

CRYSTAL STRUCTURE PREDICTION AND POLYMORPHISM

Joel Bernstein

Dept. of Chemistry, Ben-Gurion University of the Negev, Beer Sheva, Israel 84120

1. ABSTRACT

to predict: to foretell, prophesy, announce beforehand
Oxford English Dictionary

Since the early days of structure determination by diffraction methods the prediction of crystal structures of small to medium-sized organic molecules from a structural formula has been a goal and a constantly recurring theme of crystallographers. The challenge is even greater when considering the additional factor of the possible existence of polymorphic forms of the material, since true predictive capability requires accounting for all crystal forms and their relative stability. While considerable progress has been made and some practitioners have even claimed that the goal has been achieved, a great deal still needs to be learned and done before one can state with any degree of confidence the ability to predict a crystal structure from a structural formula alone, much less its polymorphic forms. An even greater challenge is the prediction of solvates, including hydrates. This contribution attempts to summarize some of the current thinking on the subject.

2. INTRODUCTION

At the outset to this contribution, I want to make it clear that I am an enthusiastic believer in the importance of computational approaches to the understanding of interactions within molecular crystals and their potential for aiding in the solution of crystal structures, as well as for surveying the energy landscape for possible, even as yet unknown, structures. I have been following the development of this field for nearly forty years, even with occasional personal forays into it. I am impressed by the progress that has been made, and I enthusiastically support and encourage those who are attempting to develop and refine the field in general and its applications to polymorphs in particular. The rapidly increasing interest in polymorphism in recent years, from the purely scientific point of view to the commercial and intellectual property aspects has led to increasing (perhaps more correctly intense) interest in our ability to predict *ab initio* the existence of multiple crystal forms for a particular compound. I do think that some of the hyperbole and optimism of the past decade concerning our ability to predict even the possible existence of polymorphs was misplaced, and this paper is an attempt to present my perspective on the current situation. Although I have attempted to be objective, this minireview is the view from one person's perspective, which must per force carry some bias.

As with many aspects of chemistry, much discussion and debate surround our definition of terms, and the associated frame of reference. In this case, what do we mean by the term "to predict"? The Longman's Dictionary has a slightly different definition from that given by the OED above: "to see or describe in advance as a result of knowledge, experience, thought, etc." Using classical physics we can predict the path of a cannon shell or an interplanetary space probe. In chemistry we can often predict the outcome of a reaction or the bonding geometry around an atom. In the current context I view a prediction as the ability to see or describe in advance a crystal structure simply on the basis of its molecular structure. The key phrase in this definition is 'in advance'. For a sound scientific prediction we don't have to do the experiment to know what the result will be. It may be technologically challenging or difficult to carry out the experiment, but we expect the known scientific truths to be valid and the accepted scientific laws to be obeyed. By either dictionary definition above the notion of a prediction in scientific terms carries with it some degree of confidence, implied in the Longman's definition "as a result of knowledge, experience, thought, etc." That is, we expect the prediction to be proven.

In the context of the historical development of “predicting” crystal structures of small to medium molecules, 1994 was a particularly notable year, with two nearly simultaneous publications. One was the official announcement by Molecular Simulations Inc. (now Accelrys) in July of the availability of the “Polymorph Predictor” program [1]. The second event was the publication in October of the stimulating paper by Angelo Gavezzotti [3] entitled “Are crystal structures predictable?” to which he gave the emphatic one word opening response “No!” The dichotomy is glaring. How can there be a program for *predicting* multiple crystal structure (i.e. polymorphs) if we can’t even predict a single crystal structure? Part of the problem arises, of course, from how each author interprets of the definition of *predict*, and we will attempt to address that aspect of the question as well as making a general assessment of the current situation as seen from the perspective of this viewer.

3. SOME HISTORICAL BACKGROUND

In the current context of automated CCD diffractometers and “black box” programs for the solution, refinement and presentation of crystal structures it may be difficult to imagine (or to recall) the days when the few text books on crystal structure analysis (e.g. 4,5) contained whole chapters on “Trial and Error Methods” for solving crystal structures, more correctly for providing the initial model. In many ways these were essentially means for describing the crystal structure in advance as a result of knowledge experience, thought, etc., i.e. predicting the crystal structure. Much thought and effort went into these methods, since the cost in time and work (often on the order of months) in getting it wrong could be frustrating, if not disastrous. With the advent of powerful desktop or laptop computers armed with efficient software for structure solution (generally by direct methods) and refinement, those chapters have disappeared from the textbooks. However, the earlier need to use every means possible to arrive at a possible crystal structure spawned some pioneering work and led to considerable understanding of the nature of crystals and crystal structures.

The classic work in this area is Kitaigorodskii’s *Organic Chemical Crystallography* [6], which contains many still-valuable observations and insights on the nature of intermolecular interactions in molecular crystals. Kitaigorodskii also built a “structure seeker” for the prediction/solution of crystal structures by incorporating the principles of van der Waals radii and closest packing into a mechanical model. In the middle 1960’s many of these ideas were incorporated by (the late) Don Williams into his PCK5 program [7], various versions of which he used as an aid for solving the structures of dibenzoylmethane [8], 2,4,6-triphenylverdazyl [9] and others employed in solving the structures of some monosaccharides [10].* A

*A personal note. In 1968 I was a postdoctoral fellow in Ken Trueblood’s lab at UCLA. He was working with Don Cram on the structures of paracyclophanes. One of these that came into the lab was the hydrocarbon [2.2]*meta,para* cyclophane, which crystallized in the space group $P2_12_12_1$. In those days (before the birth of MULTAN, SHELX, etc.) such a light atom noncentrosymmetric structure was considered non-routine at best and often not solvable, and it might have been laid to rest. However, the rigid molecule is very similar to [2.2]*para,para* cyclophane which crystallizes in tetragonal space group $P4_2/mnm$, and we had been reinvestigating the structure originally reported by Kathleen Lonsdale [11]. The cell constants of the two structures were also very similar; however, whereas the *para,para* derivative had crystallographic site symmetry *mmm* (a very rare case for molecular crystals [6]), in space group $P2_12_12_1$ $Z'=1$, the *meta,para* derivative was required to have crystallographic site symmetry *1*. The four arrangements possible for the *meta* and *para* rings were generated as input into the Williams PCK5 program at the locations suggested by the tetragonal structure. One of the four produced a significantly lower packing energy than the others as a trial structure, and refined smoothly on data collected manually on a General Electric quarter-circle diffractometer. However, the exocyclic C-C single bonds were approximately 1.62Å, a value much larger than any previously published. Another set of data led to identical results, which we felt could not be published due to this anomalously long bond. Sometime later, Renault *et al.* published another structure of the compound [12, also] in $P2_12_12_1$, but with $Z'=2$, in which 3 out of 4 of the C-C single bonds exceeded 1.54, and subsequently many C-C single bonds exceeding the ‘standard’ 1.54Å have been reported.

parallel development at the time was Bill Busing's WMIN program [13], which was used by Kwick and Noordik to aid in solving the structure of 2-amino-4-methylpyridine [14].

The field has expanded rapidly over the past thirty years, with an increasing number of practitioners developing a variety of methods. At least three useful summaries with many references to the historically important and current developments have been given by Verwer and Leusen [2], Gavezzotti and Filippini [15] and Price [16]. Price et al. also compiled a useful 'atlas' of organic molecules which have been used in crystal structure prediction studies carried out by lattice energy calculations [17]. Of the ~190 structures reported in this survey 9 are denoted by the authors as having been carried out "in advance of an experimental investigation" – much in the way that traditional predictions would be made.

The title of this paper involves the confluence of two subjects – the calculation of crystal structures on the one hand and polymorphism on the other hand. Some historical developments of the former were noted above. In many ways the beginning of modern systematic study of polymorphism also can be traced to the 1960's with two landmark papers by McCrone [18, 19]. In the first of these McCrone made an oft-quoted provocative statement about the occurrence of polymorphism:

"Every compound has different polymorphic forms, and the number of forms known for a given compound is proportional to the time and energy spent in research on that compound."

In spite of the implied universalism of this statement, there are many very well researched compounds for which polymorphism has not been discovered: naphthalene, sucrose and the pigment perylene red are but a few notable examples. Statistics on the occurrence of polymorphism are hard to come by, because almost any data set is biased, and many instances of polymorphism are not properly reported (this includes false positives as well as false negatives). For instance, a survey of the Cambridge Structural Database on the keywords "polymorph," "form," "phase" and "modification" indicates that about 4% of the included structures are polymorphic. A search on "hydrates and "solvates" results in about 23% hits. This survey is biased by the fact that it includes only compounds for which the crystal structures have been reported. Perhaps closer to the other extreme, Steve Byrn recently summarized work from SSCI on over 150 compounds for which >87% had more than one form, 37% were hydrates and 31% were solvates [20]. These statistics may be biased by the fact that some samples at least were known to be polymorphic when received, or special efforts were made to identify multiple crystal forms, even if relatively unstable. Additional statistical surveys have been reported [21].

The general situation is somewhere in between these two sets of statistics: the existence of multiple crystal forms (polymorphs and/or solvates) is not surprising when found, but cannot be foreseen or expected for any compound. Essentially, this has been the state of our knowledge and understanding for many years, and given the increasing interest and importance of multiple crystal forms, it is not surprising that considerable efforts would be made to "predict" the possible existence of polymorphs in particular, and multiple crystal forms in general. One result was the announcement of the "Polymorph Predictor" program in 1994.

4. WHAT WE CAN DO; WHAT WE CAN'T DO

In the introduction to a chapter in a 1997 book entitled *Theoretical Aspects and Computer Modeling of the Molecular Solid State* edited by Angelo Gavezzotti, Robert Gdanitz, one of the authors of the Polymorph Predictor program, quoted three authorities on their view of our ability to predict crystal structures [22]. In a since oft-quoted statement John Maddox, past editor of *Nature*, wrote in 1988 [23]:

"One of the continuing scandals in the physical sciences is that it remains in general impossible to predict the structure of even the simplest crystalline solids."

Similar skepticism was voiced by F. Hawthorne [24]:

"Thus a rigorous general solution to the question of crystal structure prediction may not be forthcoming in the near future."

and Fagan and Ward [25]:

“Unfortunately, the molecular interactions in a lattice become too complicated for any human or computer to handle.”

Nevertheless, in the same opening paragraph, Gdanitz [22] concluded that:

“...quite a few articles on predicting molecule crystal structures have been published, showing that *the problem has now become a rather routine task*” (emphasis added).

Perhaps the telltale hint of the author’s true estimation of one’s predictive ability is the title of the chapter, “*Ab Initio* Prediction of Possible Crystal Structures”. The prediction of *possible* crystal structures may be routine, but the prediction of *the* crystal structure clearly is not.

A similarly over optimistic view of the situation may be found in the chapter by Verwer and Leusen [2] in stating that

“...computational methods and algorithms that allow packing simulations with many degrees of freedom,...were not developed until this decade (e.g. the 1990’s), making true ‘ab initio’ crystal structure predictions possible on the basis of molecular information alone.”

Possible, perhaps, but the rate of success still leaves a lot to be desired before it can be claimed that the problem has been solved, and crystal structures in general, or polymorphic structures in particular, can be reliably predicted. In reviewing the general situation of the predictability of polymorphs in general, Gavezzotti’s emphatic “No” in 1994 and subsequent developments, Dunitz succinctly summarized the situation in 2003. His current answer to the question is [26]:

“...still ‘No’, although at certain levels of discussion a ‘Maybe’ or even a conditional ‘Yes’, may be entertained as possible responses”.

Perhaps the greatest service to this effort of developing methods to computationally survey possible crystal structures has been performed by the Cambridge Crystallographic Data Centre, with a series of competition “blind tests.” Developers and practitioners of these methods have been invited three times since 1998 to apply their methods to attempting to predict the crystal structures of three compounds on the basis of molecular information alone. The crystal structures of the candidate compounds have been done but not published. The results of two of these blind tests have been published [27, 28]. The rules for the competition say as much about our current ability to computationally predict crystal structures as do the results of the competition. These rules were laid out, “[t]o give a reasonable chance of success within the practical limits of known computer programs*...”:

- maximum number of atoms including H atoms did not exceed 40;
- space group required to be one of the ten most frequently encountered (i.e. $P2_1/c$, $P-1$, $P2_12_12_1$, $C2/c$, $P2_1$, $Pbca$, $Pna2_1$, Cc , $Pbcn$, $C2$);
- one molecule in the asymmetric unit;**

* The ground rules for the second blind test are quoted here. They differ only slightly from the first.

** Most of the programs used for computing the lattice energy can not take into account the possibility of the molecule lying on a crystallographic special position (Wyckoff position). In a recent survey [29] it was shown that for $Z'=1$ approximately 71% of the molecules lack molecular symmetry, while those that have an inversion center retain that in more than 80% of the cases, those with a rotation axis retain it in approximately 50% of the cases, while those with mirror plane retain that element in about 30% of the cases. Most of these structures with retention of symmetry would not be included in the computed crystal structures since the programs are not capable of handling them. In some algorithms, especially those that compare the calculated powder pattern with an experimental one, the proper spatial arrangement of the molecules might be found, but the space group would not be correctly assigned. Such was the case, for

- no solvent molecules or co-crystals;
- no disorder in the determined structure;
- hydrogen atoms located experimentally in the determined structure.

In addition, the three candidate molecules were categorized by perceived increasing difficulty for prediction:

- rigid molecule with only C, H, N and O atoms, less than 25 atoms;
- rigid molecule with some less common elements (e.g. Br), less than 30 atoms;
- flexible molecule with two degrees of acyclic torsional freedom, less than 40 atoms.

What classes or features of structures are considered too difficult or too challenging for such a competition? “Medium” or “large” molecules (>~40 atoms), all of the metal-containing complexes, most of the compounds with less common elements, molecules with more than two conformational degrees of freedom, salts, solvates, and structures with more than one molecule in the asymmetric unit, or molecules lying on special crystallographic positions (inversion center, 2-fold axis, etc.). While the eight allowed space groups comprise approximately 80% of the CSD, by adopting such a restriction one is already reducing one’s chance of success by 20%. This is a considerable fraction of the compounds that are of interest to “small molecule” chemists.

What kind of success was achieved for the two blind tests for which results have been published? There were three molecules (but four structures – see below) and 11 participants in the first blind test, for a total of 44 chances of success. Each participant was allowed three choices for each solution, so actually there were 132 possible solutions. There were seven correct predictions, five of them being the first of the three choices. 7/44 is a 16% success rate.

It turned out that the behavior of one of the compounds was particularly relevant to the title of this paper. The compound is a liquid at room temperature, and was crystallized (in orthorhombic *Pbca*) *in situ* at low temperature in a capillary on a diffractometer for the data collection and structure determination, which served as the first structure of reference for the first blind test. A breakdown in the cooling system led to melting, and all subsequent attempts at crystallization yielded a monoclinic *P2₁/c* structure [29,30]. The sequence of events suggests that the monoclinic form is the more stable one, in accord with Ostwald’s Rule of Stages, and that the orthorhombic form should be classified as a disappearing polymorph [31]. In terms of the computational prediction of crystal structures, the important point here is that four out of the seven correct responses in the first blind test were for the orthorhombic structure (the less stable form) and if it had not been found initially, the number of correct predictions would be reduced to three (7% success rate). Moreover, if the monoclinic structure had initially crystallized one might never have known about the existence of polymorphs in this system.

Seventeen groups participated in the second blind test, carried out on three structures for a total of 51 possible correct answers. The structures were correctly predicted by two groups for the first molecule, by four groups for the second molecule, with no correct answers for the conformationally flexible third molecule. The success rate is therefore 6/51 or 12%. The participants did conclude that there had been significant progress on the ability to generate crystal structures for rigid molecules, and if the solution were to be found among the top six, rather than the top three, solutions the success rate would rise significantly. The failure of all participants to find a solution for the third molecule in the second round led to additional attempts to prepare a polymorphic structure which might be found among the higher ranked computed ones [32]. The authors determined that the blind test reference structure exhibited an unusual conformation, which made the calculations less likely to succeed. After determining the cell constants of more than 80 crystals they obtained a second polymorph of the material crystallized from nitromethane; however, it crystallizes with two molecules in the asymmetric unit with two different molecular conformations, so that

instance, for the original version of the “Polymorph Predictor”, which yielded a solution for benzene in space group *P2₁2₁2₁* (*Z*=4, *Z'*=1, molecules on general positions), when it should have been *Pbca* (*Z*=4, *Z'*=0.5, molecules on inversion centers).

it would not have been included in any of the possible structures for the blind test calculated at the current level of capability.

The example of acetic acid demonstrates one of the major difficulties and challenges still to be overcome. An attempt to computationally predict the crystal structure(s) of acetic acid using both the grid-search method [33] and the Monte Carlo approach [34-36] led to ~100 structures within a narrow energy range of 5 kJ/mol. With various modifications the experimental structure always had a low energy, but it was not possible to distinguish it from the many other solutions. This led the authors to suggest that “other factors need to be considered for genuine structure prediction.” In the realm of synthetic organic chemistry, a parallel situation would be to claim that one has carried out the synthesis of the desired product which indeed may be found in the reaction mixture along with ~100 other products, but we don’t know how to independently identify or isolate that desired product.

All the methods for calculating lattice energies of possible crystal structures lead to a list that must be ranked. Ideally, the real or possible polymorphic forms appear as the most stable forms in that list. Historically, success has been measured by whether a structure, or polymorphic structures can be properly accounted for, or predicted in that list and subsequently found. I would like to suggest an alternative developmental approach, which I believe has been applied to a much lesser extent, but may be more systematic and fruitful in terms of the long term development of these methods.

The polymorphic systems to be studied should be very well characterized in terms of the relative thermodynamic stability of the various forms, monotropism, enantiotropism, heats of transition, and heats of sublimation, etc., if possible. In the end, the lattice energies that are calculated can and should be benchmarked against those experimental quantities. I do not underestimate the difficulties in obtaining good quality thermodynamic data for all of these quantities, especially sublimation enthalpies, and the paucity of laboratories where such measurements are made. The least demanding benchmark for one of these calculations is that it reproduces the experimentally determined energetic ranking of the various polymorphic forms, without regard to the actual quantities involved. At the next level of stringency, the computed differences in energy between the various polymorphic forms should conform to the experimentally determined differences. As the most demanding test, the computed lattice energies should match the experimentally determined ones. Such a developmental process should result in a force field and search procedure that are at least suitable for a class of functional groups and molecules similar to those of the test molecule or group of molecules. Attempts to develop ‘universal’ methods or force fields are reminiscent of the search for the Holy Grail.

The computational “proof of concept” experiment would then be to use that force field and those computational methods to generate, *ab initio*, the crystal structures of all of the known polymorphs in the energetic order in which they are known to exist. Suitable (albeit demanding) test systems, might be a compound like ROY, for which seven polymorphs have been identified and characterized [37,38], or *p*’-methylchalcone, for which 13 polymorphic forms have been reported [39] and at least 5 structures have been determined [40].

Table 1
Crystal forms obtained for N,N'-dithiobisphthalimide [from 41]

Form	Solvent for crystallization	Space group	Z'
Solvent-free polymorphs			
(1)	CH ₃ COOC ₂ H ₅	<i>P2₁/c</i>	2
(4)	CH ₃ OH	<i>Pna2₁</i>	1
(5)	CH ₃ CN or HCON(CH ₃) ₂	<i>P2₁/c</i>	2
Polymorphs with solvent hydrogen-bonded to (I)			
(6)	CH ₃ NO ₂	<i>P2₁/c</i>	1
Solvent molecules are in isolated cavities			
(7)	C ₆ H ₅ Cl	<i>P1̄</i>	1
(8)	C ₆ H ₅ C ₂ H ₅	<i>P2₁/c</i>	2
Solvent molecules are non-localized in channels			
(2)	CH ₂ Cl ₂	<i>C2/c</i>	0.5
(3)	C ₂ H ₅ OH	<i>C2/c</i>	0.5
(9)	C ₄ H ₈ O	<i>C2/c</i>	0.5
(10)	(CH ₃) ₃ COH/H ₂ O	<i>C2/c</i>	0.5
Solvent molecules are localized in channels			
(11)	1,4-(CH ₃) ₂ C ₆ H ₄	<i>P2₁/n</i>	1
(12)	C ₆ H ₅ CH ₃	<i>P2₁/n</i>	1

This process for developing reliable computational methods does not deal with any of the simplifying limiting assumptions made for the blind tests. For crystal structure prediction those will have to be dealt with as well. Some of the challenges are represented by the variety of polymorphs and solvates reported for the relatively simple molecule N,N'-dithiobisphthalimide, summarized in Table 1 [41]. The three solvent-free polymorphs represent a challenge that might be considered possible today, but "predicting" the other four solvated motifs is way beyond our current capabilities.

The description above might leave the impression that computational methods have not progressed sufficiently to provide useful chemical and structural information. As noted at the outset, the contrary is true. In conjunction with a variety of analytical techniques, these methods can provide guidelines and confirming parameters for modeling possible crystal structures. We are still a long way from crystal structure or polymorph *prediction*, but we can learn something about what we might want to look for and obtain structural information when crystal structure analysis is difficult or not possible. The recent case history of acetaminophen illustrates this synergism.

Acetaminophen/paracetamol (e.g. Tylenol®) is the most widely used analgesic and antipyretic which has been known to be dimorphic for many years [42]. The commercially used Form I requires binders for formulation; Form II does not require binders, but transforms to Form I. Considerable effort has gone into improving the solid state properties of acetaminophen, including the search for additional crystal forms that might have improved formulation and stability properties, and these efforts included a computational search for possible new forms. Verwer and Leusen [2] completed one such study with the conclusion that

"Recalculation of the lattice energetic...established that this third structure is actually too unstable to exist."

Chemists often see the prediction of the nonexistence of a substance as a challenge to make it,* and this case was no different. In fact, at the time the statement was made there was already evidence in the

* Philosophers of science have also argued against this approach. For instance, the relevant quotation in a recent book by the father-philosopher/son-scientist, Buddhist monk team: "But not to be able to find something is no proof of its nonexistence." [43].

literature for a third form, detected by microscope hot stage methods [44-46]. Beyer et al. carried out another set of calculations on the possible crystal structure of the compound [47]. Almost simultaneously, a set of experiments based on high throughput crystallizations found evidence for a third form, which was characterized by hot stage microscopy and Raman spectroscopy [48]. The X-ray powder diffraction pattern measured by Peterson et al. [48] also closely matches one of the low energy computed structures of Beyer et al., although the latter had ruled it out as a possible structure on the basis of packing considerations. In any event, the combination of experimental and computational techniques led to significant advances in the characterization and understanding of this system, and it is this synergism, especially between computationally-derived models for crystal structures and powder diffraction [49] which promises to provide the greatest immediate and near term benefit.

5. THE CHALLENGES

What is the current status of our ability to predict crystal structures in general and polymorphs in particular? The situation was recently summarized by S.L. Price, one of the leading practitioners in the field [16]:

“We cannot yet computationally predict the possible polymorphs of any given organic molecule.”

As we have pointed out above, and as Price further notes, the computational studies can provide useful information about polymorphs that we might want to look for. They say nothing about their existence or how they might be prepared. In fact, as Dunitz has noted in the same vein [26]:

“We have a long way to go before we can think of predicting which polymorph will be obtained under any given circumstances.”

From the computational point of view it must be shown that calculated lattice energies can truly reflect or reproduce experimentally determined energetics. This will require much more work on force fields, and the development of computational tools for determining lattice energies. Good thermodynamic data are also required on polymorphic systems to provide the standards against which these calculations can be measured. It is still difficult to distinguish between the global minimum and local minima in computational studies, which means that it is not always possible to computationally determine which polymorphic form is the most stable, and how others may be thermodynamically related to it. Price has also pointed out the necessity for assessing and modeling kinetic factors in the growth of different polymorphs.

In order for the computational techniques to be truly predictive, they must be more comprehensive. Hence, on the more practical level some of the computational challenges that must be met along the way were noted above. (In the following the numbers in parentheses are the current percentages of entries in the CSD). Molecules containing only C,H,N,O comprise about %18.5 of the CSD. Provision must be made for modeling crystals with conformationally flexible molecules, molecules larger than ~40 atoms, crystals of salts, solvates (14.3% of CSD) and hydrates (10.1% of CSD), cocrystals, crystals of organometallic compounds (56%), crystals with more or less than one molecule in the asymmetric unit (35%), crystals in the ‘less common’ space groups (12.7%) – in short the technique must be generalized to cover all the possibilities that are regularly encountered in laboratory situations.

By noting our current limitations Price and Dunitz have touched on some of the exciting challenges facing those studying the chemistry of crystal forms – polymorphs and solvates. I would define them with three rather general questions which we would like to be able to answer for any specific system:

- How many different crystal forms can we expect?
- How do we go about preparing those different crystal forms?
- What will be the properties of those different crystal forms if and when I obtain them?

This contribution made an attempt to assess where we are with regard to the computational tools used to address the first question. There has indeed been significant progress since Kitaigorodskii’s pioneering

work in the area, and in closing I was reminded of Martin Luther King's inspiring remarks in a memorable speech at the Lincoln Memorial 41 years ago:

“We’ve come a long way, but we’ve got a long, long way to go.”

6. REFERENCES

- [1] Later referenced [2] as follows: Cerius² User Guide, March 1997, Molecular Simulations, Inc. 9685 Scranton Road, San Diego, CA; <http://www.msi.com/>.
- [2] P. Verwer and F.J.J. Leusen, in *Reviews in Computational Chemistry* 1998, eds. K.P. Lipkowitz and D.B. Boyd (Wiley-VCH, John Wiley and Sons, Inc., New York 1998), p 327.
- [3] A. Gavezzotti, *Acc. Chem. Res.* **27** (1994), 309.
- [4] H. Lipson and W. Cochran, *The Determination of Crystal Structures*, ed. W.L. Bragg, *The Crystalline State*, Vol 3, (Cornell University Press, Ithaca, 1966), Chapter 6.
- [5] C.W. Bunn, *Chemical Crystallography*, (Oxford University Press, Oxford, The Clarendon Press, Oxford, 2nd. Edition, 1961), p. 204
- [6] A.I. Kitaigorodskii, *Organic Chemical Crystallography*, (Consultants Bureau, New York, 1961).
- [7] D.E. Williams, *Acta Crystallogr., Sect. A* **25** (1969), 464.
- [8] D.E. Williams, *Acta Crystallogr.*, **21** (1966), 340.
- [9] D.E. Williams, *Acta Crystallogr., Sect. B* **29** (1973), 96.
- [10] P. Zugenmaier and A. Sarko, *Acta Crystallogr., Sect. B* **28** (1972), 3158.
- [11] H. Hope, J. Bernstein and K.N. Trueblood, *Acta Crystallogr. Sect. B* **28** (1972), 1733.
- [12] A. Renault, C. Cohen-Addad, J. Lajzerowicz-Bonnetaud, J.P. Dutasta and M.J. Crisp, *Acta Crystallogr., Sect. B* **38** (1987), 480.
- [13] W.R. Busing, *Acta Crystallogr., Sect. A* **28** (1972), S252.
- [14] Å. Kvik and J.H. Noordik, *Acta Crystallogr., Sect. B* **38** (1977), 2862.
- [15] A. Gavezzotti and G. Filippini, in *Theoretical Aspects and Computer Modeling of the Molecular Solid State* 1997, ed. A. Gavezzotti (John Wiley and Sons, Inc., New York 1997), p 61.
- [16] S.L. Price, *Adv. Drug Delivery Revs.*, **56** (2004), 301.
- [17] T. Beyer, T. Lewis and S. Price, *Cryst. Eng. Comm.* **44** (2001), 1.
- [18] W.C. McCrone, in *Physics and Chemistry of the Organic Solid State*, 1965 eds. D. Fox, M.M. Labes and A. Weisberger (Wiley Interscience., New York 1965), p 725.
- [19] J. Halebian and W.C. McCrone, *J. Pharm. Sci.* **58** (1969), 911.
- [20] S.R. Byrn, presented at the 35th Erice School on Crystallography, “Diversity Amidst Similarity: A Unified Approach to Polymorphs, Solvates and Phase Changes”, Erice, Sicily, June, 2004.
- [21] J.A.R.P. Sarma and G.R. Desiraju, in *Crystal Engineering. The Design and Applications of Functional Solids*, eds. K.R. Seddon and M. Zaworotko (Klewer, Dordrecht 1999), pp. 325-356; U.J. Griesser and A. Burger, “Statistical Aspects of the Occurrence of Crystal Forms Among Organic Drug Substances” Poster P09.013 at the International Union of Crystallography Congress and General Assembly, Glasgow, Scotland, Aug. 4-13, 1999. Book of Abstracts, p. 533; U.J. Griesser, *Acta Crystallogr., Sect. A* **58** Supplement (2002), C241.
- [22] R.J. Gdanitz, in *Theoretical Aspects and Computer Modeling of the Molecular Solid State* 1997, ed. A. Gavezzotti (John Wiley and Sons, Inc., New York 1997), p 18.
- [23] J. Maddox, *Nature*, **335** (1988), 201.
- [24] F.C. Hawthorne, *Nature*, **345** (1990), 297.
- [25] P.J. Fagan and M.D. Ward, *Scientific American*, **267**(1) (1992), 28.
- [26] J.D. Dunitz, *Chem. Comm.*, (2003), 545.

- [27] J.P.M. Lommerse, W.D.S. Motherwell, H.L. Ammon, J.D. Dunitz, A. Gavezzotti, D.W.M. Hofmann, F.J.J. Leusen, W.T.M. Mooij, S.L. Price, B. Schweizer, M.U. Schmidt, B.P. van Eijck, P. Verwer and D.E. Williams, *Acta Crystallogr., Sect. B* **56** (2000), 697.
- [28] W.D.S. Motherwell, J.P.M. Lommerse, H.L. Ammon, J.D. Dunitz, A. Dzyabchenko, P. Erk, A. Gavezzotti, D.W.M. Hofmann, F.J.J. Leusen, J.P.M. Lommerse, W.T.M. Mooij, S.L. Price, H. Scheraga, B. Schweizer, M.U. Schmidt, B.P. van Eijck, P. Verwer and D.E. Williams, *Acta Crystallogr., Sect. B* **58** (2002), 647.
- [29] E. Pidcock, W.D.S. Motherwell and J.C. Cole, *Acta Crystallogr., Sect. B* **59** (2003), 634.
- [30] R. Boese, private communication.
- [31] J.D. Dunitz and J. Bernstein, *Acc. Chem. Res.* **28** (1995), 193.
- [32] R.K.R. Jetti, R. Boese, J.A.R.P. Sarma, L.S. Reddy, P. Vishweshwar and G.R. Desiraju, *Angew. Chem. Int. Ed.* **42** (2003), 1963.
- [33] W.T.M. Mooij, B.P. van Eijck, S.L. Price, P. Verwer and J. Kroon, *J. Comput. Chem.* **19** (1998), 459.
- [34] H.R. Karfunkel and F.J.J. Leusen, *Speedup* **6** (1992), 43.
- [35] H.R. Karfunkel, F.J.J. Leusen and R.J. Gdanitz, *J. Comput.-Aid Mat. Des.*, **1** (1993), 177.
- [36] H.R. Karfunkel, and R.J. Gdanitz, *J. Comput. Chem.*, **13** (1992), 1171.
- [37] L. Yu, G.A. Stephenson, C.A. Mitchell, C.A. Bunnell, S.V. Snorek, J.J. Bowyer, T.B. Borchardt, J.G. Stowell and S.R. Byrn, *J. Am. Chem. Soc.* **122** (2000), 585.
- [38] L. Yu, *J. Phys. Chem.* **A106** (2002), 544.
- [39] C. Weygand and H. Baumgärtel, *Lieb. Ann.* **469**, (1929) 251.
- [40] J. Bernstein and J.-O. Henck, *Cryst. Eng.* **1** (1998), 119.
- [41] D.M.M. Farrell, C. Glidewell, J.N. Low, J.M.S. Skakle and C.M. Zakaria, *Acta Crystallogr., Sect. B* **58** (2002), 289.
- [42] G. Nichols and C.S. Frampton, *J. Pharm. Sci.* **87** (1998), 684.
- [43] J.-F. Revel and M. Ricard, *The Monk and the Philosopher*, (Thorsons, London, 1998), p. 57.
- [44] A. Burger, *Acta Pharm. Technol.* **28** (1982), 1.
- [45] P. Di Martino, P. Conflant, M. Drache, J.-P. Huvenne and A.-M. Guyot-Hermann, *J. Therm. Anal.* **48** (1997), 447.
- [46] M. Szelagiewicz, *Therm. Anal.* **57** (1999), 23.
- [47] T. Beyer, G.M. Day and S.L. Price *J. Am. Chem. Soc.* **123** (2001), 5086.
- [48] M.L. Peterson, S.L. Morissette, C. McNulty, A. Goldsweig, P. Shaw, M. LeQuesne, J. Monagle, N. Encina, J. Marchionna, A. Johnson, J. Gonzalez-Zugasti, A.V. Lemmo, S.J. Ellis, M.J. Cima and Ö. Almarsson, *J. Am. Chem. Soc.* **124** (2002), 10958.
- [49] W.I.F. David, K. Shankland, L. McCusker and Ch. Baerlocher, (eds.) *Structure Determination from Powder Diffraction Data* (Oxford University Press, Oxford, 2000).