-binding cassette (ABC) family. Structurally, these are similar to the solute carrier proteins in having many (around 12) transmembrane-spanning regions that within the membrane can form a cylindrical pore through which the transported molecule travels across the lipid bilayer membrane. As the name implies, however, among the cytoplasmic loops linking the transmembrane-spanning regions of ABC transporters are adenosine triphosphatase sites (two), and these transporters use the energy released in ATP hydrolysis to provide conformational changes that can, if necessary, drive the movement of a molecule *up* a chemical concentration gradient (i.e., from a low to a high concentration). The two major classes of ABC transporters for drugs and drug metabolites are multidrug resistance proteins (MDRs; MDR1 is also known as P-glycoprotein and abbreviated Pgp), the products of *ABCB* genes; and multidrug resistance related protein (MRP), the product of *ABCC* genes. The MRPs, particularly MRP2 (also known as the canalicular multispecific organic-anion transporter, abbreviated cMOAT), with their preference for anions, are those involved in the forced elimination of glucuronic acid and sulfate conjugates.

As the filtrate travels through the thin segment it becomes concentrated (especially at the bottom), as a result of active resorption and a countercurrent-distribution effect enabled by the recurrent and parallel arrangement of the ascending segment, the parallel orientation of the collecting duct, and the similar recurrent geometry of the associated capillaries. In the thick segment of the ascending loop of Henle, both sodium and chloride are transported actively. In the distal tubule, sodium resorption occurs in exchange for partly potassium (potassium secretion) and partly for hydronium ions. Adrenal mineralocorticoids promote distal tubular sodium resorption, and potassium and hydronium secretion. Ammonia secretion also occurs, so that the urine either may be acidified or alkalinized, according to acid-base and electrolyte requirements. Water is resorbed selectively from the distal end of the distal convoluted tubule and the collecting ducts; water resorption is under the control of the antidiuretic hormone. The collecting duct also

resorbs sodium, secretes potassium, and acidifies and concentrates the urine.

Drugs may also be resorbed in the distal tubule; the pH of the urine there is extremely important in determining the rate of resorption, in accordance with the principle of nonionic diffusion and pH partition. Amines will tend toward a greater degree of ionization the more acid the urine, so less of the amine present will be in the nonionized form and be able to diffuse back through the lipid bilayer from the urine to blood (i.e., trapped in urine and thereby destined for elimination). The effect of urinary pH on the excretion of a weak acid is the opposite: the more acid the urine, the more will be in the nonionized form and be able to diffuse back into the blood.

The urinary pH and, hence, drug excretion may fluctuate widely according to the diet, exercise, drugs, time of day, and other factors. Obviously, the excretion of weak acids and bases can be controlled partly with acidifying or alkalinizing salts, such as ammonium chloride or sodium bicarbonate, respectively. Comparative studies on potency and efficacy in humans have demonstrated the importance of controlling urinary pH. Urinary pH is important only when the drug in question is a weak acid or base of which a significant fraction is excreted. The plasma levels will change inversely to the excretory rate. For example, it has been shown clinically with quinidine that alkalinization of the urine not only decreases the urine concentration but also increases the plasma concentration and toxicity.

BILIARY EXCRETION AND FECAL ELIMINATION

Many drugs are secreted into the bile and then pass into the intestine. For metabolites generated within the liver cell, secretion

into bile canaliculus is an alternative to secreting them into the bloodstream for subsequent renal elimination. A drug that is passed into the intestine via the bile may be reabsorbed and not lost from the body. A drug conjugate entering the intestine may be deconjugated by enzymes and the parent drug reabsorbed. This cycle of biliary secretion and intestinal resorption is called enterohepatic circulation. The biliary secretory (transporter) systems greatly resemble those of the kidney tubules. The enterohepatic system may provide a considerable reservoir for a drug. If a drug is not absorbed completely from the intestine, the unabsorbed fraction will be eliminated in the feces. An unabsorbable drug that is secreted into the bile will likewise be eliminated in the feces. Such fecal elimination is called fecal excretion. Only rarely are drugs secreted into the intestine through the succus entericus (intestinal secretions), although a number of amines are secreted into gastric juice.

ALVEOLAR EXCRETION

The large alveolar area and high blood flow make the lungs ideal for the excretion of appropriate substances. Only volatile liquids or gases are eliminated from the lungs. Gaseous and volatile anesthetics essentially are eliminated completely by this route. Only a small amount of ethanol is eliminated by the lungs, but the concentration in the alveolar air is related so constantly to the blood alcohol concentration that the analysis of expired air is acceptable for legal purposes. The high aqueous solubility and relatively low vapor pressure of ethanol at body temperature account for the retention of most of the substance in the blood. Carbon dioxide from those drugs that are partly degraded also is excreted in the lungs.

Pharmacokinetics/Pharmacodynamics in Drug Development

George L. Drusano, MD

PHASE II CLINICAL TRIAL VALIDATION OF DOSE FOR
USE IN PHASE II TRIALS 755

CREATING A PHARMACODYNAMIC
RELATIONSHIP 755

PHASE II CLINICAL TRIAL VALIDATION OF DOSE FOR
USE IN PHASE III TRIALS 761

PHASE II CLINICAL TRIAL VALIDATION OF DOSE FOR USE IN PHASE II TRIALS

Drug development is a long, expensive process of discovery and preclinical development followed by clinical trials resulting in the submission of a package of data to a regulatory agency that will ultimately lead to licensure of that product for sale.

The goal of drug development is to find a dose of a drug for a specific indication that attains the desired therapeutic outcome while engendering a low probability of the patient experiencing a toxic event. Pharmacokinetics and pharmacodynamics can straightforwardly lead to attaining this goal. Indeed, in the last one to two decades there has been a marked increase in our understanding of the relationship between drug exposure and response. This is related to wider availability of the appropriate mathematical modeling methodologies. The application of these techniques in the timeline of drug development is presented in Figure 40-1.

The clearest example of employing a pharmacokinetic/pharmacodynamic approach to drug development can be seen in the area of anti-infective agents. Part of the reason for this is that these drugs are unique in that we are not attempting to dock a molecule into a receptor in the human body. Rather, the target of drug action and the site to which we are attempting to bind the drug is a receptor in the pathogen of interest. This has several important consequences.

The first is toxicity. Anti-infective targets are chosen specifically so that they have little sequence homology to similar mammalian targets. A straightforward example is the topoisomerase enzymes seen in bacteria but also in man. The fluoroquinolone antimicrobials have a 100- to 1000-fold difference in the concentrations necessary for microbiological effect relative to activity for topoisomerase targets in man.¹ In contrast, there is often a narrow therapeutic index, for example, between normal human cells and cancerous cells, meaning that oncologic chemotherapy is often (but not always) saddled with considerable toxicity.

The other important consequence is the ease with which pharmacodynamic relationships can be developed both preclinically as well as in clinical trials. The reason is that almost always (Hepatitis C is currently an exception to the rule) one can straightforwardly grow the pathogen of interest *in vitro* and determine a measure of drug exposure that will affect the growth of the pathogen in some standardized way. For example, for viruses, we can measure an EC_{50} , a drug concentration that will cause a 50% downturn in the number of rounds of replication per unit time. For bacteria, we can measure indices such as MICs or MBCs, that are defined as drug concentrations that will keep the bacteria from growing enough over an 18- to 24-hour period to cause turbidity in the growth medium (MIC) or to cause the number of bacteria to be reduced by 1000-fold over the 18–24 hour timeframe (MBC). This ability to measure the difficulty a drug will encounter inhibiting or killing different

pathogens allows the drug exposure necessary to achieve different endpoints to be normalized across pathogens. In contrast, if one were to try to develop an antihypertensive agent, the true between-patient variability in the affinity with which a drug will bind to the receptor cannot currently be measured. Certainly, in the near future, the widespread use of pharmacogenomic profiling, looking, for example, for specific single nucleotide polymorphisms (SNPs) or deletions will allow identification of patients likely to respond less well to therapy. Currently, however, this true between-patient variance in receptor affinity is completely unobserved variability.

CREATING A PHARMACODYNAMIC RELATIONSHIP

MEASURE OF DRUG EXPOSURE

The process of creating a pharmacodynamic relationship starts with the idea of linking some measure of drug exposure to the outcome of interest. There are a number of measures of drug exposure that can be employed. Some of the most common are Peak concentration, Area Under the concentration-time Curve (AUC) and Time > Threshold. Many other measures of drug exposure are possible (e.g., trough concentrations), but they are usually related in some way to those mentioned above.

The critical idea behind which of these metrics is most closely linked to a specific measure of outcome is that the shape of the concentration-time curve may have an impact on the outcome measure. For example, for agents where the range of concentration that mediates minimal effect to that which mediates maximal effect is small, Time > Threshold will be the most useful metric for exposure. That is because higher concentrations will not produce significantly more effect than moderate concentrations. The overall effect will then be maximized by maintaining the drug concentration above the level that produces the degree of effect that is required. In contrast, many drugs are quite concentration-dependent in the effect that they produce. Here, much more effect will be produced at higher concentrations with much less being produced as concentrations decline. This will produce a situation where the total drug exposure will be linked to effect and AUC will be the most useful measure of drug exposure. Peak concentrations may be seen as linked to effect when an irreversible event occurs, such as covalent binding to a receptor that only occurs above a specific concentration. Here, only peak concentrations will produce enough exposure to have the binding occur in the appropriate timeframe. This is the rarest situation seen in the development of pharmacodynamic relationships. Peak concentrations can also appear to be linked to outcome when there is a mixture of populations of sensitive and less sensitive targets present. This will be discussed in greater detail below under the topic of suppression of emergence of resistance.

For the development of such relationships for anti-infective agents, the exposure measures are normalized to the measure

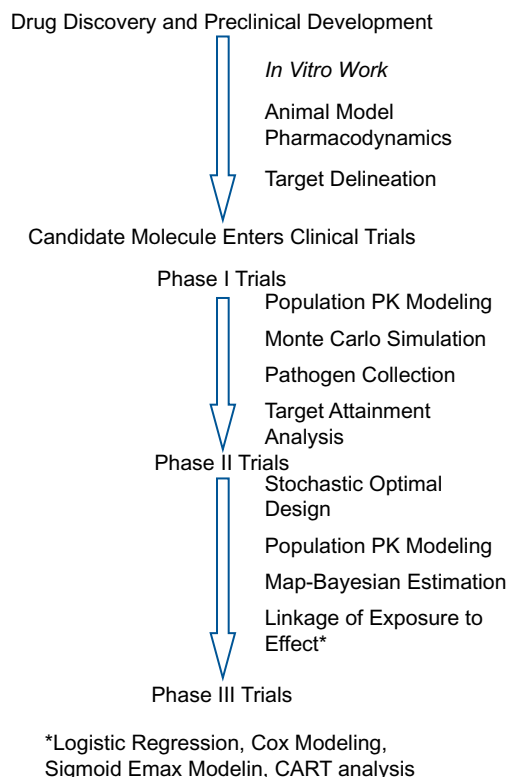


Figure 40-1. Use of pharmacodynamics in the drug development process.

of susceptibility of the pathogen to the drug being studied (e.g., MIC, EC_{50}). This produces a hybrid measure that explicitly depends on the drug exposure, but also on the pathogen being studied. So we can now measure Time > MIC, AUC/MIC ratio, or Peak Concentration/MIC ratio.

As an example of the shape of the curve having an impact on the effect developed by drug exposure, the β -lactam antibiotic imipenem/cilastatin was studied in a neutropenic mouse thigh infection model by Flückiger, Segessenmann, and Gerber.² The actual idea that the shape of the curve can affect the endpoint measured was popularized by the laboratory of Craig,³⁻⁵ but arguably the clearest demonstration was by Flückiger.² In this study, the effect of a dose of drug was determined on the number of organisms present at the primary infection site. In parallel, a second cohort of animals received the same drug dose, but on a highly fractionated basis, so that the resultant concentration-time curve had a much lower peak concentration, but remained above the MIC for a much longer time interval. The AUC/MIC ratios developed in the two cohorts were nearly identical. The results are shown in Figure 40-2. The number of organisms killed was much greater when the Time > MIC was longer. This indicates that there was no benefit derived from the high peak concentrations developed in the first group and that keeping the drug concentrations in excess of the MIC was the effect driver in this circumstance.

An example of exactly the opposite linkage can be seen with fluoroquinolone antimicrobials, as well as other agents like aminoglycosides, or the new antiMRSA agent, daptomycin. For these drugs, there is a clear relationship between drug concentration and the rate of organism kill that is engendered. In this circumstance, AUC/MIC ratio is the exposure variable most closely linked to outcome.

In a study by Louie *et al.*,⁶ a mouse thigh infection model study was performed using methodology similar to that described above, but without the massive dose fractionation. An exposure-response curve was described (Figure 40-3). On the steep part of the curve Q 24 hour, Q 12 hour, and Q 6 hour administration schedules were studied, so that the 24-hour AUC

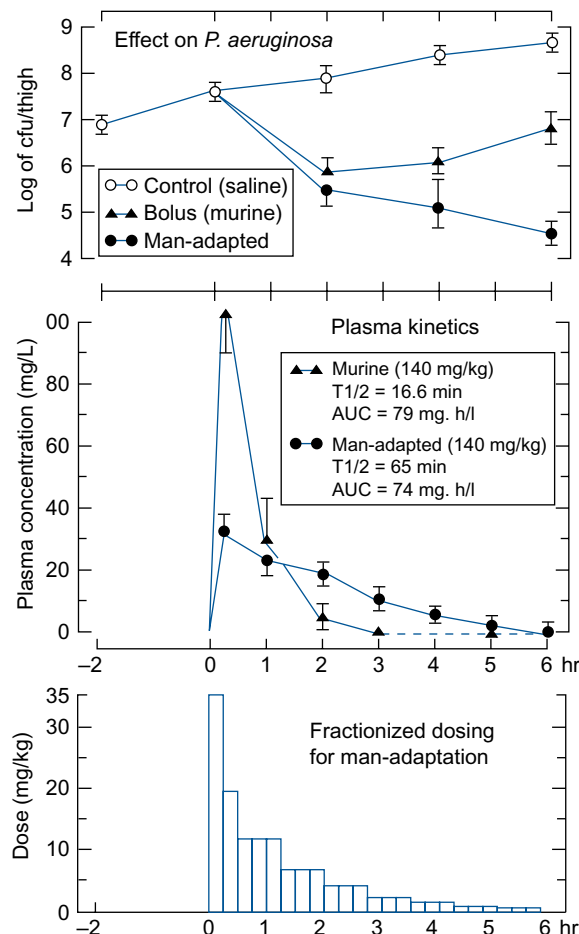


Figure 40-2. Effect of the shape of the concentration-time curve on the ability of imipenemcilastatin to kill *P. aeruginosa* in a mouse thigh infection model. (From Flückiger U, Segessenmann C, Gerber AU. *Antimicrob Agents Chemother* 1991; 35: 1905.)

was the same for each group (same AUC/MIC ratio), but that the once daily dosing group had the highest Peak concentrations (and hence Peak concentration/MIC ratio) while the Q 6 hour dosing group attained the longest Time MIC. When the results for these groups were tested for differences by analysis of variance, no differences could be discerned (Table 40-1). This indicates that for daptomycin, AUC/MIC ratio is the exposure variable most closely linked to outcome.

Table 40-1. *Staphylococcus Aureus* Densities in Thigh Muscles of Mice That Were Treated with Various Doses of Daptomycin, Administered in One, Two, or Four Divided Doses

***Staphylococcus aureus* phylococcus aureus Densities ($\log_{10}$ (CFU/g) + 1 Standard Deviation) with:**

Total Dosage (mg/kg)	1 Dose	2 Divided doses ^b	4 Divided doses ^c	P value ^a
2.5	6.54 ± 0.98	6.83 ± 0.88	6.61 ± 0.93	0.64
5.6	5.12 ± 0.66	4.96 ± 0.59	5.02 ± 0.52	0.73
15.0	3.73 ± 0.48	3.82 ± 0.55	3.68 ± 0.43	0.59

^aStatistical testing was performed by analysis of variance. A *P* value of 0.05 was considered statistically significant.

^bOne-half of the single dose was administered at 0 hour and then 12 hours later.

^cOne-quarter the single dose was administered at 0 hour and then 6, 12, and 18 hours later.

CHOOSING THE TARGET

In the examples given above, a microbiological endpoint was chosen. This is because we were dealing with animal model systems. It is important to recognize that in the drug development process, there will be a progression from a preclinical stage to the performance of clinical trials to document safety and efficacy.

Prior to the inception of clinical trials, animal model and *in vitro* systems are required for the determination of the true pharmacodynamically linked effect variable (e.g., AUC/MIC ratio, Time MIC). After the start of clinical trials, we can examine either clinical or microbiological endpoints. These endpoints can be dichotomous in nature (succeed/fail, eradication/persistence) or can be continuous (microbiological quantitative repetitive sampling from an infection site). The target that is chosen in either of these circumstances will determine, to a great degree, the dose that is required to attain the target.

In the preclinical circumstance, one can examine the ability of a regimen (dose and schedule) to kill organisms at a primary infection site (Figures 40-2 and 40-3). Alternatively, one could examine a mortality endpoint. In either circumstance, the model system could be developed so that the animal was either normal or rendered immunocompromised. In each instance, this will change the interpretation of the endpoint chosen and, in the case of an immunocompromised animal system, will require a larger drug exposure to achieve whatever endpoint is desired. The reason to perform these studies in immunocompromised animals is twofold. First, it is a conservative measure of the drug exposure required to achieve whatever target is desired. It is a more direct measure of “bug versus drug.” Second, and perhaps as important, it is much easier to find strains of pathogens that will grow in whatever model is being used. In the presence of the full immune system of the animal, many pathogens will self-clear over the period of observation in the no-treatment control group. This renders the interpretation of the experiment much more difficult.

It is also important to recognize that the animal system being employed should accurately reflect the local pharmacokinetics of the drug for the indication being sought. While a mouse thigh infection model may accurately represent the ability of the drug to kill organisms in a skin/skin structure infection, the lessons learned and exposure targets derived would not be helpful if the drug were going to be studied for a meningitis indication in clinical trials.

This also raises the issue of what endpoint will be chosen. Do we wish merely to shut off organism growth or to kill the organism at the primary infection site to a specific degree (e.g., 1 log₁₀ (CFU/g) kill, 2 log₁₀ (CFU/g) kill, 80 percent of maximal kill). Figure 40-3 demonstrates that an AUC/MIC ratio of 43.4 is required to shut off growth of the organism, whereas attaining 80 percent of the maximal bacterial kill for this strain of *Staphylococcus aureus* requires a larger exposure, with an AUC/MIC ratio of 92.4. For a relatively uncomplicated skin and skin structure infection, attaining an exposure target that will result in organism stasis is likely all that is required. However, for a complicated skin and skin structure infection, particularly if bacteremia would be likely, an exposure target that would drive some high percentage of the maximal bacterial kill (80–90 percent of maximal kill) would be more appropriate. This choice of target is crucial to the successful choice of an appropriate drug dose for clinical trial.

SUPPRESSION OF EMERGENCE OF RESISTANCE AS AN ENDPOINT

Until recently, little has been done with suppression of emergence of resistance as an endpoint for the choice of drug dose. The key idea underlying the problem of emergence of resistance is the population burden of bacterial cells at an infection site relative to the inverse of the mutational frequency to resistance. If the population burden exceeds the inverse of the

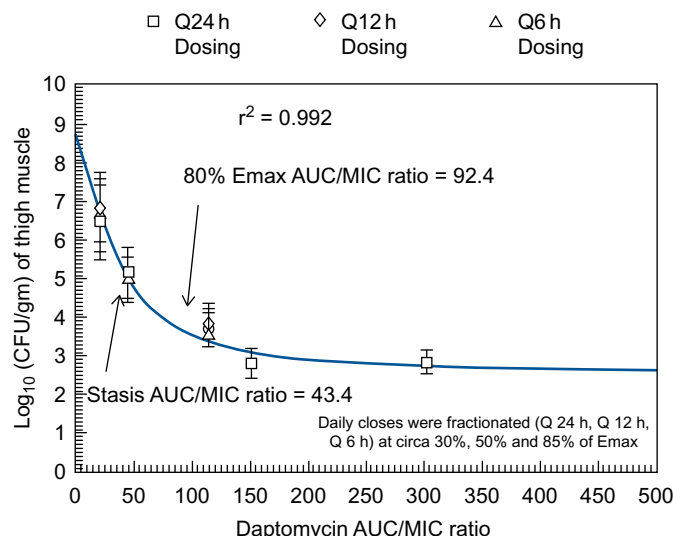


Figure 40-3. Relationship between the 24-h AUC/MIC ratio of daptomycin and log₁₀ CFU of *S. aureus* per gram of thigh muscle (mean $\pm$ 1 SD) when the total daily dose of daptomycin is given as one dose in 24h, two equally divided doses every 12h, or four equally divided doses every 6h. The total daily doses of 2.5, 5.6, and 15.0mg of daptomycin/kg resulted in AUC/MIC ratios of approximately 21, 44, and 115, respectively. The AUC/MIC ratio for a given total daily dose was similar regardless of whether a total daily dose was administered as one, two, or four equally divided doses over 24h. (From Louie A, Kaw P, Liu W *et al. Antimicrob Agents Chemother* 2001; 45: 845.)

mutation frequency, there will be a high probability (but not a certainty) that organisms bearing a resistance mechanism will already be present. The larger the burden relative to the frequency, the higher will be the probability. This means there will be multiple populations of organisms present at the time that drug therapy is initiated. It is not surprising that the population bearing a resistance mechanism will respond quite differently to the pressure of drug therapy than will the population without this mechanism. Again, as above, the example used will be from the anti-infective literature, but the lessons are clearly applicable in the realm of oncolytic chemotherapy.

In Figure 40-4 Panels A–D, the effect of different doses (including a no-treatment control) of the fluoroquinolone levofloxacin on the total and resistant populations of *Pseudomonas aeruginosa* is displayed.⁷ As the effect was determined at multiple time points, it allowed the effect of the drug concentrations on the two populations to be modeled simultaneously for all regimens.⁷ This allowed a calculation of dose to maximally amplify the resistant population as well as a dose to hold the number of clones in the resistant population steady. These doses were then studied prospectively over a longer time frame (24 versus 48 hours) as a validation of the modeling result. This is displayed in Figure 40-5.

This is the first prospective validation that a target drug exposure can be derived to suppress the amplification of mutant subpopulations. This makes the point that the choice of the exposure target is flexible and is a choice that should be made with great care if the drug is to achieve the hoped-for results when it enters clinical trials. If the toxicity profile of the drug allows it, a suppression of resistance endpoint may be wise, as it will extend the lifetime of the drug.

CHOOSING A DRUG DOSE FROM PRECLINICAL PLUS PHASE I DATA

Once the appropriate animal or *in vitro* models have been chosen and the drug studied in these systems, the linked pharmacodynamic variable will have been elucidated and the exposure target(s) identified for the indications that are to be sought for the drug. A critical issue is how to employ the pharmacodynamic

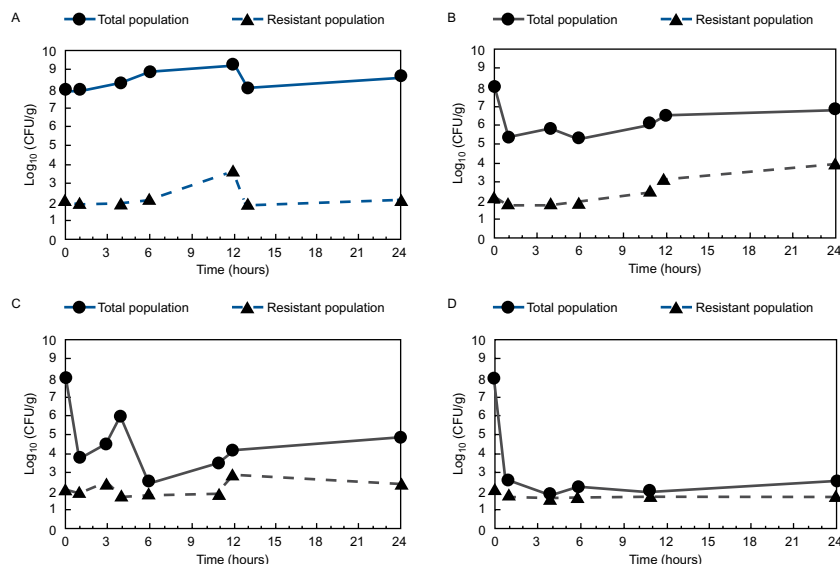


Figure 40-4. Effect of 4 drug doses of levofloxacin on the total and resistant bacterial populations of *Pseudomonas aeruginosa* over 24 hours. Drug doses were 0, 90, 215, and 600 mg/kg (Figure 40-3, panels A–D, respectively). The 90 mg/kg dose allowed amplification of the resistant population by almost $2 \text{ log}_{10} (\text{CFU/g})$. The 215 mg/kg dose allowed only minimal resistant mutant amplification. (From Jumbe N, Louie A, Leary R *et al. J Clin Invest* 2003; 112(2): 275. Reproduced with permission from the American Society for Clinical Investigation.)

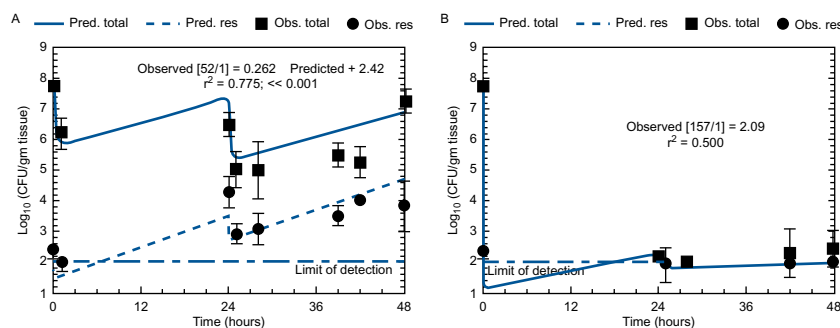


Figure 40-5. Model validation. The emergence of resistance model developed in this study was prospectively evaluated and validated by generating response predictions for doses not previously studied that would (A) encourage selection of resistance or (B) suppress emergence of resistance. An exposure of an AUC/MIC ratio of 157/1 was calculated to prevent emergence of resistance. Experiments were performed to 48, not 24 hours as in the studies performed to generate parameter estimates, using model predicted conditions. Levofloxacin dosing occurred at time 0 and at 24 hours. The lines are model predictions (not best-fit curves). (÷) represents experimental measurements of the total population. (•) represents experimental measurements of the resistant subpopulation. The model predicted changes in the resistant mutant population well at both exposures. (From Jumbe N, Louie A, Leary R *et al. J Clin Invest* 2003; 112(2): 275. Reproduced with permission from the American Society for Clinical Investigation.)

information to allow the “correct” dose of drug to be chosen for clinical trial(s).

The ability of a specific dose of drug to attain the desired exposure target is influenced by a small number of factors. These are (i) the true between-patient variability in drug pharmacokinetics; (ii) the variability in the MIC of the drug for the pathogens of interest (i.e., those pathogens that are likely causes of the diseases for which indications are being sought); (iii) the protein binding of the agent being studied. Each of these factors will be examined in turn.

PHARMACOKINETIC VARIABILITY

True between-patient variability in pharmacokinetics clearly exists. The same dose of drug will produce very different drug concentration profiles if given to a large number of patients. We can account for some portion of this variance by measuring covariates such as age, sex, weight, height, creatinine clearance, etc. However, even after examining these and other covariates, there will still be true, residual between-patient variance. Accounting for this variability plays an important role in determining the correct dose for clinical trials. The basic idea is that

whatever dose is chosen, it should be adequate to attain the desired exposure target in a very high proportion of the population of interest.

It is important to obtain an idea of the degree of variability of the drug exposure achieved when a fixed dose of drug is administered. Preston *et al.*⁸ studied 272 patients with community-acquired infections who received the fluoroquinolone antibiotic levofloxacin. In this study, all patients had a serum creatinine that was $<2.0 \text{ mg/dL}$. Figure 40-6 shows the variability in AUC achieved in these patients with a fixed dose of drug. In this relatively normal community-based population there was a >10 -fold range of AUCs observed.

It is unusual to have such a rich data set early on in drug development. The usual reality is to have anywhere between 12 and 60 volunteers who have had the drug's pharmacokinetics studied. These study subjects represent a very biased estimator of how the drug will be handled in the population of interest, most of whom will be older and, by definition, sicker than the normal volunteer population. Nevertheless, if these data can be employed to choose a drug dose, it will almost certainly be a conservative choice.

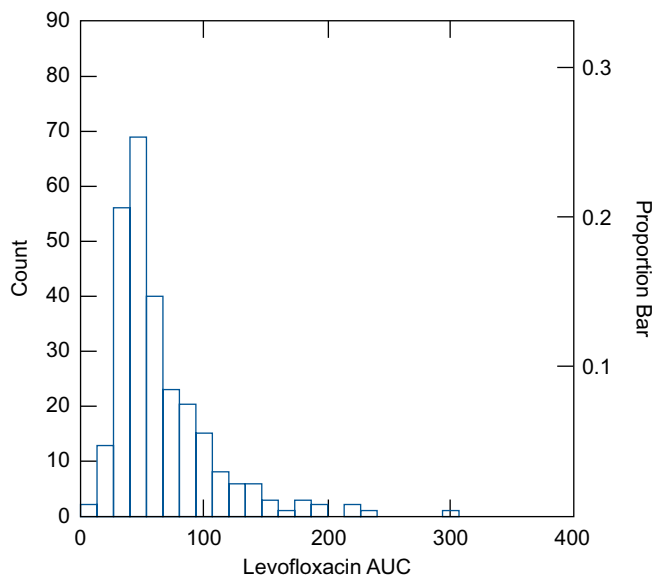


Figure 40-6. The distribution of area under the concentration–time curve (AUC) values for 272 patients receiving 500 mg of levofloxacin for the therapy of community-acquired infections.

Monte Carlo simulation is a mathematical technique that allows prior knowledge of the central tendency of a parameter and its distribution to be employed to set up a sampling distribution. That is, we can take a large number of samples (e.g., 1000, 10000, 30000) from the “known” prior distribution of parameter values. These can be used to calculate Peak concentrations, AUC values, or Time > Threshold. It provides the opportunity to perform large clinical trials “*in silico*.” These measures of drug exposure then can be employed to determine how often a specific drug dose will produce an exposure that will attain the exposure target value for a specific MIC.

MICROBIOLOGICAL VARIABILITY

Usually the pathogen(s) of interest for a specific indication have a broad range of values. It is obvious that for whatever range of drug exposures are achieved by a specific dose, it will be more difficult to achieve the exposure target for an organism with a higher MIC value.

Luckily, the determination of the range of MIC values for organisms that are causative pathogens for the indications being studied is relatively straightforward to obtain.

PROTEIN BINDING

As a rule, only non-protein bound drug is active. This has been most clearly seen in two studies, one of bacteria⁹ and one of HIV.¹⁰

Merriken *et al.*⁹ examined a group of isoxazoly penicillins for their activity against *Staphylococcus aureus*. There were seven molecules studied. Each was chosen because it had the same MIC for the challenge strain of *Staphylococcus* and had very similar pharmacokinetics, but had very different protein binding that ranged from 30% bound to >97% bound. When examined in a mouse model, the effect of the drug was related to the free fraction of the drug concentration (Figure 40-7).

For HIV, Bilello *et al.*¹⁰ examined a highly bound (ca 90%) HIV-1 protease inhibitor and examined the impact of protein binding *in vitro*. In a transitive logic set of experiments, the impact in free fraction of the major drug binding protein for this agent (α -1 acid glycoprotein) was determined. In Figure 40-8A, it is demonstrated that increasing the amount of binding protein over the physiologic range of 0.5–1.5 mg/mL decreases the free fraction in a quantitative manner. It was also demonstrated (Figure 40-8B) that lower free amounts of drug were associated with less cell-associated drug. Finally, the lower cell-associated

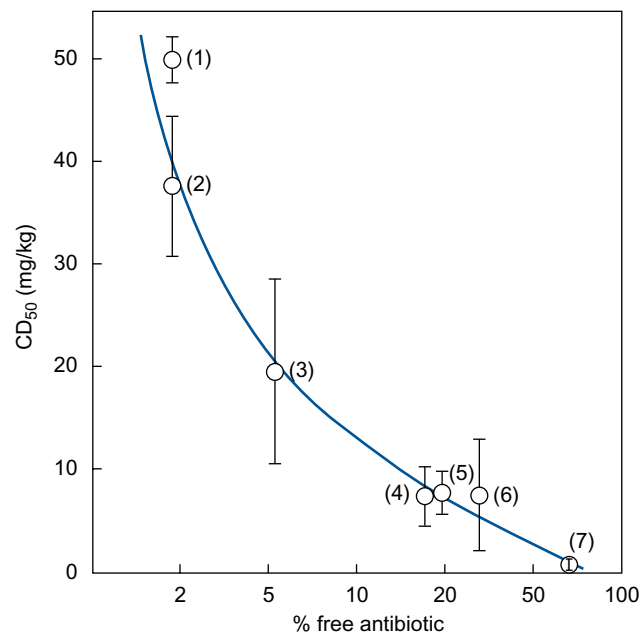


Figure 40-7. The effect of protein binding on the microbiological activity of seven isoxazoly penicillins as determined in a mouse model of staphylococcal intraperitoneal infection. (From Merriken DJ, Briant J, Rolinson GN. *J Antimicrob Chemother* 1983; 11:233.)

amounts of drug were shown to produce less antiviral effect, in a quantitative fashion (Figure 40-8C).

Clearly, then, protein binding has a major, quantitative effect on drug effect, particularly in microbiological systems. There are instances when protein binding has a less than anticipated impact on effect or where it appears that there is actually no effect of binding on activity. It is likely (but unproven) that in these instances drug is taken onto its binding effect site from its protein-binding site because of a major difference in K_d for the drug for the two receptors. Even so, when developing a drug and picking an effect target, it is wise to understand the drug binding characteristics and to ascertain the impact of drug binding on the effect that is desired. This is particularly important preclinically, as there may be considerably different binding seen in animal versus man.

INTEGRATING THE SOURCES OF VARIABILITY

Once we have chosen an exposure target, studied pharmacokinetic variability, MIC (or $EC_{50/95}$) variability and the impact of protein binding, the question remains as to how to use this preclinical information to choose dose(s) for clinical trials.

Our group^{11–14} developed the use of Monte Carlo simulation to integrate these disparate sources of variability. In the paradigm developed, a population pharmacokinetic study is performed. The measure of central tendency (usually but not necessarily the mean parameter values) and dispersion (full or major diagonal covariance matrix) are employed to generate a Monte Carlo simulation for specific drug doses. These generated values are then corrected for protein binding, as a function of the measured impact of binding on effect. Then the fraction of the simulated population that attains the desired exposure target is determined for different values of MIC (or $EC_{50/95}$). In so doing, clear breakpoint values for pathogen drug susceptibility can be determined for a specific dose of drug. Because we have information regarding the distribution of MIC (or $EC_{50/95}$) values, the overall response of the population can be determined by taking a weighted average over the product of the range of target attainment rates and MIC (or $EC_{50/95}$) values. An example of this technique validated with a prospective study was performed with an HIV-1 non-nucleoside reverse transcriptase inhibitor [NNRTI].¹⁴ The target agreed upon was keeping trough

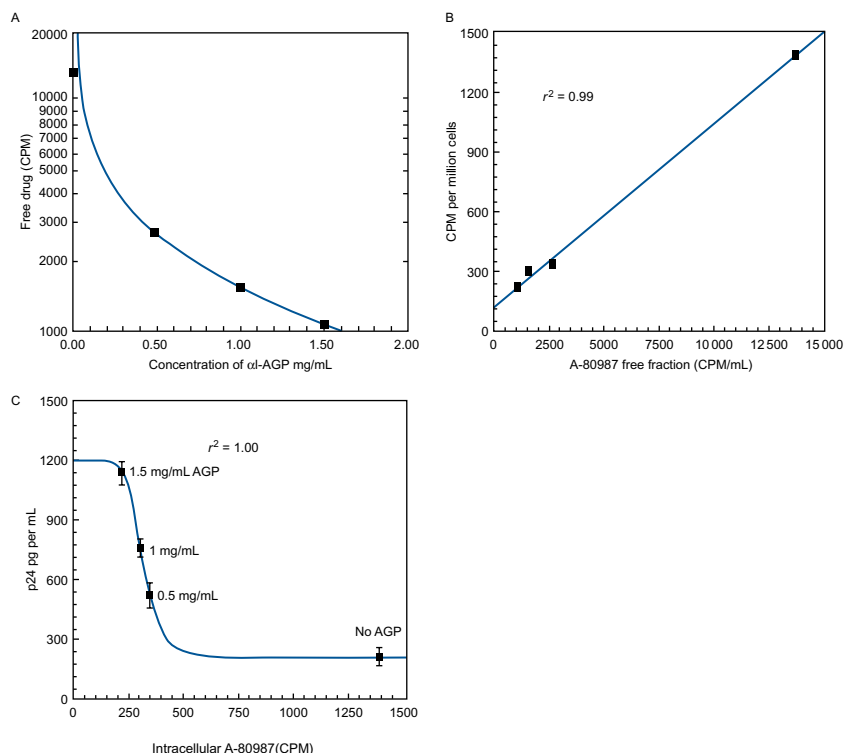


Figure 40-8. Effect of protein binding on the antiviral activity of a HIV-1 protease inhibitor. In (A), increasing amounts of the major binding protein (α -1 acid glycoprotein) results in a decreasing amount of unbound drug. In (B), the decrease in unbound drug is shown to be associated with decreased cellular penetration. In (C), it is shown that decreased amounts of intracellular drug are associated with decreased antiviral effect. Together, these experiments demonstrate that protein binding has a major impact on virological activity. (From Merriken DJ, Briant J, Rolinson GN. *J Antimicrob Chemother* 1983; 11:233.)

free drug concentrations above the EC_{90} of HIV. Preclinical study demonstrated that addition of purified binding proteins to the medium increased the EC_{50} value by 7.6-fold. In addition, other preclinical studies demonstrated that there was approximately a 10-fold change in drug concentration between the EC_{50} and the EC_{90} . Therefore, a 7.6-fold adjustment was

made to the simulated trough concentrations of the drug for doses of 50 mg, 100 mg, and 200 mg. Finally, preliminary preclinical data indicated that different wild-type HIV isolates all had EC_{50} values less than 10 nM. In Figure 40-9A, the target attainment rate by EC_{50} values (not determined in the presence of binding proteins) is displayed. It is clear that if the EC_{50} value

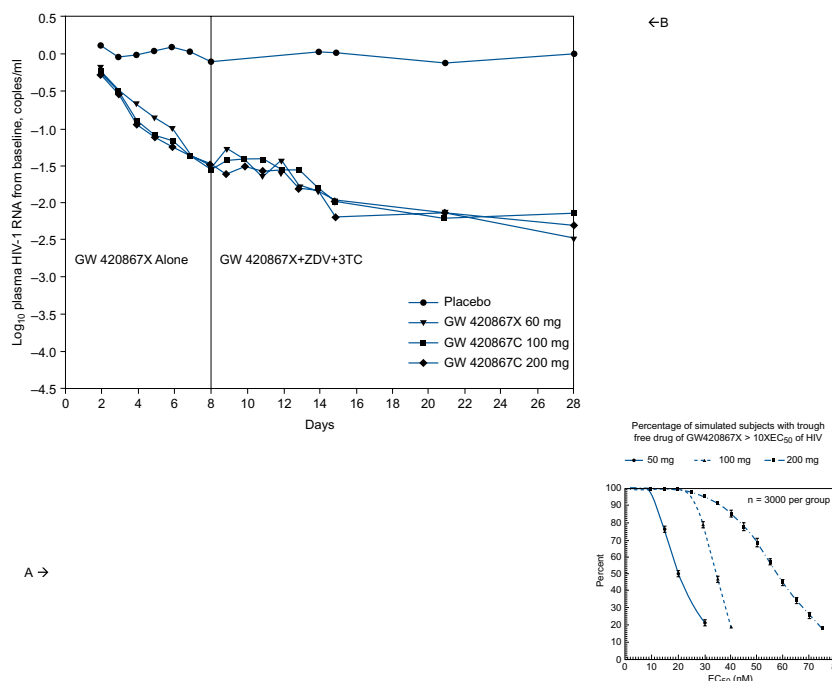


Figure 40-9. In (A), three different doses of an experimental nonnucleoside reverse transcriptase inhibitor keep free drug in excess of the EC_{90} of HIV for the whole dosing interval to the same extent, as long as the EC_{50} is less than 10 nM. As is demonstrated in (B), this leads to equipotent antiretroviral activity in a Phase I/II randomized, double-blind clinical trial.

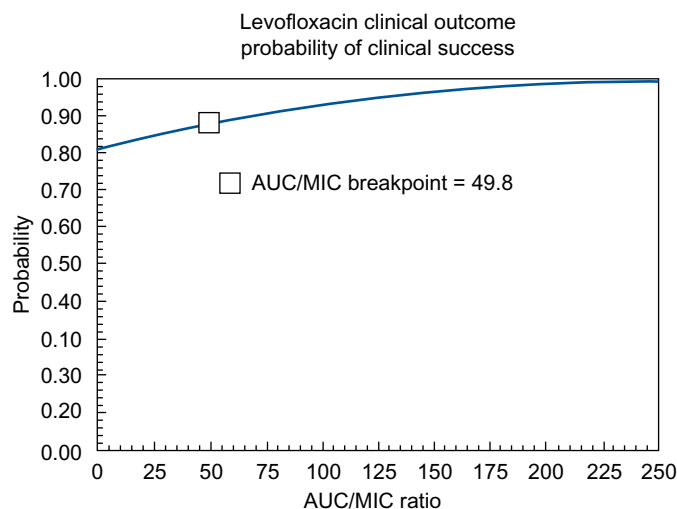


Figure 40-10. The relationship between the AUC/MIC ratio of levofloxacin and the probability of a good clinical outcome in 134 patients receiving this fluoroquinolone for community-acquired infections. Classification and regression tree (CART) analysis identified a breakpoint of a total drug AUC/MIC ratio of 50 (free drug AUC/MIC ratio of 35).

does not exceed 10 nM, there will be no difference in the target attainment rates for the three doses evaluated.

This analysis resulted in the prediction that no difference in viral load decline would be observed in a clinical trial. This agent was examined in a 60 patient Phase I/II randomized, double-blind trial in which 15 HIV-infected patients were each given one of the doses examined in the Monte Carlo simulation, and 15 patients served as a no-treatment control in this short term trial. The results are displayed in Figure 40-9B. On the morning of day 8 of the trial (the end of monotherapy), the median viral load decline ranged from 1.48 to 1.52 $\log_{10}$ (copies/mL), prospectively validating the predictions.

Other Monte Carlo simulation-based trial predictions have also been prospectively validated.^{12,13} This technique where the sources of variability are quantified and integrated is a useful guide to determining drug doses for evaluation in the Phase I/II environment.

PHASE II CLINICAL TRIAL VALIDATION OF DOSE FOR USE IN PHASE III TRIALS

Given the extensive number of patients studied and the possibility for harm if an incorrect dose is chosen for Phase III clinical trial evaluation, it is important to validate the dose(s) chosen by the use of Monte Carlo simulation. In order to do this, it is critical to pay attention to the same sources of variability as in the integration of the preclinical information.

It is important, therefore, to have an index of drug exposure for each patient participating in the analysis, a pathogen identified, and an MIC determined to the drug being employed. Finally, it is important to decide upon an endpoint (e.g., clinical success/failure, organism eradication/persistence).

The first prospective, multicentered trial of this type was published by Preston and colleagues.¹⁵ Patients were enrolled in 22 centers for the therapy of community-acquired infections (respiratory tract infections, skin and skin structure infections, and urinary tract infections) with a fluoroquinolone antimicrobial (levofloxacin). A sampling scheme was derived using a stochastic D-optimal sampling technique. There were 272 patients who had pharmacokinetic data collected. These data were analyzed employing a nonparametric population modeling technique (NPME II program of Schumitzky and Jelliffe). Individual estimates of exposure were calculated for each patient by obtaining patient-specific pharmacokinetic parameter values employing Maximum *A-posteriori* Probability (MAP) Bayesian

parameter estimation. Of these patients, there were 134 patients with a documented outcome and identified pathogen that had a levofloxacin MIC. For clinical outcome, both Peak/MIC ratio as well as AUC/MIC ratio could be linked to outcome. As there has been considerable preclinical data linking AUC/MIC ratio to fluoroquinolone effect, this relationship will be presented in Figure 40-10. The breakpoint AUC/MIC value, determined by Classification and Regression Tree (CART) analysis, was 49.8. This was for total drug. The free drug value would be 34.9. A mouse thigh infection model developed in our laboratory examined levofloxacin for the therapy of *Streptococcus pneumoniae*.¹⁶ The value for an organism kill of 1 $\log_{10}$ (cfu/g) was 29.4 for a free drug AUC/MIC ratio. The free drug AUC/MIC value associated with a good outcome determined in a separate clinical study by Ambrose *et al.* only for *Streptococcus pneumoniae* for two fluoroquinolones (levofloxacin and gatifloxacin) was 27.2–33.7.¹⁷ All these determinations are in excellent concordance.

What is clear is that it is possible to identify targets for desired drug action preclinically and to bridge between animal and man employing Monte Carlo simulation techniques. Further, it is also clear that these findings are robust. They have been validated in clinical trials. The paradigm for clinical validation is simple. It is set forth in Table 40-2. It is important to obtain good individual-patient estimates of their pharmacokinetic parameter values in order to perform clinical pharmacodynamic analysis. While population modeling allows good estimates of population mean parameter values, the precision with which the values are determined for an individual patient after the MAP-Bayesian step depend explicitly on how much information is present in the samples that have been obtained for *that* patient. This problem can be solved without undue patient invasion (i.e., minimizing the numbers of samples) by employing stochastic optimal design techniques.^{18–21} Once the patients have been studied, the population values are best attained using population modeling techniques, with patient-specific values determined through MAP-Bayesian estimation. Exposure variables can then be normalized (in the case of anti-infective agents) to some measure of the degree of susceptibility of that patient's pathogen to the drug in question (Peak/MIC ratio, AUC/MIC ratio or Time MIC). These normalized exposure variables can then be linked to the probability of a good outcome (clinical/microbiological) through use of logistic regression. Breakpoints can be sought through use of CART analysis. If the outcome is a time-to-event, Kaplan–Meier analysis (when a breakpoint is available) or Cox proportional hazards analysis (for a continuous variable) can be employed. If the outcome is a continuous variable (e.g., viral load determination), some variant of a sigmoid *E*_{max} effect model can be employed to link exposure to effect.

Table 40-2. Paradigm for the Development of Exposure–Response Relationships

1. Decide on an endpoint
2. Make potency measurements of the cells to be inhibited/killed (MIC/EC₅₀, etc.)
3. Obtain drug exposure estimates for patients in these trials
 - a. Stochastic optimal sampling design
 - b. Population pharmacokinetic modeling
 - c. MAP-Bayesian parameter determinations for individual patient exposure estimates
4. Decide on an analytical tool for endpoint analysis (examples only)
 - a. Sigmoid *E*_{max} analysis for a continuous endpoint
 - b. Logistic regression for dichotomous/polytomous outcomes
 - c. Cox proportional hazards modeling (or a fully parametric variant) for time-to-event data
 - d. Classification and regression tree (CART) analysis for breakpoint determination

Determination of a pharmacodynamically linked variable preclinically with an exposure target combined with a target attainment analysis from Phase I/II data will allow identification of a dose for Phase III trials. Validation of this outcome in a (relatively) small Phase II trial will provide confidence that the dose chosen for large, Phase III clinical trial investigation is optimal for the effect target desired. This will maximize the speed of drug development and minimize drug failure.

REFERENCES

1. Albertini S *et al.* Genotoxicity of 17 gyrase-and four mammalian topoisomerase II-poisons in prokaryotic and eukaryotic test systems. *Mutagenesis* 1995; 10: 343–51.
2. Flückiger U *et al.* Integration of pharmacokinetics and pharmacodynamics of imipenem in a human-adapted mouse model. *Antimicrob Agents Chemother* 1991; 35: 1905–10.
3. Gerber AU *et al.* Selection of aminoglycoside-resistant variants of *Pseudomonas aeruginosa* in an in vivo model. *J Infect Dis* 1982; 146: 691–7.
4. Vogelman B *et al.* Correlation of antimicrobial pharmacokinetic parameters with therapeutic efficacy in an animal model. *J Infect Dis* 1988; 158: 831–47.
5. Leggett JE *et al.* Comparative antibiotic dose-effect relations at several dosing intervals in murine pneumonitis and thigh-infection models. *J Infect Dis* 1989; 159: 281–92.
6. Louie A *et al.* Pharmacodynamics of daptomycin in a murine thigh model of *Staphylococcus aureus* infection. *Antimicrob Agents Chemother* 2001; 45: 845–51.
7. Jumbe N *et al.* Application of a mathematical model to prevent in vivo amplification of antibiotic-resistant bacterial populations during therapy. *J Clin Invest* 2003; 112: 275–85.
8. Preston SL *et al.* Levofloxacin population pharmacokinetics and creation of a demographic model for prediction of individual drug clearance in patients with serious community-acquired infection. *Antimicrob Agents Chemother* 1998; 42: 1098–104.
9. Merriken DJ, Briant J, Rolinson GN. Effect of protein binding on antibiotic activity in vivo. *J Antimicrob Chemother* 1983; 11: 233–8.
10. Bilello JA *et al.* Human serum alpha 1 acid glycoprotein reduces uptake, intracellular concentration, and antiviral activity of A-80987, an inhibitor of the human immunodeficiency virus type 1 protease. *Antimicrob Agents Chemother* 1996; 40: 1491–7.
11. FDA Anti-Infective Drug Products Advisory Committee. October 15, 1998.
12. Drusano GL *et al.* Use of preclinical data for selection of a phase II/III dose for evernimicin and identification of a preclinical MIC breakpoint. *Antimicrob Agents Chemother* 2001; 45: 13–22.
13. Drusano GL *et al.* Hollow-fiber unit evaluation of a new human immunodeficiency virus type 1 protease inhibitor, BMS-232632, for determination of the linked pharmacodynamics variable. *J Infect Dis* 2001; 183: 1126–9.
14. Drusano GL *et al.* Rational dose selection for a nonnucleoside reverse transcriptase inhibitor through use of population pharmacokinetic modeling and Monte Carlo simulation. *Antimicrob Agents Chemother* 2002; 46: 913–6.
15. Preston SL *et al.* Pharmacodynamics of levofloxacin: a new paradigm for early clinical trials. *J Am Med Assoc* 1998; 279: 125–9.
16. Tumbe NL *et al.* 40th Inter-science Conference on Antimicrobial Agents and Chemotherapy, Abstract 291, 2000.
17. Ambrose PG *et al.* Pharmacodynamics of fluoroquinolones against *Streptococcus pneumoniae* in patients with community-acquired respiratory tract infections. *Antimicrob Agents Chemother* 2001; 45: 2793–7.
18. Drusano GL *et al.* An evaluation of optimal sampling strategy and adaptive study design. *Clin Pharmacol Therapeut* 1988; 44: 232–8.
19. Drusano GL *et al.* A prospective evaluation of optimal sampling theory in the determination of the steady-state pharmacokinetics of piperacillin in febrile neutropenic cancer patients. *Clin Pharmacol Therapeut* 1989; 45: 635–41.
20. Drusano GL *et al.* Optimal sampling theory: effect of error in a nominal parameter value on bias and precision of parameter estimation. *J Clin Pharmacol* 1994; 34: 967–74.
21. Tam VH, Preston SL, Drusano GL. Optimal sampling schedule design for populations of patients. *Antimicrob Agents Chemother* 2003; 47: 2888–91.

- AAG see α -acid-glycoproteins
 Abbreviated New Drug Application (ANDA) 97, 487
 aberrant observation rejections, 162
 abnormal osmoticity, 281
 abraxane, 661
 absolute density, 113
 absolute filters, 241
 absorption
 see also absorption, distribution...;
 drug absorption
 administration routes, 742
 adjuvant effects, 745
 bioavailability, 52, 53, f52, f53
 buccal administration, 742
 blood flow, 744
 concentration factors, 743
 dissolution rates, 744, 745
 Donnan effect, 745
 factors affecting, 743
 frequency, 188
 electromagnetic radiation, 301
 emulsions, 448
 inhalation, 743
 oral administration, 742
 parenteral administration, 742
 pharmacokinetics, 715, 721, 755
 rate, 744, 745
 rectal administration, 742
 rubber closure plastics, 508
 sublingual administration, 742
 topical administration, 743, 745
 vehicle effects, 745
 absorption, distribution, metabolism and excretion (ADME), 739
 active transport, 741
 biotransformation, 748
 buccal administration, 742
 diffusion, 346, 739, 742
 dipolarity, 742
 electrochemical diffusion, 740
 exocytosis, 741
 facilitated diffusion, 741
 hypodermic administration, 742
 inhalation, 743
 injection administration, 743
 intramuscular administration, 743
 intravenous administration, 743
 ionic diffusion, 740
 lipid solubility, 741
 membrane structure and properties, 739
 nonionic diffusion, 740, 742
 oral administration, 742
 parenteral administration, 742
 partition coefficients, 741
 passive transport, 740
 phagocytosis, 741
 physicochemical factors, 741
 pinocytosis, 741
 polarity, 742
 preclinical testing, 94
 properties, 739
 rectal administration, 742
 simple nonionic diffusion, 740
 subcutaneous administration, 743
 topical administration, 743
 transport, 740
 acacia, 384, 685, 686
 academic community, drug development/discovery, 83
 acceptability limits, 619
 acceptance criteria
 assays, 11
 enantiomeric purity, 14
 identity tests, 15
 impurities, 12
 residual solvents, 13
 acceptance sampling, 165, f166
 Accogel machine, 608
 accuracy
 weighing/measuring, 106, 110
 statistics, 135, f135
 sterile preparation compounding, 533
 acesulfame potassium, 686
 acetones, 231, t230
 acetonium ions, 261
 acetylation, 752
 acid catalysis, 308
 acid-base concepts/theory
 Brønsted-Lowry theory, 260
 dissociation, 261, 262
 ionic solutions, 261, 626
 polyprotic electrolyte ionization, 263
 α -acid-glycoproteins (AAG/ α -AGP), 326, 719
 acidic solutions, 263
 acidity constants, 262
 acids, 438
 aqueous solutions, 440
 ionic solutions, 262
 acridine, 274
 action mechanisms, solvents, 231
 action-effect context, drug action, 727
 activation energy, 307, f308
 active pharmaceutical ingredients (API), 65
 active transport, 746
 activity
 coefficients, 235, 236, 270
 colligative properties, 235
 ionic solutions, 275
 activity-solubility relationship, 215
 actuators, 643
 Adagen, 660
 addition, calculations in, 116
 additives, 554, 583
 ADE *see* adverse drug events
 adhesional forces, 331
 adiabatic compression, 208
 adipic acid crystals, 331, f331
 adjusted therapeutic range
 adjustment methods, tonicity, 284
 ADME *see* absorption, distribution, metabolism and excretion
 administration
 absorption factors, 52
 aerosol dosage forms, 649
 drug absorption, 742
 parenteral preparations, 498, 499
 adsorption
 bases, 571
 dissolution, 72
 energetic, 334
 fatty acids, 338, t338
 film molecules, 336, f336, f338
 glass containers, 506
 insoluble monomolecular films, 336, f336
 interfacial phenomena
 molecules, 334, f335
 monomolecular films, 336, f336, f338
 mixed films, 338
 solid surfaces, 338
 soluble films, 336
 solutions, 336, 338
 monomolecular films, 381
 particle size, 418
 pyrogens, 509
 retention, 485
 solid surfaces, 338
 adverse drug events (ADE), 5, 490, 537
 Advisory Committees, 97
 aerosols, 633
 absorption, distribution, metabolism and excretion, 743
 aluminum containers, 642, 645, f646
 applications, 644
 barrier-type systems, 637, 638
 cascade impactors, 648, f647
 chlorofluorocarbons, 633, 634, 639
 component design, 648
 compressed gases, 637, 642
 containers, 490, 642, 645, f646
 continuous spray valves, 642, f642, f643
 definitions, 633
 developments, 649
 dispersions, 644
 emulsions, 645
 evaluation, 647
 filling equipment, 643, f643, f644
 fluorinated hydrocarbons, 639
 foams, 636, 637, f637
 formulation, 644, 648
 glass containers, 642
 hydrocarbons, 639, 641
 hydrochlorofluorocarbons, 639, 641
 hydrofluorocarbons, 639, 641, 645
 inhalation therapy, 633
 liquefied gases, 633, 634, 639
 metal containers, 642
 metering valves, 646, f646
 operation modes, 633
 packaging, 643, f643
 particle size distribution, 647, f647
 piston systems, 637, f638
 powders, 644
 propellants, 639, 645
 respiratory system, 633
 suspension formulation, 377
 testing, 647
 three-phase systems, 635
 topical products, 633, 635, 644, 647
 two-layer systems, 635
 two-phase systems, 633
 valves, 642, 645
 affinity, receptor theories, 733
 agar, 685, 686
 agglomeration, 419, 459
 aggregation
 emulsions, 388, 389, 453
 powders, 425
 agitation
 agitators, 454
 droplet formation, 379
 intensity, 71
 syrup preparation, 444
 agonists, 732, 733, 734, 735
 AHFS *see* American Hospital Formulary Service
 air cleaning, 511
 air displacement, 523
 air pistons, 596, f598
 air quality, 533, 535, 538
 air removal, 477
 air systems, 489
 air-water concentration, 669, t670
 albumin, 719
 Albuterol, 639
 alcohol
 action mechanism, 231
 emulsion formulation ingredients, 451
 liquid preparations, 438
 nonaqueous solutions, 446, 448
 non-sterile liquids, packaging, 675
 preservative excipients, 685
 solution solubility, 230, t230
 solvents, 436
 strength (proof spirit), 127
 alginic acid, 462, 683, 684, 687
 alkaloids, 662
 alligation method, 126
 alpha levels, 142
 alternative hypothesis, 142
 aluminum, 110, f110, 523, 524
 aluminum hydroxide, 350, 463
 aluminum subacetate solution, 443
 alveolar excretion, 753
 amantadine hydrochloride syrup, 445
 Amenemhet, 393
 American Drug Index, 3
 American Hospital Formulary Service (AHFS), 4
 American Society of Health-System Pharmacists (ASHP)
 secondary research literature, 1
 amides, 312
 γ -aminobutyric acid (GABA), 732, 735, 736
 ammonium chloride, 268
 ammonium compounds, 439
 amorphicity, 67
 amorphous mixtures, 199
 amorphous solids, 195, f196
 amorphous states, 420, 424, f420
 amphoteric agents, 339
 amphiphilic interactions, 345
 amphiprotic substance, 261
 ampholyte solutions, 268
 Amphotec, 457
 Amphotericin B, 457
 ampoule sealing, 523
 analysis of variance (ANOVA), 151, 153, 156, t158
 analytical considerations
 see also medicinal analytical testing
 chemical development, 88
 pharmacokinetics, 705
 quality assurance and control, 32
 testing, 9
 wavelength selection, 273
 anatomy
 gastrointestinal tract, 625
 rectal administration routes, 574
 skin, 565, f566
 topical dosage forms, 574, 575
 vaginal administration routes, 574, 575
 ANDA *see* Abbreviated New Drug Application
 Andreasen pipettes, 416
 Anemometers, 348
 angle of repose, 422, f423
 animals
 drug development, 756
 testing, 32
 topical dosage forms, 569
 anionic substances, 339, 382, t339
 anions, 257
 ANOVA *see* analysis of variance
 antagonists, 732, 733, 734, 735
 antiadherents, 684
 antibiotics
 drug discovery/development, 85
 metal ion complex formation, 319
 biotechnology, 662
 potency, 728
 biotechnology
 drug development, 655, 659
 milestones, 653
 organic synthesis, 664
 sequence-specific peptide cleavage, 664

- antilogs, 116
 antimicrobials, 497, 579
 antioxidants, 437, 497, 683, t437
 antisense DNA, 660
 antiviral agents, 87
 API *see* active pharmaceutical ingredients
 apothecary weights, 104, f104, 105, 110, 117
 apparent density, 113
 apparent dissociation constants, b264, 264, 272, t273
 apparent rate constants, 305
 apparent volume of distribution, 708
 appendages, skin, 565, 572
 approval
 pharmaceutical product stability, 37, 38
 approved product changes, 98
 approximations
 pharmaceutical product stability, 45
 weights-measures relationships, 106, t106
 aprotic solvents, 262
 APV Gaulin homogenizer, 456
 aqueous solubility, 275
 aqueous solutions
 diluted acids, 440
 dosage forms, 439
 douches, 440
 ear preparations, 442, t442
 enemas, 440
 gargles, 441
 gels, 446
 honeys, 445
 irrigation preparations, 443
 juices, 441
 mucilages, 446
 nasal preparations, 441, t442
 preparation, 443
 sweet substances, 443
 syrops, 443, 444, t444
 Archimedes, 393
 arcsine transformation, 174
 area under the concentration-time curve (AUC)
 bioequivalence testing, 51, 55, 56, 57
 drug development, 755, f757, f759, f761
 trapezoidal rule, 55, t55
 arithmetic means, 411
 Arrhenius equation, 46, 307, 397, f307
 Arrhenius plots, 314, 315
 asbestos-cellulose layers, 484
 aseptic processes/techniques
 aseptic areas, 510, 513, 514
 capability demonstrations, 490
 closures, 490
 environments, 489
 equipment, 489
 excipients, 490
 facilities, 489
 gowning, 490
 manual systems, 489
 operating procedures, 490
 personnel, 490
 sterile preparations, 469, 488, 533, f488
 testing, 490
 ventilation, 489
 ASHP *see* American Society of Health-System Pharmacists
 assays, 10, 11, 59
 assessments
 bioequivalence testing, 54
 associate constants, 320
 associated colloid interactions, 344, 361
 association constants, 216
 associations, information resources, 6
 ASTM permeation, 668, t669
 atomic theory, 177, 178
 atomizers, 633
 Attapulgit, 356
 attrition, 427
 A-type phase-solubility diagrams, 321, f321
 AUC *see* area under the concentration-time curve (AUC)
 Aufbau principle, 179
 autoclaving, 524, f525
 autoclaving, 536
 automation, 78, 79, 536
 autoprotolysis, 262
 autoxidation, 313
 auxiliary components
 solid dosage forms, 673
 non-sterile liquids, 676
 auxiliary emulsifying agents, 384, t383
 average studies, bioequivalence testing, 57, 58
 population studies, 58
 Avogadro's law, 205
 avoidupois weights, 104, 110, f104
 axenic shoot-tip cultures, 663
 axons, 736
 Bacillus Spp., 473, 479, 486
 bacteria, 664
 Bacterial Endotoxins Test (BET), 16, 528
 bacteriostatic water for injection, 703
 bag-in-bag aerosol systems, 637
 balances
 construction, 106
 protection, 109
 testing, 108
 weighing pharmaceuticals, 106
 ball mills, 427, 460, f428
 ball type instruments, 409
 Balmer series, 178
 barbiturates, 312
 barriers
 parenteral preparation manufacture, 511, 513, 569
 bases, 262, 308, 576
 basket-rack assembly, 602
 basket-stirred-flask test, 63, 73
 batch-to-batch reproducibility, 619
 Baxter Healthcare Corporation, 533
 BCS *see* Biopharmaceutical Classification System
 Beclomethosone Propionate, 639
 bed density, 587
 Beer's law plot, f167
 bentonite, 350, 356, 384, 687
 benzalkonium chloride, 687
 benzene, 182, 330, f183
 benzyl benzoate lotion, 464
 bestselling drugs, a666
 BET *see* Bacterial Endotoxins Test; Brunauer, Emmett and Teller
 Betaseron, 659
 beverages, 465
 beyond-use dating (BUD), 37, 45, f45, 533, 535
 BHA *see* butylated hydroxyanisole
 BHT *see* butylated hydroxytoluene
 bias
 statistics, 135, f135
 biliary excretion, 753
 bimodal size distributions, f413
 binary liquids in liquids, 225, f225
 binders
 compressed tablets, 613, t613
 dissolution, 68
 excipients, 683
 tablets, 583
 binding
 see also complex formation
 constants
 B-type phase-solubility diagrams, 322, f321
 determination, 320
 kinetic methods, 322
 NMR spectrometry, 323
 solubility method, 320
 ultraviolet spectroscopy, 323, f323
 distribution, 747
 drug action, 735
 states of matter, 189
 storage, 537
 Bingham bodies, 398
 binomial distribution, 138, 141, t138, t139
 bioavailability
 biotechnology challenges, 659
 coarse dispersions, 389
 concepts, 51
 dissolution, 51, 52, 63
 emulsions, 390
 pharmaceutical analysis, 51
 rectal topical dosage forms, 575
 bioburden methods, 472
 bioequivalence testing, 51, 53
 blood level-time curves, 54, 55, 56, f55
 concepts, 51
 criteria, 56
 cross-study comparison pitfalls, 58
 data evaluations, 54
 determination methods, 53
 evaluations, 54
 food studies, 57
 high-fat test meals, 57, t57
 individual studies, 58
 minimization, 54
 peak-height concentration, 51, 54, 56, 57
 pharmacokinetics, 53, 56, 57
 plasma level concentrations, 52
 population studies, 57, 58
 study conditions, 59
 study design, 58
 subject population studies, 58
 biohazards, 477
 Biological Abstracts (BIOSIS Previews) database, 3
 biological experiment sampling statistics, 136
 biological fluids
 viscoelasticity, 401
 biological incompatibility, t38
 biological indicator method, 472, 473, t474
 Biological License Applications (BLA), 29
 biological polymer solutions, 237
 biopharmaceutic extended-release systems, 625
 Biopharmaceutical Classification System (BCS), 75
 biosimilars, 660
 biotechnology, 651, 653
 bioterrorism, 661
 bone marrow cells, 658
 cell receptor molecular features, 658
 challenges, 658
 clone selection, 655
 compounding, 651
 definitions, 82
 delivery, 31, 660
 deoxyribonucleic acid, 654, 656
 dosage forms, 651
 drug development, 654
 drug discovery, 81
 ethics, 664
 fingerprinting, 657
 forensics, 657
 gene cloning, 662
 gene diagnostics, 656
 gene manipulation, 659
 gene silencing, 661
 gene therapy, 656, 661, f656
 genetic engineering, 662
 genetically modified bacteria, 664
 germ line gene therapy, 657
 hairy root cultures, 662, f663
 hybridoma production, 655, f655
 immune system, 653
 immunologic concerns, 659
 leukemia, 657
 manufacturing, 651, 664
 mergers, 662
 microarrays, 656
 milestones, 651, t652
 molecular techniques, 657
 monoclonal antibodies, 655
 moral issues, 664
 nanotechnology, 661
 neuropharmacology, 664
 off-target gene silencing, 661
 oligonucleotide technologies, 660, 664
 organic synthesis, 664
 peptides, 655
 pharmaceutical challenges, 658
 pharmaceutical manufacturing, 664
 pharmacist's role, 658
 pharmacogenomics, 656
 pharmacognosics, 662
 pharmacological applications, 664
 pharmacokinetics, 659
 plant products, 662
 plant tissue culture, 662
 product characterization, 31
 proteins, 658
 quality control, 31
 radioisotopes, 655
 recombinant DNA, 654, 660, 664
 recombinant proteins, 658, 660, 663
 research, 81
 reverse genetics, 657
 RNA interference, 661
 secondary plant products, 662, f663
 specificity, 661
 stem cell therapy, 661
 terrorism, 661, 662
 vaccination, 653, 654
 biotransformation
 chemical reaction, 751
 drug metabolism, 749, f749
 endoplasmic reticulum, 748
 enzymes, 750
 liver, 748
 metabolism, 748, 749, 751
 oxidation, 751
 phase, I, 751, 752
 reduction, 751
 types, 751
 bivalent, 260
 BLA *see* Biological License Applications
 bleeding, 41
 see also blood...
 blending powders, 429
 blister cards/packs, 673
 blood
 flow, 401
 pressure, 282, f729
 body surface area (BSA), 119
 Bohr, Niels Henrik David
 Bohr model, 178, f178
 boiling curves, 190
 boiling points, 234, 257, 284
 bond vibrations, 188
 bonding
 complex formation, 317
 coordination complexes, 317
 covalent coordination, 317
 delocalization, 181
 intermolecular forces, 318
 isomers, 184
 metal complexes, 318
 metallocenes, 183, f184
 molecular structure, 179, 181
 resonance, 182
 boric acid, 687
 borosilicate glass containers, 504, t505
 bottle centrifuges, 247
 bottle sealing, 523
 boundaries
 colloidal dispersions, 343
 heat flow/work, 207
 bovine pancreatic glucagon, 254
 Bowman's capsule, 753
 Boyle's law, 205
 BP *see* British Pharmacopeia

- bracketing, sterilization, 481, f482
 Bravais lattice, 194, f194
 breakdown, emulsions, 379
 breaking forces, 18
 breathing, 208
 British Pharmacopeia (BP), 9
 Brookfield viscometers, 41
 Brownian motion, 346
 Brunauer, Emmett and Teller (BET), 418
 Brunner, Emil, 63
 BSA *see* body surface area (BSA)
 bubble method, 608
 bubble points, 487
 buccal administration, 582
 Büchner funnels, 244
 BUD *see* beyond-use dating
 buffer areas, 535, f535
 buffer capacity, 272, 675, 677
 buffers, 237, 269, 314
 buildings, quality assurance and control, a34
 bulk powders, 432
 emulsions, 435, 448
 extracts, 465
 liquid preparations, 435
 solutions, 435
 suspensions, 458
 bulk properties, 419
 Bunsen pumps, 245
 Burgers' model, 403, f403
 butylated hydroxyanisole (BHA), 683, 688
 butylated hydroxytoluene (BHT), 688
 butylparaben, 688
 caffeine suppression, 663
 caking, 242, 372, 376
 calamine lotion, 464
 calcium carbonate, 684
 calcium stearate, 685, 688
 calcium sulfate, 688
 calculations in pharmaceuticals
 common fractions, 115
 concentration, 125
 conversions, 117
 decimal fractions, 115
 density, 113
 dilution, 125
 dimensional analysis, 120
 division, 117
 dosage calculations, 119, a132
 enlarging formulas, 121
 equivalents, 119
 exponents, 115
 formula enlarging, 121
 fractions, 115
 household equivalents, 119
 logarithms, 115
 mass, 113
 mathematical principles, 114
 measures, 101, 105
 measures-weights relationships, 105
 measuring pharmaceuticals, 114
 metrology, 101
 milliequivalents, 128
 mixing different strengths, 126
 multiplication, 117
 orthography, 103
 parts per million, 125
 parts per thousand, 122
 percentages, 122
 powers, 115
 problem-solving methods, 119
 proof spirit, 127
 ratio strengths, 122
 reading, 103
 reducing formulas, 121
 roots, 115
 saturated solutions, 128
 significant figures, t115
 solutions, 125
 specific gravity, metrology, 113
 stock solutions, 125
 strengths, 126
 subtraction, 117
 temperature, 128
 trituration, 122, 126
 weighing pharmaceuticals, 114
 weight, 101, 101, 113
 weights and measures relationships, 105
 cAMP, G protein-coupled receptors, 736
 can-in-can systems, 638
 CAP *see* cellulose acetate phthalate
 capacity factor, 208, t208
 capillarity, 333, f333, f334
 capillary electrophoresis (CE), 274, 354
 capillary gel electrophoresis (CGE), 355
 capillary instruments, 406, f406
 capping
 hard gelatin capsules, 605
 tablets, 599
 caps
 color, 556, t556
 liners, 670
 capsules, 581, 604
 cleaning, 607
 compression filling, 606
 containers, 668
 dosage unit uniformity, 607, t608
 encapsulation, 604, 608, f608
 extemporaneous preparations, 461, 605
 filling, 606, 608, t606
 gelatin, 40, 604, f604
 machine filling methods, 606
 plate filling method, 606
 vacuum filling, 606
 weight variation, 608
 captopril, 188, f189
 carbamazepine, 189, f189, 721
 carbomers, 462, 463, 464, 688, 684
 carbon atoms, 181
 carboxylate ions, 312
 carboxymethyl ethylcellulose (CMEC), 615
 carboxymethylcellulose (CMC), 463, 684, 689
 caregivers, 537
 carnauba wax, 689
 Carnot cycles, 212
 carrageenan, 684, 689
 Casarett and Doull's Toxicology: The Basic Science of Poisons, 5
 cascade impactors, 414, f415
 CAT *see* cellulose acetate trimellitate
 catalysts, 305, 308
 cataplasms, 572
 cationic substances, 383, 339
 CBER *see* Center for Biologics Evaluation and Research
 CCRIS *see* Chemical Carcinogenesis Research Information System
 CD *see* circular dichroism
 CCD *see* countercurrent distribution
 CDRH *see* Center for Devices and Radiological Health
 CE *see* capillary electrophoresis; complexation efficiency
 celecoxib, 421, f421
 cell biology, 86
 cell constant, 258
 cell cytoskeleton, 653
 cell membrane, 739
 cellulose, 690
 cellulose acetate phthalate (CAP), 615, 689
 cellulose acetate trimellitate (CAT), 615
 cellulose derivatives, 355, 367
 cellulosic tablet binders, 584
 Center for Biologics Evaluation and Research (CBER), 667
 Center for Devices and Radiological Health (CDRH), 667
 Center for Veterinary Medicine (CVM), 667
 central atom multiple bonding, 181, f182, f183
 Central Limit Theorem (CLT), 139, 141
 centralization, PCA statistics, 172
 centrifugal forces, 415
 centrifugation, 242, 246, f246, f247
 density separation, 247, f247
 filtration, 248
 isopycnic separation, 247
 rate-zonal separation, 247
 rotors, 248, f248
 sedimentation, 246, 247
 tube selection, 249
 Cephalosporins, 254
 certification, 32
 cetyl alcohol, 690
 CFC *see* chlorofluorocarbons
 CGE *see* capillary gel electrophoresis
 cGMP *see* current Good Manufacturing Practices; cyclic guanosine-3',5'-monophosphate
 chalk, 245
 changes, states of matter, 189
 characteristics
 compressed tablets, 600
 characterizing solid states of matter, 199, f200, f201, f202
 charcoals, 245
 charge stabilization, 359
 charge, gas sterilization, 481
 Charles' law, 205
 chelates, 317, 318, f319
 chemical bonding, 735
 Chemical Carcinogenesis Research Information System (CCRIS), 6
 chemical condensation, 357
 chemical equilibrium, 260
 chemical equivalence, 51, 53
 chemical failures, 39
 chemical incompatibility, 42
 chemical kinetics, 301
 collisions, 301, 307
 complex chemical reactions, 306
 complexing agents, 314
 component interactions, 313
 consecutive chemical reactions, 307
 definitions, 301
 dielectric constants, 311
 drug stabilization, 314
 electrolytes, 311
 enzyme catalysis, 310, f310
 first-order reactions, 303, 306, 308, f303, f313
 general acid/base catalysis, 310
 heat, 307
 hydrolysis, 311
 instability, 313
 mechanisms, 301
 oxidation, 313
 photochemical decomposition, 313
 pseudo-order reactions, 305
 racemization, 313
 reaction rates, 302
 reversible chemical reactions, 306
 second-order reactions, 304, 306
 simultaneous chemical reactions, 306
 testing drug stability, 314
 third-order reactions, 305
 chemical laboratory sampling statistics, 136
 chemical potential, 213
 chemical properties
 fluorinated hydrocarbons, 640
 chemical reactions
 pharmaceutical product stability, 42
 reverse solubility, 219
 solubility, 219, 223
 chemical stability, 378, 388
 chemical structure, 382
 chemotherapy
 garbing, 534
 chilling, 595
 chiral centres, 186, f186, f187
 chiral drug pharmacokinetics, 721
 Chi-square tests, 148, t148
 chlorine dioxide, 482
 chlorofluorocarbons (CFC), 635, t635
 chlorpheniramine maleate (CPM), f630, f631
 cholesterol, 384
 chromatography
 see also individual methods
 complex analysis, 326
 drug stability, 314
 gel filtration, 253
 circular dichroism (CD), 187
 cisplatin, 318
 citation databases, 3
 citric acid, 685
 Clapeyron equation, 214
 clarification methods, 252, 520
 Clark's mathematical formulation, 733
 Clark's Rule, 119
 classification
 colloidal dispersions, 344, t344
 medicinal analytical testing, 10
 ointments, 571
 solutions, 439
 surfactants, 339
 therapeutic equivalence, 60
 Clausius-Clapeyron equation, 215
 clays, 244
 clean air, 533
 clean rooms
 functional production areas, 510, f510
 maintenance, 514
 materials support areas, 510
 parenteral preparations, 510, t511
 cleaning
 compounding facilities, 535
 containers, 517
 facilities, 535
 handling containers, 518
 machinery, 518
 treatment cycles, 518
 clearance
 clinical pharmacokinetics
 drug distribution, 710, 711, 714, 717
 pharmacokinetics
 protein binding, 719
 clinical equivalence, 551
 clinical experiment sampling statistics, 136
 clinical laboratories, osmometry/osmoticity, 280
 clinical scientists, 83
 clinical trials
 conduct, 137
 design, 137
 dosage validation, 761
 drug development, 89, 755
 planning, 97
 Clinical Trials Website (www.clinical-trials.gov), 97
 clinicians, drug discovery/development, 83
 Clostridia, 473
 closures
 bottles, 670, t670
 cleaning procedures, 518
 integrity tests, 530
 liquid preparations, 675
 parenteral preparations, 504, 506
 quality assurance and control, a34
 CLT *see* Central Limit Theorem
 CMA *see* critical material attributes
 CMC *see* carboxymethylcellulose; critical micelle concentration
 CMEC *see* carboxymethyl ethylcellulose
 coagulation
 colloidal dispersions, 352, 353
 coalescence, 380, f380, 385
 coarse crushers, 427

- coarse dispersions
 - deflocculation, 372, 373, 376, f376
 - emulsifying agents, 378
 - emulsions
 - examples, t372
 - pharmaceutics, 378
 - phase combinations, 378, f387
 - physical stability, 388
 - quality control testing, 390
 - formulation, 374, 377, t377
 - interfacial properties, 371
 - methodology, 371
 - pharmaceutics, 371
 - sedimentation volumes, 374, f376
 - sedimentation, 373
 - settling, 372
 - surface free energy, 371
 - surface potential, 372
- coatings, 611
 - compressed tablets, 581
 - dosage forms, 611
 - equipment, 616
 - excipients, 684
 - fluid-bed processing, 618, f618
 - pans, 616
 - procedures, 616
 - process evolution, 611
 - Quality by Design, 619, 620
 - quality control, 619
 - recent trends, 620
 - sprays, 616, 620
 - stability-testing programs, 619
 - technology trends, 620
- coaxial-cylinder instruments, 407, f408
- cobalt ionizing gamma irradiation, 483
- cocoa butter, 576
- co-crystals, 197
- cohesion, 423
- cohesional forces, 331
- colation, 253
- cold compression, 577
- cold plasma, 206
- colligative properties
 - deviation from ideality, 235
 - electrolytes, 233, 235, 257
 - freezing points, 234
 - ideal behavior, 234
 - osmolality determination, 284
 - osmotic pressure, 232
 - phase equilibria, 232
 - practical applications, 235
 - vapor pressure, 233, f234
- collodion membranes, 484
- colloidal dispersions
 - applications, 364
 - comminution, 343, t343
 - condensation methods, 357
 - continuous phases, 344
 - cross-linked polymers, 365, 369
 - delivery systems, 365
 - diffusion coefficient, 346
 - dispersed phase interactions, 344
 - dispersed states, 344
 - dispersion methods, 356
 - double layers, 350, f351
 - drug-polymer conjugates, 367
 - dynamic light scattering, 348
 - electric charge origins, 350
 - electric double layers, 350, f351
 - electric properties, 350
 - electrokinetic phenomena, 351, 354
 - electrophoresis, 354
 - electrostatic repulsion, 351, f316
 - emulsions, 362, 366, t364
 - excipients, 367
 - formation, 361, f361, f363, 365
 - gel formation, 348
 - Hofmeister series, 349
 - hydrophilic dispersions/interactions, 344, 356
 - interactions, 344
 - interfaces, 343
 - kinetics, 357, f358
 - light scattering, 347
 - liposomes, 363, 366
 - lyophilic dispersion, 344, 355
 - lyophobic dispersions, 344, 356
 - lyotropic series, 349
 - micelles, 365
 - microelectrophoresis, 354
 - microemulsions, 362, 366
 - microspheres, 367
 - nanoparticles, 367
 - optical properties, 347
 - particle shape, 345, f345, f346
 - particulates, 356, 367
 - pharmaceutical applications, 364
 - pharmaceutics, 343
 - polymers, 348, 355, 367
 - properties, 345
 - protective mechanisms, 360, f360, f361
 - radioactive isotopes, 365
 - sedimentation, 346
 - sensitization, 360, f360, f361
 - shape, 345, f345, f346
 - solubilization, 365, f366
 - specific surface area, 343, t343
 - stabilization, 359, f359
 - surface area, 343, t343
 - suspensions, 462
 - systems, 343, t343
 - viscosity, 348
 - water-soluble polymers, 355
- colloidal silicon dioxide, 463, 684
- color
 - coating, 613
 - excipients, 684
 - tablet stability, 40
- colorants
 - film coating, 614
 - hard gelatin capsules, 605
 - tablets, 585, t586
- combinations, parenteral preparations, 499
- combinatorial chemistry, 84
- commercial containers, 673
- comminution, 426, 588, f589
- compaction, 586
- comparative clinical trials, 53
- comparisons
 - instruments, 409
 - rheometers, 409
- compatibility
 - parenteral preparations, 498, 504, t505
- compendia
 - dissolution, 73
 - medicinal analytical testing, 9
 - pharmaceutical product stability, 38
- competitive antagonists, 733
- competitive inhibition, 310
- complex formation
 - cyclodextrins, 323, 325, f323, f324, t324
 - drug-protein binding, 326
 - equilibrium constants, 321
 - hard acids and bases, 318, t318
 - inclusion complexes, 320
 - metal ions, 326
 - molecular complexes, 317, 319
 - non-covalent bonding, 318
 - phase-solubility diagrams, 321
 - plasma protein binding, 326
 - solubility, 238
 - therapeutics, 327
- complexation efficiency (CE), 321
- compliance
 - viscoelasticity, 403
- components
 - mapping, 473
 - packaging, 667
 - parenteral preparations, 501
- compounding
 - areas, 513
 - asepsis, 533
 - aseptic isolators, 533
 - calculations, 119
 - compounded sterile preparations, 535
 - dating, 535
 - delivery, 537
 - environmental controls, 538
 - environmental monitoring, 538
 - excipients, 683
 - facilities, 535
 - filters, 536
 - garbing, 534
 - hygiene, 534
 - labeling requirements, 537
 - laminar air flow hoods, 533, f534
 - medication storage, 538
 - monitoring, 537
 - multi-dose vials, 536
 - National Formulary standards, 533
 - oral solid preparations, 581
 - packaging, 667
 - parenteral preparations, 495
 - personal protective equipment, 534
 - personnel responsibilities, 533
 - personnel training/qualifications, 537
 - powders, 411
 - primary engineering controls, 533, 535
 - qualifications, 537
 - quality assurance, 536
 - risk levels, 533, 535
 - single-use containers, 536
 - standard operating procedures, 533, 537, 538
 - sterile preparations, 533
 - sterilization methods, 536
 - storage, 533, 535, 538
 - tablets, 581
 - temperature monitoring, 538, f538
 - training, 537
 - USP<797>, 533
 - verification, 537
- compound-lever balances, 107
- compressed gases, t636
- compressed tablets (CT), 581, 582, 612, 614
- compressibility, powders, 422
- compression
 - coating, 612
 - machines, 595
 - molding, 577
 - stills, 501, f502
 - tablet granulation, 592
- computational methods
 - molecular structure, 188
 - osmolality, 278
 - probabilities, 140
 - rheology, 404
- concentration
 - see also* drug concentration
 - drug effect, 705, 722
 - effect relationship, 705
 - emulsifying agents, 381
 - pharmacokinetics, 705
 - ratios, 250
 - response, 722
 - time profile, 755, f756
- concentric-cylinder viscometers, 407
- condensate removal, 477
- conductance, 258
- conductivity
 - emulsions, 379
 - ionic solutions, 257
 - pure water pH, 263
 - weak electrolytes, 258
- cone-and-plate instruments, 408, f408
- confidence intervals
 - bioequivalence testing, 57
 - estimation statistics, 141, 169, f169
 - for γ , 168
- confidence limits, 168
- conformational strain, 318
- conformations, isomers, 184, f185
- conjugative biotransformations, 751, 752
- conservation of energy, 207
- constant gas volume, 419, f419
- construction
 - hypothesis, 142
 - null hypothesis, 142
 - t-test, 144
- consumption, syrups, 445
- contact angles, 332, f332, f333, 334
- contact lenses, 559, t560
- container mapping, 473
- containers
 - measuring pharmaceuticals, 111
 - unit-dose medications, 673
- contaminants/contamination, 600
- content uniformity, 602
- continuous compression filling, 606
- continuous drug inputs, 712, f712
- continuous washing, 253, f253
- contraception, 579
- contrast agents, 281
- control
 - pyrogens, 509
- control charts, 165, t165
- controlled flocculation, 375
- controlled release dosage forms, 582, 615
- convection, 428
- conventional topical dosage forms, 571
 - drug delivery, 570, 576
 - formulation, 570
 - nails, 574
 - semisolids, 571
 - transdermal drug delivery evolution, 573
- conversions
 - conversion tables, 103
 - metric systems, 118
 - rules, 118
- coordinate bonding, 183, f184
- coordination complexes, 317
- coordination numbers, 318
- corrections, non-Newtonian fluid rheology, 409
- correlation statistics, 170, f170, f171
- cosolvents, 237, 424
- cotton filters, 243
- cottonseed oil, 691
- coulomb interactions, 180
- Coulter Counters, 416, f417
- countercurrent distribution (CCD), 250, f251, f252
- counterions, 350
- coupling receptors, 732
- covalent bonding, 180, f181
- covariance matrices, 172
- covariance, 134
- CPM *see* chlorpheniramine maleate
- cracking/cracks, 506, 616
- creaming, 347, 388, 453
- creams, 573, 676
- creep, 403
- crimping machines, f524
- critical bulk manufacturing precautions, 496
- critical material attributes (CMA), 79
- critical micelle concentration (CMC), 362
- croscarmellose sodium, 684, 691
- crospovidone, 691
- cross-linking, 40
- crushers, 427, f427
- crushing strength, 587, 601
- cryoprotectants, 497
- crystal lattice, 194, f194
- crystalline solids, 193, f193, 194, f194
- crystallization
 - colloidal dispersions, 357
 - molecular aggregation, 425
 - particle properties, 419
 - states of matter, 195
- CT *see* compressed tablets
- Cube Root Law, 63
- cumulative dose frequency curves, 730
- cups, packaging, 674
- current Good Manufacturing Practices (cGMP)
 - see also* Good Manufacturing Practices
- current status
 - drug development, 81

- curved surface pressure differences, 334
- cutaneous biotransformation, 568
- CVM see Center for Veterinary Medicine
- CYP see cytochrome P450
- cycle development, sterilization, 471
- validation, 481
- cyclic guanosine-3',5'-monophosphate (cGMP), 737
- cyclodextrins, 691
- cytochrome P450 (CYP), 735, 748, 749
- Dalton's atomic theory, 177
- Dalton's law, 205
- Danckwerts' dissolution model, 63
- dapsone hydroxylamine, 723
- daptomycin dosage, 757
- darbepoietin, 659
- Darcy equation, 242
- data considerations
- collection, 135
 - designing experiments, 135
 - reduction, 134, 173
 - statistics, 133
 - transformations, 173
 - visualization, 133
- de Broglie wavelength, 178
- death, microbials, 469
- Deborah number, 401
- debye-Huckel theory, 260
- decantation, 253
- decarboxylation, 44
- decimal notation, 102
- decision making, 141
- decoctions, 465
- decoloration/decolorization, 253
- deflocculation, 357
- deformation, rheology, 394
- degree of flocculation, 374
- degrees of freedom, 192
- delamination, 505
- delayed-release dosage forms
- analytical testing, 18
 - enteric-coatings, 623, t624
 - film coating, 615
 - oral modified release, 623
 - polymeric-coatings, 623, t624
 - product performance, 19
 - product quality, 17, 18
 - universal analytical tests, 18
- delivery systems
- emulsions, 572
 - USP classification, 571
- denervation supersensitivity, 731
- density
- gradient centrifugation, 247
 - suspensions, 458
 - weight, 113
- deoxyribonucleic acid (DNA), 651, f652, 653, 737
- Department of Health and Human Services (DHHS), 536
- depth filters, 241, 242, 245, f243
- depyrogenation tunnels, 480, f480
- Derjaguin, Landau, Verwey and Overbeek (DLVO) theory, 351, 353
- dermis, 565, f566
- description tests, 10, 15
- design
- crossover, statistics, 156
 - experiments, 151
 - packaging, 667
- Design Space, 30
- detection/detector errors, 73
- detergents, 376
- determination
- osmolality, 284
 - particle size, 413
- dewaxing, 252
- dextrates, 691
- dextrin, 684, 691
- DHHS see Department of Health and Human Services
- diagnostics
- contrast agents, 281
- dialysis, 253
- dichlorotetraamminecobalt (III) chloride, 318
- dielectric constants, 230, 377
- dietary supplements, 78
- difference factor, dissolution, 78
- differential centrifugation, 246
- differential scanning calorimetry (DSC), 201
- differentiating solvents, 262
- diffusion, 253
- extended-release systems, 627, t627
 - Pick's law, 253
 - powder mixing, 428
 - rate of solution, 220
 - tests, 487, 520
 - topical skin dosage forms, 568
- difluoroethane, 692
- digestion, 465
- dilatancy, 204, 397
- diluents, 68, 583, 684
- dilution
- calculations, 122
 - tests, 379
- dimensionless Deborah number, 401
- direct acting reversible agents, 723
- disc centrifuges, 247
- discharge actuations, 647
- disintegrants/disintegration
- tablets, 18, 19
 - dissolution, 66, 76
 - colloidal dispersions, 356
 - tablets, 585, 602, f602
- excipients, 684
- disk instruments, 408
- dispense/dispensing
- tablets, 582
- dispersion
- see also coarse dispersions
 - Einstein's law of viscosity, 404
 - forces, 318
 - formulation, 374
 - models, 404
- dissertations, 2
- dissociation
- constants, 262, 264, t264
 - electrolytes, 259
- dissolution
- definitions, 64
 - dosage forms, 70, 75
 - drug physicochemical properties, 66
 - drug production formulation, 67
 - fluid media characteristics, 72
 - hardness, 171
 - humidity, 72
 - mathematics, 64
 - modified release dosage forms, 78
 - oral cavity drugs, 75
 - paddle methods, 74
 - particle size, 66
 - permeability, 75
 - powders, 75, 424
 - profile comparisons, 78
 - reciprocating cylinders, 63, 74
 - rotating cylinders, 74
 - rotating disk systems, 66
 - shaft guiding, 70
 - solubility, 66, 75
 - solutions, 220, 439, f220
 - sorption, 72
 - stationary disk systems, 66
 - suppositories, 77
 - tablets, 18, 19, 40, 603, f603
 - temperature, 71
 - testing, 68
 - topical products, 77
 - validation, 79
- dissolved gases, 79
- Distek system, f79
- distillation, 501
- distribution
- see also absorption, distribution...;
 - drug distribution; Student's t-test
 - coefficients, 250
 - Donnan effect, 746
 - electrochemical methods, 746
 - elimination, 746
 - nonequilibrium, 748
 - passive transport, 746
 - pH partition principle, 746
 - pharmaceutical product stability, 37
 - quality assurance and control, a34
 - specification, 134
 - water for injection, 502
- divided powders, 432
- divided solids, 451
- DLVO see Derjaguin, Landau, Verwey and Overbeek
- DMF see Drug Master Files
- DNA see deoxyribonucleic acid
- documentation, 31, a35
- Doppler anemometers, 348
- dosage
- see also drug delivery
 - absorption factors, 52
 - dependency, 720, t720, 721
 - dosage calculations, 119
 - drug action, 727, 728, 730
 - effect relationships, 728
 - equivalents, 106
 - forms
 - dissolution, 52
 - dusting powders, 432
 - film coating, 612, 614, 616, f620
 - liquefied gases, 634
 - medicinal analytical testing, 15, 17
 - microencapsulation coating, 612
 - modified release, 615, 623, 624
 - molded tablets, 581, 582, 600
 - nonaqueous solutions, 446
 - ophthalmic preparations, 281
 - oral modified release, 623
 - Pellegrini coating pans, 617
 - pharmaceutical product stability, 40
 - preservatives, 556, t557
 - problems with film coating, 616
 - quality assurance and control, 32
 - side-vented coating pans, 617, 620
 - sterility assurance, 469
 - sterilization processes, 469
 - sugar coating, 611, 612, 616, f617
 - terminal sterilization, 469
 - uniformity, 16, 17 - frequency
 - drug action-response relationships, 730 - intensity curves, 729, f729
 - Monte Carlo simulations, 761
 - protein binding, 721
 - salicylic acid, 721, f721
 - selection
 - drug development, 759
 - HIV protein binding, 759, f760
 - pathogen variability, 759
 - pharmacokinetic variability, 758, 759
 - time-dependency, 720, t720
 - uniformity, 647
- dosage selection, 758, 759
- dosimeters, 484
- double-stranded RNA molecules, 661
- DPI see dry powder inhalers
- dressings, 573
- droppers, 112
- drug absorption
- see also absorption..., 275, 715, 716
- drug action, 734
- concepts, 727
 - drug effects, 727, 734
 - drug receptors, 732
 - effective dose, 729, 731
 - efficacy, 727, 727, 729
 - enzyme inhibitors, 734, 737
 - function, 755
 - G protein-coupled receptors, 736
 - kinase-linked receptors, 737
 - maximum efficacy, 728, 279, 731
 - mechanisms, 734
- median effective dose, 731
- multiple effects, 727, 730
- occupation theory, 732
- pharmacodynamics, 727
- pharmacokinetics, 727
- potency, 727
- receptors, 729
- mechanisms, 734
 - regulation, 738
 - structure, 735
 - target types, 734
 - voltage-sensitive channels, 735
- relative potency, 728, 730
- response variations, 731
- selectivity, 727, 728, 730, 731
- tyrosine kinase-linked receptors, 737
- drug activity, 747
- drug approvals, 95, 96
- drug concentration
- drug response, 722
 - modeling, 706
 - multiple dose administration, 713, f713
 - statistical moments, 714
 - time profile, 705
 - vs. effect relationship, 705, f706
- drug delivery
- see also dosage...
 - extended-release systems, 624, f624
 - oral modified release, 623
 - topical dosage forms, 566, 569
 - skin, 570
 - suppositories, 575, 576, 577
- drug development, 81, 755, f756
- drug discovery, 81
- exposure measures, 755, t761
 - immunology, 87
 - information sciences, 88
 - instrumentation, 88
 - Investigational New Drug applications, 89
 - microbiology, 85
 - molecular biology, 86
 - New Drug Application, 89
 - organic chemistry, 85
 - pharmacodynamics, 755
 - Phase I studies, 89
 - Phase II studies, 89
 - Phase III studies, 89
 - pre-clinical development, 755, 757
 - protein binding, 759, f759, f760
 - relationship creation, 755
 - research, 89
 - sample collections, 84
 - target selection, 757
 - toxicology, 87
 - twenty-first century research, 81
- drug discovery
- Lipinski Rule of 5, 84
 - molecular diversity, 84
 - rational design, 84
 - research, 81
 - structure-based design, 84
- drug distribution
- see also distribution
 - absorption, distribution, metabolism and excretion, 746
 - drug inputs, 708
 - elimination, 708, 709, 710, 714, 717, f709
 - half-lives, 709
 - hepatic clearance, 717
 - intravenous injection, 708, 710
 - organ-specific clearance, 717
 - pharmacokinetics, 708
 - renal clearance, 717, 718
 - steady-state volume of distribution, 711, 714
- drug incorporation into ointments, 572
- drug inputs, pharmacokinetics, 713
- drug interactions, 319, 320
- Drug Master Files (DMF), 668
- Drug Price Competition and Patent Term Restoration Act of 1984, 56

- drug properties, 52
 drug resistance, 732
 drug stability, 275
 drug stabilization, 314
 drug transport, 708
 drug-excipient interactions, 319
 drug-protein binding, 326
 dry comminution, 426
 dry goods sterilization, 474
 dry granulation, 589
 dry heat sterilization, 478, 479, f479, 524
 dry heating, 536, 537
 dry powder aerosols, 633
 dry powder inhalers (DPI), 42, 432
 dry sieving, 414
 drying, states of matter, 191
 DSC *see* differential scanning calorimetry
 Duncan's New Multiple Range Procedure, 153, t155
 Dunnett's Procedure, 155, t156
 duration, pharmacologic effect, 723
 D-values, 285, 470, 479, f470
 dwell periods, 479
 dyes, 379, 613, 614
 E values, tonicity, 285
 eccentricity, stirring devices, 70
 effect vs. concentration relationship, 705
 effervescent tablet manufacture/compounding, 582
 efficacy, 479
 eigenvalues, 172
 eigenvectors, 172
 elastic solids, 395, t395
 elasticity, 393, t395
 electrical potential, 380
 electrically modulated transdermal delivery systems, 573
 electrolytes, 257
 see also ionic solutions
 electrolytic conductance, 257
 equivalent conductance, 258, t258
 flocculation, 376
 ionic strength, 260
 osmotic pressure, 233
 pH, 264, 267, 275
 salts, 267
 pharmaceutical significance, 274
 proton-balance equations, 266, b267
 salting, 349
 species concentration, 266
 electron angular momentum, 178
 electron beams, 483, 484
 electron filling, 179, f178, f197
 electron microscopy, 414
 electron orbits, 178
 electronic balances, 108, f108
 electronic configurations, 179
 electronic mortar and pestle, 429
 electronic structure, 177
 electrostatics, 317, 388
 electrotransport, 573
 elemental impurities, 15
 elevation, osmolality determination, 284
 elimination
 absorption, 739
 emollients, 684
 empirical equations, 402
 empty capsule preparation, 605
 emptying
 stomach, 744
 emulsifiers
 see also emulsions
 desirable properties, 380
 emulsifying agents, 381, t381
 excipients, 684
 film formation, 380
 finely dispersed substances, 384
 formulation, emulsions, 385
 hydrophile-lipophile balance, 385, t385, t386, t387
 interfacial monolayers, 385
 interfacial tension, 380
 mechanism of action, 381
 mixed substances, 387
 multimolecular films, 382
 natural substances, 384
 non-ionic substances, 383
 particle films, 382
 rheology, 381
 stabilization, 388
 synthetic substances, 382
 emulsions
 see also emulsifiers
 formulation, 450, 572
 inversion, 38
 preparation, 451
 properties, 452
 stability, 41
 viscosity, t381
 EN *see* enteral nutrition
 enantiomers, 14, 184, f186, 721
 Encyclopedia of Common Natural Ingredients Used in Food, Drugs, and Cosmetics, 5
 endotoxins, 528, f528
 endpoints
 drug development, 757
 resistance emergence/suppression, 757, f758
 energetics
 interfacial forces, 329
 surface tension, 330, t330
 tension, 329, 330, 334, f329, f330
 wetting, 332, f332, f333
 energy changes, 228
 energy levels/states, 178, 179
 English weights and measures, 101, 105, t105
 enteral hyperalimentation, 282
 enteral nutrition (EN), 282
 enthalpy, 210, 214, 228
 entropy, 211, 212, 228
 entry, absorption, distribution, metabolism and excretion, 739, f740
 Environmental Protection Agency (EPA), 639
 environments monitoring, 30
 enzymes, 310, 326, 732
 EPA *see* Environmental Protection Agency
 epidermis, 565, f566
 epimerization, 44
 epinephrine, 639
 EPO *see* erythropoietins
 equilibrium
 free energy, pharmaceuticals, 214
 thermodynamics, 207
 equipment
 freeze-drying, 528
 insect control
 powder mixing, 429
 qualification, 472
 quality assurance and control, a34
 equivalence in sterilization, 487
 equivalence, concepts, 51
 equivalent conductance, 258, t258
 equivalent weights-measures relationships, 105, 106, t106
 ER *see* extended-release
 β -errors, 142
 erythrocytes, 284
 erythromycin, 59, f59
 erythropoietins (EPO), 658, 659
 EST *see* expressed sequence tags
 esters, 311, 439
 ethanol, 230
 ethers, t636
 ethyl oleate, 692
 ethylcellulose, 692
 ethylene oxide (ETO), 480, 481
 European Pharmacopeia (Ph. Eur.), 9
 eutectics, 198, f198
 evacuation, gas sterilization, 481
 evaluation
 therapeutic equivalence, 60
 evolving there
 transdermal drug delivery, 573
 exact tests, statistics, 161
 excipients, 684, 685
 excretion
 see also alveolar excretion; biliary excretion; renal excretion, 739, 753
 exhaust, gas sterilization, 481
 exposure
 gas sterilization, 481
 expressed proteins, 652
 expressed sequence tags (EST), 653
 expressing solubility, 219
 expression
 immiscible liquids, 249
 separation methods, 249
 extended-release (ER) dosage forms
 analytical testing, 19
 dissolution, 78
 dosage, 625
 drug formulation, 625
 drug properties, 625
 film coating, 615
 gastrointestinal tract, 625, t626, t627
 hydrophilic matrices, 628, t628, t629
 inert matrices, 628
 ion-exchange resins, 628
 kinetics, 267, t627
 matrix-based systems, 627
 oral dosage forms, 627
 osmotic pump systems, 630, f631
 physicochemical properties, 625
 pump systems, 630, f631
 release rate, 627
 reservoir systems, 629, f630
 risk management, 625
 solubility, 625
 universal analytical tests, 19
 wax matrices, 627
 extractables, packaging/containers, 505, 679
 extraction ratio, 714
 extracts, 465
 extruders, 587, 589
 Exubera, 660
 eyes, 281
 f distribution, 149
 facilities
 environmental controls, 535
 set-up, 535
 facsimile transmission
 falling ball instruments, 408
 fatty acids, 383
 FDA *see* Food and Drug Administration
 FDAMA *see* Food and Drug Administration Modernization Act of 1997
 feathering, 716
 fed bioequivalence studies, 57
 Feret's diameters, 411
 ferrous sulfate syrup, 445
 fiber glass filters, 244
 Fick's law of diffusion, 220, 740
 filling
 functional production rooms, 510
 parenteral preparations, 511, 521, f513, f521
 soft elastic capsules, 607
 film coatings
 recent technology trends, 620
 sustained release dosage forms, 615
 tablet manufacturing and compounding, 581
 filters
 filter bias, 79
 filter presses, 249
 filtration sterilization, 484, f484, f485, f486
 filtration, 241, 242
 sterilization-in-place, 478
 filtration
 flow rates, 242
 fritted glass filters 244
 funnels 244, f244
 glass-wool filters, 244
 Gooch crucibles, 244
 historical overviews, 484
 integrity testing, 484, 486, f486, f487
 kieselguhr, 244
 magnesium carbonate, 245
 materials, 484
 mathematics, 242
 media, 242
 membranes, 241, 243, 484, f243
 parenteral preparations, 520, f520
 pharmacokinetics, 718
 powdered substances, 244
 rapid apparatus, 245
 retention mechanisms, 485
 retention, 242
 separation methods, 241, f241
 siliceous earth, 244
 sintered-glass filters, 244, f244
 solution stability, 41
 sterilization, 484
 talc, 244
 validation, 486
 volatile liquids, 244, f245
 fine grinding mills, 427
 finished preparation checks/tests, 533
 finished products, terminal sterilization, 475
 first-order reactions, 708, f707
 first-order stability data, 47
 first-pass effects, 718
 fixed oils, 496
 fixed-angle rotors, 249
 flavoprotein monooxygenases (FMO), 749
 floating compression, 597
 flocculation, 373, 376, 378, 397
 Florentine Receiver, 249, f249
 flow, 393, 421
 fluid energy mills, 428, 459, f428, f461
 fluid-bed drying, 429, 590
 fluidization, 589
 fluocinolone acetate, 570, f570
 fluorinated hydrocarbons, 640
 fluoroquinolones, f758
 FMO *see* flavoprotein monooxygenases
 foam valves, 643
 foams, 637
 Food and Drug Administration (FDA)
 packaging, manufacturing and compounding, 668
 product stability, 38, 45
 quality assurance and control, 29
 therapeutic equivalence, 60
 Food and Drug Administration Modernization Act of 1997 (FDAMA), 32
 foods, 663
 formation
 ionic solutions, 274, t275
 formulation
 coarse dispersions, 374
 freeze-drying, 528
 ointments, 571
 parenteral preparations, 497, t498
 pharmaceutical product stability, 37, 42
 solution aerosols, 644
 suspensions, 459, 644
 topical dosage forms, 569
 forskolin, 734
 fraction ionized, 271, f271
 fractional prefixes, 103, f104
 Frankenfoods, 663
 Franz diffusion cells, 568, f569
 free energy
 colloidal dispersions, 343
 fugacity, 213
 partitioning behavior, 216
 protein binding, 216
 solids, 331
 thermodynamics, 213
 freeze-drying
 parenteral preparations, 525, f526
 powders, 426, f426
 process rates, 527
 freezing points, 284

- frequency distributions, 134, 730, t135
friability, 18, 40, 587, 601, f601
Fried's Rule, 119
fructose, 692
fumaric acid, 692
function
 quality assurance control, 29
fusion, 577
future directions
 medicinal analytical testing, 19
GABA *see* γ -aminobutyric acid (GABA)
gamma radiation, 483
Gans, Richard, 347
gap analysis, 533
garbing, 534
gas adsorption, 419
gas sterilization, 480, 481, 482
gases
 liquid solution solubility, 227
 solutions solubility, 226
 states of matter, 205
GDP *see* guanosine diphosphate
gelatin, 611
 attributes/characteristics, 693
 emulsifying agents, 384
 emulsion, 452
 enteric film coatings, 615
 suspensions, 463
 tablet binders, 583
gelatinous precipitates, 369
gelation, 369
gelling agents, 462, 572
gels, 344, 458, 462
Gelsinger, Jesse, 654, 661
generic drugs, 660
genetic engineering, 654, 663
genetic tests, 82
GENE-TOX, 6
Geobacillus, sterilization, 473
Geomatrix technology, 629, f629
GFR *see* glomerular filtration rate
Gibbs free energy, 213
Gibbs' phase rule, 192
Gibbs-Helmholtz equation, 215
glass containers, 47, 504, 506, 677
glass measures, 111
glass transition temperatures, 199
glidants, 424, 584
global warming effect, 639
glomerular filtration rate (GFR), 719
gloves, 534
glucuronic acid, 752
glutathione (γ -glutamylcysteinyl glycine), 752
glycerin, 230, 436
glyceryl esters, 383
glyceryl monostearate, 693
GMP *see* Good Manufacturing Practices
gold, 365
Good Manufacturing Practices (GMP)
 see also current Good
 Manufacturing Practices, 32
gowning, 510, 514, 515, f515
graduated glass measures, 106, 111, f111
grain apothecary, 104
grain avoirdupois, 104
granulation, 69, 587
granulator top-spray process, 619
gravity filtration, 242
grinding mills, 427
growth
 microbials, 469
guanosine diphosphate (GDP), 732
guanosine triphosphate (GTP), 732, 737
guar gum, 693
gums, tablet binders, 583
hair, 565, 567
half-cycle method, 41, 472
handling
 pharmaceutical product stability, 37
 powders, 428
hand-shaping suppositories, 577
hard goods sterilization, 474
hardness, 40, 601
Haworth Projection models
hazardous compounding, 534, 536
HC *see* hydrocarbons
HCFC *see* hydrochlorofluorocarbons
heat
 capacity, 190, 201, 208
 emulsion preparation, 451, 452
 flow, 212
 pyrogens, 509
 syrup preparation, 444
heat of solution, 220, f221, 222, 228
heating states of matter, 189, f190
helium pycnometry, 418
Helmholtz free energy, 213
heme units, 318, f318
Henderson-Hasselbalch equation, 746
HEPA *see* high efficiency particulate arrestor
hepatic clearance, 717, 718
hepatic metabolism, 752
hepatic transporter activity, 720
heptafluoropropane, 645
Herschel-Bulkley model, 403
hexane, 332
HFA *see* hydrofluoroalkanes
HFC *see* hydrofluorocarbons
high efficiency particulate arrestor (HEPA) filters, 479, 511, 512, 513
high pressure syringes/pumps, 242
high-energy water molecule release, 318
high-speed rotary machines, 597
high-throughput screening (HTS), 84
history
 coatings, 611
 measures/metrolgy, 101
 molecular structure and properties, 177
 rheology, 393, 394
 weights and measures, 101
Hixson-Crowell equation, 63
holding, pharmaceutical product stability, 37
homogenizers, 455
Hopper flow rate, 422, f424
hormones, 653
host cells, 658
HPC *see* hydroxypropyl cellulose
HPMCAS *see* hydroxypropyl methylcellulose acetate succinate
HPMCP *see* hydroxypropyl methylcellulose phthalate
HP β CD *see* 2-hydroxypropyl- β -cyclodextrin
HTS *see* high-throughput screening
human factor of carefulness, 112
human gamma-glutamylcysteine synthetase, 254
humidification, gas sterilization, 481
humidity, 604
Hund's rule, 179
hydration, 257, 349
hydraulic presses, 249, 250
hydroalcoholic solvents, 230
hydrocarbons (HC), 571, t636
hydrochloric acid, 693
hydrochlorofluorocarbons (HCFC), 640, t636
hydrofluoroalkanes (HFA), 639, 645
hydrofluorocarbons (HFC), t636, 640
hydrogen bonding, 231, 318
hydrogen peroxide, 482
hydrolysis, 43, 305
hydronium ion concentration, 263, 264, 265, b265, t266
hydrophilic-lipophilic balance, 450, 451, t450
hydrophobic interactions, 318
hydroxypropyl cellulose (HPC), 694
hydroxypropyl methylcellulose acetate succinate (HPM-CAS), 615
hydroxypropyl methylcellulose phthalate (HPMCP), 615
2-hydroxypropyl- β -cyclodextrin (HP β CD), 325, 326
hygiene, 534
hyperalimentation, 282
hyperosmoticity, 281, 282
hyperresponsive response variations, 731
hypersensitivity, 731
hyporeactivity, 731
hyporesponsive response variations, 731
hyposensitive response variations, 731
hypotonic parenterals, 281
hypromellose, 694
hysteresis, rheology, 399
ICH *see* International Conference on Harmonization
ideal gas, 207, 210, f210
identity tests, 10, 15, 678
idiosyncrasy, drug action, 732
IDR *see* intrinsic dissolution rate
imipenem cilastatin, 756, f756
immediate release dosage forms, 17
immiscible liquids, 249
immune system biotechnology, 653
immunogenicity, 663
impaction, 414
impurities, 10, 12, t12, 436
in vitro procedures, 54, 75, 757
incompatibility
 pharmaceutical product stability, 42
indirect acting reversible agents, 723, f724
induction
 enzymes, 749
inference statistics, 141
infinite dilution, 259, t259
infinite gap instruments, 407
infusion, 715
ingredient considerations, 507, t507
inhalation
 see also metered-dose inhalers, 447, 660, 677
inhibition, 311
injection molding, 577
injections, 16, 496, 501
in-process fluids, 476
inserts, compressed tablets, 582
inspections, 508
instantaneous drug distribution/inputs, 708, 710, 713
instrumentation
 see also equipment, t405, 599
insulin, 660
Integrated Risk Information System (IRIS), 6
integrity tests, 486, f486, f487
intensity, pharmacologic effect, 722, 723, 724
interfacial film theory, 449
interfacial phenomena
 molecules, 329, 339, f339
 pharmaceutics, 329
 pressure differences, 334
 solid surfaces, 334, f335
 vapor adsorption, 334, f335
 work, 330, 331, 333
interionic compatibility, 44
interionic interference, 259
intermediate grinders, 427
intermittent compression filling, 606
intermolecular forces, 189
internal pressure, 236
International Conference on Harmonization (ICH), 31
International Journal of Pharmaceutical Compounding, 533
International Organization for Standardization (ISO), 511, 535, t535
intradermal injection, 743
intrinsic activity, 733
intrinsic dissolution rate (IDR), 424
intrinsic dissolution, 65
intrinsic efficacy, 733
inverse sine transformation, 174
iodine, 319, 443
ion product of water, 263
ionic bonds, 180, f180
ionic liquids, 275
ionic solutions, 274
ionization, 259
ion-suppression chromatography
ipeacac syrup, 445
IRIS *see* Integrated Risk Information System
ISO *see* International Organization for Standardization
isoelectric points, 269
isolation/isolators, 489, 513, 524
isomers, 184, f184, f185
isomotic, 233
isopropyl alcohol, 230
isopropyl myristate, 694
isopropyl, 436
isothermal compression, 210
isotonicity, 277, 285
Japanese Pharmacopeia (JP), 9
kaolin, 356, 694
Kauzmann temperature, 196, f196
Keesom forces, 205
Kelvin equation, 358
kidneys, 753
kilograms, definitions, 103
kinetic energy, 352
kinetics, 38, 722
 see also chemical kinetics
Koppel and Spiro's buffer capacity, 272
Kozeny-Carman relationship, 242
labeling, a34, 498, 530
laboratory controls, a35
laboratory media sterilization, 476
laboratory techniques, solutions of solids in liquids, 224
lactic acid, excipients, 695
lactose, 695
LAL *see* Limulus Amebocyte Lysate
laminar flow, 395, f396
lamination, 616
lanolin, 695
Laplace equation, 334
large quantity measurements, 110
laws
 law of mass action, 222, 302
 rules, 116
layered tablets, 581
Le Chatelier's principle, 260
leaching, 507, 508
leakage rates, 647
lecithins, 384, 695
level of significance, 142
Levofloxacin, 757, f758, 761
lidstock, packaging, 674
lifestyle drugs, 82
ligands, 735
light microscopy, 414
Light Obscuration Particle Count Test, 16
light transmission, 670, 679, f671, f672
light-houses, 653
Limulus Amebocyte Lysate (LAL) test, 528
line fitting, 168
linear measures, 104, 106, t104, t105
liniments, 448
lipid-to-water partition coefficients, 747
Liposomes, 254, 456
liquid biological waste, 477
liquid crystals, 202
liquid effects, 111
liquid measures, 105, 106, f106
liquid models, 402
liquids
 liquid solution solubility, 225
 moist heat sterilization, 475
 packaging, 674
 permeation, f669, t676
 sealed containers, 475
 solid solutions, 219
 solubility, 219
 solvents, 435

- liquids (continued)
 - states of matter, 189, 204
 - sterilization, 480, 482
 - viscosity, 204
- load mapping, 473
- load-elongation curves, rheology, 395
- loading, gas sterilization, 481
- location considerations, 134
- logarithms, 173
- London forces, 205
- lotions, 254, 464
- low temperature comminution, 426
- lubricants, 584
- lungs, 208, 212
- lyotropic liquid crystals, 202, f203

- M&A *see* merger and acquisitions
- mAb *see* monoclonal anti-bodies
- macromolecular agents, 658
- Magnesium Aluminum Silicate NF, 356
- magnesium stearate, 695
- maintenance
 - sterile preparation compounding, 533
- maltodextrin, 683, 695
- manned aseptic processing, 489
- mannitol, 696
- manometers, 41
- manufacture/manufacturing
 - soft elastic capsules, 607
- markets
 - sterilization, 487
- Master Production Batch Records (MPBR), 31
- materials
 - film coating, 614, 616
 - packaging, 667, 670, 679, t669, t670
 - parametric release, 488
 - rheology, 403
 - sterilization, 471
- mathematics
 - rate of dissolution, 64
 - receptor theories, 733
- Maxwell elements, 403, f403
- MCC *see* microcrystalline cellulose
- MDI *see* metered-dose inhalers
- mean
 - absorption time, 715
 - analysis of variance, 153
 - diameters, 411
 - ionic activity coefficient, 260
 - location parameters, 134
 - residence time, 715
 - transit time, 715
- measurements/measures
 - definition, 101
 - medicine droppers, 112
 - meniscus, 110
 - metallic measures, 111
 - metric systems, 102, 104, 105, t104
 - metrology, 110
 - pipettes, 106, 112
 - rheology, 405
 - significant figures, 114
 - small quantities, 112
 - systems, 101
 - teaspoons, 112
- MEC *see* minimum effective concentration
- mechanical mixers, 454
- mechanically modulated transdermal delivery systems, 574
- media
 - fill, 516, t516
- median, 134
- Medical Dictionary for Drug Regulatory Affairs (MedDRA)
- medicinal analytical testing
 - monographs, 10
 - oral dosage forms, 17
 - Pharmacopeial Conventions, 9
 - product performance, 16, 17, 18, 19
 - Regulatory Agencies, 9
 - release testing, 10
 - residual solvents, 13
- stability testing, 10
- standards, 9
- tablets, 17
- taxonomy, 15
- United States Pharmacopeial Convention, 9
- universal tests, 10, 17, 18
- MedDRA *see* Medical Dictionary for Drug Regulatory Affairs
- MEKC *see* micellar electrokinetic chromatography
- melt molding, 577
- melting point temperature, 195, f196
- membranes, 739
- Mendeleev's periodic table, 177
- mercury, 418
- merger and acquisitions (M&A), 90, t90
- metabolic engineering, 653
- metabolism, 721, 739
- metal containers, 48
- metallic conductance, 257
- meta-reactions, 732
- meter, weights and measures, 103
- metered-dose inhalers (MDI)
 - advantages, 633
 - nonaqueous solutions, 447
 - operation modes, 633
 - plastic-bag aerosol systems, 637
 - stability, 42
 - three-phase systems, 635
 - two-layer systems, 635
 - two-phase systems, 633
 - applications, t644
- method of residuals, 710, f710, f711
- methylation, 752
- methylcellulose, 463
- methylparaben, 696
- metric systems, 102, 103, 104, 109, t104
- metrology, 106, 113
- MIC ratio, 755, 756, f757, f761
- micellar electrokinetic chromatography (MEKC), 355
- microbials, 18, 469
- microcrystalline cellulose (MCC), 369, 583
- Microfluidizers, 456
- miconization, dissolution, 66
- micro-propagation, 663
- microscopy, 413
- Mie, Gustav, 347
- Milk of Magnesia, 465
- milliliter, 103
- Millipore's Waters Chromatography Division, 78
- mills, 459, f460, f461
- mineral oils, 696
- minimum effective concentration (MEC), 54
- minimum toxic concentration (MTC), 55
- minimum weighable quantities (MWQ), 110
- minoxidil, 570
- mixing emulsions, 454, f454
- MO *see* molecular orbitals
- mode
 - statistics, 134
- modeling/models
 - pharmacokinetics, 706
 - rheology, 402
 - viscoelasticity, 403
- modified transporters, 720
- modulated temperature differential scanning calorimetry (MTDSC), 202
- moist heat, 474
- moisture retention, 589
- molding suppositories, 577
- molecular complexes, 319
- molecular filtration, 255
- molecular orbitals (MO), 180
- molecular properties, 177
- molecular structure, 177
 - multiple chiral centres, 187, f187
 - polymorphism, 189
 - visualizing forces, 188
- molecular transport, 567
- molecular weight cutoff (MWCO), 255
- monoclonal anti-bodies (mAb), 31, 653, 664
- mortar and pestle, 452, 460
- motility, 744
- mottling, 616
- MPBR *see* Master Production Batch Records
- MTC *see* minimum toxic concentration
- MTDSC *see* modulated temperature differential scanning calorimetry
- multi-compressed tablets, 581
- multilayer rotary machines, 598
- multiple component solids, 197, f198
- multiple effect stills, 502, f503
- multiple prefixes, weight, 103, f104
- multispecies solute transport, 222
- mutant population suppression, 757
- MWCO *see* molecular weight cutoff
- MWQ *see* minimum weighable quantities

- NADPH *see* nicotinamide adenine dinucleotide phosphate
- nails, 566, 569,
 - ungual drug delivery, 569
- nanotechnology, 82
- nasal preparations, 42, 633
- natural emulsifiers, 451
- natural polymers, 355
- Natural Standard and Natural Medicines Comprehensive Database, 5
- NCE *see* new chemical entities
- ncogenes, 664
- nebulizers, 633
- negative thixotropy, 400
- nematic liquid crystals, 204, f204
- neonatal enteral medication, 281
- neonatal therapy, 281
- nephelometers, 347
- Nernst, Walther Hermann, 250
- net chemical reactions, 302
- Neulasta, 660
- neutral solutions, 263
- new chemical entities (NCE), 568
- Newton's laws, 393, 395
- Newtonian fluids, 395
- Next Generation Impactor (NGI), 415, f416
- nicotinamide adenine dinucleotide phosphate (NADPH), 749
- nicotinic receptor complex, 736
- nitrofurantoin, 56, f56
- nitrogen, 696
- nonaqueous solutions, 448
- noncompetitive antagonists, 733
- non-electrolytes, 232
- non-ionic surfactants, 341
- non-linear pharmacokinetics, 720
- non-Newtonian fluids, 397, 401
- nonparametric tests, 159
- non-polar solvents, 231
- nonrecoverable deformation, 403
- normal distribution, 134, 138, 139
- normal probability distribution, 138, 139
- normalization, 172
- Norton capsule machine, 608
- nucleation, 357
- nucleophiles, 312
- nucleotides, 651
- nutrition, 282
- occlusion, 567
- oil-in-water (O/W) emulsions, 379, 383, 385, 387
- oil-in-water-oil (O/W/O) emulsions, 379
- ointments, 571, 572
- oleic acid, 696
- oleyl alcohol, 697
- omeprazole, 735
- Oncaspar, 660
- one-compartment models, 708, f709
- one-sided tests, 144
- operation
 - single-punch machines, 595
- oral dosage forms, 326
- orbital labeling, 179, f179
- organic solvents, 611
- organization
 - quality assurance, a34
- orientation
 - colloidal dispersion adsorption, 359
 - interfacial phenomena, 330
 - oriented-wedge theory, 449
- oscillatory tests, 401
- osmolality, 278, 284
- osmosis/osmoticity, 277
 - parenteral medication, 281
- pediatric therapy, 281
- solutions, 232
- therapeutic considerations, 278
- tonicity, 277, 281
- total peripheral medication, 281
- undesirable effects, 281
- water, 436
- Ostwald ripening, 358, 459
- Ostwald-Freundlich equation, 358
- overkill methods, 472
- oxidants, 261
- oxidation, 42
- oxygen scavengers, 673
- ozone, 639
- packaging
 - liquid preparations, 437
 - parenteral preparations, 500, 508, 677, f504
 - permeation, 668, 669, 673, 679, f674
 - plastics, 677
 - quality assurance and control, 30, 678
 - standards, 667
 - transdermal products, 677
- pairing, 145
- pan-coating processes, 611
- paraffin, 697
- parallel layer flow, 394
- parallel plate instruments, 408
- parameters, statistical moments, 714
- parametric release, 488
- parenteral preparations
 - additional substances, 496
 - general considerations, manufacturing and compounding, 499
- particulate matter, 529
- personnel, 514
- preparation, 499
- production procedures, 517
- quality, 500, 528
- rubber closures, 506, 518
- safety tests, 530
- sealing, 523
- solutes, 496
- sterility tests, 528
- sterilization, 524
- storage, 530
- uniqueness, 495
- vehicles, 496, 499
- water, 496
- water-miscible vehicles, 496
- partial agonists, 733
- particles
 - characterization, powders, 411
 - colloidal dispersions, 343
 - density, 418
 - dispersion, 376
 - shape, 418
 - size
 - colloidal dispersions, 358
 - powders, 423, 424
 - reduction, 459
 - screening, 414
 - sedimentation, 415
 - particulates, 201
 - partition coefficients, 250
 - partitioning distribution, 747
 - partitioning effects, 388
 - parts sterilization, 474

- Pascalian fluids, 395
 pastes, 572
 patents, 660
 pathogens, 756
 Pauli Exclusion Principle, 179
 PBR *see* Production Batch Records
 PCA *see* Principal Component Analysis
 PCS *see* photon correlation spectroscopy
 peanut oil, 697
 pearl coatings, 611
 pectin, 697
 peeling, drug absorption, 716
 PEG *see* polyethylene glycols
 Pegasys, 660
 PEG-Intron, 660
 PEGylation, 660
 penetration
 topical dosage enhancers, 567, 568
 penetrometers, 409
 pentylenetetrazole response, 730, f731
 peptides, 653
 peptization, 356
 peracetic acid, 482
 percentage strength, 439
 percolation, 445
 percutaneous absorption, 566, 567
 periodic law, 177
 permeability, 54
 permeation, 508, 566
 personalized medicine, 82
 petrolatum, 697
 pH
 hydrolysis, 311
 sodium acetate, b268
 strong acids, 267, 264
 suspensions, 459
 weak acids, 267
 weak bases, 268
 zwitterions, 269
 pharmaceutical analysis
 product stability, 37
 quality assurance and control, 29
 testing, 9, 254, 280, 281, 325
 pharmacists, 658
 pharmacodynamics
 absorption, distribution, metabolism and excretion, 739
 bioequivalence testing, 53
 pharmacokinetics, 705
 pharmacoepidemiology, 84
 pharmacokinetics
 absorption, distribution, metabolism and excretion, 739
 orders of reactions, 706
 parenteral preparations, 498
 protein binding, 721
 reaction orders, 706
 renal clearance, 719
 renal excretion, 722
 steroechemical considerations, 721
 systematic availability, 715, 718
 transporters role in modifications, 720
 two compartment models, 710, f710
 volume of distribution, 708, 711, 714
 zero-order reactions, 707, f707
 pharmacology
 drug administration, 705, f706
 magnitude, 722
 reversible agents, 723, f724
 phase rule, 224
 phase solubility, 224
 phase transitions, 190
 phases, states of matter, 189, 190, 191, 192
 phenobarbital, 447
 phenols, 697
 phenylethyl alcohol, 698
 Ph. Eur *see* European Pharmacopeia
 phonophoresis, 574
 phosphatides, 340
 phospholipase C (PLC), 737
 phosphoric acid, 698
 photochemical reactions, 44
 photon correlation spectroscopy (PCS), 348
 physical measurements, 474
 physical properties
 suspensions, 458
 physiology
 absorption, distribution, metabolism and excretion, 739
 rectal administration routes, 574
 urethral administration routes, 576
 vaginal administration routes, 574, 575
 picking, film coating problems, 616, 617
 pigments, 613, 614
 placebo interfaces, 78
 planetary mixers, 429
 plasma membranes, 739
 plasma, states of matter, 206
 plasters, 573
 plasticity, 397, f398
 plasticizers, 614
 plastics, 48, 506
 plate instruments, 408, f408
 plate processes, 608
 plate-and-frame filter press, 245
 PLC *see* phospholipase C
 plume geometry, 647
 plungers, 508
 Poiseuille equation, 242
 Poiseuille's law, 396
 polar solvents, 231
 polishing coated tablets, 614
 poloxamer, 463, 698
 polyethylene glycols (PEG), 572, 660
 polymer chains, 398, f398
 polymers, 199, 237, 614
 polymorphism, 194, f195, 420
 polyoxyethylene block copolymers, 383
 polyoxyethylene glycol esters and ethers, 383
 polyoxyethylene sorbitan fatty acid esters, 698
 polyvinyl acetate phthalate (PVAP), 615
 polyvinyl alcohol, 699
 polyvinylpyrrolidone (PVP), 376, 584
 population mean, 134
 population statistics, 133
 porcelain, 484
 porcine striatal membranes, 254
 porosity, 612
 post-conditioning, gas sterilization, 481
 postglomerular blood vessels, 753
 post-sterilization maintenance, 478
 potable water, 501, f501
 potassium iodide, t230
 potassium phenoxymethyl penicillin, 58, f58
 potassium sorbates, 699
 potency, 37
 potential energy, 352
 poultices, 572
 powders
 production methods, 425
 skin, 573
 specific surface area, 419
 spray drying, 425, f425
 storage, 432
 practical equivalents, weights-measures, 105
 precipitation, 249
 precision, 135, f135
 pre-conditioning, gas sterilization, 481
 prednisolone levels, 60, f60
 preformulation stage, 37
 preparation
 liposomes, 457
 suppositories, 576
 suspensions, 378, 459
 syrops, 444
 prescription balances, 108
 prescriptions
 solubility, solutions of solids in liquids, 224
 preservatives, 437, 579
 Principal Component Analysis (PCA), 171
 printers/printing, 614
 probability statistics, 138
 process dwell times, 473
 process sequences, 478
 procurement
 parenteral preparations, 500
 Production Batch Records (PBR), 31
 properties
 molecular properties, 177
 proportion tests, 147
 proportions, buffer ionic solutions, 271
 propylene glycol, 230, 570
 propylparaben, 699
 protein aggregation, 520
 proteins, 340, 651, 657
 protocols
 product analysis, t39
 Protolysis, 261
 pseudoplastic liquid, 204
 pulmonary clearance, 717
 pulsatile release, 582
 pumps, 245, f245
 punches, 583, 584
 purified water, 436
 purity
 see also separation
 water for injection, 504
 putty (Silly Putty), 401
 PVAP *see* polyvinyl acetate phthalate
 PVP *see* polyvinylpyrrolidone
 Pycnometry, 418, f419
 pyrogens, 509
 quality assurance (QA), 32
 quality control (QC), 163, f163, f164
 quality of life (QOL), 82
 quantitative analysis, 236
 quantum numbers, 178, 179
 questionnaires, statistics, 135
 R&D *see* research and development
 RABS *see* restricted access barrier system
 radiation
 sterilization, 483, 484, 524
 rank sum tests, 159, t160
 ranked means, 153
 rate constants, 47, 302
 rate of solution, 220
 raw film coatings, 614
 Rayleigh scattering, 347
 rDNA *see* recombinant DNA
 receptors, 732
 recombinant DNA (rDNA), 31
 re-conditioning, gas sterilization, 481
 reconstitution, medicinal analytical testing, 17
 recrystallization, 249, 254
 rectal administration, 574, 575, 576
 redistribution, nonequilibrium, 748
 refractory patients, 732
 related substance impurities, 12
 relative retention time (RRT), 15
 release
 see also delayed-release dosage forms; extended-release..., 462
 renal excretion, 753
 reproducibility, coatings, 619
 research, 83
 research and development (R&D)
 drug development/discovery, 81
 restricted access barrier system (RABS), 489
 returned drug products, a35
 reverse osmosis, 502
 rheology
 fundamentals, 393
 non-Newtonian fluids, 397
 rheometers, 405, t405, t406
 ball type instruments, 409
 rolling ball instruments, 408
 rotational instruments, 407
 slit instruments, 407
 types, 405
 tensile testers, 409
 torsion testers, 409
 viscometers, 405, t405, t406
 viscosity, 393, 395
 rheopexy, 400
 risk
 assessments, 667
 robustness, 79
 Rodac plates, 515
 roller presses, 249, 250
 rolling
 powder comminution, 427
 suppository preparation, 577
 rolling ball instruments, 408
 rotary machines, 596, f597
 rotary-drum vacuum filters, 245, f246
 rotational instruments, 407
 rotor stators, 454, f455, f456
 RRT *see* relative retention time
 rubbing alcohols, 436
 rules
 addition, 116
 division, 117
 multiplication, 117
 subtraction, 117
 Rutherford, Ernest, 177
 Rydberg constant, 178
 saccharin, 699
 salts, 193, 222
 sampling, 133, 142
 sanitization, 514
 SBE β CD *see* sulfobutylether β -cyclodextrin sodium salt
 scales, metrology, 106, f107
 SciFinder Scholar database, 3
 screens, 589, 590
 sealing, 612
 sebaceous glands, 565
 sedimentation, 252, 373, 374
 segregation, 429
 semipolar solvents, 231
 semisolids, 637
 sensitivity, 731
 separation
 see also chromatography; dialysis; diffusion; reverse osmosis
 countercurrent distribution, 250
 straining, 253
 tangential flow filtration, 255
 washing, 253, 254, f253
 Sephadex gels, 254
 serum osmoticity, 281
 shape
 hard gelatin capsules, 605
 tablets, 581
 shear
 rheology, 394
 mixers, 587, f589
 powder mixing, 428
 rheology, 402
 strain, 394
 stress, 395, 399, 401
 thickening, 397
 thinning fluids, 397, 399, f397, f399
 shelf-lives, 619
 shell labeling, 179
 shellac-based seal coats, 612
 shikonin, 662
 shRNA *see* synthetic or expressed small hairpin RNA
 sieve retention, 485
 sign tests, 161
 signed-rank sum test, 159
 significance level selection, 142
 silicon dioxide, 369
 silver iodide, 350, 352, 352, f351
 single-beam equal-arm balances, 106, f107
 single-punch machines, 595
 sink-condition equation, 65
 size issues, 604
 SIP *see* sterilization-in-place
 skin, 565, 567
 skin-flap methods, 569

- slit instruments, 407
- Slit-To-Agar (STA) quantitative air samplers, 515
- slopes
 - comparison statistics, 170
- sludge clarification, 252
- smoothing compressed tablets, 613
- soap emulsifying agents, 382
- soda-lime glass containers, , 504
- sodium chloride equivalent method, 285
- sodium chloride, 257, t230
- sodium lauryl sulfate, 700
- sodium metabisulfite, 700
- sodium starch glycolate, 700
- sodium stearyl fumarate, 700
- solid dispersions, 238
- solids
 - states of matter, 189, 193, 195, 197
 - in liquids, 219, 222
 - in solids, 227, f227, f228
 - solubility, 227
- solid-wall coating pans, 617
- sols, 344
- solubility
 - bioequivalence testing, 54
 - colloidal dispersions, 343, 349, 358
 - descriptive terms, t220
 - determination, 224
 - fluorinated hydrocarbons, 640
 - parenteral preparations, 498
 - rules, 232, t232
 - solutions, 219
 - surfactants, 238
 - suspensions, 458
 - ternary liquids in liquids, 226, f227
 - thermodynamics, 228
- solutes
 - countercurrent distribution separation, 250
 - pyrogen sources, 509
 - solution pharmaceuticals, 219
 - solutions, 228
- solutions
 - compressed tablets, 582
 - liquids in liquids, 225
 - phase equilibria, 219, 229
 - solution solubility, 225
 - solvents, 219
 - stability, 41
- solvent/solvation, 197
- solvents
 - film coating, 614
 - freezing points, 284
 - immiscible liquid separation, 250
 - topical skin dosage forms, 567
 - types, 229
- somatic cells, 661
- Somavert, 660
- sonophoresis, 574
- sorbic acid, 685, 700
- sorbitan esters, 700
- space sprays, 634
- spare receptors, 733
- spectroscopy
 - see also individual techniques, 188, 200
- spheronization, 593, 594
- spiral twist press, 249
- spontaneous changes, 211
- spontaneous reactions, 301
- spray drying, 594
- spray patterns, 647
- sprays
 - compressed gas aerosols, 637
- spread parameters, 134
- spreading coefficients, 331
- Square Meter Surface Area Method, 119
- square-root transformation, 174
- STA see Slit-To-Agar
- stability
 - microemulsions, 453
 - soft elastic capsules, 607
- suspensions, 41
- tablets, 40
- standards
 - packaging, 668
- starches, 701
- states of matter, 195, f196, 203, f203
- statistics
 - straight lines, 166
 - statistical inference, 141
 - stem-and-leaf plots, 133
 - two-sided tests, 144
 - variance, 134
 - Z transformation, 140
 - t-test, 145, 170
- steam sterilization, 474, f474, f475
- stearic acid, 701
- stearyl alcohol ,701
- stereoisomers, 735
- sterile stabilization, 376
- sterile preparations, 533
- sterility, 488
- sterilization, 469, 471, f472, 477
- sterilization-in-place (SIP), 477, 478
- stirring element alignment, 71
- Stokes' law, 248, 373
- storage, 663
- strain, 394, 395
- stratum corneum, 566, 567, 743
- stratum, 565
- stress
 - relaxation, 587
 - rheology, 394, 395
 - transmissions, 587
 - viscoelasticity, 401, 402, 403
- stripping, drug absorption, 716
- structure
 - molecular structure, 177
- structured vehicles
 - suspension formulation, 374
- Student's t-test, 46
- subcoatings, 612, t613
- subsidence, suspensions, 374
- sucralose, 701
- sugar coatings, 581, 619
- sulfamethoxazole hydroxylamine, 723, t723
- sulfobutylether β -cyclodextrin sodium salt (SBE β CD), 326, t325
- sulfonates, 383
- sun screens, 566
- supersensitivity, 731
- suppositories, 576
- surface area, 744
- surface tension, 332, f332, f333, 449
- surface-coating spray, 634, f634
- surfaces, colloidal dispersions, 343
- surfactants, 339, 441, 568
- surroundings, thermodynamic definition, 207
- suspensions, 372, 376, 462, 463
- swinging-bucket rotors, 248
- synthetic emulsifiers, 451
- synthetic or expressed small hairpin RNA (shRNA), 661
- system, thermodynamic definition, 207
- systems
 - weights, 101
- tablets
 - dissolution, 52
 - syrup preparation, 445
 - thickness, 601, f601
 - validation, 603
 - wet granulation, 587
- tack, compressed tablets, 612
- tapped density measuring devices, 422, f423
- tartaric acid, 702
- taste masking/problems, 444
- technetium Tc-99m, 365
- temperature
 - colloidal dispersions, 349
- states of matter, 190
- van't Hoff equation, 320
- vapor pressure, 190, f191
- terminal disposition constant, 711
- testing/tests, 529
- tetrafluoroethane, 645
- theophylline, 447
- therapeutics, 42, 51, 487, 731
- thermodynamics, 262
- thermograms, 202
- thiopental, 748
- thixotropy, 399, f399
- Thomson, J.J., 177
- threshold dose, 729
- thrombophlebitis, 282
- time
 - of peak concentration, 51, 55
 - time-dependency of
 - non-Newtonian fluids, 398
- tin plated steel containers, 642
- tip-seals, 523
- titanium dioxide, 702
- titrations, 271
- tolerance
 - drug action-response variations, 732
- tonicity, 285, 497
- top selling drugs, t90
- topical dosage forms
 - delivery, 569, 576
 - dosage forms, 565
 - emulsions, 448
 - skin, 565, 566
 - transdermal drug delivery, 567
 - urethral administration routes, 574
- torsion balances, 108
- Total Product Quality (TPQ), 30
- Total Quality Management (TQM), 30
- toxicity, 755
- Toxline database, 3
- TP see triple point
- TPQ see Total Product Quality
- TQM see Total Quality Management
- trace elements, 437
- tragacanth, 463, 702
- transdermal drug delivery, 566, 567
- transference number, 259
- transgenic cell cultures, 663
- transition state theory, 308
- tray drying, 589
- triacetin, 702
- triethanolamine, 702
- triple point (TP), 191, f191
- trisulfapyrimidines oral suspension, 460
- trituration, 378
- Troy weight, 102, 104, f104
- tubular secretion, 753
- Turbidity, 347
- Tussionex Pennkinetic system, 462
- two independent sample t-tests, 145
- Tyndall effect, 347
- UDP-glucuronosyltransferase-2B7 (UGT2B7), 748
- ultracentrifuge, 41
- ultraviolet (UV) lamps, 514
- unequal-arm balances, 107
- univalent ionic strength, 260
- uniform dosage metered-dose inhalers, 647
- uniformity problems, 616
- unit cell characteristics, 194
- United Kingdom (UK), 101
- United States Pharmacopeia (USP), 74, 667
- univalent ions, 260
- universal analytical tests, 11, 17
- USP see United States Pharmacopeia
- UV see ultraviolet
- vaccinia virus, 653
- validation
 - enantiomeric purity, 14
- identity tests, 15
- pharmaceutical analysis, 11
- residual solvents, 14
- valve-bag aerosol systems, f638
- van der Waals equation, 205
- Van Slyke's buffer capacity, 272
- van't Hoff factor, 257
- vanillin, 703
- vapor pressure, 284, 334, t334, 640
- vapour sterilization, 480, 482
- veegum, 384
- vehicles
 - colloidal dispersions, 343
- velocity gradients, rheology, 394
- vertical rotors, 249
- vibration
 - dissolution rate effects, 71
 - powders, 414
 - states of matter, 189
- viruses, 653
- viscometers
 - capillary instruments, 406, f406
 - coaxial-cylinder instruments, 407, f408
 - falling ball instruments, 408
 - types, 405
- viscosifiers, 449, 459
- viscosity, 72
- visual appearance, solution stability, 41
- vitrification, 196
- Voigt-Kelvin elements, 403, f403
- volume-in-volume percentages, 122, 124
- water
 - attributes/characteristics, 703
 - interfacial phenomena, 332
 - phase diagrams, 190, f191, 192, f192
 - solution solubility, 229
 - tablet dosage forms, 18
 - topical skin dosage forms, 567
- water-in-oil (W/O) emulsions, 385, 387
- water-soluble dyes, 613
- water-soluble pigments, 613
- Watson, James, 651
- waxes, 611
- Weber-Fechner law, 730
- weight
 - definition, 101
 - English systems, 104
 - specific gravity, 113
 - tablet stability, 40
 - weighing pharmaceuticals, 107, 109, 114
 - weight-in-volume percentages, 122
 - weight-in-weight percentages, 122, 123
- Weissenberg effect, 402
- Weissenberg number, 402
- wetting
 - powders, 424
 - suspension preparation, 378
- Wheatstone bridge apparatus, 257, f258
- white lotion, 464
- whole genome sequencing, 653
- wine measures, 101, 102, 105, t105
- wire weights, 110, f110
- work
 - elastic solids, 395
 - hardening, 395
- Wurster process, 618
- xanthan gum, 703
- X-ray diffraction (XRD), 18
- X-ray powder diffraction, 201, f201
- X-ray techniques, states of matter, 201, f201
- Young's modulus, 393
- Young's Rule, 119
- zero-order drug absorption, 715
- zeta potentials, 351, 372