

05/172 1/183

OLARTERAISBECK


Industria y Comercio
SUPERINTENDENCIA

SUPERINTENDENCIA DE PROPIEDAD INDUSTRIAL
Indicador: 06125963 00000000 Fecha: 165
Fecha USPO: 2003-12-25 14:45:13
Invento: 611 PCT-FOTELI 373 FASE ENCI 411 PRESENTAR
Deposición: 2003 DT-5133 DE 22506 CREACIONES

PETITORIO - FORMULARIO ÚNICO DE SOLICITUD

PATENTE PCT

ENTRADA EN FASE NACIONAL

27-01-06

SOLICITUD DE:

1

☒

Patente de Invención.

☐

Patente de Modelo de Utilidad.

☐

Capítulo I.

☒

Capítulo II.

2

SOLICITANTE
(71)

Nombre: BIOREN, INC

Dirección: 100 Glen Way, Suite #1, San Carlos, California 94070, E.U.A.

Nacionalidad o Domicilio: ESTADOUNIDENSE

Lugar de Constitución: Estados Unidos de América.

Teléfono: _____ Fax: _____

E-mail: _____

IDENTIFICACIÓN

C.C.

☐

NIT

☐

C.E.

☐

Otro

☐

Cual _____

Número _____

3

REPRESENTANTE
APODERADO

Nombre: CARLOS R. OLARTE

Dirección: World Trade Center Torre A, Calle 100 No. 8A-37 Piso 10

Teléfono: 601-7700 Fax: 601-7799

E-mail: office@olarteraisbeck.com

IDENTIFICACIÓN

C.C.

☒

NIT

☐

C.E.

☐

Otro

☐

Cual _____

Número 79.782.747Bta
TP 74.295

4

INVENTOR (ES)
(72)

Nombre: ROBERTO CREA

Dirección: 700 Occidental Avenue, San Mateo, California 94402, Estados Unidos de América.

Nacionalidad o Domicilio: ESTADOUNIDENSE

5

PCT (86)

Solicitud Internacional No. PCT/US2004/020306

Fecha Junio 25, 2004

Publicación Internacional No. WO 2005/003345

Fecha Enero 13, 2005

6

Título (54) MUTAGENESIS POR REVISION

Clasificación Internacional (51) C12N 15/10

Prioridad

SI

☒

N

☐

(33) País de Origen

US

(32) Fecha

Junio 27, 2003

(31) Número de Solicitud

60/483,282

7

Comprobante de pago No. 96.849

Fecha

Diciembre 21, 2005

10

ANEXOS

☒

Comprobante de pago de la tasa de presentación de la solicitud.

☐

Documento que acredite la existencia y representación legal cuando el solicitante sea persona jurídica.

☐

Poderes, si fuere el caso.

☐

Documento de cesión del inventor al solicitante o a su causante.

☐

Declaraciones.

☒

Copia de la solicitud internacional. (Resumen, descripción, reivindicaciones, dibujos).

☒

Traducción de la solicitud internacional al castellano. (Resumen, descripción, reivindicaciones, dibujos).

☐

Copia del examen preliminar internacional efectuado por la Autoridad Internacional de Examen Preliminar (IPEA).

☐

Traducción del examen preliminar internacional efectuado por la IPEA.

☐

De ser el caso, copia del contrato de acceso.

☐

De ser el caso, documento que acredite la licencia o autorización de uso de conocimientos tradicionales de las comunidades indígenas.

☐

De ser el caso, certificado de depósito del material biológico.

☐

Arte final 12 x 12.

☐

De ser el caso, información sobre otras solicitudes de patente o títulos obtenidos en el extranjero por el mismo titular o su causante, relacionadas parcial o totalmente con la invención de esta solicitud.

☐

De ser el caso, modificaciones efectuadas a la solicitud internacional.

☒

Otros. Búsqueda Internacional

11

FIGURA CARACTERÍSTICA:

12

Solicito la concesión de la patente,

NOMBRE: CARLOS P. DIARTE

FIRMA: [Signature]

C.C. 79.782.747 Btá T.P. 74.295

Instrucciones para el diligenciamiento del presente formulario, ver reverso de esta página.

10

ANEXOS

☒

Comprobante de pago de la tasa de presentación de la solicitud.

☐

Documento que acredita la existencia y representación legal cuando el solicitante sea persona jurídica.

☐

Poderes, si fuere el caso.

☐

Documento de cesión del inventor al solicitante o a su causante.

☐

Declaraciones.

☒

Copia de la solicitud internacional. (Resumen, descripción, reivindicaciones, dibujos).

☒

Traducción de la solicitud internacional al castellano. (Resumen, descripción, reivindicaciones, dibujos).

☐

Copia del examen preliminar internacional efectuado por la Autoridad Internacional de Examen Preliminar (IPEA).

☐

Traducción del examen preliminar internacional efectuado por la IPEA.

☐

De ser el caso, copia del contrato de acceso.

☐

De ser el caso, documento que acredite la licencia o autorización de uso de conocimientos tradicionales de las comunidades indígenas.

☐

De ser el caso, certificado de depósito del material biológico.

☐

Arte final 12 x 12.

☐

De ser el caso, información sobre otras solicitudes de patente o títulos obtenidos en el extranjero por el mismo titular o su causante, relacionadas parcial o totalmente con la invención de esta solicitud.

☐

De ser el caso, modificaciones efectuadas a la solicitud internacional.

☒

Otros. Búsqueda Internacional

11

FIGURA CARACTERÍSTICA:

12

Solicito la concesión de la patente,

NOMBRE: CARLOS R. OLARTE

FIRMA: [Signature]
C.C. 79.782.747 Btá T.P. 74.295

Instrucciones para el diligenciamiento del presente formulario, ver reverso de esta página.

3

SUPERINTENDENCIA Y DEPARTAMENTO
Auditoria: 00000000 00000000
Fecha (REC): 2005-12-23 16:40:44
Trámite : 011 REC-PROYECT 273 PARA 1001 411 PRESENTAR
Expediente: 2005 D. DISEÑO DE NUEVOS PROCEDIMIENTOS

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO

NIT : 800179969-2

RECIBO OFICIAL DE CAJA : CS 96,849
FECHA : DICIEMBRE 21 DE 2005

CONSULTA

DEPOSITANTE	TIPO PAGO	BANCO	CUENTA	No. PAGOS	FECHA PAGO	Gr. PAGOS
DIARTE RAISEBER	CONSULTACION	BANCO FOLIO	090-00110	27326200	21/12/2005	3,344,000.00

C O N T A B I L I T A T O

CANT. CONTABILITADA	CONCEPTO	TOTAL CANCELADO
1 50005-01-6: SOLICITUDES	2 TRAMITES DE SOL. DE PATENTE DE IN.	443,600.00
	TOTAL :	443,600.00

SEN: CUATROCIENTOS CUARENTA Y TRES MIL PESOS

RESPONSABLE : _____

RECIBO DE CAJA APLICADO AL EXPEDIENTE No. _____

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau



INTERNATIONAL PATENT COOPERATION TREATY (PCT)

(43) International Publication Date
13 January 2005 (13.01.2005)

PCT

(10) International Publication Number
WO 2005/003345 A2

(51) International Patent Classification: C12N 15/10

(21) International Application Number:
PCT/US2004/020306

(22) International Filing Date: 25 June 2004 (25.06.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60483282 27 June 2003 (27.06.2003) US

(71) Applicant (for all designated States except US): B. CREA
& CO. (US/US); 700 Occidental Avenue, San Mateo, CA
94402 (US).

(72) Inventor; and

(75) Inventor/Applicant (for US only): CREA, Roberto
(ITA/US); 700 Occidental Avenue, San Mateo, CA 94402
(US).

(74) Agents: REMILLAND, Jane, E. et al.; Lahive & Cock-
field, LLP, 28 State Street, Boston, MA 02109 (US).

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AI, AM,
AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN,
CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,
GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE,
KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD,
MG, MK, MN, MW, MX, MY, NA, NG, NO, NZ, OM, PG,
PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM,
TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM,
ZW.

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,
FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI,
SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, ML, MR, NE, SN, TD, TO).

Published

— without international search report and to be republished
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

(84) Title: LOOK-THROUGH MUTAGENESIS

(57) Abstract: A method of mutagenesis by which a predetermined amino acid is introduced into each and every position of a
selected set of positions in a preselected region (or several different regions) of a polypeptide to produce a library of polypeptide
analogs. The method is based on the premise that certain amino acids play crucial role in the structure and function of proteins.
Libraries can be generated which contain only desired polypeptide analogs and are of reasonable size for screening. The libraries can
be used to study the role of specific amino acids in polypeptide structure and function and to develop new or improved polypeptides
such as antibodies, antibody fragments, single chain antibodies, enzymes, and ligands.

WO 2005/003345 A2

LOOK-THROUGH MUTAGENESIS

Related Applications and Information

This application claims priority to U.S. Provisional Application No. 60/483282, filed on June 27, 2003, the entire contents of which are incorporated by reference herein. The entire contents of all other patents, patent applications, and references cited throughout the following specification also are incorporated by reference herein in their entireties.

Background of the Invention

Mutagenesis is a powerful tool in the study of protein structure and function. Mutations can be made in the nucleotide sequence of a cloned gene encoding a protein of interest and the modified gene can be expressed to produce mutants of the protein. By comparing the properties of a wild-type protein and the mutants generated, it is often possible to identify individual amino acids or domains of amino acids that are essential for the structural integrity and/or biochemical function of the protein, such as its binding and/or catalytic activity. The number of mutants that can be generated from a single protein, however, renders it difficult to select mutants that will be informative or have a desired property, even if the selected mutants that encompass the mutations are solely in putatively important regions of a protein (e.g., regions that make up an active site of a protein). For example, the substitution, deletion, or insertion of a particular amino acid may have a local or global effect on the protein.

Previous methods for mutagenizing polypeptides have been either too restrictive, too inclusive, or limited to knocking out protein function rather than to gaining or improving function. For example, a highly restrictive approach is selective or site-directed mutagenesis which is used to identify the presence of a particular functional site or understand the consequences of making a very specified alteration within the functional site. A common application of site directed mutagenesis is in the study of phosphoproteins where an amino acid residue, that would ordinarily be phosphorylated and allow the polypeptide to carry out its function, is altered to confirm the link between phosphorylation and functional activity. This approach is very specific for the polypeptide and residue being studied.

Conversely, a highly inclusive approach is saturation or random mutagenesis that is designed to produce a large number of mutations encompassing all possible alterations within a defined region of a gene or protein. This is based on the principle that, by generating essentially all possible variants of a relevant protein domain, the proper arrangement of amino acids is likely to be produced as one of the randomly generated mutants. However, in practice, the vast number of random combinations of mutations

generated can prevent the capacity to meaningfully select a desired candidate because of the presence of the so-called "noise" of so many undesired candidates.

Another approach, referred to as "Walk Through" mutagenesis (see, e.g., U.S. Patent Nos. 5,830,630; 5,798,208) has been used to mutagenize a defined region of a polypeptide by synthesizing a mixture of degenerate oligonucleotides that, statistically, contain a desired set of mutations. However, because degenerate polynucleotide synthesis is employed, Walk-Through mutagenesis yields a number of undesired alterations in addition to the desired set of mutations. For example, to sequentially introduce a mutation across a defined region of only five amino acid positions, a set of over 100 polynucleotide must be made (and screened) (see, e.g., Fig. 6). Accordingly, to make and screen, for example, two or three regions becomes increasingly complex, i.e., requiring the making and screening of 200 to over 300 polynucleotides, respectively, for the presence of only 10 to 15 mutations.

In yet another approach which has been used to mutagenize proteins is alanine scanning mutagenesis, where an alanine residue is "scanned" through a portion of a protein to identify positions where the protein's function is interrupted. However, this approach only looks at loss of protein function by way of substituting a neutral alanine residue at a given position, rather than gain or improvement of function. Thus, it is not a useful approach for generating proteins having improved structure and function.

Accordingly, a need remains for a systematic way to mutagenize a protein for new or improved function.

Summary of the Invention

The invention pertains to a method of mutagenesis for the generation of novel or improved proteins (or polypeptides) and to libraries of polypeptide analogs and specific polypeptides generated by the methods. The polypeptide targeted for mutagenesis can be a natural, synthetic or engineered polypeptide, including fragments, analogs and mutant forms thereof.

In one embodiment, the method comprises introducing a predetermined amino acid into essentially every position within a defined region (or several different regions) of the amino acid sequence of a polypeptide. A polypeptide library is generated containing polypeptide analogs which individually have no more than one predetermined amino acid, but which collectively have the predetermined amino acid in every position within the defined region(s). The method can be referred to as "look-through" mutagenesis because, in effect, a single, predetermined amino acid (and only the predetermined amino acid) is substituted position-by-position throughout one or more defined region(s) of a polypeptide. Thus, the invention allows one to "look-through" the structural and functional consequences of separately substituting a

Figure 10 illustrates the integration of all six CDRs of a binding region of an antibody that have been subjected to LTM and the resultant diversity if one predetermined amino acid residue or twenty different predetermined amino acid residues are used.

5 **Figure 11** illustrates the integration of all six CDRs of a binding region of an anti-TNF single chain antibody (scFv) that have been subjected to LTM and the resultant diversity if one predetermined amino acid residue or three different predetermined amino acid residues are used.

10 **Figure 12** illustrates an arrayed library representing some of the possible polypeptide analogs of the six CDRs of an antibody binding region that can be achieved using LTM.

Figure 13 illustrates the screening of an arrayed expression library using cell-free ribosome display.

15 **Figure 14** illustrates the combinatorial chemistry explored when the binding region of an antibody variable region (*i.e.*, all six CDRs) is subjected to LTM.

Figure 15 shows the sequence of the variable region (in single chain format) of several representative anti-TNF binding molecules that can be subjected to LTM.

Figure 16 shows a schematic for carrying out the protease selection assay for screening catalytic candidates of LTM.

20 **Figure 17** shows a schematic detailing the mechanism (and advantages) of the protease selection assay when carried out in bacterial cells.

Figure 18 shows a flowchart detailing the mechanics of screening gene libraries for catalytic activity, for example, catalytic antibody activity, using either ribosome or yeast display.

25 **Figure 19** shows a schematic for carrying out LTM where *a priori* information (*e.g.*, computer modeling information) and empirical information (assay results) can be coordinated for more efficient molecule design and development. This guided approach of LTM is referred to as "guide-through" mutagenesis.

30 **Figure 20** shows a hypothetical V_H CDR3 wild-type sequence (uppermost shaded ovals) and the resulting sequences in LTM His substitution (in open ovals) library members. The individual LTM His substitutions are encoded by individual oligonucleotides, *e.g.*, oligonucleotides synthesized in a high-throughput fashion. LTM subset libraries for the other (LTM) amino acid substitutions in this CDR domain are constructed in a similar fashion.

35 **Figure 21** illustrates the generation of scFv libraries. On the top row of the x axis and the far left column of the y axes the three digits represent the 3 CDRs on each of the light and heavy chains. A "0" indicates a wild-type CDR sequence, whereas a "1" indicates an LTM mutated CDR. The number on the grid indicates the complexity of the

subset library. For example in the uppermost left corner of the matrix is a "0" where the corresponding x and y axes are "000" and "000" indicating that none of the CDR in either the V_H and V_L are LTM mutated respectively. In moving one row over from the "0" corner to the neighboring "1" grid position would be designated by x-axis "100" and y-axis "000" indicating that V_H CDR1 is mutated while all V_L CDR remain wild type. Thus, in likewise fashion, a grid numbering of "4" would mean that there are four CDRs simultaneously mutated. Initially the seven V_H and seven V_L chains (indicated by the arrows) are made using SOE-PCR. The V_H and V_L chains are then amplified and mixed and matched by mega primer to generate all the remaining V_L-V_H combinations in one step.

Detailed Description of the Invention

In order to provide a clear understanding of the specification and claims, the following definitions are provided below.

Definitions

As used herein the term "analog" refers to a variant or mutant polypeptide (or a nucleic acid encoding such a polypeptide) having one or more amino acid substitutions.

The term "binding molecule" refers to any binding molecule, including proteins, polypeptides, peptides, and small molecules, that bind to a substrate or target. In one embodiment, the binding molecule is an antibody or binding fragment thereof (e.g., a Fab fragment), single domain antibody, single chain antibody (e.g., scFv), or peptide capable of binding a ligand.

The term "defined region" refers to a selected region of a polypeptide. Typically, the defined region includes all or a portion of a functional site, e.g., the binding site of a ligand, the binding site of a binding molecule or receptor, or a catalytic site. The defined region may also include multiple portions of a functional site. For example, the defined region can include all, a portion, or multiple portions of a complementarity determining region (CDR) or a complete heavy and/or light chain variable region (VR) of an antibody. Thus, a functional site may include a single or multiple defined regions that contribute to the functional activity of the molecule.

The term "library" refers to two or more molecules mutagenized according to the method of the invention. The molecules of the library can be in the form of polynucleotides, polypeptides, polynucleotides and polypeptides, polynucleotides and polypeptides in a cell free extract, or as polynucleotides and/or polypeptides in the context of a phage, prokaryotic cells, or in eukaryotic cells.

The term "mutagenizing" refers to the alteration of an amino acid sequence. This can be achieved by altering or producing a nucleic acid (polynucleotide) capable of

encoding the altered amino acid sequence, or by the direct synthesis of an altered polypeptide using protein chemistry.

The term "polynucleotide(s)" refers to nucleic acids such as DNA molecules and RNA molecules and analogs thereof (e.g., DNA or RNA generated using nucleotide analogs or using nucleic acid chemistry). As desired, the polynucleotides may be made synthetically, e.g., using art-recognized nucleic acid chemistry or enzymatically using, e.g., a polymerase. Typical modifications include methylation, biotinylation, and other art-known modifications. In addition, the nucleic acid molecule can be single-stranded or double-stranded and, where desired, linked or associated (e.g., covalently or non-covalently) to a detectable moiety.

The term "variant polynucleotide" refers to a polynucleotide encoding a corresponding polypeptide analog (or portion thereof) of the invention. Thus, variant polynucleotides contain one or more codons that have been changed to result in expression of a different amino acid.

The term "polypeptide(s)" refers to two or more amino acids joined by a peptide bond, e.g., peptides (e.g., from 2 to ~50 amino acid residues), as well as longer peptide sequences e.g., protein sequences which typically comprises amino acid sequences from as few as 50 amino acid residues to more than 1,000 amino acid residues.

The term "pooling" refers to the combining of polynucleotide variants or polypeptide analogs to form libraries representing the Look-Through mutagenesis of an entire polypeptide region. The molecules may be in the form of a polynucleotide and/or polypeptide and may coexist in the form of a sublibrary, as molecules on a solid support, as molecules in solution, and/or as molecules in one or more organisms (e.g., phage, prokaryotic cells, or eukaryotic cells).

The term "predetermined amino acid" refers to an amino acid residue selected for substitution at each position within a defined region of a polypeptide to be mutagenized. This does not include position(s) within the region that already (e.g., naturally) contain the predetermined amino acid and, thus, which need not be substituted with the predetermined amino acid. Accordingly, each polypeptide analog generated in accordance with the present invention contains no more than one "predetermined amino acid" residue in a given defined region. However, collectively, the library of polypeptide analogs generated contains the predetermined amino acid at each position within the region being mutagenized. Typically, a predetermined amino acid is selected for a particular size or chemistry usually associated with the side group of the amino acid. Suitable predetermined amino acids include, for example, glycine and alanine (sterically small); serine, threonine, and cysteine (nucleophilic); valine, leucine, isoleucine, methionine, and proline (hydrophobic); phenylalanine, tyrosine, and

tryptophan (aromatic); aspartate and glutamate (acidic); asparagine, glutamine, and histidine (amide); and lysine and arginine (basic).

Detailed Description

5 The study of proteins has revealed that certain amino acids play a crucial role in their structure and function. For example, it appears that only a discrete number of amino acids participate in the binding of an antibody to an antigen or are involved in the catalytic event of an enzyme.

10 Though it is clear that certain amino acids are critical to the activity or function of proteins, it is difficult to identify which amino acids are involved, how they are involved, and what substitutions can improve the protein's structure or function. In part, this is due to the complexity of the spatial configuration of amino acid side chains in polypeptides and the interrelationship of different portions of the polypeptide that contribute to form a functional site. For example, the interrelationship between the six
15 CDRs of the variable heavy and light chain regions of an antibody contribute to the antigen or ligand-binding pocket.

Previous mutagenesis methods, such as selective (site-directed) mutagenesis and saturation mutagenesis, are of limited utility for the study of protein structure and function in view of the enormous number of possible variations in complex
20 polypeptides. This is especially true given that desirable combinations are often accompanied by the presence of vast amounts of undesirable combinations or so-called noise.

The method of this invention provides a systematic, practical, and highly accurate approach for evaluating the role of particular amino acids and their position,
25 within a defined region of a polypeptide, in the structure or function of the polypeptide and, thus, for producing improved polypeptides.

1. Selecting a Defined Region

In accordance with the present invention, a defined region or regions within a
30 protein are selected for mutagenesis. Typically, the regions are believed to be important to the protein's structure or function (see, e.g., Fig. 1). This can be deduced, for example, from what structural and/or functional aspects are known or can be deduced from comparing the defined region(s) to what is known from the study of other proteins, and may be aided by modeling information. For example, the defined region can be one
35 that has a role in a functional site, e.g., in binding, catalysis, or another function. In one embodiment, the defined region is a hypervariable region or complementarity determining region (CDR) of an antigen binding molecule. In another embodiment, the defined region is a portion of a complementarity determining region (CDR). In other

embodiments, two or more defined regions; e.g., CDRs or portions thereof, are selected for mutagenesis.

2. Selecting a Predetermined Amino Acid Residue

5 The amino acid residue chosen for substitution within the defined region(s) is generally selected from those known to be involved in the structure or function of interest. The twenty naturally occurring amino acids differ with respect to their side chain. Each side chain is responsible for chemical properties that make each amino acid unique. For the purpose of altering binding or creating new binding affinities, any of the
10 twenty naturally occurring amino acids generally can be selected. Thus, previous methods of mutagenesis, which created vast numbers of analogs for every substitution, were impractical for evaluating the effect on protein binding of substitution each of the twenty amino acids. In contrast, the methods of the present invention creates a practical number of analogs for each amino acid substitution and, thus, allows for the evaluation
15 of a greater variety of protein chemistries within a segregated region or regions of a protein.

In contrast to protein binding, only a subset of amino acid residues typically participate in enzymatic or catalytic events. For example, from the chemical properties of the side chains, only a selected number of natural amino acids preferentially
20 participate in catalytic events. These amino acids belong to the group of polar and neutral amino acids such as Ser, Thr, Asn, Gln, Tyr, and Cys, the group of charged amino acids, Asp and Glu, Lys and Arg, and especially the amino acid His. Other polar and neutral side chains are those of Cys, Ser, Thr, Asn, Gln and Tyr. Gly is also considered to be a borderline member of this group. Ser and Thr play an important role
25 in forming hydrogen bonds. Thr has an additional asymmetry at the beta carbon, therefore only one of the stereoisomers is used. The acid amide Gln and Asn can also form hydrogen bonds, the amido groups functioning as hydrogen donors and the carbonyl groups functioning as acceptors. Gln has one more CH₂ group than Asn which renders the polar group more flexible and reduces its interaction with the main chain.
30 Tyr has a very polar hydroxyl group (phenolic OH) that can dissociate at high pH values. Tyr behaves somewhat like a charged side chain; its hydrogen bonds are rather strong.

Neutral polar acids are found at the surface as well as inside protein molecules. As internal residues, they usually form hydrogen bonds with each other or with the
35 polypeptide backbone. Cys can form disulfide bridges.

Histidine (His) has a heterocyclic aromatic side chain with a pK value of 6.0. In the physiological pH range, its imidazole ring can be either uncharged or charged, after taking up a hydrogen ion from the solution. Since these two states are readily available,

His is quite suitable for catalyzing chemical reactions. It is found in most of the active centers of enzymes, for example, serine proteases.

Asp and Glu are negatively charged at physiological pH. Because of their short side chain, the carboxyl group of Asp is rather rigid with respect to the main chain. This may be the reason why the carboxyl group in many catalytic sites is provided by Asp and not by Glu. Charged acids are generally found at the surface of a polypeptide.

In addition, Lys and Arg are found at the surface. They have long and flexible side chains presenting multiple rotamers of similar energies. In several cases, Lys and Arg take part in forming internal salt bridges or they help in catalysis. Because of their exposure at the surface of the polypeptide, Lys is a residue more frequently recognized by enzymes that either modify the side chain or cleave the peptide chain at the carbonyl end of Lys residues.

While the side group chemistry of an amino acid can guide the selection of a predetermined amino acid residue, the lack of a desired side group chemistry can be a criterion for excluding an amino acid residue for use as the predetermined amino acid. For example, sterically small and chemically neutral amino acids, such as alanine, can be excluded from Look-Through mutagenesis for lacking a desired chemistry.

2. Synthesizing Polypeptide Analog Libraries

In one embodiment, a library of polypeptide analogs is generated for screening by synthesizing individual oligonucleotides that encode the defined region of the polypeptide and have no more than one codon for the predetermined amino acid. This is accomplished by incorporating, at each codon position within the oligonucleotide either the codon required for synthesis of the wild-type polypeptide or a codon for the predetermined amino acid. This differs from the oligonucleotides produced in saturation mutagenesis, random mutagenesis, or walk-through mutagenesis in that, for each oligonucleotide, only one mutation, as opposed to multiple mutations is made.

The oligonucleotides can be produced individually and then mixed or pooled as desired. When the codon of the wild type sequence and the codon for the predetermined amino acid are the same, no substitution is made.

Accordingly, the number of amino acid positions within the defined region will determine the maximum number of oligonucleotides made. For example, if five codon positions are altered with the predetermined amino acid, then five polynucleotides plus one polynucleotide representing the wild-type amino acid sequence are synthesized. Two or more regions can simultaneously be altered.

The mixture of oligonucleotides for generation of the library can be synthesized readily by known methods for DNA synthesis. The preferred method involves use of solid phase beta-cyanoethyl phosphoramidite chemistry. See U.S. Pat. No. 4,725,677.

For convenience, an instrument for automated DNA synthesis can be used containing specified reagent vessels of nucleotides. The polynucleotides may also be synthesized to contain restriction sites or primer hybridization sites to facilitate the introduction or assembly of the polynucleotides representing, e.g., a defined region, into a larger gene context.

The synthesized polynucleotides can be inserted into a larger gene context of the polypeptide being mutagenized by using standard genetic engineering techniques. For example, the polynucleotides can be made to contain flanking recognition sites for restriction enzymes. See Crea, R., U.S. Pat. No. 4,888,286. The recognition sites are designed to correspond to recognition sites that either exist naturally or are introduced in the gene proximate to the DNA encoding the region. After conversion into double stranded form, the polynucleotides are ligated into the gene by standard techniques. By means of an appropriate vector (including, e.g., phage vectors, plasmids) the genes can be introduced into a cell-free extract, phage, prokaryotic cell, or eukaryotic cell suitable for expression of the mutant polypeptides.

In cases where the amino acid sequence of the polypeptide to be mutagenized is known or where the DNA sequence is known, gene synthesis is a possible approach. For example, partially overlapping polynucleotides, typically about 20-60 nucleotides in length can be designed. The internal polynucleotides are then phosphorylated annealed to their complementary partner to give a double-stranded DNA molecule with single-stranded extensions useful for further annealing. The annealed pairs can then be mixed together and ligated to form a full-length double-stranded molecule (see, e.g., Fig. 5). Convenient restriction sites can be designed near the ends of the synthetic gene for cloning into a suitable vector. The full-length molecules can be cleaved with those restriction enzymes and ligated into a suitable vector. Convenient restriction sites can also be incorporated into the sequence of the synthetic gene to facilitate introduction of mutagenic cassettes.

As an alternative to synthesizing polynucleotides representing the full-length double-stranded gene, polynucleotides which partially overlap at their 3' ends (i.e., with complementary 3' ends) can be assembled into a gapped structure and then filled in with a suitable polymerase to make a full length double-stranded gene. Typically, the overlapping polynucleotides are from 40-90 nucleotides in length. The extended polynucleotides are then ligated. Convenient restriction sites can be introduced at the ends and/or internally for cloning purposes. Following digestion with an appropriate restriction enzyme or enzymes, the gene fragment is ligated into a suitable vector. Alternatively, the gene fragment can be blunt end ligated into an appropriate vector.

In these approaches, if convenient restriction sites are available (naturally or engineered) following gene assembly, the degenerate polynucleotides can be introduced

subsequently by cloning the cassette into an appropriate vector. Alternatively, the degenerate polynucleotides can be incorporated at the stage of gene assembly. For example, when both strands of the gene are fully chemically synthesized, overlapping and complementary degenerate polynucleotides can be produced. Complementary pairs will anneal with each other.

When partially overlapping polynucleotides are used in the gene assembly, a set of degenerate nucleotides can also be directly incorporated in place of one of the polynucleotides. The appropriate complementary strand is synthesized during the extension reaction from a partially complementary polynucleotide from the other strand by enzymatic extension with a polymerase. Incorporation of the degenerate polynucleotides at the stage of synthesis also simplifies cloning where more than one domain or defined region of a gene is mutagenized.

In another approach, the gene of interest is present on a single stranded plasmid. For example, the gene can be cloned into a phage vector or a vector with a filamentous phage origin of replication that allows propagation of single-stranded molecules with the use of a helper phage. The single-stranded template can be annealed with a set of degenerate polynucleotides representing the desired mutations and elongated and ligated, thus incorporating each analog strand into a population of molecules that can be introduced into an appropriate host (Sayers, J. R. et al., *Nucleic Acids Res.* 16: 791-802 (1988)). This approach can circumvent multiple cloning steps where multiple domains are selected for mutagenesis.

Polymerase chain reaction (PCR) methodology can also be used to incorporate polynucleotides into a gene. For example, the polynucleotides themselves can be used as primers for extension. In this approach, polynucleotides encoding the mutagenic cassettes corresponding to the defined region (or portion thereof) are complementary to each other, at least in part, and can be extended to form a large gene cassette using a polymerase, e.g., using PCR amplification.

The size of the library will vary depending upon the length and number of regions and amino acids within a region that are mutagenized. Preferably, the library will be designed to contain less than 10^{15} , 10^{14} , 10^{13} , 10^{12} , 10^{11} , 10^{10} , 10^9 , 10^8 , 10^7 , and more preferably, 10^6 polypeptide analogs or less.

The description above has centered on the mutagenesis of polypeptides and libraries of polypeptides by altering the polynucleotide that encodes the corresponding polypeptide. It is understood, however, that the scope of the invention also encompasses methods of mutagenizing polypeptides by direct synthesis of the desired polypeptide analogs using protein chemistry. In carrying out this approach, the resultant polypeptides still incorporate the features of the invention except that the use of a polynucleotide intermediate is eliminated.

For the libraries described above, whether in the form of polynucleotides and/or corresponding polypeptides, it is understood that the libraries may be also attached to a solid support, such as a microchip, and preferably arrayed, using art recognized techniques.

4. Expression and Screening Systems

Libraries of polynucleotides generated by any of the above techniques or other suitable techniques can be expressed and screened to identify polypeptide analogs having desired structure and/or activity. Expression of the polypeptide analogs can be carried out using any suitable expression display system known in the art including, but not limited to, cell-free extract display systems (e.g., ribosome display and arrayed (e.g., microarrayed or macroarrayed) display systems), bacterial display systems, phage display systems, prokaryotic cells, and/or eukaryotic cells (e.g., yeast display systems).

In one embodiment, the polynucleotides are engineered to serve as templates that can be expressed in a cell free extract. Vectors and extracts as described, for example in U.S. Patent Nos. 5,324,637; 5,492,817; 5,665,563, can be used and many are commercially available. Ribosome display and other cell-free techniques for linking a polynucleotide (i.e., a genotype) to a polypeptide (i.e., a phenotype) can be used, e.g., Profusion™ (see, e.g., U.S. Patent Nos. 6,348,315; 6,261,804; 6,258,558; and 6,214,559).

Alternatively, the polynucleotides of the invention can be expressed in a convenient *E. coli* expression system, such as that described by Flockthun and Skerra. (Flockthun, A. and Skerra, A., *Meth. Enzymol.* 178: 476-515 (1989); Skerra, A. et al., *Biotechnology* 9: 273-278 (1991)). The mutant proteins can be expressed for secretion in the medium and/or in the cytoplasm of the bacteria, as described by M. Better and A. Horwitz, *Meth. Enzymol.* 178: 476 (1989). In one embodiment, the single domains encoding VH and VL are each attached to the 3' end of a sequence encoding a signal sequence, such as the *ompA*, *phoA* or *pelB* signal sequence (Lei, S. P. et al., *J. Bacteriol.* 169: 4379 (1987)). These gene fusions are assembled in a dicistronic construct, so that they can be expressed from a single vector, and secreted into the periplasmic space of *E. coli* where they will refold and can be recovered in active form. (Skerra, A. et al., *Biotechnology* 9: 273-278 (1991)). For example, antibody heavy chain genes can be concurrently expressed with antibody light chain genes to produce antibody or antibody fragments.

In still another embodiment, the polynucleotides can be expressed in eukaryotic cells such as yeast using, for example, yeast display as described, e.g., in U.S. Patent Nos. 6,423,538; 6,331,391; and 6,300,065. In this approach, the polypeptide analogs of the library are fused to a polypeptide that is expressed and displayed on the surface of the yeast. Other eukaryotic cells for expression of the polypeptides of the invention can also

be used such as mammalian cells, for example myeloma cells, hybridoma cells, or Chinese hamster ovary (CHO) cells. Typically, the polypeptide analogs when expressed in mammalian cells are designed to be expressed into the culture medium, or expressed on the surface of such a cell. The antibody or antibody fragments can be produced, for example, as entire antibody molecules or as individual VH and VL fragments, Fab fragments, single domains, or as single chains (sFv) (see Huston, J. S. et al., Proc. Natl. Acad. Sci. USA 85: 5879-5883 (1988)).

The screening of the expressed polypeptide analogs (or polypeptides produced by direct synthesis) can be done by any appropriate means. For example, binding activity can be evaluated by standard immunoassay and/or affinity chromatography and catalytic activity can be ascertained by suitable assays for substrate conversion. Screening of the polypeptide analogs of the invention for proteolytic function can be accomplished using a standard hemoglobin plaque assay as described, for example, in U.S. Patent No. 5,798,208.

5. Computer Modeling-Assisted Look Through Mutagenesis

The look-through mutagenesis of the invention may also be conducted with the benefit of structural or modeling information concerning the polypeptide analogs to be generated, such that the potential for generating analogs having the desired improved function is increased. The structural or modeling information can also be used to guide the selection of predetermined amino acid to introduce into the defined regions. Still further, actual results obtained with the polypeptide analogs of the invention can guide the selection (or exclusion) of subsequent polypeptides to be made and screened in an iterative manner. Accordingly, structural or modeling information can be used to generate initial subsets of polypeptide analogs for use in the invention, thereby further increasing the efficiency of generating improved polypeptides.

In a particular embodiment, in silico modeling is used to eliminate the production of any polypeptide analog predicted to have poor or undesired structure and/or function. In this way, the number of polypeptide analogs to be produced can be sharply reduced thereby increasing signal-to-noise in subsequent screening assays. In another particular embodiment, the in silico modeling is continually updated with additional modeling information, from any relevant source, e.g., from gene and protein sequence and three-dimensional databases and/or results from previously tested analogs, so that the in silico database becomes more precise in its predictive ability (Fig. 19).

In yet another embodiment, the in silico database is provided with the assay results of previously tested polypeptide analogs and categorizes the analogs, based on the assay criterion or criteria, as responders or nonresponders, e.g., as polypeptide analogs that bind well or not so well or as being enzymatic/catalytic or not so

enzymatic/catalytic. In this way, the guide-through mutagenesis of the invention can equate a range of functional response with particular structural information and use such information to guide the production of future polypeptide analogs to be tested.

Accordingly, the method is especially suitable for screening antibody or antibody fragments for a particular function, such as binding affinity (e.g., specificity), stability (e.g., half life) and/or effector function (e.g., complement activation and ADCC).

Accordingly, mutagenesis of noncontiguous residues within a region can be desirable if it is known, e.g., through in silico modeling, that certain residues in the region will not participate in the desired function. The coordinate structure and spatial interrelationship between the defined regions, e.g., the functional amino acid residues in the defined regions of the polypeptide, e.g., the predetermined amino acid(s) that have been introduced, can be considered and modeling. Such modeling criteria include, e.g., amino acid residue side group chemistry, atom distances, crystallography data, etc. Accordingly, the number of polypeptide analogs to be produced can be intelligently minimized.

In a preferred embodiment, one or more of the above steps are computer-assisted. The method is also amenable to be carried out, in part or in whole, by a device, e.g., a computer driven device. Accordingly, instructions for carrying out the method, in part or in whole, can be conferred to a medium suitable for use in an electronic device for carrying out the instructions. In sum, the methods of the invention are amendable to a high throughput approach comprising software (e.g., computer-readable instructions) and hardware (e.g., computers, robotics, and chips).

6. Exploring the Combinatorial Chemistry of Multiple Defined Regions

The present invention provides the important advantage of allowing for evaluation by mutagenesis of several different regions or domains of a polypeptide simultaneously. This can be done using the same or a different predetermined amino acid within each region, enabling the evaluation of amino acid substitutions in conformationally related regions, such as the regions that upon folding of the polypeptide, are associated to make up a functional site (e.g., the binding site of an antibody or the catalytic site of an enzyme). This, in turn, provides an efficient way to create new or improved functional sites.

For example, as depicted in Fig. 14, the six CDRs of an antibody that make up the unique aspects of the antigen binding site (Pv region), can be mutagenized simultaneously, or separately within the VH or VL chains, to study the three dimensional interrelationship of selected amino acids in this site. In one embodiment, the combinatorial chemistry of three or more defined regions are systematically explored using look-through mutagenesis, and preferably six defined regions, for example, the six

CDRs of an antibody heavy and light chain variable region. For performing look-through mutagenesis on a CDR, typically 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or more amino acid positions are altered.

Accordingly, the present invention opens up new possibilities for the design of many different types of novel and improved polypeptides. The method can be used to improve upon an existing structure or function of a protein. For example, a binding site for an antibody or antibody fragment can be introduced or affinity for a pre-existing antigen, effector function and/or stability improved. Alternatively, the introduction of additional "catalytically important" amino acids into a catalytic domain of an enzyme can be performed resulting in a modified or enhanced catalytic activity toward a substrate. Alternatively, entirely new structures, specificities or activities may be introduced into a polypeptide. De novo synthesis of enzymatic activity can be achieved as well. The new structures can be built on the natural or consensus "scaffold" of an existing protein by mutating only relevant regions by the method of the invention.

7. Look-Through Mutagenesis for Making New or Improved Antibodies

The method of this invention is especially useful for modifying antibody molecules. As used herein, antibody molecules or antibodies refers to antibodies or portions thereof, such as full-length antibodies, Fv molecules, or other antibody fragments, individual chains or fragments thereof (e.g., a single chain of Fv), single chain antibodies, and chimeric antibodies. Alterations can be introduced into the variable region and/or into the framework (constant) region of an antibody. Modification of the variable region can produce antibodies with better antigen binding properties, and, if desired, catalytic properties. Modification of the framework region can also lead to the improvement of chemo-physical properties, such as solubility or stability (e.g., half life), which are especially useful, for example, in commercial production, bioavailability, effector function (e.g., complement activation and/or ADCC) and binding affinity (e.g., specificity) for the antigen. Typically, the mutagenesis will target the Fv region of the antibody molecule, i.e., the structure responsible for antigen-binding activity which is made up of variable regions of two chains, one from the heavy chain (VH) and one from the light chain (VL). Once the desired antigen-binding characteristics are identified, the variable region(s) can be engineered into an appropriate antibody class such as IgG, IgM, IgA, IgD, or IgE.

8. Look-Through Mutagenesis for Making / Improving Catalytic / Enzymatic Polypeptides

The method of the invention also is particularly suited to the design of catalytic proteins, particularly catalytic antibodies. Presently, catalytic antibodies can be

prepared by an adaptation of standard somatic cell fusion techniques. In this process, an animal is immunized with an antigen that resembles the transition state of the desired substrate to induce production of an antibody that binds the transition state and catalyzes the reaction. Antibody-producing cells are harvested from the animal and fused with an immortalizing cell to produce hybrid cells. These cells are then screened for secretion of an antibody that catalyzes the reaction. This process is dependent upon the availability of analogues of the transition state of a substrate. The process may be limited because such analogues are likely to be difficult to identify or synthesize in most cases.

The method of the invention provides a different approach that eliminates the need for a transition state analogue. By the method of the invention, an antibody can be made catalytic by the introduction of suitable amino acids into the binding site of an immunoglobulin (Fv region). The antigen-binding site (Fv) region is made-up of six hypervariable (CDR) loops, three derived from the immunoglobulin heavy chain (H) and three from the light chain (L), which connect beta strands within each subunit. The amino acid residues of the CDR loops contribute almost entirely to the binding characteristics of each specific monoclonal antibody. For instance, catalytic triads (comprising of amino acid residues serine, histidine, and aspartic acid) modeled after serine proteases can be created in the hypervariable segments of the Fv region of an antibody with known affinity for the substrate molecule and screened for proteolytic activity of the substrate.

In particular, the method of the invention can be used to produce many different enzymes or catalytic antibodies, including oxidoreductases, transferases, hydrolases, lyases, isomerases and ligases. Among these classes, of particular importance will be the production of improved proteases, carbohydrases, lipases, dioxygenases and peroxidases. These and other enzymes that can be prepared by the method of the invention have important commercial applications for enzymatic conversions in health care, cosmetics, foods, brewing, detergents, environment (e.g., wastewater treatment), agriculture, tanning, textiles, and other chemical processes. These include, but are not limited to, diagnostic and therapeutic applications, conversions of fats, carbohydrates and protein, degradation of organic pollutants and synthesis of chemicals. For example, therapeutically effective proteases with fibrinolytic activity, or activity against viral structures necessary for infectivity, such as viral coat proteins, can be engineered. Such proteases could be useful anti-thrombotic agents or anti-viral agents against viruses such as, for example, HIV, rhinoviruses, influenza, or hepatitis. In the case of oxygenases (e.g., dioxygenases), a class of enzymes requiring a co-factor for oxidation of aromatic rings and other double bonds, industrial applications in biopulping processes, conversion of biomass into fuels or other chemicals, conversion of waste water contaminants,

bioprocessing of coal, and detoxification of hazardous organic compounds are possible applications of novel proteins.

The present invention is further illustrated in the following examples, which
5 should not be construed as limiting.

Exemplification

Throughout the examples, the following materials and methods were used unless
10 otherwise stated.

Materials and Methods

In general, the practice of the present invention employs, unless otherwise indicated, conventional techniques of chemistry, molecular biology, recombinant DNA
15 technology, PCR technology, immunology (especially, e.g., antibody technology), expression systems (e.g., cell-free expression, phage display, ribosome display, and ProfusionTM), and any necessary cell culture that are within the skill of the art and are explained in the literature. See, e.g., Sambrook, Fritsch and Maniatis, *Molecular Cloning: Cold Spring Harbor Laboratory Press* (1989); *DNA Cloning*, Vols. 1 and 2,
20 (D.N. Glover, Ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait, Ed. 1984); *PCR Handbook Current Protocols in Nucleic Acid Chemistry*, Beaucage, Ed. John Wiley & Sons (1999) (Editor); *Oxford Handbook of Nucleic Acid Structure*, Neidle, Ed., Oxford Univ Press (1999); *PCR Protocols: A Guide to Methods and Applications*, Innis et al., Academic Press (1990); *PCR Essential Techniques: Essential Techniques*, Burke, Ed.,
25 John Wiley & Son Ltd (1996); *The PCR Technique: RT-PCR*, Siebert, Ed., Eaton Pub. Co. (1998); *Antibody Engineering Protocols (Methods in Molecular Biology)*, 510, Paul S., Humana Pr (1996); *Antibody Engineering: A Practical Approach (Practical Approach Series, 169)*, McCafferty, Ed., Irl Pr (1996); *Antibodies: A Laboratory Manual*, Harlow et al., C.S.H.L. Press, Pub. (1999); *Current Protocols in Molecular Biology*, eds. Ausubel et al., John Wiley & Sons (1992); *Large-Scale Mammalian Cell Culture Technology*, Lubiniecki, A., Ed., Marcel Dekker, Pub., (1990). *Phage Display: A Laboratory Manual*, C. Barbas (Ed.), CSHL Press, (2001); *Antibody Phage Display*, P O'Brien (Ed.), Humana Press (2001); Border et al., Yeast surface display for screening combinatorial polypeptide libraries, *Nature Biotechnology*, 15(6):553-7 (1997); Border et al.,
30 Yeast surface display for directed evolution of protein expression, affinity, and stability, *Methods Enzymol.*, 328:430-44 (2000); ribosome display as described by Pluckthun et al. in U.S. Patent No. 6,348,315, and ProfusionTM as described by Szostak et al. in U.S. Patent Nos. 6,258,558; 6,261,804; and 6,214,553.

EXAMPLE 1**LOOK-THROUGH MUTAGENESIS OF THREE DEFINED REGIONS IN AN ANTIGEN BINDING MOLECULE**

5 In this example, the look-through mutagenesis of three CDRs of an antibody to improve binding and proteolysis of a substrate is described.

In particular, the "look-through" mutagenesis of three complementarity determining regions (CDRs) of a monoclonal antibody is performed. CDR1, CDR2, and CDR3 of the heavy chain variable region (VH) are defined regions selected for look-
10 through mutagenesis. For this embodiment, the predetermined amino acids selected are the three residues of the catalytic triad of serine proteases, Asp, His and Ser. Asp is selected for VH CDR1, His is selected for VL CDR2, and Ser is selected for VH CDR3. The selection of these three predetermined amino acids allows for the use of a convenient protease assay in order to detect when the three residues are positioned
15 correctly to exhibit a functional activity, i.e., proteolysis of a test substrate.

An exemplary antibody, MCPC 603, is recognized as a good model for investigating binding and catalysis because the antibody binding region has been well characterized. The amino acid sequence and DNA sequence of the MCPC 603 VH and VL regions are publicly available (see, e.g., Rudikoff, S. and Potter, M., *Biochemistry*
20 13: 4033 (1974); Pluckthun, A. et al., *Cold Spring Harbor Symp. Quant. Biol.*, Vol. LII: 105-112 (1987)). The CDRs for the MCPC 603 antibody have been identified as shown in Fig. 2. In the heavy chain, CDR1 spans amino acid residue positions 31-35, CDR2 spans positions 50-69, and CDR3 spans positions 101-111. In the light chain, the amino acid residues of CDR1 are 24-40, CDR2 spans amino acids 55-62, and CDR3 spans
25 amino acids 95-103.

The design of the oligonucleotides for look-through mutagenesis in the CDRs of MCPC 603 is such that polypeptide analogs result having the amino acid sequence for CDR1, CDR2, and CDR3 as shown in Figs. 3-5. It is understood that the oligonucleotides synthesized can be larger than the CDR regions to be altered to facilitate
30 insertion into the target construct as shown in Fig. 7. A single chain antibody format is chosen for convenience in subsequent expression and screening steps. The oligonucleotides can be converted into double-stranded chains by enzymatic techniques (see e.g., Oliphant, A. R. et al., 1986, *supra*) and then ligated into a restricted plasmid as shown in Fig. 8. The restriction sites can be naturally occurring sites or engineered
35 restriction sites.

Polynucleotides encoding corresponding polypeptide analogs can be expressed in any of the convenient expression systems described herein and screened using the serine protease assay described, e.g., in U.S. Patent No. 5,798,208 (see also, Figs. 16-17).

24

Briefly, the expressed polypeptide analogs are exposed to a test substrate and examined for proteolysis of the test substrate. The amount of proteolysis, revealed by a zone of clearance of the substrate, indicates a polypeptide analog having the desired functional activity.

EXAMPLE 2

LOOK-THROUGH MUTAGENESIS OF SIX DEFINED REGIONS IN AN ANTIGEN BINDING MOLECULE

In this example, the look-through mutagenesis of all six CDRs of an antibody to improve binding and proteolysis of a substrate is described.

In particular, a "look-through" mutagenesis of all six of the hypervariable regions or complementarity determining regions (CDRs) of the above mentioned model antibody (MCPC 603) is performed. In this example, "look-through" mutagenesis is carried out from two to three times with a different amino acid in a given region or domain. For example, Asp, Ser and His are sequentially walked-through the heavy and light chains as shown in Fig. 10.

Mutagenesis of noncontiguous residues within a region can be desirable if it is known, or if one can deduce, that certain residues in the region will not participate in the desired function. In addition, the number of analogs can be minimized. Other considerations in selecting the predetermined amino acid and the particular positions to be altered are that the residues must be able to hydrogen bond with one another. This consideration can impose a proximity constraint on the variants generated. Thus, only certain positions within the CDRs may permit the amino acids of the catalytic triad to interact properly. Thus, molecular modeling or other structural information can be used to enrich for functional variants.

In this case, known structural information was used to identify residues in the regions that may be close enough to permit hydrogen bonding between Asp, His and Ser, as well as the range of residues to be mutagenized. Roberts et al. have identified regions of close contact between portions of the CDRs (Roberts, V. A. et al., Proc. Natl. Acad. Sci. USA 87: 6654-6658 (1990)). This information together with data from the x-ray structure of MCPC 603 is used to select promising areas of close contact among the CDRs targeted for mutagenesis. This type of look-through mutagenesis guided by structural / modeling information can be referred to as "guide-through" mutagenesis.

Look-through mutagenesis is carried out as illustrated in Fig. 10 where each CDR is subjected to one predetermined amino acid resulting 8×10^5 polypeptide analogs and if all twenty amino acids are explored, 5×10^{13} polypeptide analogs.

Polynucleotides can be expressed in any of the convenient expression systems described herein and screened using the serine protease described, e.g., in U.S. Patent No.

5,798,208. Briefly, the expressed polypeptide analogs are exposed to a test substrate and examined for proteolysis of the test substrate. The amount of proteolysis, revealed by a zone of clearance of the substrate, indicates a polypeptide analog having the desired functional activity.

5

EXAMPLE 3

LOOK-THROUGH MUTAGENESIS OF ANTI-TNF BINDING MOLECULES TO IMPROVE FUNCTION

In this example, the look-through mutagenesis of an anti-TNF antibody to improve binding is described.

In particular, the "look-through" mutagenesis of all six of the hypervariable regions or complementarity determining regions (CDRs) of two different anti-TNF antibodies is performed. Anti-TNF antibodies have general application in the treatment of immune disease in patients having inappropriate levels of the ligand TNF (tumor necrosis factor). Two commercially available anti-TNF antibodies exist. For convenience in performing look-mutagenesis and subsequent screening, the variable light and heavy chain regions (see SEQ ID NOs: 2-4) of these antibodies were converted to a single chain format using a poly Gly-Ser linker (see Fig. 15). The defined regions selected for look through mutagenesis with a predetermined amino acid are identified by the presence of a black bar as shown in Fig. 15. These defined regions correspond to the CDRs of the single chain antibodies.

Polynucleotides representing the six CDR regions and sufficient flanking regions to allow for the assembly of the polynucleotides into the complete single chain sequence shown in Fig. 15 are synthesized as described herein. Predetermined amino acid residues are selected for each CDR region and separately and sequentially introduced into each amino acid position throughout the six CDR regions. The polynucleotides are further engineered to serve as templates capable of supporting the transcription of a corresponding RNA transcript that can then be translated into a polypeptide using ribosome display.

Polynucleotides encoding corresponding polypeptide analogs are expressed using a cell-free transcription and translation extracts. The RNA transcripts are covalently linked to a detectable moiety such as a fluorescent moiety. Alternatively, the sequence of interest can be fused in frame to a fluorescent moiety such as green fluorescent protein (GFP) to allow for the convenient detection of expression and normalization of binders versus non-binders. Preferably, the polynucleotides encoding the polypeptide analogs are at least partially arrayed, e.g., expressed in a well that has multiple polypeptide analogs but can be readily deconvoluted. Accordingly, each well contains a subset of polypeptide analogs now linked to the corresponding transcript linked to a fluorescent moiety using

ribosome display. The well is probed with target ligand, i.e., TNF and assayed for polypeptide-analogs that bind and with what binding affinity as compared to wild-type polypeptide.

Polypeptide analogs that bind better than the wild-type polypeptide are then re-engineered into a full-length IgG antibody format for parallel testing with the analogous commercial antibody for improved binding using standard techniques.

EXAMPLE 4

LOOK-THROUGH MUTAGENESIS OF ANTIBODIES AGAINST BOTULINUM NERVE TOXIN SEROTYPE B (BoNT/B) AND BOTULINUM NERVE TOXIN SEROTYPE A (BoNT/A) TO IMPROVE FUNCTION

In this example, improved antibodies against botulinum nerve toxin serotype B (BoNT/B) and botulinum nerve toxin serotype A (BoNT/A) are generated using look-through mutagenesis (LTM) to improve function. The LTM approach is based on creating single mutations per CDR throughout the binding pocket, based on a subset (the LTM set) of the 20 amino acids that explore size, charge, hydrophobicity, and hydrogen bonding characteristics. The criteria for selecting the amino acids in the LTM set are discussed below.

1. Antibody Purification and Gene Sequencing and Single Chain (scFv) Design

Murine antibody fragments (Fab) with binding affinities to BoNT/B can be obtained as described in Emanuel *et al.* (1996) *Journal of Immunological Methods* 193: 189-197. The *BotFab 5* (SEQ ID NOs:10 and 11, light and heavy chain polypeptides, respectively), the *BotFab 20* (SEQ ID NOs:12 and 13, light and heavy chain polypeptides, respectively) or the *BotFab 22* (SEQ ID NOs:14 and 15, light and heavy chain polypeptides, respectively) antibodies can be used. The foregoing sequences are described in U.S. Patent No. 5,932,449, the contents of which are incorporated herein by reference. The murine BoNT/B antigen (full-length toxin, light chain, and/or heavy chain) can be obtained from Metabiolabs, Inc., WI.

The anti-BoNT/A antibodies described in Pless *et al.* (2001) *Infect. Immun.* 69:570 can be used with the objective of improving their affinities by at least one order of magnitude. The BoNT/A heavy chain binding domain (BoNT/A Hc) antigen can be obtained from Metabiolabs, Inc., WI.

The V_L and V_H fragments of the antibody(ies) are cloned and sequenced using standard molecular biology techniques. The variable regions of the molecules are amplified by the polymerase chain reaction (PCR) and linked with a poly-Gly-Ser linker (typically SGGGGSGGGGSGGGGS (SEQ ID NO:7)) to generate single chain antibodies (scFv). A poly-His tag (HHHHHH (SEQ ID NO:8)) and a myc tag

(EQKLISEEDL (SEQ ID NO:9)) are also appended to the C-terminus of the genes to facilitate purification and detection. These molecules are displayed on any of the well known technologies (e.g., yeast, bacterial or phage based technologies) and tested for their ability to bind BoNT/B or BoNT/A.

The monovalent scFv version of whole antibodies provide for a good format to undertake neutralization studies and are known to generally reproduce the binding mechanism of the whole molecule, with the exception of the avidity effects displayed by multivalent molecules.

2. Antibody Improvement Using Look Through Mutagenesis (LTM) and Gene Design

For the Look Through Mutagenesis (LTM), the following nine amino acids and their representative functional characteristics are chosen: Alanine and Leucine (aliphatics), Serine (hydroxyl group), Aspartic Acid (acidic) and Glutamine (amide), Lysine and Histidine (basic), Tyrosine (aromatic), Proline (hydrophobic). These amino acids display adequate chemical diversity in size, charge, hydrophobicity, and hydrogen bonding ability to provide meaningful initial information on the chemical functionality needed to improve antibody properties. The choices are also based on the frequency of occurrence of these amino acids in the CDRs of antibodies. For example, given between tyrosine and phenylalanine to represent amino acids with aromatic side chains, the former is chosen because of its significantly higher preponderance in antibody CDRs and its ability to hydrogen bond. LTM is initially employed to identify specific amino acids and chemical properties that are beneficial for binding, neutralization and/or any additional properties desired in the final antibody. However, LTM is not limited to these nine amino acids or nine total amino acids. LTM analysis can be performed with any combination of amino acids and the LTM subset can be as high as 18 amino acids (1 short of saturation mutagenesis).

2.1 LTM Oligonucleotide Synthesis

In the primary LTM analysis, the goal is to explore each amino acid's side chain contribution to the overall binding affinity within each CDR. To efficiently generate meaningful diversity, each of the LTM subset amino acid is targeted at every single position within the CDR sequence with only one substitution per CDR. Thus, each individual oligonucleotide encodes for only a single CDR mutation. For instance, in order to do LTM Histidine mutagenesis on a hypothetical eleven amino acid V_H CDR3 domain, 11 oligonucleotides are synthesized that encode for all 11 possible single Histidine mutations (see Figure 20). Such an analysis tests the effects of having a bulky amide in each position in the CDR. Therefore to generate a V_H CDR3 LTM library,

only 99 oligonucleotides are synthesized for the LTM analysis (9 LTM amino acids x 11 V_H CDR3 positions). Oligonucleotide sequences are tested for inadvertent stop codons, wild-type duplication, inefficient codon usage, hairpins, loops, and other secondary structures using publicly available software.

2.2 Degenerate oligonucleotide synthesis for combinations of beneficial LTM mutations

In employing systematic LTM replacement of individual amino acids within a CDR, the preference of chemical functionalities at each position is uncovered. In order to combine all the mutations from the LTM selections and explore possible additivity and energetic synergy between these, degenerate oligonucleotides encoding for these mutations and the wild-type sequence are synthesized. This degenerate pool of oligonucleotides with 1.6×10^4 variants is subsequently used to generate a second generation library that explores the additive nature of these substitutions.

2.3 Computer Assisted Oligonucleotide Design, Library, and Results Database

Software coupled with automated custom-built DNA synthesizers enables rapid oligonucleotide synthesis. The first step involves deciding which target amino acids will be incorporated into the CDRs. The software determines the codon preference (e.g., yeast, bacterial or phage codon preferences, depending on the chosen system) needed to introduce the targeted amino acids and also eliminates any duplication of the wild-type sequence that may be generated by this design process. It then analyzes for potential stop codons, hairpins, and loop structures or for other problematic sequences that are subsequently corrected prior to synthesis. The completed LTM design plan is then sent to the DNA synthesizer, which performs an automated synthesis of the oligonucleotides. In this manner, the oligonucleotides needed to create the libraries can be rapidly generated.

An electronic database may store information of all LTM oligonucleotide sequences synthesized, details of the scFv library CDR substitutions, and binding assay results for a target antigen. Archival of the oligonucleotide data allows for rational iterations of the design strategies, and re-use of reagent oligonucleotides.

3. Overall LTM Strategy

LTM is initially employed to identify specific amino acids and chemical properties that are beneficial for binding, neutralization and/or any additional properties desired in the final antibody. It also quickly identifies the regions that do not tolerate any mutagenesis without significant loss in affinity (or any physical property being selected). Therefore, not only is the LTM analysis a good methodology to explore the chemical

requirements at each position in all CDRs of an antibody, but it also rapidly determines the amino acids absolutely required for antigen binding. After the identification of beneficial mutations by LTM, combinatorial mutagenesis schemes may be employed that firstly incorporate all these disparate amino acid mutations to generate multiply mutated CDRs. Additionally or alternatively, "Walk Through" mutagenesis (WTM) may be used to probe the effects of multiple mutations of the same amino acid in a CDR (as described, for example, in U.S. Patent Nos: 5,830,650; 5,798,208).

4. LTM scFv Libraries

The LTM techniques described above are used to create pools of oligonucleotides with mutations in a single CDR of the light or heavy chain. These oligonucleotides are synthesized to include some of the surrounding framework to facilitate the overlap and hybridization during PCR. These pools of oligonucleotides are utilized to generate all possible V_L and V_H chains in which there are mutations in single, double, and triple CDRs (single, double, and triple combinations of CDR1, 2, and 3) using single overlap extension PCR (SOE-PCR) (as described by Horton *et al.* (1989) *Gene* 77:61-68). SOE-PCR is a fast and simple method for combining DNA fragments that does not require restriction sites, restriction endonucleases, or DNA ligase. In SOE-PCR two regions in the gene are first amplified by PCR using primers designed so that the PCR products share a complementary sequence at one end. Under PCR conditions, the complementary sequences hybridize, forming an overlap. The complementary sequences then act as primers, allowing extension by DNA polymerase to produce a recombinant molecule.

For example, to create the pool of V_H chains in which both CDR-H1 and CDR-H2 are mutated and CDR-H3 is wild-type, which is denoted as "110" (1 denotes a mutant CDR and 0 denotes a wild-type CDR), the CDR-H1 mutant genes are used as templates and SOE-PCR is conducted to link the CDR-H2 oligonucleotides to generate the doubly mutated pool (Figure 21). Considering that each CDR may be either wild-type or mutant, there are seven possible combinations (depicted by arrows in Figure 21) for each of the pools of V_L and V_H chains (not including the wild-type molecule "000"). Combining the seven V_L and eight V_H pools creates 63 V_L - V_H non-wild-type combinations (scFvs), (Figure 21). Each of the 64 V_L - V_H combinations (including the wild-type sequence) is termed as a "subset" of the whole LTM scFv library ensemble. An scFv library ensemble is created for each amino acid selected for substitution. The number of amino acid sequences represented within each subset library depends on the length of the CDR, the amino acid sequence within the CDR, and the LTM oligonucleotide design strategy.

5. High-Throughput Library Screening and Improved Antibody Selection

A variety of methods are available for antibody expression and display. These include bacteriophage, *Escherichia coli*, and yeast. While each of these methods has been used for antibody improvement, the yeast display system affords several advantages (Boder and Wittrup (1997) *Nat. Biotechnol.* 15:553-557). Yeast can readily accommodate library sizes up to 10^7 , with 10^3 - 10^5 copies of each antibody being displayed on each cell surface. Yeast cells are easily screened and separated using flow cytometry and fluorescence-activated cell sorting (FACS) or magnetic beads. Yeast also affords rapid selection and re-growth. The eukaryotic secretion system and glycosylation pathways of yeast allow for a much larger subset of scFv molecules to be correctly folded and displayed on the cell surface than prokaryotic display systems. Yeast display coupled with directed evolution has been used to increase the K_D of an scFv antibody fragment for fluorescein to 48 fM, two orders of magnitude stronger binding than any previously reported monovalent ligand (Boder *et al.* (2000) *PNAS* 97: 10701-10705). The display system utilizes the α -agglutinin yeast adhesion receptor to display proteins on the cell surface. The proteins of interest, in the present case, anti-BoNT/B scFv LTM libraries or anti-BoNT/A scFv LTM libraries, are expressed as fusion partners with the Aga2 protein. These fusion proteins are secreted from the cell and become disulfide linked to the Aga1 protein, which is attached to the yeast cell wall (see Invitrogen, pYD1 Yeast Display product literature). In addition, there are carboxyl terminal tags included which may be utilized to monitor expression levels and/or normalize binding affinity measurements.

Streptavidin coated magnetic beads (Spherotech) are used to screen and select for antibodies that bind to BoNT/B or BoNT/A with high affinity. This methodology employs high affinity antibody binding to biotinylated antigen which then binds to streptavidin coated beads in order to select for the yeast clones (Yeung and Wittrup (2002) *Biotechnol. Prog* 18:212-220) and Feldhaus *et al.* (2003) *Nature Biotech.* 21:163-170). The BoNT/B or BoNT/A polypeptide (MetabioLogics) is biotinylated using standard protocols (Pierce) and screening is performed using methods well known in the art. Using equilibrium and kinetic based selection, antibodies with improved affinities are selected for from these libraries. The efficacy of each round of selection is monitored by analytical FACS (FACScan). In addition, relative binding affinities of individual molecules displayed on the yeast surface are measured by titrating the antigen. This allows for rapid identification of molecules with improved affinity. The scFv clones are then sequenced to identify beneficial mutations.

6. Generation of Soluble Antibodies for BIAcore Affinity Measurements

Antibodies of interest are sub-cloned into soluble expression systems (*Pichia pastoris* and/or *E. coli*) and soluble protein is generated. There are several commercially available vectors and cell lines for soluble antibody expression, including those from Invitrogen (e.g., pPIC9 for *P. pastoris*) and Novagen (pET20b for periplasmic expression in *E. coli*). These systems are routinely used to generate soluble single chain or full-length antibodies. The *P. pastoris* expression system (Invitrogen) routinely produces 1-5 mg per liter of soluble purified scFv. Purification of proteins is facilitated by the presence of a His-tag at the C-terminus of the molecule, in the case of single chains or by protein A or protein G columns for full-length antibodies. Soluble single chain and full-length antibodies are generated to obtain BIAcore affinity kinetic rate measurements. This step is necessary as high affinity scFv molecules on the yeast clone cell surface must be verified as a cell free soluble molecule.

Soluble single chain and full-length antibodies generated in the foregoing manner may be used to obtain BIAcore affinity measurements, as well as in the neurite outgrowth assay or the mouse lethality assay described below.

7. Screening of Selected Antibodies for *In Vivo* or *In Vitro* Neutralization

A cell-based assay using the neurite outgrowth of primary chick neurons as an indicator of BoNT intoxication and, thus, as a means to quantify toxin neutralization may be used to screen the selected antibodies. Preliminary experiments using dorsal root ganglion explant cultures from fertilized chicken eggs have indicated that there was less axonal outgrowth from explants treated with BoNT/A than the controls.

Alternatively, a chick ciliary ganglion-iris muscle neuromuscular junction assay (as described in Lomoth *et al.* (1990) *Neuroscience Letters* 113:211-216) may be used.

The mouse lethality assay (MLA, as described by, for example, Schantz and Kautter (1978) *J. Assoc. Off. Anal. Chem.* 61:96-99) is another well known and accepted *in vivo* method for testing for BoNT neutralization. The test involves the intraperitoneal injection of approximately 0.5 ml of sample preparations of BoNT/B or BoNT/A with and without the antibody into 20 to 30 gm, white ICR strain mice. Lethality due to respiratory failure is noted over 1-4 days. Quantification of neutralization requires serial dilutions of the MAbs with varying levels of mouse BoNT/B or BoNT/A LD₅₀. The MLA may be used to determine the neutralizing ability of the optimized antibodies identified by the *in vitro* screening.

The neutralizing ability of antibodies can also be measured *in vitro* by the mouse protection assay (MPA, Goeschel *et al.* (1997) *Exp. Neurol.* 147:96-102). In the MPA the left phrenic nerve, together with the left hemidiaphragm, is excised from the mouse.

The phrenic nerve is then continuously electrostimulated in a tissue bath. Purified antibodies are incubated with BoNT/B or BoNT/A and added to the tissue bath. Toxin induced paralysis is defined as a 50% reduction in the initial muscle twitch.

5

Equivalents

10 Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

Claims

1. A method of generating a library of polypeptide analogs in which a predetermined amino acid appears at each position in a defined region of the polypeptide comprising:
 - 5 selecting a defined region of the amino acid sequence of the polypeptide;
 - determining an amino acid residue to be substituted at each amino acid position within the defined region;
 - synthesizing individual polynucleotides encoding the defined region, the polynucleotides collectively representing possible variant polynucleotides according to
 - 10 the following criteria:
 - i) each polynucleotide containing at each codon position in the defined region, either a codon required for the amino acid residue of the polypeptide or a codon for the predetermined amino acid residue, and
 - ii) each polynucleotide containing no more than one codon for the predetermined
 - 15 amino acid residue,
 - thereby generating a library of polynucleotides in which the predetermined amino acid residue appears at each amino acid position within the defined region.
2. The method of claim 1, wherein the polynucleotides are pooled together.
- 20 3. The method of claim 1, wherein two or more defined regions within the polypeptide are mutagenized.
4. The method of claim 3, wherein the same predetermined amino acid is selected for
- 25 substitution within each of the two or more defined regions.
5. The method of claim 3, wherein different predetermined amino acids are selected for substitution within each of the two or more defined regions, respectively.
- 30 6. The method of any one of the previous claims, wherein the defined region or defined regions comprises a functional domain of the polypeptide.
7. The method of claim 6, wherein the functional domain is selected from the group consisting of an antibody binding site, an antibody framework region, an antibody
- 35 effector region, a receptor binding site, and a catalytic site.

8. The method of claim 7, wherein the antibody binding site or portion thereof comprises a CDR domain selected from the group consisting of CDR1, CDR2, CDR3, CDR4, CDR5, CDR6 and a combination thereof.
9. The method of claim 7, wherein the antibody framework region comprises a domain selected from the group consisting of FR1, FR2, FR3, FR4 and a combination thereof.
10. The method of claim 7, wherein the antibody effector region comprises a domain selected from the group consisting of a complement binding site and an Fc binding region.
11. The method of claim 1, wherein the predetermined amino acid residue is selected from the group consisting of Ser, Thr, Asn, Gln, Tyr, Cys, His, Glu, Asp, Lys, Arg, Ala, Gly, Ile, Leu, Met, Phe, Pro, Trp and Val.
12. The method of any one of the previous claims, wherein the defined region comprises at least about 3 to 40 amino acids.
13. The method of any one of the previous claims, wherein the polynucleotides are synthesized as expression library.
14. The method of any one of the previous claims, wherein the polynucleotides are synthesized using enzymatic means.
15. The method of any one of the previous claims, wherein the polynucleotides are synthesized using polymerase chain reaction.
16. The method of any one of the previous claims, wherein the expression library is selected from the group consisting of a phage display library, a ribosome/polysome display library, a yeast display library, a bacterial display library and an arrayed library.
17. A library of polypeptide analogs prepared by the method of claim 1.
18. A method of identifying a polypeptide having a desired structure or function comprising:
selecting a defined region of the amino acid sequence of the polypeptide;
determining an amino acid residue to be substituted at each amino acid position within the defined region;

synthesizing polynucleotides encoding the defined region, the polynucleotides collectively representing possible variant polynucleotides according to the following criteria:

- 5 i) each polynucleotide containing at each codon position in the defined region, either a codon required for the synthesis of the amino acid residue of the polypeptide or a codon for one of the predetermined amino acid residue, and
- ii) each polynucleotide containing no more than one codon for the predetermined amino acid residue,
- 10 thereby generating an expression library containing the polynucleotides; expressing the expression library to produce polypeptide analogs; and screening the polypeptide analogs to select for a polypeptide having a desired structure or function.

15 19. The method of claim 18, wherein the method further comprises the step of identifying the polynucleotide that encodes the selected polypeptide analog.

20 20. The method of claim 18, wherein the screening comprises, contacting a polypeptide with a target substrate, the polypeptide being associated with the polynucleotide encoding the polypeptide, the polynucleotide further comprising a detectable moiety, such that a variant polypeptide capable of binding a target substrate is detected and thereby identified as encoded by the polynucleotide.

25 21. The method of claim 20, wherein the detectable moiety is selected from the group consisting of a fluorescent moiety, a UV moiety, and a visible light absorbing moiety.

22. The method of claim 20, wherein the detectable moiety is selected from the group consisting of a biotin moiety, a GST moiety, and a His tag moiety.

30 23. The method of claim 20, wherein the polynucleotide is associated with the polypeptide analog using ribosome display.

24. The method of claim 18, wherein two or more defined regions within the polypeptide are mutagenized.

35 25. The method of claim 24, wherein the same predetermined amino acid is selected for substitution within each of the two or more defined regions.

26. The method of claim 24, wherein different predetermined amino acids are selected for substitution within each of the two or more defined regions.
27. The method of claim 18, wherein the polypeptide is a single chain antibody (sFVs).
28. The method of claim 18, wherein the defined region comprises a functional domain of the polypeptide.
29. The method of claim 18, wherein the defined region comprises a CDR or portion thereof selected from the group consisting of CDR1, CDR2, CDR3, CDR4, CDR5, CDR6 and a combination thereof.
30. The method of claim 18, wherein the defined region is an antibody framework region comprising a domain selected from the group consisting of FR1, FR2, FR3, FR4 and a combination thereof.
31. The method of claim 18, wherein the defined region is an antibody effector region comprising a domain selected from the group consisting of a complement binding site and an Fc binding region.
32. The method of claim 18, wherein the predetermined amino acid residue is selected from the group consisting of Ser, Thr, Asn, Gln, Tyr, Cys, His, Glu, Asp, Lys, Arg, Ala, Gly, Ile, Leu, Met, Phe, Pro, Trp and Val.
33. A library of polynucleotides encoding polypeptide analogs comprising one or more defined regions wherein a predetermined amino acid residue is substituted at each amino acid position within the defined region, the polynucleotides collectively representing all possible variants according to the following criteria:
- i) each polynucleotide contains at each codon position in the defined region, either a codon required for the amino acid residue of the polypeptide or a codon for the predetermined amino acid residue, and
 - ii) each polynucleotide contains no more than one codon for the predetermined amino acid residue.
34. The library of claim 33, wherein two or more defined regions within the polypeptide are mutagenized.

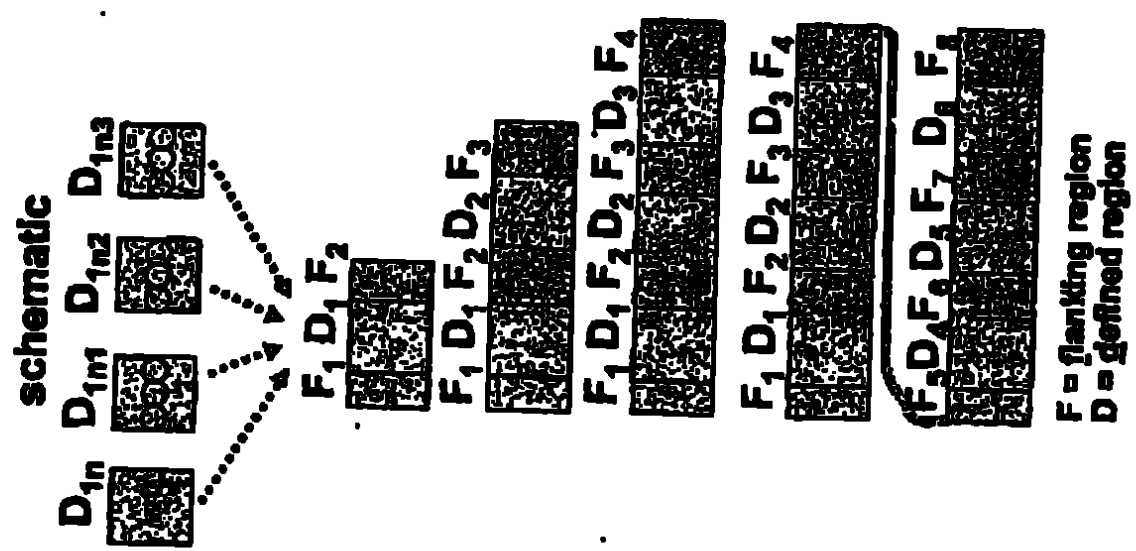
35. The library of claim 33, wherein the same predetermined amino acid is selected for substitution within each of the two or more defined regions.
36. The library of claim 33, wherein different predetermined amino acids are selected for substitution within each of the two or more defined regions.
37. The library of claim 33, wherein the library is an expression library.
38. The library of claim 33, wherein the expression library is selected from the group consisting of a phage display library, a ribosome/polysome display library, a yeast display library, a bacterial display library and an arrayed library.
39. The library of claim 33, wherein the polynucleotides further comprise one or more transcriptional regulatory elements.
40. The library of claim 39, wherein the polynucleotides, when transcribed and translated in vitro, are associated with the polypeptides encoded by the corresponding polynucleotides.
41. The library of claim 40, wherein the polynucleotides are associated with the polypeptide using ribosome/polysome display.
42. The library of claim 41, wherein the polynucleotides comprise RNA.
43. The library of claim 42, wherein the polynucleotides further comprise a detectable moiety.
44. The library of claim 43, wherein the detectable moiety comprises a fluorescent moiety.
45. The library of claim 33, wherein the library comprises at least 10^4 different polynucleotides.
46. The library of claim 33, wherein the library comprises at least $45 \cdot 10^{12}$ different polynucleotides.
47. The library of claim 33, wherein the polypeptide encodes a binding polypeptide.

48. The library of claim 47, wherein the binding polypeptide is selected from the group consisting of a heavy-chain variable-region (V_H), a light chain variable region (V_L), and a single chain antibody (sFv).
- 5 49. The library of claim 33, wherein the polynucleotide encodes an enzyme.
50. The library of claim 33, wherein the polynucleotide encodes an enzyme inhibitor.
51. The library of claim 33, wherein the polynucleotide encodes a catalytic polypeptide.
- 10 52. The library of claim 33, wherein the library is immobilized on a solid support.
53. The library of claim 33, wherein the solid support is a microchip.
- 15 54. The library of claim 33, wherein the library is an arrayed library.
55. A microchip comprising an array of immobilized polynucleotides according to the library of claim 33.
- 20 56. A polypeptide analog identified according to the library of claim 33, wherein the polypeptide binds to a target molecule and comprises a binding region selected from the group consisting of a heavy chain variable region (V_H), a light chain variable region (V_L), and a single chain antibody (sFv).
- 25 57. The molecule of claim 56, wherein the specified target molecule is TNF α .
58. The molecule of claim 56, wherein the polypeptide has at least 70% identity to the amino acid sequence of SEQ ID NO:1, 2, 3, 4, 5 or 6.
- 30 59. A method of identifying a subset of polypeptide analogs having a desired structure or function comprising:
- selecting a defined region of the amino acid sequence of the polypeptide;
- determining an amino acid residue to be substituted at each amino acid position within the defined region;
- 35 synthesizing polynucleotides encoding the defined region, said polynucleotides collectively representing possible variant polynucleotides according to the following criteria:

- i) each polynucleotide containing at each codon position in the defined region, either a codon required for the synthesis of the amino acid residue of the polypeptide or a codon for the predetermined amino acid residue, and
- 5 ii) each polynucleotide containing no more than one codon for the predetermined amino acid residue,
- thereby generating an expression library containing the polynucleotides; exposing the expression library to conditions under which the library is expressed;
- 10 screening the expressed library to identify a polypeptide having a desired structure or function;
- comparing the structure or function of the polypeptide as compared to a control criterion, wherein a polypeptide that corresponds or exceeds the control criterion is categorized as a responder and a polypeptide that fails the control criterion is categorized
- 15 as a nonresponder;
- categorizing responders and nonresponders in a database; and
- querying the database to determine the sequence of a subset polypeptides to be synthesized.
- 20 60. The method of claim 59, wherein one or more of the above steps is computer-assisted.
61. The method of claim 59, wherein the control criterion is selected from the group consisting of binding affinity, stability and effector function.
- 25 62. The method of claim 59, wherein the control criterion is catalytic activity on a specified substrate.
63. The method of claim 59, wherein the polypeptide is selected from the group consisting of a heavy chain variable region (V_H), a light chain variable region (V_L), and a single chain antibody (sFv).
- 30 64. A medium suitable for use in an electronic device having instructions for carrying out one or more steps of the method of claim 59.
- 35 65. A device for carrying out one or more steps of the method of claim 59.

Ex mply Look-Through Mutag nesis (LTM) Library Formats

structural features (examples)	functional activity (assays)
D region library wild type (wt) and sequential alterations	
1 D region (ligand)	binding to receptor (binding or cell response assay)
2 D regions (ligand or catalytic site)	binding to receptor or catalysis (binding, cell response, or catalysis assay)
3 D regions (VL, VH, single domain, or catalytic site)	binding to receptor or catalysis (binding, cell response, or catalysis assay)
> 3 D regions (aFv, antibody (e.g., IgG), or catalytic site)	binding to receptor or catalysis (binding, cell response, or catalysis assay)

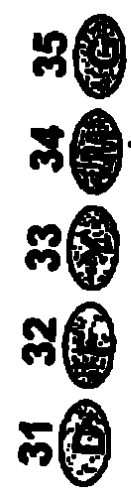
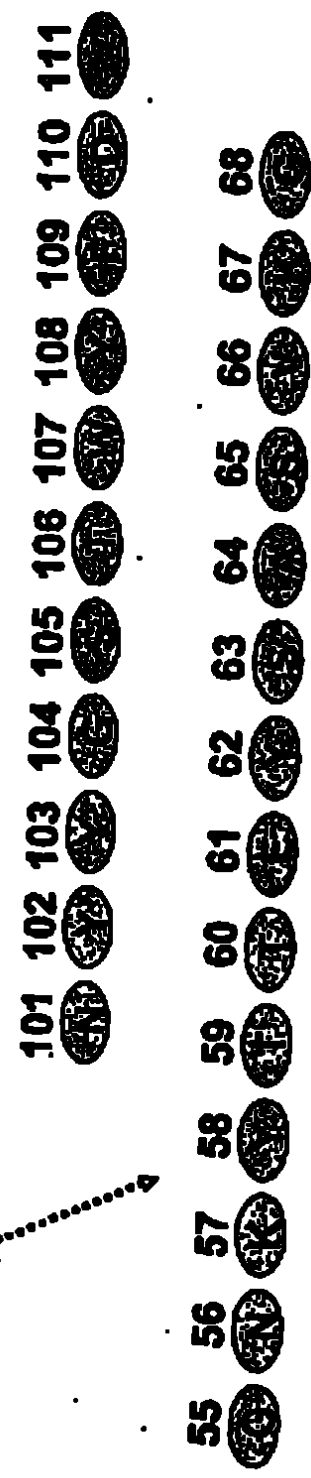


Identification of 3 Defined Regions (D1-D3) in a Synthetic Gene (sFv) for Look-Through Mutagenesis (LTM)

heavy chain variable region (V_H) linker light chain variable region (V_L)

CDR1 CDR2 CDR3 CDR1 CDR2 CDR3

D1 D2 D3 D4 D5 D5



**Sequential LTM of a First Defined Region (D1)
(V_H CDR1 amino acid positions 31-35)**

31	32	33	34	35

Sequential LTM of a Second Defined Region (D2) (V_H CDR2 amino acid positions 55-68)

[illegible]

Sequential LTM of a Third Defined Region (D3) (V_H CDR3 amino acid positions 101-111)

[illegible]

WALK-THROUGH MUTAGENESIS (WTM)

V_H CDR1 Asp (D) walk-through

31 32 33 34 35


GAC TTC TAC ATG GAG 1 oligo
 GA G GAT T run yielding
 1 4 2 7 2 27 = 128
 amino acids oligo comb.

   
 112 amino acid sequences

% desired
 amino acid (D) 100 25 50 12.5 50
 % undesired
 amino acid 0 75 50 87.5 50

LOOK-THROUGH MUTAGENESIS (LTM)

V_H CDR1 Asp (D) look-through

31 32 33 34 35


GAC TTC TAC ATG GAG

    
 5 oligo runs yielding
 5 oligo combinations
 5 amino acid sequences

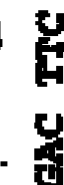
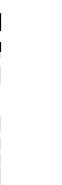
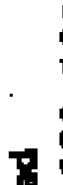
GAC GAC TAC ATG GAG

GAC TTC GAC ATG GAG

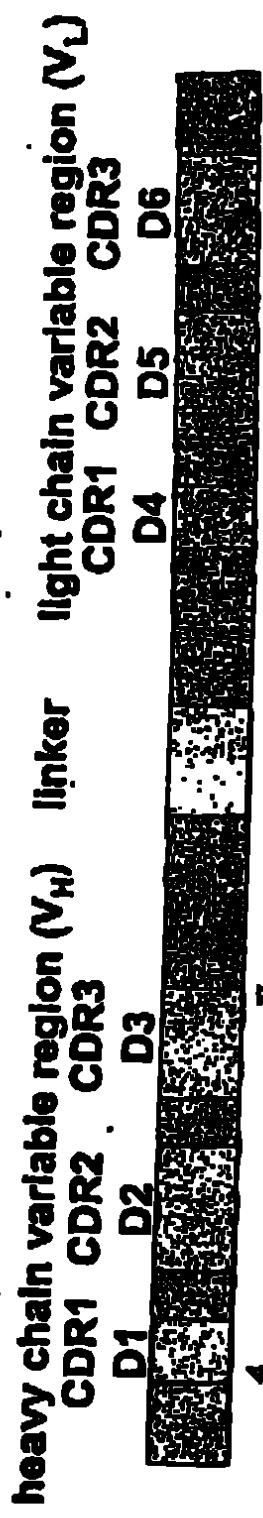
GAC TTC TAC GAC GAG

GAC TTC TAC ATG GAG

% desired
 amino acid (D) 100 100 100 100 100
 % undesired
 amino acid 0 0 0 0 0

Integration of LTM of Defined Regions D1-D3 into a Synthetic Gene (sFv)



7/21

101 102 103 104 105 106 107 108 109 110 111

101 102 103 104 105 106 107 108 109 110 111

101 102 103 104 105 106 107 108 109 110 111

Exemplary Assembly of Defined Regions into a Synthetic Gene (sFv)



convert single-stranded nucleotides to double-stranded



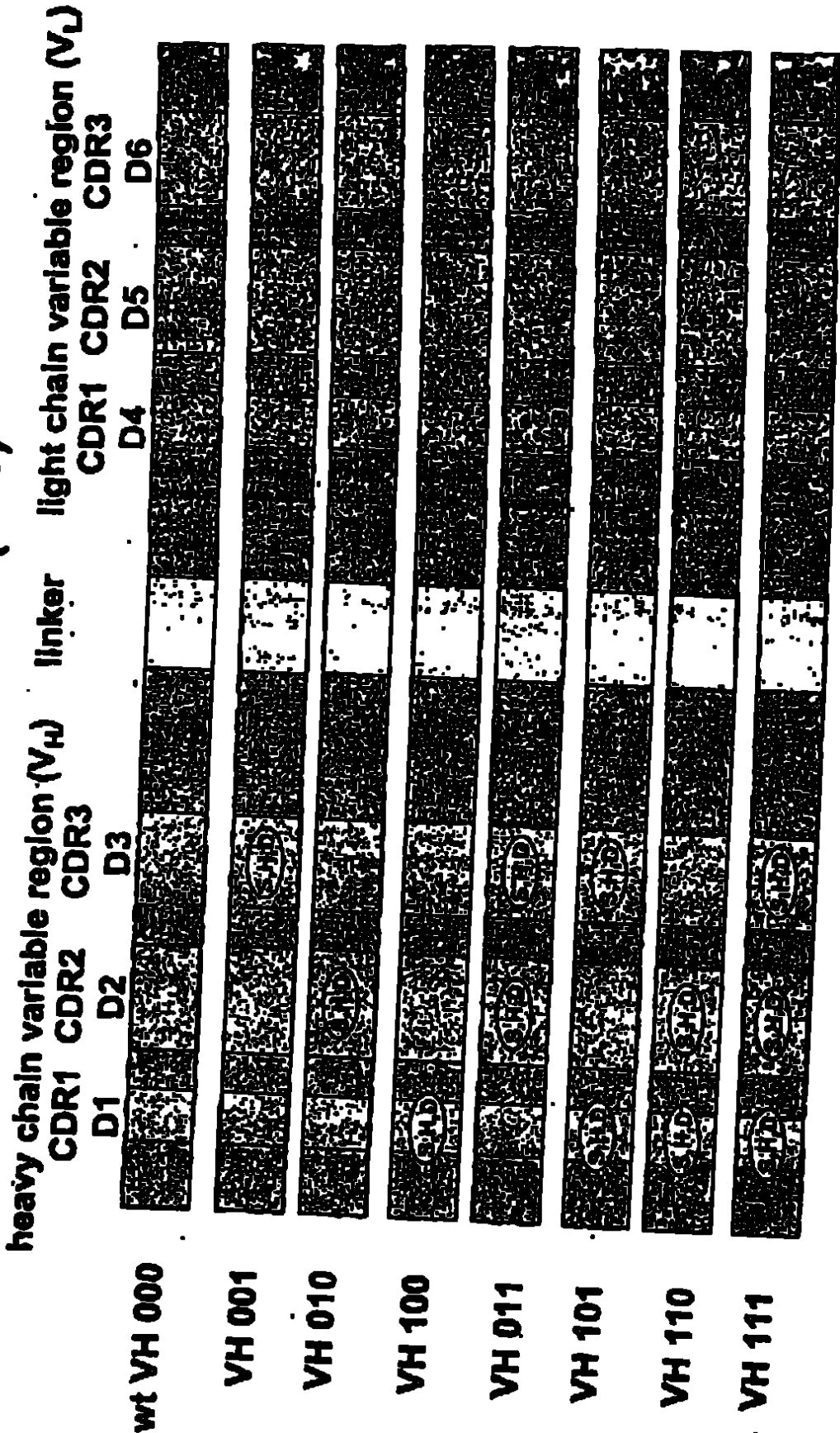
assemble w/ PCR / ligation



heavy chain variable region (V_H) linker light chain variable region (V_L)



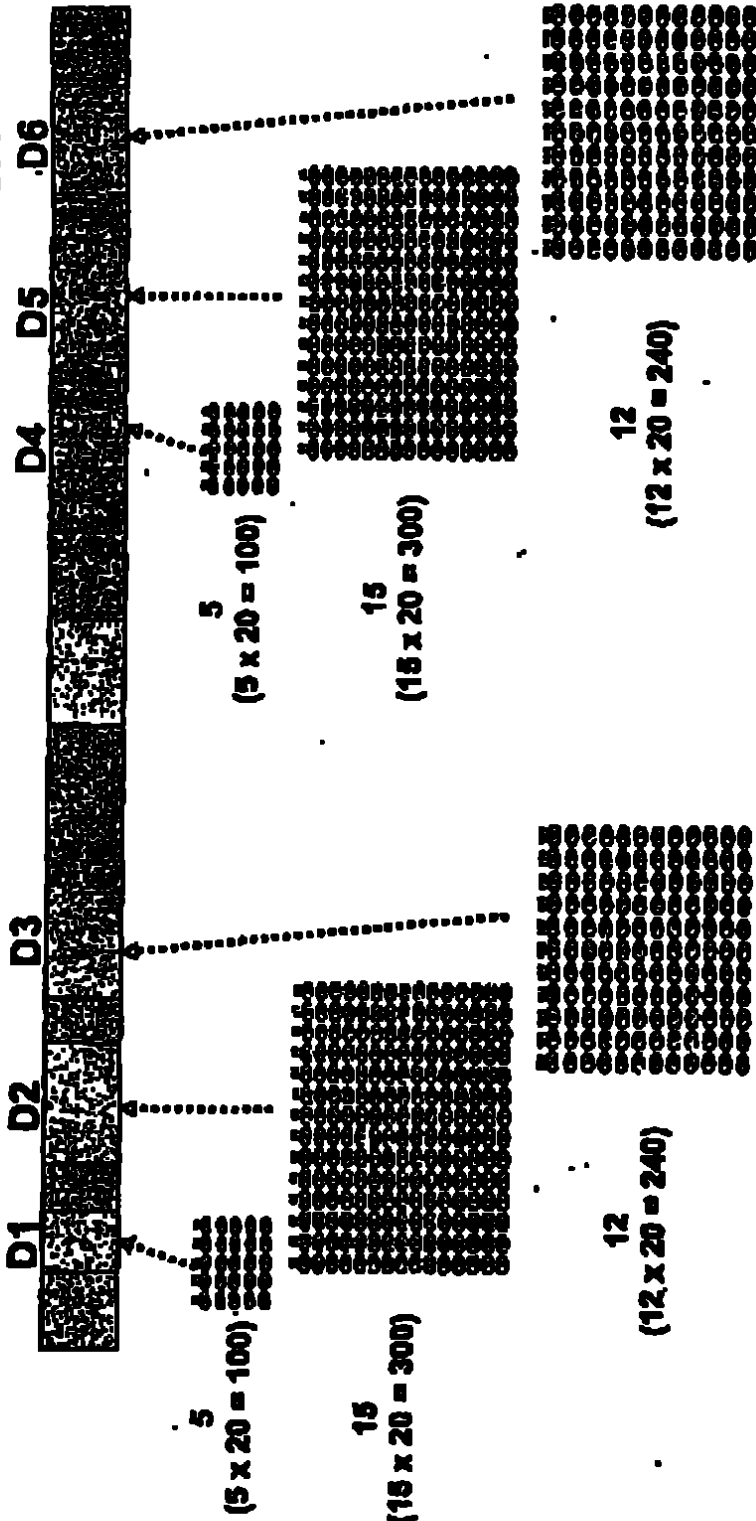
Exemplary Diversity Formats of Defined Regions D1-D3 in a Synthetic Gene (sFv)



Exemplary Diversity of Defined Regions D1-D6 in a Synthetic Gene (sFv) Using LTM

heavy chain variable region (V_H) linker light chain variable region (V_L)

CDR1 CDR2 CDR3 CDR1 CDR2 CDR3

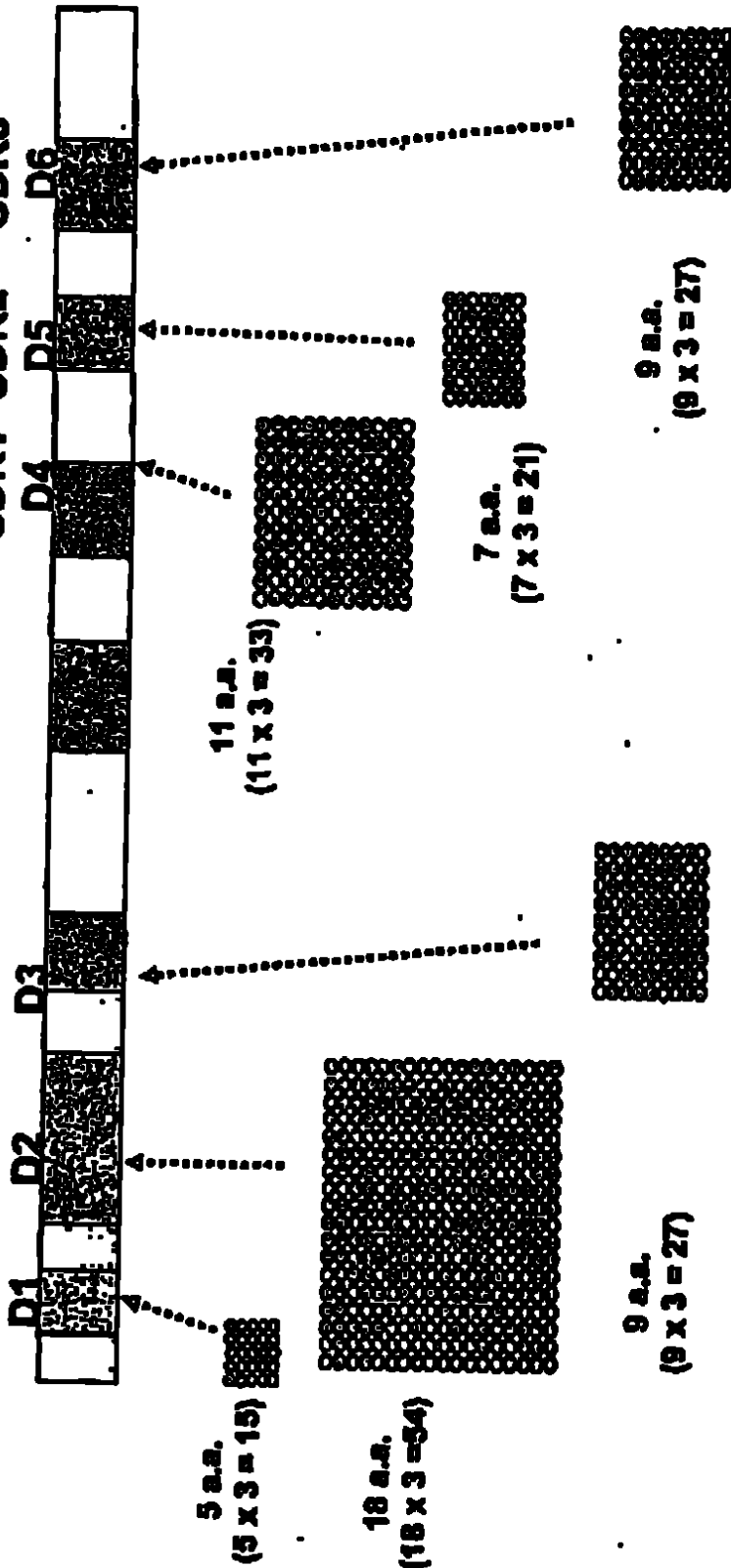


total diversity = V_H CDR1 x V_H CDR2 x V_H CDR3 x V_L CDR1 x V_L CDR2 x V_L CDR3
(1 amino acid) 5 x 15 x 12 x 5 x 15 x 12 = 8 x 10⁵

total diversity = V_H CDR1 x V_H CDR2 x V_H CDR3 x V_L CDR1 x V_L CDR2 x V_L CDR3
(20 amino acids) 100 x 300 x 240 x 100 x 300 x 240 = 5 x 10¹³

Exemplary Diversity of Defined Regions D1-D6 in a Synthetic Gene (a-TNF sFv) Using LTM

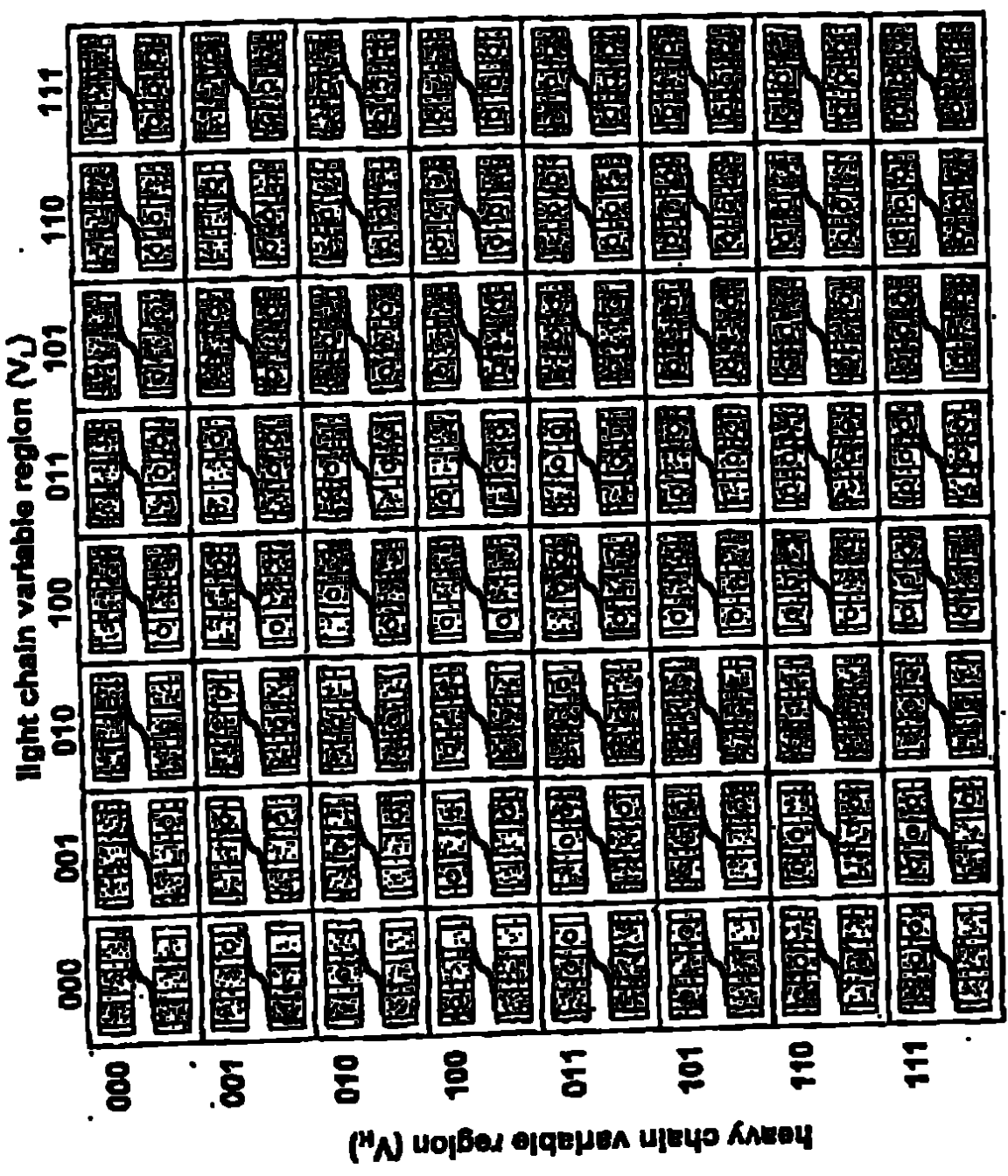
heavy chain variable region (V_H) linker light chain variable region (V_L)
CDR1 CDR2 CDR3 CDR1 CDR2 CDR3



total diversity = V_H CDR1 x V_H CDR2 x V_H CDR3 x V_L CDR1 x V_L CDR2 x V_L CDR3
(1 amino acid) 5 x 15 x 12 x 5 x 15 x 12 = 8 x 10⁵

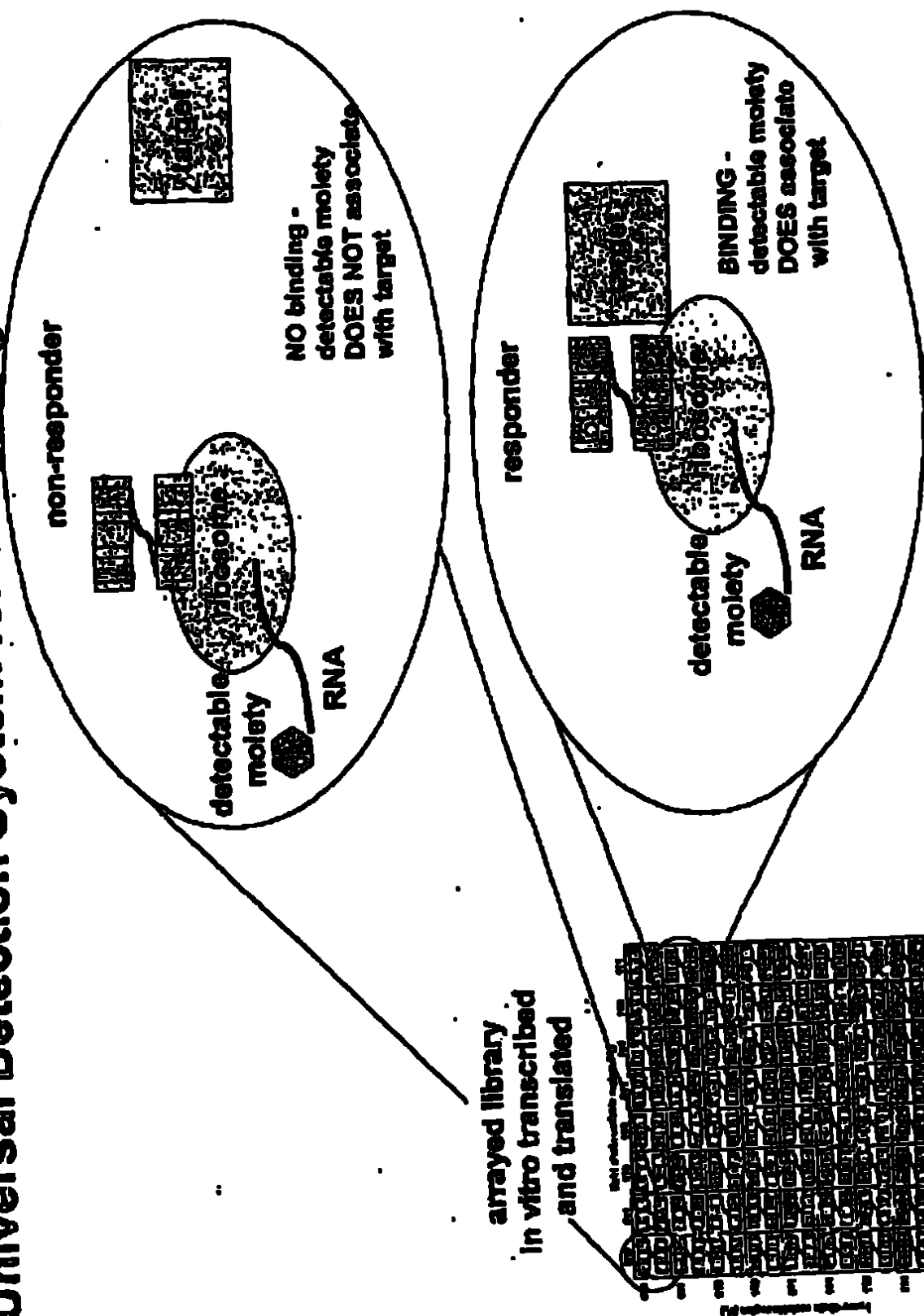
total diversity = V_H CDR1 x V_H CDR2 x V_H CDR3 x V_L CDR1 x V_L CDR2 x V_L CDR3
(3 amino acids) 15 x 54 x 27 x 33 x 21 x 27 = 4.1 x 10⁸
Asp, Ser, His

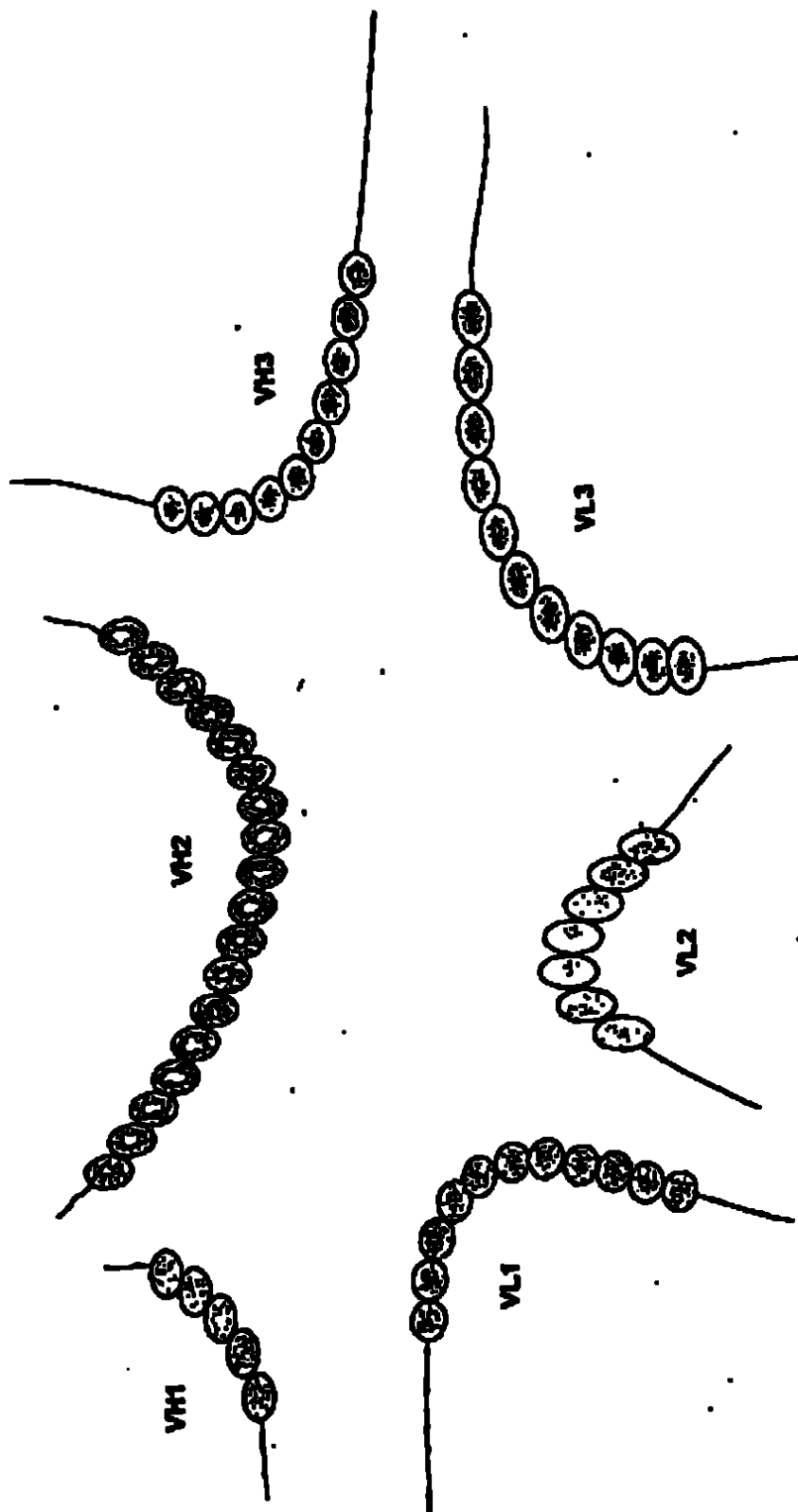
Arrayed Diversity Formats of a Synthetic Gene (sfv) Using LTM



52

Universal Detection System for Screening an Arrayed Library



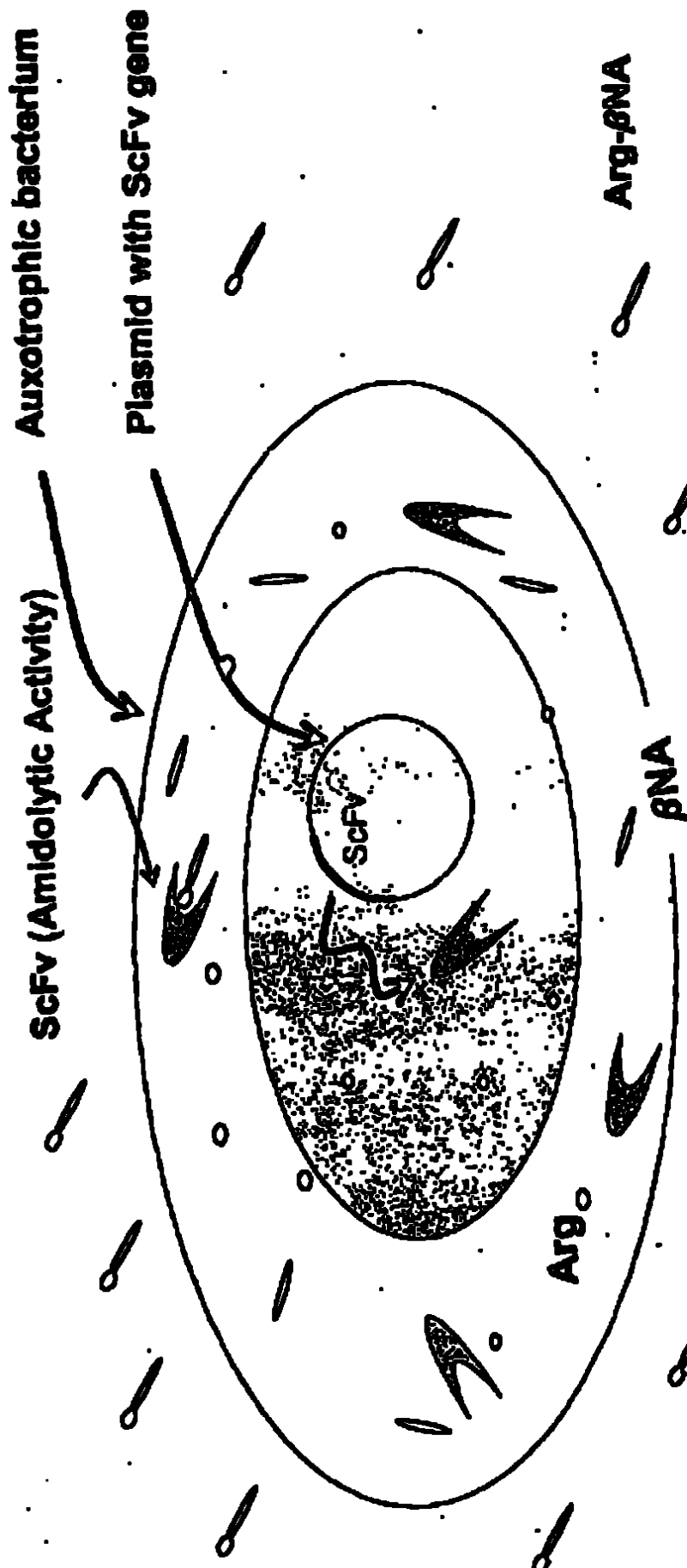


CA2					
VH1:	5 A.A.	5 oligos	5 gene analogs	VL1: 11 A.A.	11 oligos
VH2:	18 A.A.	18 oligos	18 gene analogs	VL2: 7 A.A.	7 oligos
VH3:	9 A.A.	9 oligos	9 gene analogs	VL3: 9 A.A.	9 oligos

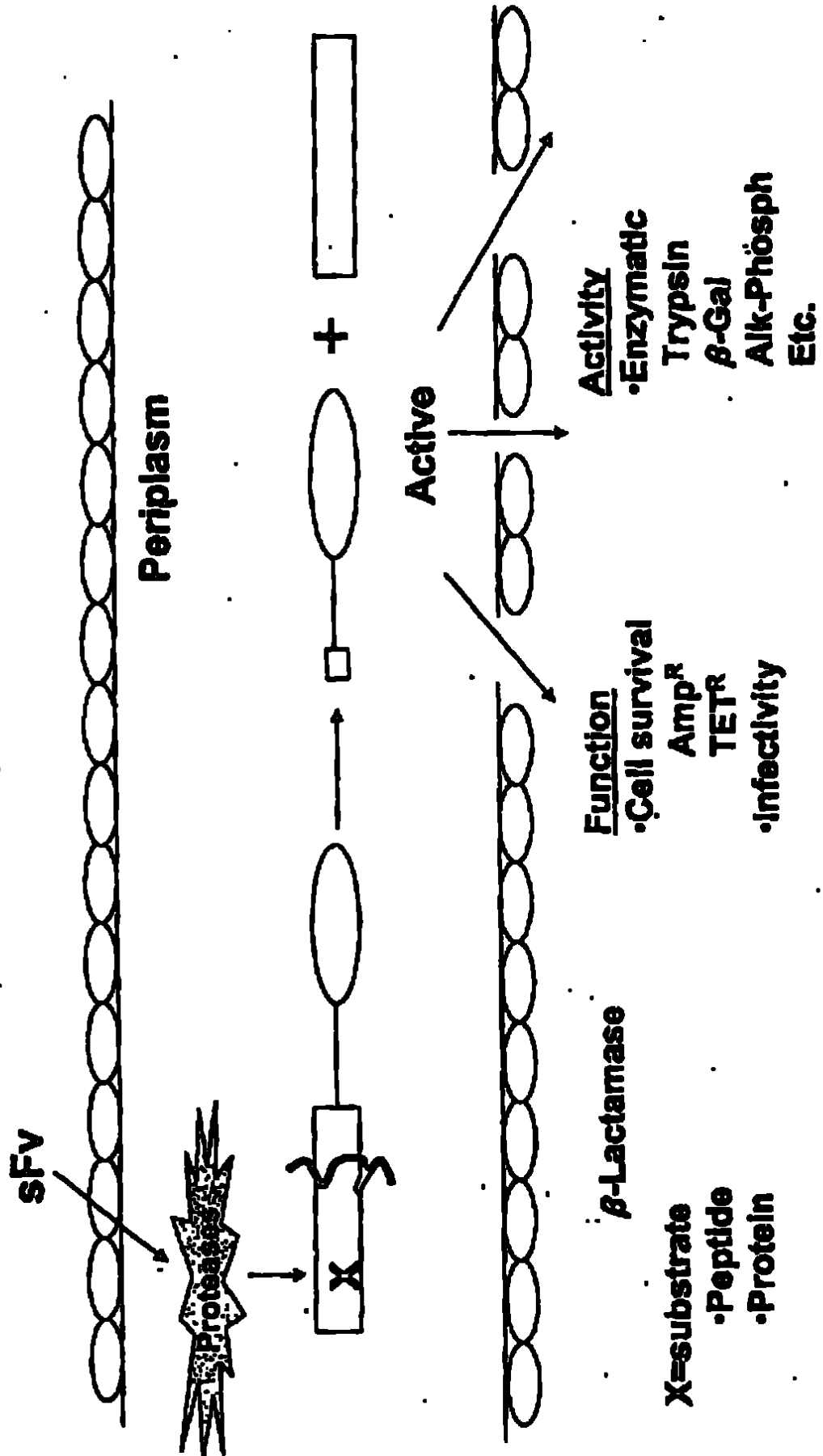
1521

7615
 7616
 7617
 7618
 7619
 7620
 7621
 7622
 7623
 7624
 7625
 7626
 7627
 7628
 7629
 7630
 7631
 7632
 7633
 7634
 7635
 7636
 7637
 7638
 7639
 7640
 7641
 7642
 7643
 7644
 7645
 7646
 7647
 7648
 7649
 7650
 7651
 7652
 7653
 7654
 7655
 7656
 7657
 7658
 7659
 7660
 7661
 7662
 7663
 7664
 7665
 7666
 7667
 7668
 7669
 7670
 7671
 7672
 7673
 7674
 7675
 7676
 7677
 7678
 7679
 7680
 7681
 7682
 7683
 7684
 7685
 7686
 7687
 7688
 7689
 7690
 7691
 7692
 7693
 7694
 7695
 7696
 7697
 7698
 7699
 7700
 7701
 7702
 7703
 7704
 7705
 7706
 7707
 7708
 7709
 7710
 7711
 7712
 7713
 7714
 7715
 7716
 7717
 7718
 7719
 7720
 7721
 7722
 7723
 7724
 7725
 7726
 7727
 7728
 7729
 7730
 7731
 7732
 7733
 7734
 7735
 7736
 7737
 7738
 7739
 7740
 7741
 7742
 7743
 7744
 7745
 7746
 7747
 7748
 7749
 7750
 7751
 7752
 7753
 7754
 7755
 7756
 7757
 7758
 7759
 7760
 7761
 7762
 7763
 7764
 7765
 7766
 7767
 7768
 7769
 7770
 7771
 7772
 7773
 7774
 7775
 7776
 7777
 7778
 7779
 7780
 7781
 7782
 7783
 7784
 7785
 7786
 7787
 7788
 7789
 7790
 7791
 7792
 7793
 7794
 7795
 7796
 7797
 7798
 7799
 7800
 7801
 7802
 7803
 7804
 7805
 7806
 7807
 7808
 7809
 7810
 7811
 7812
 7813
 7814
 7815
 7816
 7817
 7818
 7819
 7820
 7821
 7822
 7823
 7824
 7825
 7826
 7827
 7828
 7829
 7830
 7831
 7832
 7833
 7834
 7835
 7836
 7837
 7838
 7839
 7840
 7841
 7842
 7843
 7844
 7845
 7846
 7847
 7848
 7849
 7850
 7851
 7852
 7853
 7854
 7855
 7856
 7857
 7858
 7859
 7860
 7861
 7862
 7863
 7864
 7865
 7866
 7867
 7868
 7869
 7870
 7871
 7872
 7873
 7874
 7875
 7876
 7877
 7878
 7879
 7880
 7881
 7882
 7883
 7884
 7885
 7886
 7887
 7888
 7889
 7890
 7891
 7892
 7893
 7894
 7895
 7896
 7897
 7898
 7899
 7900
 7901
 7902
 7903
 7904
 7905
 7906
 7907
 7908
 7909
 7910
 7911
 7912
 7913
 7914
 7915
 7916
 7917
 7918
 7919
 7920
 7921
 7922
 7923
 7924
 7925
 7926
 7927
 7928
 7929
 7930
 7931
 7932
 7933
 7934
 7935
 7936
 7937
 7938
 7939
 7940
 7941
 7942
 7943
 7944
 7945
 7946
 7947
 7948
 7949
 7950
 7951
 7952
 7953
 7954
 7955
 7956
 7957
 7958
 7959
 7960
 7961
 7962
 7963
 7964
 7965
 7966
 7967
 7968
 7969
 7970
 7971
 7972
 7973
 7974
 7975
 7976
 7977
 7978
 7979
 7980
 7981
 7982
 7983
 7984
 7985
 7986
 7987
 7988
 7989
 7990
 7991
 7992
 7993
 7994
 7995
 7996
 7997
 7998
 7999
 8000

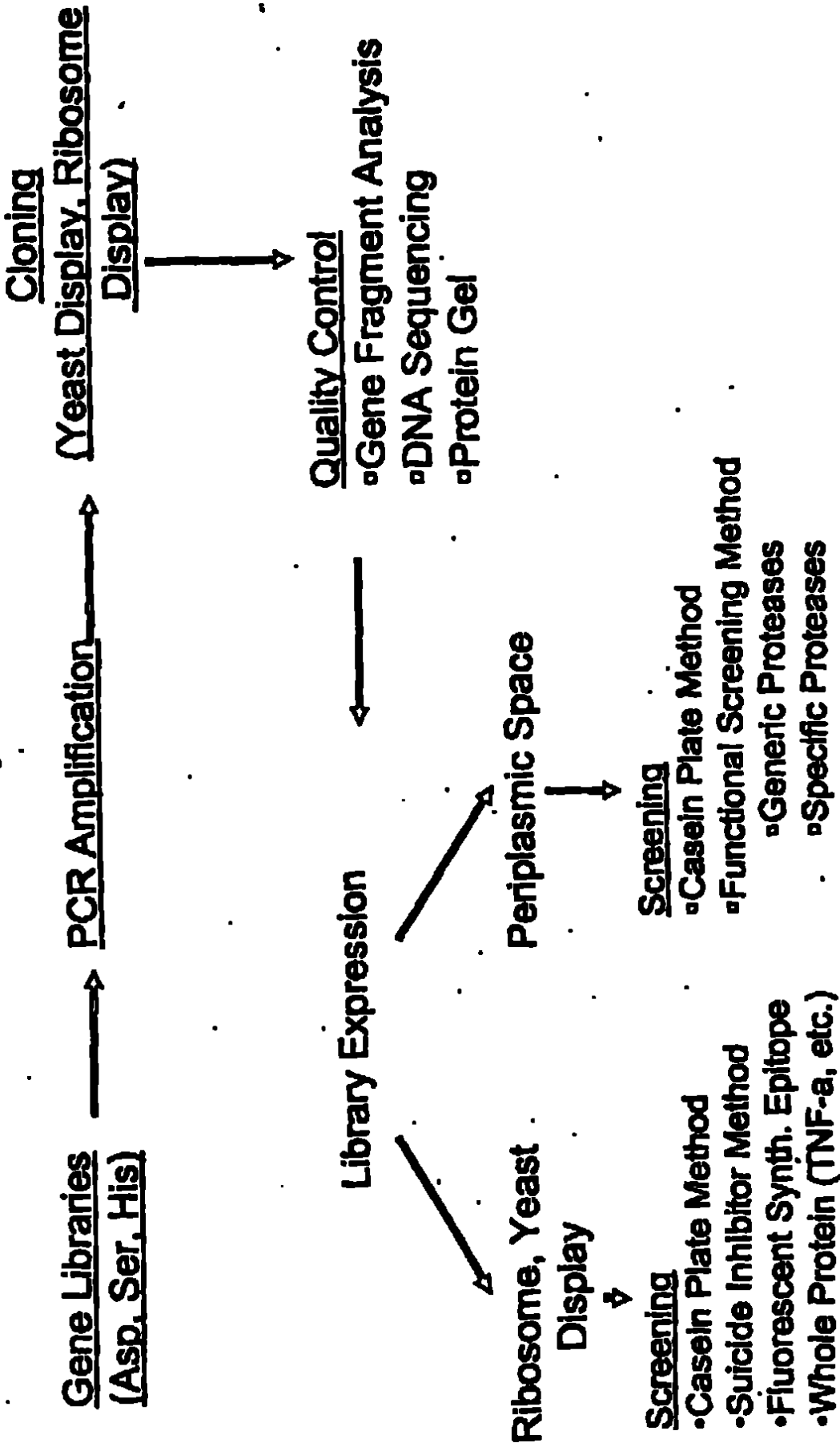
PROTEASE SELECTION ASSAY

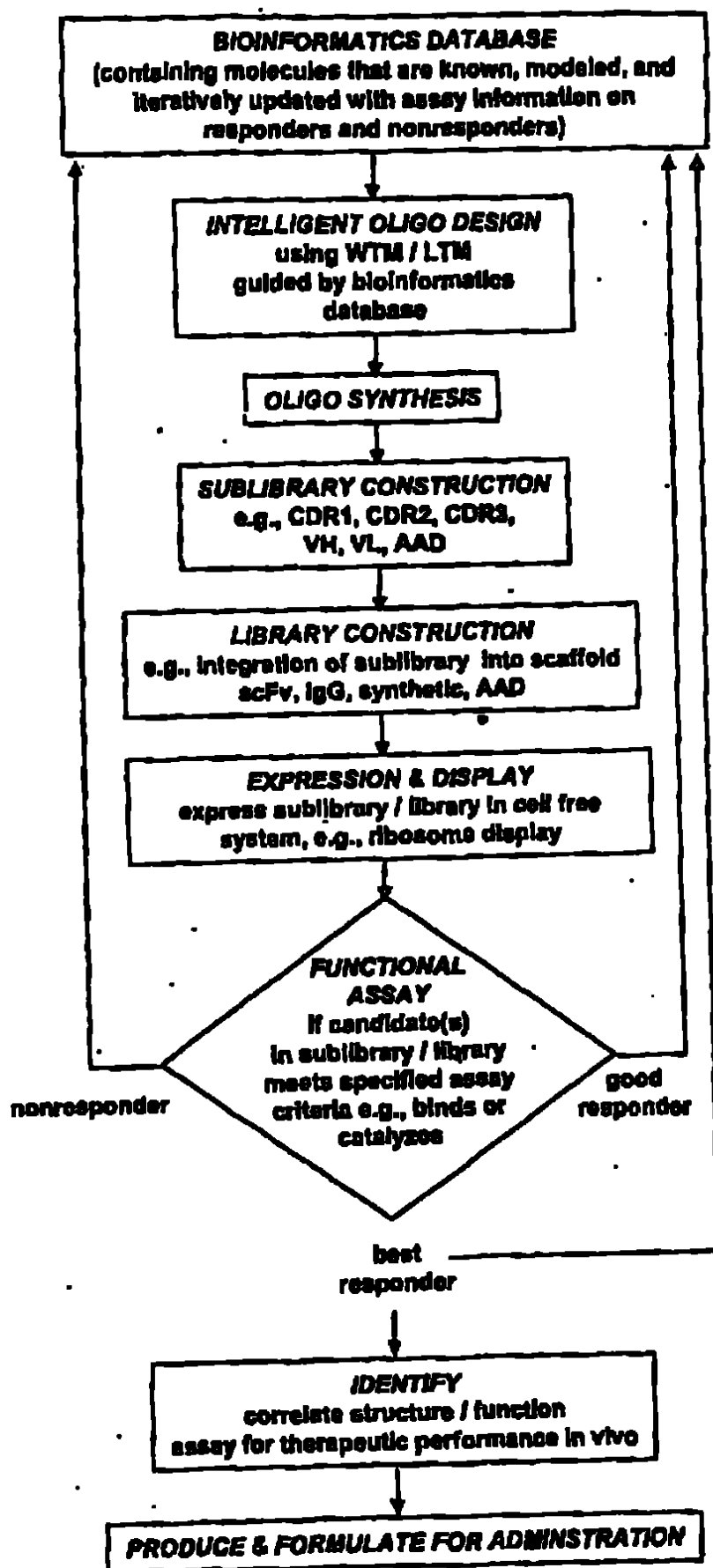


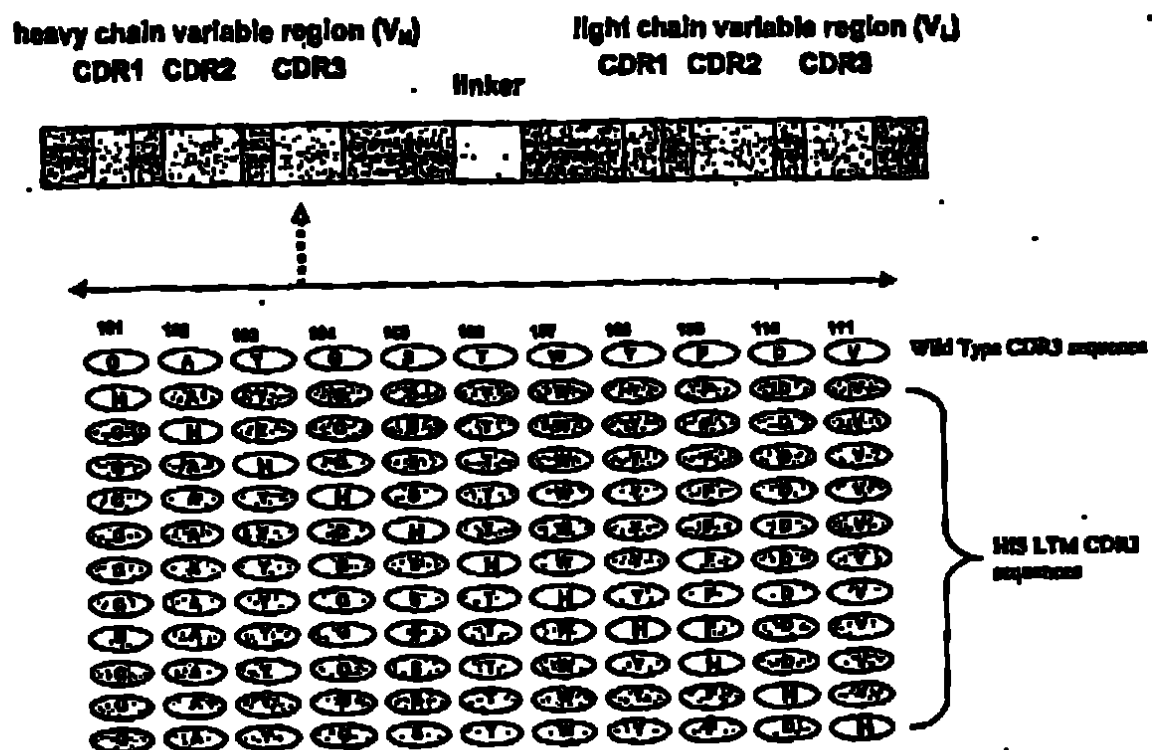
Bacterial Periplasmic Space

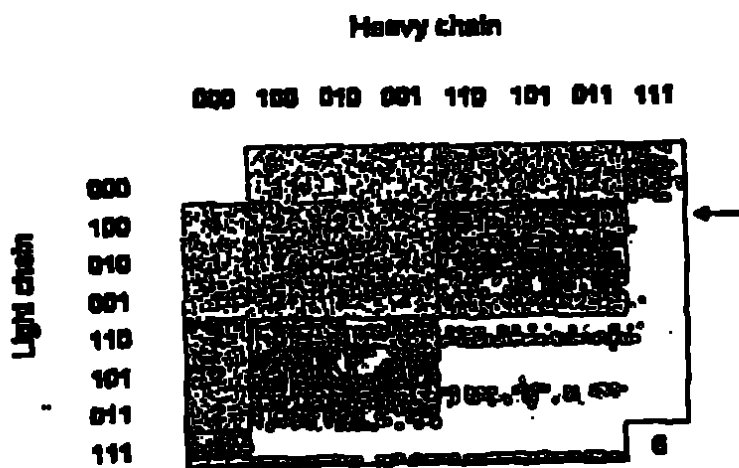


**Catalytic Abs – Proteases
Summary of LTM Process**









SEQUENCE LISTING

<110> R. Crea & Co., et al.

<120> LOOK-THROUGH MUTAGENESIS

<130> BRI-0017C

<150> 60/483,282

<151> 2003-06-27

<160> 13

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 121

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 1

Gln Val Gln Leu Val Gln Ser Gly Ala Glu Val Val Lys Pro Gly Ser
 1 5 10 15
 Ser Val Lys Val Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Asp Tyr
 20 25 30
 Asn Val Asp Trp Val Lys Gln Ala Pro Gly Gln Gly Leu Gln Trp Ile
 35 40 45
 Gly Asn Ile Asn Pro Asn Asn Gly Gly Thr Ile Tyr Asn Gln Lys Phe
 50 55 60
 Lys Gly Lys Gly Thr Leu Thr Val Asp Lys Ser Thr Ser Thr Ala Tyr
 65 70 75 80
 Met Gln Leu Ser Ser Leu Thr Ser Gln Asp Thr Ala Val Tyr Tyr Cys
 85 90 95
 Ala Arg Ser Ala Phe Tyr Asn Asn Tyr Glu Tyr Phe Asp Val Trp Gly
 100 105 110
 Gln Gly Thr Thr Val Thr Val Ser Ser
 115 120

<210> 2

<211> 119

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 2

Glu Val Lys Leu Glu Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Gly
 1 5 10 15
 Ser Met Lys Leu Ser Cys Val Ala Ser Gly Phe Ile Phe Ser Asn His
 20 25 30
 Trp Met Asn Trp Val Arg Gln Ser Pro Glu Lys Gly Leu Glu Trp Val
 35 40 45
 Ala Glu Ile Arg Ser Lys Ser Ile Asn Ser Ala Thr His Tyr Ala Glu
 50 55 60

WO 2003/003345

PCT/US2004/020306

Ser Val Lys Gly Arg Phe Thr Ile Ser Arg Asp Asp Ser Lys Ser Ala
 65 70 75 80

Val Tyr Leu Gln Met Thr Asp Leu Arg Thr Glu Asp Thr Gly Val Tyr
 85 90 95

Tyr Cys Ser Arg Asn Tyr Tyr Gly Ser Thr Tyr Asp Tyr Trp Gly Gln
 100 105 110

Gly Thr Thr Leu Thr Val Ser
 115

<210> 3
 <211> 114
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Synthetic construct

<400> 3
 Asp Ile Met Met Thr Gln Ser Pro Ser Thr Leu Ser Ala Ser Val Gly
 1 5 10 15
 Asp Arg Val Thr Ile Thr Cys Lys Ser Ser Gln Ser Leu Leu Tyr Ser
 20 25 30
 Asn Asn Gln Lys Asn Tyr Leu Ala Trp Tyr Gln Gln Lys Pro Gly Gln
 35 40 45
 Ala Pro Lys Leu Leu Ile Ser Trp Ala Ser Thr Arg Glu Ser Gly Val
 50 55 60
 Pro Ser Arg Phe Ile Gly Ser Gly Ser Gly Thr Glu Phe Thr Leu Thr
 65 70 75 80
 Ile Ser Ser Leu Gln Pro Asp Asp Val Ala Thr Tyr Tyr Cys Gln Gln
 85 90 95
 Tyr Tyr Asp Tyr Pro Trp Thr Phe Gly Gln Gly Thr Lys Val Glu Ile
 100 105 110

Lys Arg

<210> 4
 <211> 107
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Synthetic construct

<400> 4
 Asp Ile Leu Leu Thr Gln Ser Pro Ala Ile Leu Ser Val Ser Pro Gly
 1 5 10 15
 Glu Arg Val Ser Phe Ser Cys Arg Ala Ser Gln Phe Val Gly Ser Ser
 20 25 30
 Ile His Trp Tyr Gln Gln Arg Thr Asn Gly Ser Pro Arg Leu Leu Ile
 35 40 45
 Lys Tyr Ala Ser Glu Ser Met Ser Gly Ile Pro Ser Arg Phe Ser Gly
 50 55 60
 Ser Gly Ser Gly Thr Asp Phe Thr Leu Ser Ile Asn Thr Val Glu Ser
 65 70 75 80
 Glu Asp Ile Ala Asp Tyr Tyr Cys Gln Gln Ser His Ser Trp Pro Phe
 85 90 95
 Thr Phe Gly Ser Gly Thr Asn Leu Glu Val Lys
 100 105

WO 2005/003345

PCT/US2004/020306

<210> 5
 <211> 16
 <212> PRT
 <213> Artificial Sequence

<220>
 <221> Synthetic construct

<400> 5
 Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Ser
 1 5 10 15

<210> 6
 <211> 6
 <212> PRT
 <213> Artificial Sequence

<220>
 <221> Synthetic construct

<400> 6
 His His His His His His
 1 5

<210> 7
 <211> 10
 <212> PRT
 <213> Artificial Sequence

<220>
 <221> Synthetic construct

<400> 7
 Glu Gln Lys Leu Ile Ser Gln Glu Asp Leu
 1 5 10

<210> 8
 <211> 236
 <212> PRT
 <213> Artificial Sequence

<220>
 <221> Synthetic construct

<400> 8
 Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala
 1 5 10 15
 Ala Gln Pro Ala Met Ala Asp Ile Gln Met Thr Gln Ser Pro Ala Ser
 20 25 30
 Leu Ser Ala Ser Val Gly Glu Thr Val Thr Ile Thr Cys Arg Ala Ser
 35 40 45
 Gly Asn Ile His Asn Tyr Leu Ala Trp Tyr Gln Gln Lys Gln Gly Lys
 50 55 60
 Ser Pro Gln Leu Leu Val Tyr Asn Ala Lys Thr Leu Ala Asp Gly Val
 65 70 75 80
 Pro Ser Arg Phe Ser Gly Ser Gly Ser Gly Thr Gln Tyr Ser Leu Lys
 85 90 95

Ile Asn Ser Leu Gln Pro Glu Asp Phe Gly Ser Tyr Tyr Cys Gln His
 100 105 110
 Phe Trp Ser Thr Pro Trp Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile
 115 120 125
 Lys Arg Ala Asp Ala Ala Pro Thr Val Ser Ile Phe Pro Pro Ser Ser
 130 135 140
 Glu Gln Leu Thr Ser Gly Gly Ala Ser Val Val Cys Phe Leu Asn Asn
 145 150 155 160
 Phe Tyr Pro Lys Asp Ile Asn Val Lys Trp Lys Ile Asp Gly Ser Glu
 165 170 175
 Arg Gln Asn Gly Val Leu Asn Ser Trp Thr Asp Gln Asp Ser Lys Asp
 180 185 190
 Ser Thr Tyr Ser Met Ser Ser Thr Leu Thr Leu Thr Lys Asp Glu Tyr
 195 200 205
 Glu Arg His Asn Ser Tyr Thr Cys Glu Ala Thr His Lys Thr Ser Thr
 210 215 220
 Ser Pro Ile Val Lys Ser Phe Asn Arg Asn Glu Cys
 225 230 235

<210> 9

<211> 254

<212> FRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 9

Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala
 1 5 10 15
 Ala Gln Pro Ala Met Ala Gln Val Gln Leu Gln Gln Ser Gly Ala Gln
 20 25 30
 Leu Val Lys Pro Gly Ala Ser Val Lys Leu Ser Cys Thr Ala Ser Gly
 35 40 45
 Phe Asn Ile Lys Asp Thr Phe Met His Trp Val Lys Gln Arg Pro Glu
 50 55 60
 Gln Gly Leu Gln Trp Ile Gly Arg Ile Asp Pro Ala Asn Gly Asn Thr
 65 70 75 80
 Glu Tyr Asp Pro Lys Phe Gln Gly Lys Ala Thr Ile Thr Ala Asp Thr
 85 90 95
 Ser Ser Asn Thr Val Asn Leu Gln Leu Ser Ser Leu Thr Ser Glu Asp
 100 105 110
 Thr Ala Val Tyr Tyr Cys Ala Ser Gly Gly Glu Leu Gly Phe Pro Tyr
 115 120 125
 Trp Gly Gln Gly Thr Leu Val Thr Val Ser Ala Ala Lys Thr Thr Pro
 130 135 140
 Pro Ser Val Tyr Pro Leu Ala Pro Gly Ser Ala Ala Gln Thr Asn Ser
 145 150 155 160
 Met Val Thr Leu Gly Cys Leu Val Lys Gly Tyr Phe Pro Glu Pro Val
 165 170 175
 Thr Val Thr Trp Asn Ser Gly Ser Leu Ser Ser Gly Val His Thr Phe
 180 185 190
 Pro Ala Val Leu Gln Phe Asp Leu Tyr Thr Leu Ser Ser Ser Val Thr
 195 200 205
 Val Pro Ser Ser Thr Trp Pro Ser Gln Thr Val Thr Cys Asn Val Ala
 210 215 220
 His Pro Ala Ser Ser Thr Lys Val Asp Lys Lys Ile Val Pro Arg Asp
 225 230 235 240
 Cys Thr Ser Gly Gly Gly Gly Ser His His His His His
 245 250

<210> 10
<211> 236
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 10
Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Ala
1 5 10 15
Ala Gln Pro Ala Met Ala Asp Ile Gln Met Thr Gln Ser Pro Ala Ser
20 25 30
Leu Ser Ala Ser Val Gly Glu Thr Val Thr Ile Thr Cys Arg Ala Ser
35 40 45
Gly Asn Ile His Asn Tyr Leu Ala Trp Tyr Gln Gln Lys Gln Gly Lys
50 55 60
Ser Pro Gln Leu Leu Val Tyr Asn Ala Lys Thr Leu Ala Asp Gly Val
65 70 75 80
Pro Ser Arg Phe Ser Gly Ser Gly Ser Gly Thr Gln Tyr Ser Leu Lys
85 90 95
Ile Asn Ser Leu Gln Pro Glu Asp Phe Gly Ser Tyr Tyr Cys Gln His
100 105 110
Phe Trp Ser Thr Pro Trp Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile
115 120 125
Lys Arg Ala Asp Ala Ala Pro Thr Val Ser Ile Phe Pro Pro Ser Ser
130 135 140
Glu Gln Leu Thr Ser Gly Gly Ala Ser Val Val Cys Phe Leu Asn Asn
145 150 155 160
Phe Tyr Pro Lys Asp Ile Asn Val Lys Trp Lys Ile Asp Gly Ser Glu
165 170 175
Arg Gln Asn Gly Val Leu Asn Ser Trp Thr Asp Gln Asp Ser Lys Asp
180 185 190
Ser Thr Tyr Ser Met Ser Ser Thr Leu Thr Leu Thr Lys Asp Glu Tyr
195 200 205
Glu Arg His Asn Ser Tyr Thr Cys Glu Ala Thr His Lys Thr Ser Thr
210 215 220
Ser Pro Ile Val Lys Ser Phe Asn Arg Asn Glu Cys
225 230 235

<210> 11
<211> 254
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 11
Met Lys Tyr Leu Leu Pro Thr Ala Ala Val Gly Leu Leu Leu Leu Ala
1 5 10 15
Ala Gln Pro Ala Met Ala Glu Val Gln Leu Gln Gln Ser Gly Ala Glu
20 25 30
Leu Val Lys Pro Gly Ala Ser Val Lys Leu Ser Cys Thr Ala Ser Gly
35 40 45
Phe Asn Ile Lys Asp Thr Phe Met His Trp Val Lys Gln Arg Pro Glu
50 55 60

Gln Gly Leu Gln Trp Ile Gly Arg Ile Asp Pro Ala Asn Gly Asn Thr
69 70 75 80
Glu Tyr Asp Pro Lys Phe Gln Gly Lys Ala Thr Ile Thr Ala Asp Thr
85 90 95
Ser Ser Asn Thr Val Asn Leu Gln Leu Ser Ser Leu Thr Ser Glu Asp
100 105 110
Thr Ala Val Tyr Tyr Cys Ala Ser Gly Gly Gln Leu Gly Phe Pro Tyr
115 120 125
Trp Gly Gln Gly Thr Leu Val Thr Val Ser Ala Ala Lys Thr Thr Pro
130 135 140
Pro Ser Val Tyr Pro Leu Ala Pro Gly Ser Ala Ala Gln Thr Asn Ser
145 150 155 160
Met Val Thr Leu Gly Cys Leu Val Lys Gly Tyr Phe Pro Glu Pro Val
165 170 175
Thr Val Thr Trp Asn Ser Gly Ser Leu Ser Ser Gly Val His Thr Phe
180 185 190
Pro Ala Val Leu Gln Ser Asp Leu Tyr Thr Leu Ser Ser Ser Val Thr
195 200 205
Val Pro Ser Ser Thr Trp Pro Ser Glu Thr Val Thr Cys Asn Val Ala
210 215 220
His Pro Ala Ser Ser Thr Lys Val Asp Lys Lys Ile Val Pro Arg Asp
225 230 235 240
Cys Thr Ser Gly Gly Gly Gly Ser His His His His His His
245 250

<210> 12

<211> 236

<212> PKT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 12

Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala
1 5 10 15
Ala Gln Pro Ala Met Ala Asp Ile Gln Met Thr Gln Ser Pro Ala Ser
20 25 30
Leu Ser Ala Ser Val Gly Glu Thr Val Thr Ile Thr Cys Arg Ala Ser
35 40 45
Gly Asn Ile His Asn Tyr Leu Ala Trp Tyr Gln Gln Lys Gln Gly Lys
50 55 60
Ser Pro Gln Leu Leu Val Tyr Asn Ala Lys Thr Leu Ala Asp Gly Val
65 70 75 80
Pro Ser Arg Phe Ser Gly Ser Gly Ser Gly Thr Gln Tyr Ser Leu Lys
85 90 95
Ile Asn Ser Leu Gln Pro Gln Asp Phe Gly Ser Tyr Tyr Cys Gln His
100 105 110
Phe Trp Ser Thr Pro Trp Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile
115 120 125
Lys Arg Ala Asp Ala Ala Pro Thr Val Ser Ile Phe Pro Pro Ser Ser
130 135 140
Gln Gln Leu Thr Ser Gly Gly Ala Ser Val Val Cys Phe Leu Asn Asn
145 150 155 160
Phe Tyr Pro Lys Asp Ile Asn Val Lys Trp Lys Ile Asp Gly Ser Glu
165 170 175
Arg Gln Asn Gly Val Leu Asn Ser Trp Thr Asp Gln Asp Ser Lys Asp
180 185 190
Ser Thr Tyr Ser Met Ser Ser Thr Leu Thr Leu Thr Lys Asp Glu Tyr
195 200 205

Glu Arg His Asn Ser Tyr Thr Cys Glu Ala Thr His Lys Thr Ser Thr
 210 215 220
 Ser Pro Ile Val Lys Ser Phe Asn Arg Asn Glu Cys
 225 230 235

<210> 13
 <211> 254
 <212> PRT
 <213> Artificial Sequence

<400> 13
 Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala
 1 5 10 15
 Ala Gln Pro Ala Met Ala Glu Val Gln Leu Gln Gln Ser Gly Ala Glu
 20 25 30
 Leu Val Lys Pro Gly Ala Ser Val Lys Leu Ser Cys Thr Ala Ser Gly
 35 40 45
 Phe Asn Ile Lys Asp Thr Phe Met His Trp Val Lys Gln Arg Pro Glu
 50 55 60
 Gln Gly Leu Glu Trp Ile Gly Arg Ile Asp Pro Ala Asn Gly Asn Thr
 65 70 75 80
 Glu Tyr Asp Pro Lys Phe Gln Gly Lys Ala Thr Ile Thr Ala Asp Thr
 85 90 95
 Ser Ser Asn Thr Val Asn Leu Gln Leu Ser Ser Leu Thr Ser Glu Asp
 100 105 110
 Thr Ala Val Tyr Tyr Cys Ala Ser Gly Gly Glu Leu Gly Phe Pro Tyr
 115 120 125
 Trp Gly Gln Gly Thr Leu Val Thr Val Ser Ala Ala Lys Thr Thr Pro
 130 135 140
 Pro Ser Val Tyr Pro Leu Ala Pro Gly Ser Ala Ala Gln Thr Asn Ser
 145 150 155 160
 Met Val Thr Leu Gly Cys Leu Val Lys Gly Tyr Phe Pro Glu Pro Val
 165 170 175
 Thr Val Thr Trp Asn Ser Gly Ser Leu Ser Ser Gly Val His Thr Phe
 180 185 190
 Pro Ala Val Leu Gln Ser Asp Leu Tyr Thr Leu Ser Ser Ser Val Thr
 195 200 205
 Val Pro Ser Ser Thr Trp Pro Ser Glu Thr Val Thr Cys Asn Val Ala
 210 215 220
 His Pro Ala Ser Ser Thr Lys Val Asp Lys Lys Ile Val Pro Arg Asp
 225 230 235 240
 Cys Thr Ser Gly Gly Gly Gly Ser His His His His His
 245 250

MUTAGÉNESIS POR REVISIÓN

Solicitudes relacionadas e información

Esta solicitud reivindica prioridad a la solicitud provisional de Estados Unidos N° 60/483282, presentada el 27 de junio de 2003, cuyo contenido entero se incorpora en esta memoria descriptiva como referencia. El contenido entero de todas las otras patentes, solicitudes de patentes, y referencias citadas a lo largo de la siguiente memoria descriptiva se incorporan también como referencia en su totalidad.

10

Antecedentes de la invención

La mutagénesis es una herramienta útil en el estudio y función de estructura de proteínas. La mutagénesis se puede realizar en la secuencia de nucleótidos de un gen clonado que codifica una proteína de interés y el gen modificado se puede expresar para producir mutantes de la proteína. Comparando las propiedades de la proteína de tipo salvaje y los mutantes generados, a menudo es posible identificar aminoácidos individuales o dominios de aminoácidos que son esenciales para la integridad estructural y/o función bioquímica de la proteína, tal como su unión y/o actividad catalítica. Sin embargo, el número de mutantes que se puede generar a partir de la proteína individual hace difícil seleccionar mutantes que serán informativos o tienen la propiedad deseada, incluso si los mutantes seleccionados que abarcan las mutaciones están únicamente en regiones supuestas importantes de una proteína (por ejemplo, regiones que conforman un sitio activo de una proteína). Por ejemplo, la sustitución, supresión, o inserción de un aminoácido particular

25

puede tener un efecto local o global sobre la proteína.

Los procedimientos anteriores para la mutagenización de polipéptidos han sido o bien demasiado restrictivos, demasiado incluso, o limitados para inactivar la función de las proteínas en lugar de ganar o mejorar la función. Por ejemplo, un planteamiento altamente restrictivo es de mutagénesis selectiva o dirigida al sitio que se usa para identificar la presencia de un sitio funcional particular o entender las secuencias de realización de una alteración muy especificada dentro del sitio funcional. Una aplicación común de mutagénesis dirigida al sitio en el estudio de las fosfoproteínas en las que un resto aminoácido, que ordinariamente se fosforilaría y permitiría que el polipéptido llevara a cabo su función, se altera para confirmar la unión entre la actividad de fosforilación y funcional. Este planteamiento es muy específica para el polipéptido y resto que se está estudiando.

Recíprocamente, un planteamiento altamente inclusivo es la saturación o mutagénesis al azar que se diseña para producir un gran número de mutaciones que abarcan todas las alteraciones posibles dentro de una región definida de un gen o proteína. Esto se basa en el principio de que, generando esencialmente todas las posibles variantes de un dominio proteico relevante, la disposición apropiada de aminoácidos a producir como uno de los mutantes generados al azar. Sin embargo, en la práctica, el vasto número de combinaciones al azar de mutaciones generadas pueden prevenir la capacidad de seleccionar de manera significativa un candidato deseado debido a la presencia del llamado "ruido" de de esta forma muchos candidatos indeseados.

Otro planteamiento, denominado mutagénesis de "espera" (véase, las patentes números 5.830.650; 5.798.208) se ha usado para mutagenizar una

región definida de un polipéptido sintetizando una mezcla de oligonucleótidos degenerados que, estadísticamente, contienen un conjunto deseado de mutaciones. Sin embargo, debido a la síntesis de polinucleótidos degenerados, la mutagénesis de paseo por produce un número de alteraciones no deseadas
5 además del conjunto deseado de mutaciones. Por ejemplo, para introducir secuencialmente una mutación a través de una región definida de cinco posiciones de aminoácidos solamente, se puede preparar un conjunto de alrededor de 100 polinucleótidos (y seleccionar) (véase, por ejemplo, la figura 6). De acuerdo a lo anterior, para preparar y seleccionar, por ejemplo, dos o
10 tres regiones que llegan a ser complejo creciente, es decir, que requiere la preparación y selección de 200 a hasta 300 polinucleótidos, respectivamente, para la presencia de 10 a 15 mutaciones solamente.

Todavía en otro planteamiento que se ha usado para mutagenizar proteínas es la mutagénesis de exploración de alanina, en la que un resto
15 alanina se "explora" a través de una porción de una proteína para identificar posiciones en las que la función de proteínas está interrumpida. Sin embargo, este planteamiento solamente mira la pérdida de la función de proteínas por medio de la sustitución de un resto alanina neutro en una posición dada, en lugar de ganar o mejorar la función. De este modo, no es un planteamiento útil
20 para generar proteínas que han mejorado la estructura y función.

De acuerdo con lo anterior, permanece una necesidad de una forma sistemática de mutagenizar una proteína para una función nueva o mejorada.

Sumario de la Invención

25 La invención se refiere a un procedimiento de mutagénesis para la

generación de proteínas novedosas o mejoradas (o polipéptidos) y a librerías de análogos de polipéptidos y polipéptidos específicos generados por los procedimientos. El polipéptido dirigido para mutagénesis puede ser un polipéptido natural, sintético o modificado por ingeniería genética, que incluye
5 fragmentos, análogos y las formas mutantes de los mismos.

En una realización, el procedimiento comprende la introducción de un aminoácido predeterminado esencialmente en cada posición dentro de una región definida (o varias regiones diferentes) de la secuencia de aminoácidos de un polipéptido. Una librería de polipéptidos se genera de manera que
10 contenga análogos de polipéptidos que no tienen individualmente más de un aminoácido predeterminado, pero que colectivamente tienen el aminoácido predeterminado en cada posición dentro de la (s) región (es) definida (s). El procedimiento se puede denominar "revisión" de mutagénesis debido a que, en efecto, un aminoácido individual, predeterminado (y solamente el aminoácido
15 predeterminado) se sustituye posición por posición a lo largo de una (s) región (es) más definida (s) de un polipéptido. De este modo la invención permite "revisar" las consecuencias estructurales y funcionales de la sustitución de manera separada de un aminoácido predeterminado en cada posición de aminoácido dentro de una región definida del polipéptido por lo tanto
20 segregando una química de proteínas específica a la región definida sin ninguna interferencia o "ruido" de la generación de análogos de polipéptidos no deseados (es decir análogos que contienen sustituciones de aminoácidos distintas de las que siguen al esquema de "revisión") véase por ejemplo la figura 6.

25 De acuerdo con lo anterior, la presente invención permite la evaluación

--

sistemática altamente eficaz y precisa del papel de un cambio de aminoácidos específico en una o más regiones definidas de un polipéptido. Esto llega a ser particularmente importante cuando la evaluación (mediante mutación) de dos o más regiones definidas, de manera que el número de análogos de polipéptidos requeridos se incrementa en gran medida y, de este modo, también se incrementa la presencia de análogos no deseados. La presente invención obvia este problema mediante la eliminación completa de análogos no deseados y, de este modo, el potencial de que cualquier cambio en la estructura o función de proteína observada son el resultado de cualquier cosa sino la sustitución del aminoácido predeterminado. De este modo, el efecto de la segregación de una química de proteínas específica a incluso regiones múltiples con una proteína se puede estudiar con alta precisión y eficacia. De manera importante, esto incluye el estudio de cómo la mutagénesis puede efectuar la interacción de tales regiones, por lo tanto mejorando la estructura y función global de la proteína.

En una realización particular de la invención, la librería de los análogos de polipéptidos se genera y selecciona sintetizando primero los polinucleótidos individuales que codifican una región o regiones de un polipéptido en las que, de manera colectiva, los polinucleótidos representan todos los polinucleótidos posibles variantes de acuerdo con el criterio de revisión descrito en esta memoria descriptiva. Los polinucleótidos variantes se expresan, por ejemplo, usando transcripción y traducción *in vitro* y/o usando una tecnología de representación, tal como representación de ribosomas, representación d fagos, representación bacteriana, representación de levaduras, representación en serie o cualquier otro sistema de representación adecuada conocido en la

técnica.

Los polipéptidos expresados se rastrean y seleccionan después usando ensayos funcionales, tales como ensayos de unión o ensayos enzimáticos/catalíticos. En una realización, los polipéptidos se expresan en asociación con el polinucleótido que codifica el polipéptido, por lo tanto permitiendo la identificación de la secuencia de polinucleótidos que codifica el polipéptido. Todavía en otra realización, los polipéptidos se sintetizan directamente usando química de proteínas.

De este modo, la presente invención proporciona un procedimiento de mutagénesis que se puede usar para generar librerías de análogos de polipéptidos que son de un tamaño práctico para la selección, en parte, debido a que las librerías están desprovistas de cual polipéptido análogo no deseado o también llamado ruido. El procedimiento se puede usar para estudiar el papel de aminoácidos específicos en la estructura y función de polipéptidos y para desarrollar polipéptidos nuevos o mejorados tales como anticuerpos, fragmentos de unión o análogos de los mismos, anticuerpos de cadenas individuales, anticuerpos catalíticos, enzimas y ligandos. Además, el procedimiento se puede realizar con el beneficio de una información *a priori*, por ejemplo, mediante modelado por ordenador, que se puede usar para seleccionar un subconjunto inicial de análogos de polipéptidos a producir y estudiar usando mutagénesis por "revisión".

Otras ventajas y aspectos de la presente invención serán fácilmente evidentes a partir de la descripción y ejemplos siguientes.

25

Breve descripción de las figuras

La figura 1 ilustra las regiones definidas de manera ejemplar (o regiones D) que se pueden examinar usando mutagénesis por revisión (LTM) y ensayos funcionales para identificar análogos de polipéptidos deseados a partir de tales regiones D.

5 La figura 2 ilustra el uso de LTM entres regiones definidas en una región variable de anticuerpo (es decir, CDR1, CDR2, y CDR3 de una región variable de cadena pesada de anticuerpo). La región variable de la cadena ligera se puede explorar de manera similar, o bien sola o en combinación con la región variable de la cadena pesada. Para conveniencia en los ensayos de
10 selección posterior, la región variable de la cadena pesada se puede explorar en el contexto de un anticuerpo de la cadena individual (sFv) como se muestra.

La figura 3 ilustra el uso de LTM en una región definida (es decir, posiciones 31 - 25 de CDR1) de una región variable de la cadena pesada.

La figura 4 ilustra el uso de LTM en una región definida (es decir,
15 posiciones 55 - 68 de CDR2) de una región variable de la cadena pesada.

La figura 5 ilustra el uso de LTM en una región definida (es decir, posiciones 101 - 111 de CDR3) de una región variable de la cadena pesada.

La figura 6 ilustra las ventajas de LTM comparada con la mutagénesis por espera. La LTM de una región representativa definida, es decir CDR1 de
20 una región variable de cadena pesada de anticuerpo, da como resultado la alteración secuencial de cada posición de aminoácidos mediante la región definida, sin la introducción de ningún residuo de aminoácidos no deseados o también llamado ruido.

La figura 7 ilustra la integración de tres regiones definidas de una
25 proteína de nuevo en el contexto de las proteínas siguiendo la mutagénesis por

revisión, en particular, la integración de las tres CDR de una región variable de la cadena pesada en un formato de anticuerpo de una sola cadena siguiendo la mutagénesis por revisión.

La figura 8 ilustra el uso de la reacción en cadena de la polimerasa (PCR) para construir regiones definidas de una cadena pesada y ligera de anticuerpo sometida a LTM en un contexto de gen mayor.

La figura 9 ilustra de manera ejemplar la diversidad de formatos de cada una de las CDR en una región variable de anticuerpo para obtener un sitio catalítico que comprende, por ejemplo, una serina, histidina, y/o ácido aspártico y como se pueden disponer.

La figura 10 ilustra la integración de las seis CDR de una región de unión de un anticuerpo que se ha sometido a LTM y la diversidad resultante si se usa un resto de aminoácido predeterminado de veinte restos de aminoácidos predeterminados diferentes.

La figura 11 ilustra la integración de las seis CDR de un anticuerpo de cadena individual anti - TNF (sFv) y que se ha sometido a LTM y la diversidad resultante si se usa un resto de aminoácido predeterminado de tres restos de aminoácidos predeterminados diferentes.

La figura 12 ilustra una librería ordenada que representa alguno de los posibles análogos de polipéptidos de las seis CDR de una región de unión de anticuerpo que se puede lograr usando LTM.

La figura 13 ilustra la selección de una librería de expresión ordenada usando representación de ribosomas libres.

La figura 14 ilustra la química combinatoria explorada cuando la región de unión de una región variable de anticuerpo (es decir, las seis CDR) se

somete a LTM.

La figura 15 muestra la secuencia de la región variable (en formato de cadena individual) de varias moléculas de unión anti - TNF representativas que están sometidas a LTM.

- 5 La figura 16 muestra un esquema para llevar a cabo el ensayo de selección de proteasa para candidatos catalíticos de selección de LTM.

La figura 17 muestra un esquema que detalla el mecanismo (y ventajas) del ensayo de selección de proteasa cuando se lleva a cabo en células bacterianas.

- 10 La figura 18 muestra un diagrama de flujo que detalla la mecánica de librerías génicas de selección para actividad catalítica, por ejemplo, actividad de anticuerpo catalítica, que usa representación o bien de ribosomas o de levaduras.

- La figura 19 muestra un esquema para llevara cabo LTM cuando se puede coordinar una información (por ejemplo, información de molde por ordenador) y empírica (resultados de ensayos) para in diseño y desarrollo de molécula más eficaz. Este planteamiento guiado de LTM se denomina mutagénesis "a través de guía"
- 15

- La figura 20 muestra una secuencia de tipo salvaje hipotética V_H CDR3 (óvalos sombreados en la parte superior) y las secuencias resultantes en los miembros de la librerías de la sustitución His por LTM (en óvalos en blanco). Las sustituciones LTM His individuales están codificadas por oligonucleótidos individuales, por ejemplo, oligonucleótidos sintetizados de una manera de alto rendimiento. Las librerías de subconjuntos de LTM para las otras sustituciones de aminoácidos (LTM) en este dominio de CDR se construyen de una manera
- 20
- 25

similar.

La figura 21 ilustra la generación de librerías scFv. En la línea de la parte superior del eje x y la columna alejada de la parte izquierda del eje x los tres dígitos representan las 3 CDR sobre cada una de las cadenas ligeras y pesadas. Un "0" indica una secuencia de CDR de tipo salvaje, mientras que un "1" indica una CDR mutada por LTM. El número sobre la cuadrícula indica la complejidad de la librería subconjunto. Por ejemplo en el extremo superior izquierdo de la matriz es un "0" cuando los ejes x e y correspondientes son "000" y "000" que indican que ninguna de las CDR en o bien el V_H y V_L están respectivamente mutados. Al mover una fila desde el extremo "0" a la posición de la cuadrícula vecina "1" estaría diseñada por el eje a "100" y "000" en el eje y que indica que V_H CDR1 está mutada mientras que V_L CDR permanece de tipo salvaje.

De este modo, de manera similar, una numeración de cuadrícula de "4" significaría que existen una numeración de cuadrícula e "4" significaría que existen cuatro CDR mutadas simultáneamente. Inicialmente las siete cadena V_H y V_L (indicadas por las flechas) se preparan usando SOE - PCR. Las cadenas V_H y V_L se amplifican después y se mezclan y se emparejan mediante un cebador mega para generar las combinaciones V_H y V_L restantes en una etapa.

Descripción detallada de la invención

Con el fin de prevenir un entendimiento claro de la memoria descriptiva y reivindicaciones, se proporcionan a continuación las siguientes definiciones.

Definiciones

Como se usa en esta memoria descriptiva el término "análogo" se refiere a un polipéptido variante o polipeptídico (o un ácido nucleico que codifica tal polipéptido) que tiene una o más sustituciones de aminoácidos.

5 El término "molécula de unión" se refiere a cualquier molécula de unión, incluyendo proteínas, polipéptidos, péptidos, y pequeñas moléculas, que se unen a un sustrato o diana. En una realización, la molécula de unión es un anticuerpo o fragmento de unión (por ejemplo, un fragmento Fab), anticuerpo de dominio individual, anticuerpo de cadena individual (por ejemplo, sFv), o
10 péptido capaz de unirse a un ligando.

El término "región definida" se refiere una región seleccionada de un polipéptido. Típicamente, la región definida incluye todo o una porción de un sitio funcional, por ejemplo, el sitio de unión de un ligando, el sitio de unión de una molécula de unión o receptor, o un sitio catalítico. La región definida
15 también puede incluir porciones múltiples de un sitio funcional. Por ejemplo, la región definida puede incluir todas, o porciones múltiples de una región determinante de complementariedad (CDR) o una región variable (VR) de cadena pesada y/o ligera completa de un anticuerpo. De este modo, el sitio funcional puede incluir una región individual o múltiple definida que contribuya a
20 la actividad funcional de la molécula.

El término "librería" se refiere a dos o más moléculas mutagenizadas de acuerdo con el procedimiento de la invención. Las moléculas de la librería pueden estar en la forma de polinucleótidos, polipéptidos, polinucleótidos y polipéptidos, polinucleótidos y polipéptidos en un extracto libre de células, o
25 como polinucleótidos y/o polipéptidos en el contexto de un fago, células

procarióticas, o en células eucarióticas.

El término "mutagenización" se refiere a la alteración de una secuencia de aminoácidos. Esto se puede lograr alterando o produciendo un ácido nucleico (polinucleótido) capaz de codificar la secuencia de aminoácidos alterados, o mediante la síntesis directa de un polipéptido alterado usando química de proteínas.

El término "polinucleótido (s)" se refiere a ácidos nucleicos tales como moléculas de ADN y moléculas de ARN y análogos de los mismos (por ejemplos, ADN o ARN generado usando análogos de nucleótidos o usando química de ácidos nucleicos). Según se desea, los polinucleótidos se pueden preparar de manera sintética, por ejemplo, usando química de ácidos nucleicos reconocidos en la técnica o enzimáticamente usando, por ejemplo, una polimerasa. Las modificaciones típicas incluyen mutilación, biotinilación, y otras modificaciones conocidas en la técnica. Además, la molécula de ácido nucleico puede ser de cadena sencilla o de cadena doble, cuando se desea, unida o asociada (por ejemplo, de manera covalente o no covalente) a un resto detectable.

El término "polinucleótido variante" se refiere a un polinucleótido que codifica un análogo de polipéptido correspondiente (o porción del mismo) de la invención. De este modo, los polinucleótidos variantes contienen uno o más codones que se han cambiado para dar como resultado la expresión de un aminoácido diferente.

El término "polipéptido (s)" se refiere a dos o más aminoácidos unidos mediante un enlace peptídico, por ejemplo, péptidos (por ejemplo, de a aproximadamente 50 restos aminoácidos), así como secuencias de péptidos

más largos por ejemplo, deficiencias de proteínas que comprenden típicamente secuencias de aminoácidos de cómo poco 50 restos aminoácidos a más de 1000 restos aminoácidos.

El término "combinación" se refiere a la combinación de variantes de polipéptidos o análogos de polipéptidos para formar librerías que representan la mutagénesis por revisión de una región polipeptídica entera. Las moléculas pueden estar en la forma de un polinucleótido y/o polipéptido y puede coexistir en la forma de una sublibrería, como moléculas de un soporte sólido, como moléculas en solución, y/o como moléculas en uno o más organismos (por ejemplo, fago, células procarióticas, o células eucarióticas).

El término "aminoácido predeterminado" se refiere a un resto aminoácido seleccionado para sustitución en cada posición dentro de una región definida o polipéptido a mutagenizar. Esto no incluye posición (es) dentro de la región que ya (por ejemplo, naturalmente) contiene el aminoácido predeterminado y, de este modo, que necesita no estar sustituida con el aminoácido predeterminado. De acuerdo con lo anterior, cada análogo de polipéptido generado de acuerdo con la presente invención contiene no más de un resto "aminoácido predeterminado" en una región determinada dada. Sin embargo, de manera colectiva, la librería de análogos de proteína generados contiene el aminoácido predeterminado en cada posición de la región que se está mutagenizando. Típicamente, un aminoácido predeterminado se selecciona para un tamaño particular o química usualmente asociado con el grupo lateral del aminoácido. Los aminoácidos predeterminados adecuados incluyen, por ejemplo, glicina y alanina (estéricamente pequeño); serina, treonina, y cisteína (nucleófila); valina, leucina, isoleucina, metionina, y prolina (hidrófoba); fenilalanina, tirosina,

y triptófano (aromático); aspartato y glutamato (ácido); asparagina, glutamina, e histidina (amida); y lisina y arginina (básica).

Descripción detallada

5 El estudio de proteínas ha revelado que ciertos aminoácidos juegan un papel crucial en su estructura y función. Por ejemplo, parece que solamente un número de aminoácidos participan en la unión de un anticuerpo a un antígeno o están implicados en el caso catalítico de una enzima.

Aunque está claro que ciertos aminoácidos son críticos para la actividad
10 o función de proteínas, es difícil identificar qué aminoácidos están implicados, cómo están implicados, y qué sustituciones pueden mejorar la estructura o función de proteína. En parte, esto se debe a la complejidad de la configuración espacial de cadenas laterales de aminoácidos en polipéptidos y la interrelación de porciones diferentes del polipéptido que contribuye a formar un sitio
15 funcional. Por ejemplo, la interrelación entre las seis CDR de las regiones variables de las cadenas pesadas y ligeras de un anticuerpo contribuyen al bolsillo de unión a antígeno o ligando.

Los procedimientos de mutagénesis previa, tales como mutagénesis selectiva (dirigida al sitio) y mutagénesis por saturación, son de utilidad limitada
20 para el estudio de estructura de proteínas y función en vista del número enorme de posibles variantes en polipéptidos complejos. Esto es especialmente cierto dado que combinaciones deseables están a menudo acompañadas por la presencia de vastas cantidades de combinaciones no deseables o también llamadas ruido.

25 El procedimiento de esta invención proporciona un planteamiento

sistemático, práctica, y altamente exacto de evaluación del papel de los aminoácidos particulares y su posición, dentro de una región definida de un polipéptido, en la estructura o función del polipéptido y, de este modo, para producir polipéptidos mejorados.

5 1. Selección de una región definida

De acuerdo con la presente invención, una región o regiones definidas dentro de una proteína se seleccionan por mutagénesis. Típicamente, las regiones se cree que son importantes para la estructura o función de las proteínas (véase, por ejemplo, la figura 1). Esto se puede deducir, por ejemplo, a partir de los que los aspectos estructurales y funcionales se conocen o se pueden deducir de la comparación de la (s) región (es) definidas que se conocen a partir del estudio de otras proteínas, y se pueden ayudar mediante modelado de la información. Por ejemplo, la región definida puede ser una que tienen un papel en un sitio funcional, por ejemplo, en catálisis de unión, u otra función. En una realización, la región definida es una región hipervariable o región determinante de complementariedad (CDR) de una molécula de unión a antígeno. En otra realización, la región definida es una porción de una región determinante de complementariedad (CDR). En otra realización, dos o más regiones definidas, por ejemplo, CDR o porciones de la misma, se seleccionan para mutagénesis.

20 2. Selección de un residuo de aminoácidos predeterminado

El resto de aminoácidos elegido para sustitución dentro de la (s) región (es) definida (s) generalmente se selecciona entre las conocidas que están implicadas en la estructura o función de interés. Los veinte aminoácidos de origen natural difieren con respecto a su cadena secundaria. Cada cadena

secundaria es responsable de propiedades químicas que hacen cada aminoácido único. Para el propósito de alteración de la unión o creación de nuevas afinidades de unión, se puede seleccionar en general cualquiera de los aminoácidos de origen natural. De este modo, los procedimientos previos de mutagénesis, que crean vastos números de análogos para cada sustitución, no eran prácticos para evaluar el efecto sobre la unión de proteína de sustitución de cada uno de los veinte aminoácidos. Por el contrario, los procedimientos de la presente invención crean un número práctico de análogos para cada una de las sustituciones de los aminoácidos y, de este modo, permite la evaluación de una diversidad de químicas de proteínas dentro de una región o regiones de una proteína.

En contraste con la unión de proteínas, solamente un subconjunto de restos de aminoácidos participan típicamente en episodios enzimáticos no catalíticos. Por ejemplo, a partir de las propiedades químicas de las cadenas secundarias, solamente un número seleccionado de aminoácidos naturales participan preferentemente en episodio catalíticos. Estos aminoácidos pertenecen al grupo de aminoácidos polares y neutros tales como Ser, Thr, Gln, Tyr, y Cys, el grupo de aminoácidos cargados, Asp y Glu, Lys y Arg, y especialmente el aminoácido His. Otras cadenas secundarias polares y neutras son las de Cys, Ser, Thr, Asn, Gln y Tyr. Gly se considera también que es un miembro límite de este grupo. Ser y Thr juegan un papel importante en la formación de enlaces hidrógeno. Thr tiene una asimetría adicional en el carbono beta, por lo tanto solamente se usa uno de los estereoisómeros. El aminoácido Gln y Asn también puede formar enlaces hidrógeno, los grupos amido que funcionan como dadores de hidrógeno y los grupos carbonilo que

funcionan como aceptores. Gln tiene un grupo más CH₂ que Asn que hace que el grupo polar sea más flexible y reduce su interacción con la cadena principal. Tyr tiene un grupo hidroxilo muy polar (OH fenólico) que se puede disociar a valores de pH altos. Tyr se comporta de alguna manera como una cadena . 5 secundaria cargada; sus enlaces hidrógeno son bastante fuertes.

Los ácidos polares neutros se encuentran en la naturaleza así como en el interior de las moléculas de las proteínas. Como residuos internos, usualmente forman enlaces de hidrógeno con cada otro o con la estructura principal del polipéptido. Cys puede formar puentes disulfuro.

10 Histidina (His) tiene una cadena lateral aromática heterocíclica con un valor de pK de 6,0. En el intervalo de pH fisiológico, su anillo imidazol puede estar o bien no cargado o cargado, después de recoger un ion hidrógeno de la solución. Ya que estos dos estados están fácilmente disponible, His es completamente adecuado para catalizar reacciones químicas. Se encuentra en 15 la mayoría de los centros activos de las enzimas, por ejemplo, serina proteasas.

Asp y Glu están cargados negativamente pH fisiológico. Debido a su corta cadena secundaria, el grupo carboxilo de Asp es bastante rígido con respecto a la cadena principal. Esta puede ser la razón por la que el grupo 20 carbonilo en muchos sitios catalíticos se proporciona mediante Asp y no mediante Glu. Los ácidos cargados se encuentran generalmente en la superficie de un polipéptido.

Además, Lys y Arg se encuentran en la superficie. Tienen cadenas secundarias largas y flexibles que presentan múltiples rotámeros o energías 25 similares. En varios casos, Lys y Arg toman parte en la formación de puentes

de sal interna o su ayuda en catálisis. Debido a su exposición a la superficie del polipéptido, Lys es un resto reconocido más frecuentemente por enzimas que o bien modifican la cadena secundaria o escinden la cadena peptídica en el extremo carbonilo de los restos Lys.

- 5 Mientras que la química de grupo secundario de un aminoácido puede guiar la selección de un resto de aminoácido predeterminado, la carencia de una química de grupo secundario deseado puede ser un criterio para excluir un resto aminoácido para uso como el aminoácido predeterminado. Por ejemplo, los aminoácidos estéricamente pequeños y químicamente neutros, tales como alanina, se pueden excluir de la mutagénesis por revisión para carecer de una química deseada.
- 10

3. Síntesis de librerías de análogos de polipéptidos

- En una realización, una librería de análogos de polipéptidos se genera para seleccionar mediante síntesis de oligonucleótidos individuales que codifican la región definida del polipéptido y no tienen más de un codón para el aminoácido predeterminado. Esto se lleva a cabo incorporando en cada posición de codón dentro del oligonucleótido o bien el codón requerido para la síntesis de polipéptido de tipo salvaje o un codón para el aminoácido predeterminado. Esto se diferencia de los oligonucleótidos producidos en mutagénesis por saturación, mutagénesis al azar, o mutagénesis por revisión porque, para cada oligonucleótido, solamente se realiza una mutación, opuesta a mutaciones múltiples.
- 15
- 20

- Los oligonucleótidos se pueden producir de manera individual y después mezclarse o combinarse como se desee. Cuando el codón de la secuencia de
- 25

tipo salvaje y el codón para el aminoácido son los mismos, no se realiza ninguna sustitución.

De acuerdo con lo anterior, el número de posiciones de aminoácidos dentro de la región definida determinará el número máximo de oligonucleótidos preparados. Por ejemplo, si cinco posiciones de codón están alteradas con el aminoácido predeterminado, entonces se sintetizan cinco oligonucleótidos más un polinucleótido que representa la secuencia de aminoácidos de tipo salvaje. Pueden estar alteradas simultáneamente dos o más regiones.

La mezcla de oligonucleótidos para la generación de la librería se puede sintetizar fácilmente mediante procedimientos conocidos para la síntesis de ADN. El procedimiento preferido implica el uso de química de beta - cianoetil fosforamidita de fase sólida. Véase la patente de Estados Unidos N° 4.725.877. Por conveniencia, se puede usar un instrumento para la síntesis de ADN automática que contenga recipientes de reactivo específico de nucleótidos. Los polinucleótidos también se pueden sintetizar para que contengan sitios de restricción o sitios de hibridación de cebadores para facilitar la introducción o ensamblaje de los polinucleótidos que representan, por ejemplo, una región definida en un contexto de genes mayor.

Los polinucleótidos sintetizados se pueden insertar en un contexto de genes más grande del polipéptido que se está mutagenizando usando técnicas de ingeniería genética convencional. Por ejemplo, los polinucleótidos se pueden preparar para que contengan sitios flanqueantes de reconocimiento para las enzimas de restricción. Véase, Crea, R., patente de estados Unidos N° 4.888.286. Los sitios de reconocimiento se diseñan para que correspondan a sitios de reconocimiento que o bien existen de manera natural o se introducen

en el gen próximo al ADN que codifica la región. Después de la conversión en la forma de cadena doble, los polinucleótidos se ligan en el gen mediante técnicas convencionales. Mediante un vector apropiado (que incluye, por ejemplo, vectores de fagos, plásmidos) los genes se pueden introducir en un
5 extracto sin células, células procarióticas, o células eucarióticas adecuadas para la expresión de los polipéptidos mutantes.

En los casos en los que se conoce la secuencia de aminoácidos del polipéptido a mutagenizar o cuando se conoce la secuencia de ADN, la síntesis de genes en un posible planteamiento. Por ejemplo, polinucleótidos solapados
10 parcialmente, típicamente se pueden diseñar aproximadamente 20 - 60 nucleótidos de longitud. Los polinucleótidos internos se fosforilan después hibridados a su pareja complementaria para proporcionar una molécula de ADN de doble cadena con extensiones de cadena sencilla útiles para hibridación adicional. Los pares hibridados se pueden después mezclar juntos y ligarse
15 para formar una molécula de doble cadena de longitud completa (véase, por ejemplo, la figura 8). Los sitios de restricción convenientes se pueden diseñar cerca de los extremos del gen sintético para la clonación en un vector adecuado. Las moléculas de longitud completa se pueden escindir con las enzimas de restricción y ligarse en un vector adecuado. Los sitios de restricción
20 convenientes también se pueden incorporar en la secuencia del gen sintético para facilitar la introducción de módulos mutagénicos.

Como alternativa para la síntesis de polinucleótidos que representan el gen de doble cadena de longitud completa, los polinucleótidos que se superponen parcialmente en sus extremos 3' (es decir, con extremos 3'
25 complementarios) se pueden ensamblar en una estructura abierta y después

cargarse con una polimerasa adecuada para preparar un gen de doble cadena de longitud completa. Típicamente, los polinucleótidos de superposición son entre 40 - 90 nucleótidos de longitud. Los polinucleótidos extendidos se ligan después. Los sitios de restricción convenientes se pueden introducir en los extremos y/o internamente para propósitos de clonación. Después de la digestión con una enzima o enzimas de restricción apropiada (s), el fragmento de gen se liga en un vector adecuado. Como alternativa, el fragmento de gen puede estar ligado en el extremo como en un vector apropiado.

En estos planteamientos, si están disponibles sitios de restricción convenientes (de manera natural o modificados por ingeniería genética) después del ensamblaje de genes, los polinucleótidos degenerados se pueden introducir posteriormente mediante clonación del módulo en un vector apropiado. Como alternativa, los polinucleótidos degenerados se pueden incorporar en la fase de ensamblaje de genes. Por ejemplo, cuando ambas cadenas del gen se sintetizan completamente químicamente, se puede producir polinucleótidos degenerados complementarios. Los pares complementarios se hibridarán entre sí.

Cuando se usan polinucleótidos parcialmente superpuestos en el ensamblaje de genes, un conjunto de nucleótidos degenerados también se puede incorporar directamente en lugar de uno de los polinucleótidos. La cadena complementaria apropiada se sintetiza durante la reacción de extensión a partir de un polinucleótido complementario parcialmente de la otra cadena mediante extensión enzimática con una polimerasa. La incorporación de los polinucleótidos degenerados en la fase de síntesis también simplifica la clonación cuando se mutageniza más de un dominio o región definida de un

gen.

En otro planteamiento, el gen de interés está presente en un plásmido de cadena sencilla. Por ejemplo, el gen se puede clonar en un vector de fago o un vector con un origen de fago filamentoso de replicación que permite la propagación de moléculas de cadena sencilla con el uso de un fago auxiliar. El molde de cadena sencilla se puede hibridar con un conjunto de polinucleótidos degenerados que representan las mutaciones deseadas y alargarse y ligarse, incorporando de esta manera cada cadena análoga en una población de moléculas que se pueden introducir en un huésped apropiado (Sayers, J. R. y col., Nucleic Acids Res. 16: 791 - 802 (1988)). Este planteamiento puede evitar muchas etapas de clonación cuando los dominios múltiples se seleccionan para mutagénesis.

La metodología de reacción en cadena de la polimerasa (PCR) también se puede usar para incorporar polinucleótidos en un gen. Por ejemplo, los propios polinucleótidos se pueden usar como cebadores para extensión. En este planteamiento, los polinucleótidos que codifican que codifican los módulos mutagénicos que corresponden a la región definida (o porción de la misma) son complementarios entre sí, al menos en parte, y se pueden extender para formar un módulo de genes grande que usa una polimerasa, por ejemplo, usando amplificación de PCR.

El tamaño de la librería variará dependiendo de la longitud y número de regiones y aminoácidos dentro de una región que se mutageniza. Preferiblemente, la librería se diseñar para que contenga menos de 10^{15} , 10^{14} , 10^{13} , 10^{12} , 10^{11} , 10^{10} , 10^9 , 10^8 , 10^7 y más preferiblemente, 10^5 análogos de polipéptidos o menos.

La descripción anterior se ha centrado en la mutagénesis de polipéptidos y librerías de polipéptidos alterando el polinucleótido que codifica el polipéptido correspondiente. Sin embargo, se entiende que el alcance de la invención también abarca procedimientos de mutagenización de polipéptidos mediante al
5 síntesis directa de los análogos de polipéptidos deseados usando química de proteínas. Cuando se lleva a cabo este planteamiento, los polipéptidos resultantes todavía incorporan las características de la invención excepto que se elimina el uso de un polinucleótidos.

Para las librerías descritas anteriormente, bien en la forma de
10 polinucleótidos y/o polipéptidos correspondientes, se entiende que las librerías también pueden estar unidas a un soporte sólido, tal como un microchip, y preferiblemente ordenadas, usando técnicas reconocidas en la técnica.

4. Sistemas de expresión y selección

Las librerías de polinucleótidos generadas mediante cualquiera de las
15 técnicas anteriores u otras técnicas adecuadas se pueden expresar y seleccionar para que identifiquen análogos de polipéptidos que tienen la estructura y/o actividad deseada. La expresión de los análogos de polipéptidos se puede llevar a cabo usando cualquier sistema de representación de expresión adecuada conocida en la técnica incluyendo, pero no limitada a,
20 sistema de representación de extractos sin células (por ejemplo, representación de ribosomas y sistemas de representación ordenada (por ejemplo, microordenadas o macroordenadas), sistemas de representación bacteriana, sistemas de representación de fago, células procarióticas, y/o células eucarióticas (por ejemplo, sistemas de representación de levaduras).

25 En una realización, los polinucleótidos se modifican por ingeniería

genética para que sirvan como moldes que se pueden expresar en un extracto sin células. Los vectores y extractos como se describen, por ejemplo, en las patentes de Estados Unidos N° 5.324.837; 5.492.817; 5.665.563, se pueden usar y muchos están comercialmente disponibles. Se puede usar la
5 representación de ribosomas y otras técnicas sin células para la unión a un polinucleótido (por ejemplo, un genotipo) a un polipéptido (es decir, un fenotipo) por ejemplo, Profusion™ (véase por ejemplo, las patentes de Estados Unidos números 6.348.315; 6.21.804; 6.258.558; y 6.214.553).

Como alternativa, los polinucleótidos de la invención se pueden expresar
10 en un sistema de expresión conveniente de *E. coli*, tal como el descrito por Pluckthum y Skerra. (Pluckthum, A. y Skerra, A., *Meth. Enzymol.* 178: 476 - 515 (1989); Skerra, A. y col., *Biotechnology* 9: 273 - 178 (1991)). Las proteínas mutantes se pueden expresar para la secreción en el medio y/o en el citoplasma de la bacteria, como describen M. Better y A. Horwitz, *Meth.*
15 *Enzymol.* 178: 476 (1989). En una realización, los dominios individuales que codifican VH y VL están cada uno unidos al extremo 3' de una secuencia que codifica una secuencia señal, tal como la secuencia señal ompA, phoA o pelB (Lei, S. P. y col., *J. Bacteriol.* 169: 4379 (1987). Estas fusiones de genes se ensamblan en una construcción dicistrónica, de manera que se pueden
20 expresar a partir de un vector individual, y secretarse en el espacio periplásmico de *E. coli* cuando se repliegan y se pueden recuperar en una forma activa, (Skerra, A. y col., expresada simultáneamente con los genes de las cadenas ligeras de anticuerpos que producen anticuerpos y fragmentos de anticuerpos.

25 Todavía en otra realización, los polinucleótidos se pueden expresar en

células eucarióticas tales como el uso de levaduras, por ejemplo, representación de levaduras como se describe por ejemplo, en las patentes de Estados Unidos números 6.423.538; 6.331.391; y 6.300.065: En este planteamiento, los análogos de polipéptidos de al librería se condensan a un

5 polipéptido que se expresa y se representa sobre la superficie de la levadura. También se pueden usar otras células eucarióticas para la expresión de los polipéptidos de la invención, tales como células de mamíferos, por ejemplo células de mieloma, células de hibridoma, o células de ovarios de hámster chino (CHO). Típicamente, los análogos de polipéptidos cuando se expresan

10 en células de mamíferos se diseñan para que se expresen en el medio de cultivo, o expresarse sobre la superficie de tal célula. El anticuerpo o fragmento de anticuerpos se puede producir, por ejemplo, una molécula de anticuerpo entera o como fragmentos VH y VL individuales, fragmentos Fab, dominios individuales, o como cadenas individuales (sFv) (véase, Huston, J. S. y col.,

15 Proc. Natl. Acad. Sci. USA 85: 5879 - 5883 (1988)).

La selección de los análogos de polipéptidos expresados (o polipéptidos producidos mediante síntesis directa) se pueden hacer mediante cualquier medio apropiado. Por ejemplo, la actividad de unión se puede evaluar mediante inmunoensayo convencional y/o cromatografía de afinidad y actividad catalítica

20 se puede determinar mediante ensayos adecuados para la conversión de sustratos. La selección de los análogos de polipéptidos de la invención para la función proteolítica se puede llevar a cabo usando un ensayo en placa de hemoglobina convencional como se describe, por ejemplo, en la patente de Estados Unidos N° 5.798.208.

5. Mutagénesis por revisión asistida por modelado por ordenación

La mutagénesis por revisión de la invención también se puede llevar a cabo con el beneficio de información estructural o modelado concerniente a los análogos de polipéptidos a generar, de manera que el potencial para generar análogos que tienen la función deseada mejorada se incrementa. La información estructural o de modelado se pueda también usar para guiar la selección de aminoácidos predeterminado para introducir en las regiones definidas. Todavía adicionalmente, los resultados reales obtenidos con los análogos polipeptídicos de la invención pueden guiar la selección (o exclusión) de polipéptidos posteriores a preparar y seleccionar de una manera interactiva. De acuerdo con lo anterior, la información estructural o de modelado se puede usar para generar subconjuntos iniciales de análogos de polipéptidos para uso en la invención, por lo tanto incrementando adicionalmente la eficacia de generación de polipéptidos mejorados.

En una realización particular, el modelado en silico se usa eliminar la producción de cualquier análogo de polipéptido predicho que tiene un estructura pobre o indeseada y/o función. De esta forma, el número de análogos de polipéptidos a producirse se puede reducir notablemente incrementando por lo tanto la señal a ruido en ensayos de selección posteriores. En otra realización particular, el modelado silico se actualiza continuamente con modelado adicional, a partir de cualquier fuente relevante, por ejemplo, a partir de una secuencia de genes y de proteínas y la base de datos de tres dimensiones y/o resultados de los análogos ensayados previamente, de manera que la base de datos en silico se hace más precisa en su capacidad predictiva (figura 19).

Todavía en otra realización, la base de datos en silico se proporciona con los resultados de ensayo de los análogos de polipéptidos previamente ensayados y clasifica los análogos, basándose en el criterio o criterios de ensayo, como efectores o no efectores, por ejemplo, como análogos de polipéptidos que se unen bien o no tan bien o por ser enzimática /catalítica o no ser tan enzimática/catalítica. De esta forma la mutagénesis mediante guía de la invención puede igualar un intervalo de respuesta funcional con información estructural particular y usarse tal como información para guiar la producción de análogos de polipéptidos futuros a ensayar. De acuerdo con lo anterior el procedimiento es especialmente adecuado para seleccionar anticuerpos o fragmentos de anticuerpos para una función particular, tal como afinidad de unión (por ejemplo, especificidad), estabilidad (por ejemplo, vida media) y/o función efectora (por ejemplo, activación de complementos y ADCC). De acuerdo con lo anterior, la mutagénesis de restos no contiguos dentro de una región puede ser deseable si se conoce, por ejemplo, mediante un modelado en silico que ciertos restos en la región no participarán en la función deseada. La estructura coordinada e interrelación espacial entre las regiones deseadas, por ejemplo, los restos aminoácidos funcionales en las regiones definidas del polipéptido, por ejemplo, el (los) aminoácido (s) predeterminado (s) que se han introducido, se pueden considerar y modelar. Tales criterios de modelado, por ejemplo, química de grupos secundarios de aminoácidos, distancias atómicas, datos de cristalografía, etc. De acuerdo con lo anterior, el número de análogos de polipéptidos a producir se pueden minimizar de manera inteligente.

En una realización preferida, uno o más de las etapas anteriores están

asistidas por ordenador. El procedimiento también es capaz de llevarse a cabo, en parte o en conjunto, mediante un dispositivo dirigido por ordenador. De acuerdo con lo anterior, las instrucciones para llevar a cabo el procedimiento, en parte o en conjunto, se pueden conferir a un medio adecuado para uso en un dispositivo electrónico para llevar a cabo las instrucciones. En total, los procedimientos de la invención son capaces de un planteamiento de alto rendimiento que comprende software (por ejemplo, instrucciones legibles de ordenador) y hardware (por ejemplo, los ordenadores, robots y chips).

6. Exploración de la química combinatoria de regiones múltiples
definidas

La presente invención proporciona la importante ventaja de permitir la evaluación por mutagénesis de varias regiones o dominios diferentes de un polipéptido simultáneamente. Esto se puede hacer usando el mismo o diferente aminoácido diferente predeterminado dentro de cada región, permitiendo la evaluación de sustituciones de aminoácidos en regiones relacionadas de manera conformacional, de forma que las regiones que tras el plegamiento del polipéptido, están asociadas para preparar un sitio funcional (por ejemplo, el sitio de unión de un anticuerpo o el sitio catalítico de una enzima). Esto, a su vez, proporciona una forma eficaz de crear sitios funcionales nuevos o mejorados.

Por ejemplo, como se muestra en la figura 14, las seis CDR de un anticuerpo que preparan los aspectos únicos del sitio de unión de antígeno (región Fv), se puede mutagenizar simultáneamente, o separadamente en las cadenas VH o VL, para estudiar las interrelaciones de tres dimensiones de aminoácidos seleccionados en este sitio. En una realización, la química

combinatoria de las tres o más regiones se exploran de manera sistemática usando mutagénesis por revisión, y preferiblemente seis regiones definidas, por ejemplo las seis CDR de una región variable de la cadena pesada y ligera de anticuerpo. Para realizar esta mutagénesis por revisión sobre una CDR,
5 típicamente 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, o más posiciones de aminoácidos se alteran.

De acuerdo con lo anterior, la presente invención abre nuevas posibilidades para el diseño de tipos muy diferentes de polipéptidos nuevos y mejorados. El procedimiento se puede usar para mejorar una estructura o
10 función existente de una proteína. Por ejemplo, se puede introducir un sitio de unión para un anticuerpo o fragmento de anticuerpo o afinidad para un antígeno preexistente, función efectora y/o estabilidad mejorada. Como alternativa, la introducción de aminoácidos "catalíticamente importantes" adicionales en un dominio catalítico de una enzima se puede
15 realizar dando como resultado una actividad modificada o potenciada hacia un sustrato. Como alternativa, las estructuras completamente enteras, las especificidades o actividades se pueden introducir en un polipéptido. La síntesis de novo de actividad enzimática se puede lograr también. Las nuevas estructuras se pueden edificar sobre la "estructura" natural o de consenso de
20 una proteína existente mediante mutación de regiones solamente relevantes mediante el procedimiento de la invención.

7. Mutagénesis por revisión para preparar anticuerpos nuevos o mejorados

El procedimiento de esta invención se usa específicamente para modificar moléculas de anticuerpos. Como se usa en esta memoria descriptiva, las moléculas de anticuerpos o anticuerpos se refiere a anticuerpos o porciones de los mismos, tales como anticuerpos de longitud completa, moléculas Fv, u
5 otros fragmentos de anticuerpos, cadenas individuales o fragmentos de los mismos (por ejemplo, una sola cadena de Fv), anticuerpos de cadena individual, y anticuerpos quiméricos. Se pueden introducir alteraciones en la región variable y/o en la región marco conservada (constante) de un anticuerpo. La modificación de la región variable puede producir anticuerpos
10 con mejores propiedades de unión de antígeno, y, si se desea, propiedades catalíticas. La modificación de la región marco conservada también puede conducir a la mejorar de las propiedades químico - físicas, tales como solubilidad o estabilidad (por ejemplo, vida media), que son especialmente útiles, por ejemplo, en producción comercial, biodisponibilidad, función efectora (por
15 ejemplo, activación de complementos y/o ADCC) y afinidad de unión (por ejemplo, especificidad) para el antígeno. Típicamente, al mutagénesis dirigirá la región Fv de la molécula de anticuerpo, es decir, la estructura responsable para la actividad de unión de antígeno que se compone de regiones variables de dos cadenas, una de la cadena pesada (VH) y una de la cadena ligera (VL). Una
20 vez que las características de unión de antígeno se identifican, la (s) región (es) variable (s) se pueden modificar por ingeniería genética en una clase de anticuerpo apropiada tal como IgG, IgM, IgA, IgD, o IgE.

— —

8. Mutagénesis por revisión para preparar/mejorar polipéptidos
25 catalíticos/enzimáticos

El procedimiento de la invención también es particularmente adecuada para el diseño de proteínas catalíticas, anticuerpos particularmente catalíticas. Actualmente, los anticuerpos catalíticos se pueden preparar mediante una adaptación de técnicas de fusión de células somáticas convencionales. En este
5 procedimiento, un animal se inmuniza con un antígeno que parece el estado de transición del sustrato deseado para inducir la producción de un anticuerpo que une el estado de transición y cataliza la reacción. Las células que producen anticuerpos se recogen a partir del animal y se condensan con una célula de inmortalización para producir células híbridas. Después las células se
10 seleccionan para la secreción de un anticuerpo que cataliza la reacción. Este procedimiento depende de la disponibilidad de análogos del estado de transición de un sustrato. El procedimiento puede estar limitado debido a tales análogos son probablemente que sean difíciles de identificar o sintetizar en la mayoría de los casos.

15 El procedimiento de la invención proporciona un planteamiento diferente que elimina la necesidad de un análogo del estado de transición. Mediante el procedimiento de la invención, un anticuerpo se puede hacer catalítico mediante la introducción de aminoácidos adecuados en el sitio de unión de una inmunoglobulina (región Fv). El sitio de unión de antígeno (Fv) se compone de
20 seis bucles hipervariables (CDR), tres derivadas de la cadena pesada de inmunoglobulina (H) y tres de la cadena ligera (L), que conecta cadenas beta con cada subunidad. Los restos de aminoácidos de los bucles CDR contribuyen casi enteramente a las características de cada anticuerpo monoclonal específico. Por ejemplo, las triadas catalíticas (constituidas por restos de
25 aminoácidos serina, histidina, y ácido aspártico) modeladas después de serina

proteasas se pueden crear en los regiones hipervariables de la región Fv de un anticuerpo con afinidad conocida para la molécula de sustrato y seleccionada para actividad proteolítica del sustrato.

En particular, el procedimiento de la invención se puede usar para
5 producir muchas enzimas diferentes o anticuerpos catalíticos, incluyendo
óxidorreductasas, transferasas, hidrolasas, liasas, isomerasas y ligasas. Entre
estas clases, de particular importante será la producción de proteasas,
carbohidrasas, lipasas, dioxigenasas y peroxidasas mejoradas. Éstas y otras
enzimas que se pueden preparar mediante el procedimiento de la invención
10 tienen aplicaciones comerciales importantes para las conversiones enzimáticas
en asistencia sanitaria, cosméticos, alimentos, elaboración de cerveza,
detergentes, ambiente (por ejemplo, tratamiento de aguas residuales),
agricultura, curtido, textiles, y otros procedimientos químicos. Estos incluyen,
pero no se limitan a, aplicaciones diagnósticas y terapéuticas, conversiones de
15 grasas, carbohidratos y proteína, degradación de contaminantes orgánicos y
síntesis de agentes químicos. Por ejemplo, las proteasas terapéuticamente
eficaces con actividad fibrinolítica, o actividad contra estructuras virales
necesarias para infectividad, tal como proteínas de revestimiento viral, se
pueden modificar por ingeniería genética. Tales proteasas pueden ser útiles
20 agentes anti - trombolíticos o agentes anti - virales contra virus tales como, por
ejemplo, VIH, rinovirus, influenza, o hepatitis. En el caso de oxigenasas (por
ejemplo, dioxigenasas), una clase de enzimas que requieren un co - factor para
oxidación de anillos aromáticos y otros dobles enlaces, aplicaciones
industriales en procesos de formación de biopulpa, conversión de biomásas en
25 combustibles u otros agentes químicos, conversión de contaminantes de aguas

residuales, bioprocesamiento de carbón, y detoxificación de compuestos orgánicos peligrosos son posibles aplicaciones de proteínas novedosas.

La presente invención se ilustra adicionalmente en los siguientes ejemplos, que no se deben considerar como limitantes.

5

Ejemplificación

A lo largo de los ejemplos, se usaron los siguientes ejemplos y procedimientos salvo que se especifique otra cosa.

10

Materiales y procedimientos

En general, la práctica de la invención emplea, salvo que se indique otra cosa, técnicas convencionales de química, biología molecular, tecnología de ADN recombinante, tecnología de PCR, inmunología (especialmente, por ejemplo, tecnología de anticuerpos), sistemas de expresión (por ejemplo, expresión sin células, representación de fagos, representación de ribosomas, y Profusión™), y cualquier cultivo de células necesario que están dentro de la habilidad en la técnica y se explican en la bibliografía. Véase, por ejemplo, Sambrook, Fritsch y Maniatis, *Molecular Cloning: Cold Spring Harbour Laboratory Press* (1989); *DNA Cloning*, Volúmenes 1 y 2, (D. N. Glover, Ed. 1985); *Oligonucleotides Synthesis* (M. J. Gait, Ed. 1984); *PCR Handbook* *Current Protocols in Nucleic Acid Chemistry*, Beaucage, Ed. John Wiley y Sons (1999) (Editor); *Oxford Handbook of Nucleic Acid Structure*, Neidle, Ed., Oxford Univ. Press (1999); *PCR Protocols: A Guide to Methods and Applications*, Innis y col., Academic Press (1990); *PCR Essential Techniques: Essential Techniques*, Burke, Ed., John Wiley y Sons Ltd (1996); *The PCR Technique*:

15

20

25

RT - PCR, Siebert, Ed, Eaton Pub. Co. (1998); *Antibody Engineering Protocols* (Methods in Molecular Biology), 510, Paul, S., Humana Pr (1996); *Antibody Engineering: A Practical Approach* (Practical Approach Series, 169), McCafferty, Ed., Irl Pr (1996); *Antibodies: A Laboratory Manual*, Harlow y col., C. S. H. L. Press, Pub. (1999); *Current Protocols in Molecular Biology*, eds, Ausubel y col., John Wiley y Sons (1992); *Large - Scale Mammalian Cell Culture Technology*, Lubieniecki, A., Ed., Marcel Dekker, Pub., (1990). *Phage Display: A Laboratory Manual*, C. Barbas (Ed.), CSHL Press, (2001); *Antibody Phage Display*, P. O'Brien (Ed.), Humana Press (2001); Border y col., Yeast surface display for screening combinatorial polypeptide libraries, *Nature Biotechnology*, 15 (8): 553 - 7 (1997); Border y col.; Yeast surface display for directed evolution of protein expression, affinity, and stability, *Methods Enzymol.*, 328: 430 - 44 (2000); representación de ribosomas como describen Pluchthum y col., en la Patente de Estados Unidos Nº 6.348.315, y Profusión™ como describen Szoatak y col., Patente de Estados Unidos Números 6.258.558; 6.261.804; y 6.214.553.

EJEMPLO 1

MUTAGÉNESIS POR REVISIÓN DE LAS TRES REGIONES DEFINIDAS RN UNA MOLÉCULA DE UNIÓN DE ANTÍGENO

En este ejemplo, se describe a mutagénesis por revisión de tres CDR de un anticuerpo para mejorar la unión y proteólisis de un sustrato.

En particular, se realiza la mutagénesis "por revisión" de tres regiones determinantes de complementariedad (CDR) de un anticuerpo monoclonal.

CDR1, CDR2, y CDR3 de la región variable de la cadena pesada (VH) son regiones definidas seleccionadas para la mutagénesis por revisión. Para esta realización, los aminoácidos predeterminados seleccionados son los tres restos de la tríada catalítica de las serina proteasas, Asp, His y Ser. Asp se selecciona para VH CDR1, His se selecciona para VL CDR2, y Ser se selecciona para VH CDR3. La selección de tres aminoácidos predeterminados permite el uso de un ensayo de proteasa conveniente con el fin de detectar cuando los tres restos están posicionados correctamente para mostrar una actividad funcional, es decir, proteólisis de un sustrato de ensayo.

Un anticuerpo ejemplar, MCPC 603, se reconoce como un buen modelo para investigar la unión y catálisis debido a que la región de unión de anticuerpo se ha caracterizado bien. La secuencia de aminoácidos y la secuencia de ADN del MCPC 603 VH y regiones VL están públicamente disponibles (véase, por ejemplo, Rudikoff, S. y Potter, M., *Biochemistry* 13: 4033 (1974); Plukthun, A. y col., *Cold Spring Harbor Symp. Quant. Biol.*, Vol. LII: 105 - 112 (1987)). Las CDR para el anticuerpo MCPC 603 se han identificado como se muestra en la figura 2. En la cadena pesada, las posiciones 31 - 35 de los residuos de aminoácidos de los tramos CDR1, las posiciones 50 - 69 de los tramos CDR2, y las posiciones 101 - 111 de los tramos CDR3. En la cadena ligera, los restos de aminoácidos de CDR1 son 24 - 40, los aminoácidos 55 - 62 de los tramos CDR2, y los aminoácidos 95 - 103 de los tramos de CDR3.

El diseño de los oligonucleótidos para la mutagénesis por revisión en las CDR de MCPC 603 es tal que los análogos de polipéptidos resulta que tienen la secuencia de aminoácidos para CDR1, CDR2, y CDR3 como se muestran en

las figuras 3 - 5. Se entiende que los oligonucleótidos sintetizados pueden ser mayores que las regiones CDR a alterar para facilitar la inserción en la construcción diana como se muestra en la figura 7. Un formato de anticuerpo de cadena individual se elige por conveniencia en la expresión posterior y etapas de selección. Los oligonucleótidos se pueden convertir en cadenas de doble cadena mediante técnicas enzimáticas (véase, por ejemplo, Oliphant, A. R. y col., 1986, anteriormente) y después se ligan a un plásmido restringido como se muestra en la figura 8. Los sitios de restricción pueden ser sitios de origen natural o sitios de restricción modificados por ingeniería genética.

Los polinucleótidos que codifican los análogos de polipéptidos se pueden expresar en cualquiera de sistemas de expresión convenientes descritos en esta memoria descriptiva y seleccionarse usando el ensayo de la serina proteasa descrito, por ejemplo, en la patente de Estados Unidos Nº 5.798.208 (véase también las figuras 16 - 17). En resumen, los análogos de polipéptidos expresados se exponen a un sustrato de ensayo y examinarse para proteólisis del sustrato de ensayo. La cantidad de proteólisis, revelado por una zona de eliminación del sustrato, indica un análogo de polipéptido que tiene la actividad funcional deseada.

20

EJEMPLO 2

MUTAGÉNESIS POR REVISIÓN DE SEIS REGIONES DEFINIDAS EN UNA MOLÉCULA DE UNIÓN A ANTÍGENOS

En este ejemplo, se describe la mutagénesis por revisión de los seis CDR de un anticuerpo para mejorar la unión y proteólisis de un sustrato.

25

En particular, se realiza una mutagénesis por "revisión" de las seis regiones hipervariable o regiones determinantes de complementariedad (CDR) del anticuerpo de modelo mencionado anteriormente (MCPC 603). En este ejemplo, la mutagénesis por "revisión" se lleva a cabo a partir de dos a tres veces con un aminoácido diferente en una región o dominio dada. Por ejemplo, Asp., Ser e His están secuencialmente en el camino de espera de las cadenas pesadas y ligeras como se muestra en la figura 10.

La mutagénesis de los residuos continuos dentro de una región puede ser deseable si se conoce, o si se puede deducir, que ciertos residuos en la región no participará en la función deseada. Además, el número de análogos se puede minimizar. Otras consideraciones en la selección del aminoácido predeterminado y las posiciones particulares a alterar son que los residuos pueden ser capaces de hidrogenar enlaces entre sí. Esta consideración puede imponer una restricción de proximidad sobre las variantes deseadas. De este modo, solamente ciertas posiciones dentro de las CDR pueden permitir que los aminoácidos de la triada catalítica interaccione de manera apropiada. De este modo, el modelo molecular u otra información estructural se puede usar para enriquecer variantes funcionales.

En este caso, la información estructural conocida se usó para identificar restos en las regiones que pueden estar suficientemente cerca para permitir el enlace hidrógeno entre Asp, His y Ser, así como el intervalo de restos a mutagenizar. Roberts y col., han identificado regiones de estrecho contacto entre porciones de las CDR (Roberts, V. A. y col., Proc. Natl. Acad. Sci. USA 87: 6654 - 6658 (1990)). Esta información junto con los datos de la estructura de rayos x de MCPC 603 se usa para seleccionar áreas prometedoras de

estrecho contacto entre las CDR dirigidas para mutagénesis. Este tipo de mutagénesis por revisión guiada por información estructural/de modelo se puede denominar como mutagénesis "a través de guía".

La mutagénesis por revisión se lleva a cabo como se ilustra en la figura 5 10 en la que cada CDR se somete a un aminoácido resultante predeterminado que produce los análogos de polipéptidos 8×10^5 y si se exploran los veinte aminoácidos, análogos de polipéptidos 5×10^{13} .

Los polinucleótidos se pueden expresar si cualquiera de los sistemas de expresión convenientes descritos en esta memoria descriptiva y seleccionarse 10 usando la serina proteasa descrita, por ejemplo, en la patente de Estados Unidos N° 5.788.208. En resumen, los análogos de polipéptidos expresados se exponen a un sustrato de ensayo y se examinan para evaluar proteolisis del sustrato de ensayo. La cantidad de proteolisis, revelada mediante una zona de eliminación del sustrato, indica un análogo de polipéptido que tiene la actividad 15 funcional deseada.

EJEMPLO 3

MUTAGÉNESIS POR REVISIÓN DE MOLÉCULAS DE UNIÓN A ANTI - TNF PARA MEJORAR LA FUNCIÓN

20

En este ejemplo, se describe la mutagénesis por revisión de un anticuerpo anti - TNF para mejorar la unión.

En particular, se realiza la mutagénesis por "revisión" de las seis regiones hipervariables o regiones determinantes de complementariedad 25 (CDR) de dos anticuerpos anti - TNF. Los anticuerpos anti - TNF tienen

aplicación general en el tratamiento de una enfermedad inmune en pacientes que tienen niveles inapropiados del ligando TNF (factor de necrosis tumoral). Existen dos anticuerpos anti - TNF disponibles comercialmente. Para conveniencia en la realización de mutagénesis por revisión y posterior
5 selección, las regiones variables de cadena ligera y pesada (véase las SEQ ID números 2 - 4) de estos anticuerpos se convirtieron en un formato de cadena individual usando un engarce poli Gly - Ser (véase la figura 15). Las regiones definidas seleccionadas por mutagénesis por revisión con una aminoácido predeterminado se identifican mediante la presencia de una barra negra como
10 se muestra en la figura 15. Estas regiones definidas corresponden a las CDR de los anticuerpos de las cadenas individuales.

Los polinucleótidos que representan las seis regiones CDR y regiones flanqueantes suficientes para permitir el ensamblaje de los polipéptidos en la secuencia de cadena individual completa mostrada en la figura 15 se sintetizan
15 como se describe en esta memoria descriptiva. Los restos de aminoácidos predeterminados se seleccionan para cada región CDR y separadamente y secuencialmente introducidos en cada posición de aminoácidos a través de las seis regiones CDR. Los polinucleótidos se modifican por ingeniería genética adicionalmente para que sirvan como moldes capaces de soportar la
20 transcripción de una transcripción de ARN correspondiente que se pueden después traducir en un polipéptido usando representación de ribosomas.

Los polinucleótidos que codifican los análogos de polinucleótidos correspondientes se expresan usando una transcripción *in vitro* sin células y extractos de traducción. Las transcripciones de ARN están unidas covalentemente a un
25 resto detectable tal como un resto fluorescente. Como alternativa, la secuencia

de interés se puede condensar en marco a un resto fluorescente tal como una proteína fluorescente verde (GFP) para permitir la detección conveniente de expresión y normalización de aglutinantes frente a no aglutinantes. Preferiblemente, los polinucleótidos que codifican los análogos de polipéptidos están al menos parcialmente ordenados, por ejemplo, expresados en un pocillo que tienen múltiples análogos de polipéptidos pero se pueden fácilmente desplegar. De acuerdo con lo anterior, cada pocillo contiene un subconjunto de análogos de polipéptidos unidos ahora a la transcripción correspondiente unida a un resto fluorescente usando representación de ribosomas. El pocillo se sonda con ligando diana, es decir TNF y se ensaya para análogos de polipéptidos que se unen y con los que la afinidad de unión comparada al polipéptido de tipo salvaje.

Los análogos de polipéptidos que se unen mejor que los polipéptidos de tipo salvaje se vuelven a modificar después por ingeniería genética en un formato de anticuerpo IgG de longitud completa para ensayo paralelo con los análogos de anticuerpos comerciales para unión mejorada usando técnicas convencionales.

EJEMPLO 4

**MUTAGÉNESIS POR REVISIÓN DE ANTICUERPOS CONTRA EL
SEROTIPO B DE TOXINA NERVIOSA DE BOTULINUM (BoNT/B) Y
SEROTIPO A DE TOXINA NERVIOSA DE BOTULINUM (BoNT/A) PARA
MEJORAR LA FUNCIÓN**

En este ejemplo, anticuerpos mejorados contra el serotipo Bb de toxina

nerviosa de botulinum (BoNT/B) y serotipo a de toxina nerviosa de botulinum (BoNT/A) se generan usando mutagénesis por revisión (LTM) para mejorar la función. El planteamiento LTM se basa en la creación de mutaciones sencillas por CDR a través del bolsillo de unión, basándose en un subconjunto (el conjunto de LTM) de los 20 aminoácidos que exploran las características de tamaño, carga, hidrofobicidad y unión por hidrógeno. Los criterios para seleccionar los aminoácidos en el conjunto de LTM se describen más adelante.

1. Purificación de anticuerpos y secuenciación de genes y diseño de cadenas individuales (scFv)

Los fragmentos de anticuerpos murinos (Fab) con afinidad de unión a BoNT/B se pueden obtener como se describe en Emanul y col., (1996) *Journal of Immunological Methods* 193: 189 - 197. Se pueden usar los anticuerpos BotFab 20 (SEQ ID números 12 y 13, polipéptidos de cadena ligera y pesada, respectivamente), el BotFab 20 (SEQ ID Números 14 y 15, polipéptidos de cadena ligera y pesada, respectivamente). Las secuencias foráneas se describen en la patente de Estados Unidos N° 5.932.449, cuyo contenido se incorpora en esta memoria descriptiva como referencia. El antígeno BoNT/B (toxina de longitud completa, cadena ligera, y/o cadena pesada) se puede obtener a partir de Metabionics, Inc., WI.

Los anticuerpos anti - BoNT/A descritos en Pless y col., (2001) *Infect. Immun.* 69: 570 se pueden usar con el objetivo de mejorar sus afinidades mediante al menos un orden de magnitud. El antígeno de dominio de unión de cadena pesada BoNT/A (BoNT/A) se puede obtener a partir de Metabionics,

Unc., WI.

Los fragmentos V_L y V_H del (de los) anticuerpo (s) se clonan y se secuencias usando técnicas de biología molecular convencionales. Las regiones variable de las moléculas se amplifican mediante la reacción en

5 cadena de la polimerasa (PCR) y se unen con un engarce poli - Gly - Ser (típicamente SGGGSGGGGSGGGGS (SEQ ID N° 7)) para generar anticuerpos de cadena individual (csFv). Una etiqueta poli - His (HHHHHH (SEQ ID N° 8)) y una etiqueta myc (EQKLISEEDL (SEQ ID n° 9)) también se anexan al extremo C de los genes para facilitar la purificación y detección.

10 Estas moléculas se muestran en cualquiera de las tecnologías bien conocidas (por ejemplo, tecnologías basadas en levaduras, bacterias o fagos) y se ensayaron para evaluar su capacidad de unirse a BoNT/B o BoNT/A.

La versión monovalente de scFv de anticuerpos completos proporcionan un buen formato para acometer estudios de mutagénesis y se conocen para

15 reproducir generalmente el mecanismo de unión de la molécula completa, con la excepción de los efectos de avidéz mostrados por moléculas multivalentes.

2. Mejora de anticuerpos usando mutagénesis por revisión (LTM) y diseño de genes

20 Para la mutagénesis por revisión (LTM), se eligen los siguientes nueve aminoácidos y sus características funcionales representativas: Alanina y Leucina (alifáticos), Serina (grupo hidroxilo), Ácido aspártico (ácido) y Glutamina (amida), Lisina e Histidina (básicos), Tirosina (aromático), Prolina (hidrófobo). Estos aminoácidos muestran diversidad química adecuada en

25 tamaño, carga, hidrofobicidad, y unión por hidrógeno para proporcionar

información inicial significativa sobre la funcionalidad química necesaria para mejorar las propiedades de anticuerpo. Las elecciones también se basan en la frecuencia de aparición de sus aminoácidos en las CDR de los anticuerpos. Por ejemplo, dado entre tirosina y fenilalanina para representar aminoácidos con cadenas secundarias aromáticas, el anterior se elige debido a su preponderancia significativamente más alta en las CDR de anticuerpos y su capacidad a enlace por hidrógeno. LTM se emplea inicialmente para identificar aminoácidos específicos y propiedades químicas que son beneficiosas para unión, neutralización y/o cualesquiera propiedades adicionales deseadas en el anticuerpo final. Sin embargo, LTM no se limita a estos nueve aminoácidos o nueve aminoácidos totales. El análisis de LTM se puede realizar con cualquier combinación de aminoácidos y el subconjunto de LTM puede ser tan alto como 18 aminoácidos (1 corto de mutagénesis por saturación).

2.1 Síntesis de oligonucleótidos de LTM

En el análisis primario de LTM, el objetivo es explorar cada contribución de la cadena de aminoácidos a la afinidad de unión global dentro de cada CDR. Para generar eficazmente la diversidad significativa, cada uno de los aminoácidos del subconjunto de LTM se dirige en cada posición individual dentro de la secuencia de CDR con una sustitución solamente por CDR. De este modo, cada oligonucleótido individual codifica solamente una mutación de CDR individual. Por ejemplo, con el fin de realizar la mutagénesis de histidina de LTM sobre un dominio V_H CDR3 de once aminoácidos hipotéticos, se sintetizan 11 oligonucleótidos que codifican 11 mutaciones posibles individuales (véase al figura 20). Tal análisis ensaya los efectos de tener una

amida voluminosa en cada posición en la CDR. Por lo tanto para generar una librería de V_H CDR3 LTM, solamente se sintetizan 99 oligonucleótidos para el análisis de LTM (9 aminoácidos de LTM x 11 posiciones de V_H CDR3). Las secuencias de oligonucleótidos se ensayan para codones de parada
5 accidentales, duplicación de tipo salvaje, uso de codón ineficaz, horquillas, bucles, y otras estructuras secundarias que usan software disponible públicamente.

2.2 Síntesis de oligonucleótidos degenerados para las combinaciones de 10 **mutaciones de LTM beneficiosas**

Al emplear el reemplazo de LTM sistemático de aminoácidos individuales dentro de un CDR, la preferencia de las funcionalidades químicas en cada posición está sin cubrir. Con el fin de combinar todas las mutaciones de las selecciones de LTM y explorar la posible aditividad y sinergia energética
15 entre éstas, se sintetizan los oligonucleótidos degenerados que codifican estas mutaciones y la secuencia de tipo salvaje. La combinación degenerada de oligonucleótidos con $1,6 \times 10^4$ variantes se usa posteriormente para generar una segunda librería de generación que explora la naturaleza aditiva de estas sustituciones.

20

2.3 Diseño de oligonucleótidos asistido por ordenador, librería, y base de **datos de los resultados.**

El software acoplado con sintetizadores de ADN de construcciones de costumbre automáticos permite la síntesis rápida de oligonucleótidos. La
25 primera etapa implica la decisión de qué aminoácidos diana se incorporarán en

las CDR. El software determina la preferencia de codón (por ejemplo, preferencia de codón de levadura, bacteriano, o de fago, dependiendo del sistema elegido) necesario para introducir los aminoácidos diana y también elimina cualquier duplicación de la secuencia de tipo salvaje que se puede
5 generar mediante este procedimiento de diseño. Después se analiza para evaluar las estructuras de los codones de parada potenciales, horquillas, y otras bucles o para secuencias problemáticas que se corrigen posteriormente antes de la síntesis. El plan de diseño de LTM completo se envía después al sintetizador de ADN, que realiza una síntesis automática de los
10 oligonucleótidos. De esta manera, los oligonucleótidos necesarios para crear las librerías se pueden generar rápidamente.

Una base de datos electrónica puede almacenar información de todas las secuencias de oligonucleótidos de LTM, detalles de las sustituciones de las CDR de la librería de scFv, y resultados de los ensayos de unión para un
15 antígeno diana. El archivo de los datos de los oligonucleótidos permite las iteraciones racionales de las estrategias del diseño, y reutilización de los oligonucleótidos de reactivos.

3. Estrategia de LTM global

20 LTM se emplea inicialmente para identificar aminoácidos específicos y propiedades químicas que son beneficiosas para propiedades de unión, neutralización y/o cualquier propiedad adicional deseada en el anticuerpo final. También rápidamente identifica las regiones que no toleran ninguna mutagénesis sin la pérdida significativa de afinidad (o cualquier propiedad física
25 que se selecciona). Por lo tanto, no solamente es el análisis de LTM una buena

metodología para explorar los requerimientos químicos en cada posición en todas las CDR de un anticuerpo, pero también determina rápidamente los aminoácidos absolutamente requeridos para la unión a anticuerpo. Después de la identificación de las mutaciones beneficiosas mediante LTM, los esquemas de mutagénesis combinatoria se pueden emplear para que en primer lugar incorporen todas estas mutaciones de aminoácidos distintas para generar CDR múltiplemente mutadas. Adicionalmente o como alternativa, la mutagénesis en espera (WTM) se puede usar para sondear los efectos de mutaciones múltiples del mismo aminoácido en una CDR (como se describe, por ejemplo, en las patentes de Estados Unidos números 5.830.650; 5.798.208).

4. Librerías LTM scFv

Las técnicas de LTM descritas anteriormente se usan para crear combinaciones de oligonucleótidos con mutaciones en una CDR individual de la cadena ligera o pesada. Estos oligonucleótidos se sintetizan para que incluyan algunos de los marcos conservados circundantes para facilitar la superposición e hibridación durante PCR. Estas combinaciones de oligonucleótidos se utilizan para generar todas las posibles cadenas V_L y V_H en las que existen mutaciones en CDR individuales, dobles, y triples (combinaciones individuales, dobles, y triples de CDR1, 2, y 3) usando PCR de extensión de superposición individual (SOE - PCR) (como se describe en Horton y col., (1989) *Gene* 77: 61 - 68). SOE - PCR es un procedimiento rápido y sencillo para combinar los fragmentos de ADN que no requieren sitios de restricción, endonucleasas de restricción, o ligasas de ADN. En dos regiones SOE - PCR en el gen se amplifican primero mediante PCR usando cebadores

diseñados de manera que los productos de PCR compartan una secuencia complementaria en un extremo. Bajo estas condiciones de PCR, las secuencias complementarias se hibridizan, formando un solapamiento. Las secuencias complementarias actúan después como cebadores, permitiendo la
5 extensión mediante la ADN polimerasa para producir una molécula recombinante.

Por ejemplo, para crear la combinación de las cadenas V_H en las que tanto CDR - H1 como CDR - H2 están mutadas y la CDR - H3 es de tipo salvaje, que se llama "110" (1 significa una CDR mutante y 00 significa una
10 CDR de tipo salvaje), los genes mutantes de CDR - H1 se usan como moldes y SOE - PCR se lleva a cabo para unir los oligonucleótidos de CDR - H2 para generar la combinación doblemente mutada (figura 21). Considerando que cada CDR puede ser o bien de tipo salvaje o mutante, existen siete posibles combinaciones (mostradas mediante flechas en la figura 21) para cada una de
15 las combinaciones de las cadenas V_L y V_H (no incluyendo la molécula de tipo salvaje "000"). La combinación de las siete combinaciones V_L y V_H crea 63 combinaciones $V_L - V_H$ de tipo no salvaje (scFv), figura 21). Cada una de las 64 combinaciones V_L y V_H (incluyendo la secuencia de tipo salvaje) se denomina un "subconjunto" del conjunto de librería de LTM scFv entera. Un conjunto de la
20 librería de scFv se crea para cada aminoácido seleccionado para sustitución. El número de secuencias de aminoácidos representadas dentro de cada librería subconjunto depende de la longitud de la CDR, la secuencia de aminoácidos dentro de la CDR, y la estrategia de diseño de oligonucleótidos de LTM.

25 5. Selección de librerías de revisión y selección de anticuerpos

mejorados

Están disponibles una diversidad de procedimientos para la expresión y representación de anticuerpos. Estos incluyen bacteriófago, *Escherichia coli*, y levadura. Aunque cada uno de estos procedimientos se ha usado para mejora de anticuerpos, el sistema de representación de levaduras produce varias ventajas (Poder y Wittrup (1997) *Nat. Biotechnol.* 15: 553 - 557). La levadura se puede acomodar fácilmente tamaños de librerías de hasta 10^7 , con 10^3 - 10^5 copias de cada anticuerpo mostrándose en cada superficie celular. Las células de levaduras se pueden seleccionar fácilmente y separarse usando citometría de flujo y separación de células activadas por fluorescencia (FACS) o perlas magnéticas. Las levaduras también producen rápida selección y re - crecimiento. El sistema de secreción eucariótica y rutas de glucosilación de levadura permiten un subconjunto mucho mayor de moléculas scFv para plegarse correctamente y mostrarse sobre la superficie celular que los sistemas de representación procarionóticos. La representación de levaduras acoplada a la evolución directa se ha usado para incrementar la K_D de un fragmento anticuerpo scFv para fluoresceína a 48 fM, dos órdenes de magnitud más fuerte de unión que cualquier ligando monovalente reseñado previamente (Border y col., (2000) *PNAS* 97: 10701 - 10705). El sistema de representación utiliza el receptor de adhesión de levaduras de aglutinina a para representar proteínas sobre la superficie celular. Las proteínas de interés, en el caso presente, librerías de anti - BoNT/B scFv LTM o librerías de anti - BoNT/A scFv LTM, se expresan como parejas de fusión con la proteína Aga2. Estas proteínas de fusión se seleccionan a partir de la célula y se llegan a unir por disulfuro a la proteína Aga1, que se une a la pared de las células de levaduras

(véase Invitrogen, bibliografía del producto de representación de levaduras pYD1). Además, existen marcas terminales carboxilo incluido que se pueden utilizar para controlar los niveles de expresión y/o normalizar mediciones de afinidad de unión.

5 Las perlas magnéticas revestidas con estreptavidina (Spherotech) se usan para rastrear y seleccionar anticuerpos que se unen a BoNT/B o BoNT/A con alta afinidad. Esta metodología emplea unión de anticuerpos de alta afinidad a antígeno biotinilado que se une después a perlas revestidas con estreptavidina con el fin de seleccionar los clones de levaduras (Yeung y
10 Wittrup (2002) *Biotechnol. Prog* 18: 212 - 220) y Feldhaus y col., (2003) *Nature Biotech.* 21: 163 - 170). El polipéptido BoNT/A o BoNT/A (Metabionics) se biotinila usando protocolos convencionales (Pierce) y la selección se realiza usando procedimientos bien conocidos en la técnica. Usando selección basada en equilibrio y cinética, los anticuerpos con afinidades mejoradas se
15 seleccionan a partir de estas librerías. La eficacia de cada ronda de selección se controla mediante FACS analítica (FACSscan). Además, las afinidades de unión relativa de las moléculas individuales representadas sobre la superficie de levaduras se miden mediante valoración del antígeno. Esto permite una rápida identificación de moléculas con afinidad mejorada. Los clones de scFv
20 se secuencian después para identificar mutaciones beneficiosas.

6. Generación de anticuerpos solubles para mediciones de afinidad

BIAcore

Los anticuerpos de interés se subclonan en sistemas de expresión
25 solubles (*Picchia pastoris* y/o *E. coli*) y se genera proteína soluble. Existen

vectores comercialmente disponibles y las líneas celulares para expresión de anticuerpos solubles, que incluyen los de Invitrogen (por ejemplo, pPIC9 para *P. pastoris*) y Novagen (pET20b para expresión periplásmica en *E. coli*). Estos sistemas se usan de manera rutinaria para generar anticuerpos de cadena individual o de longitud completa solubles. El sistema de expresión de *P. pastoris* (Invitrogen) produce de manera rutinaria 1 - 5 mg por litro de scFv purificado soluble. La purificación d proteínas está facilitada por la presencia de His - tag en el extremo C de la molécula, en el caso de cadenas individuales o mediante columnas de proteina A o proteina G para anticuerpos de longitud completa. Los anticuerpos de cadena individual o de longitud completa solubles se generan para obtener mediciones de velocidad cinética de afinidad BIAcore. Esta etapa es necesaria como moléculas de scFv de alta afinidad sobre la superficie celular de clones de levaduras se debe verificar como una molécula soluble sin células.

Los anticuerpos solubles de cadena individual y de longitud completa generados de la manera anterior se puede usar para obtener mediciones de afinidad BIAcore, así como en el ensato de crecimiento hacia fuera de neuritas o el ensayo de letalidad de ratón descrito más adelante.

7. Selección de anticuerpos seleccionados para neutralización *in vivo* o *in vitro*

Un ensayo basado en células que usa el crecimiento hacia fuera de neuritas de neuronas primaras de pollo como un indicador de intoxicación d BoNT y, de esta manera, como medio de cuantificar la neutralización de toxinas se puede usar para seleccionar los anticuerpos seleccionados. Los experimentos preliminares que usan cultivos de explantes de ganglios de la

raíz dorsal de huevos de pollos fertilizados han indicado que existía menos crecimiento hacia fuera axonal de plantas tratadas con BoNT/A que los controles. Como alternativa, se puede usar un ensayo de articulación neuromuscular de músculo de gaglio - iris ciliar de pollo (como se describe en 5 Lomneth y col., (1990) *Neuroscience Letters* 113: 211 - 216).

El ensayo de letalidad de ratón (MLA, como describen, por ejemplo, Schantz y Kautter (1978) *J. Assoc. Off. Anal. Chem.* 61: 98 - 99) es otro procedimiento bien conocido y aceptado *in vivo* para ensayar la neutralización de BoNT. El ensayo implica la inyección interperitoneal de aproximadamente 10 0,5 ml de preparaciones de muestra de BoNT/B o BoNT/A con y sin el anticuerpo en 20 a 30 gramos, ratones blancos de la raza ICR. La mortandad debida a la insuficiencia respiratoria se observa durante 1 - 4 días. La cuantificación de la neutralización requiere diluciones en serie de los MABs con niveles variables de DL₅₀ de BoNT/B o BoNT/A de ratón. La MLA se puede 15 usar para determinar la capacidad de neutralización de los anticuerpos optimizados identificados en la selección *in vitro*.

La capacidad de neutralización de anticuerpos también se puede medir *in vitro* mediante el ensayo de protección de ratón (MPA, Goeschel y col., (1997) *Exp. Neurol.* 147: 98 - 102). En la MPA el nervio frénico izquierdo, junto 20 con el hemidiagrama izquierdo, se escinde del ratón.

El nervio frénico se electroestimula continuamente en un baño de tejido. Los anticuerpos purificados se incuban con BoNT/B o BoNT/A y se añaden al baño de tejido. La parálisis inducida por toxina se define como un 50% de reducción en el temblor músculo inicial.

Equivalentes

Los expertos en la técnica reconocerán, o serán capaces de averiguar usando nada más que experimentación rutinaria, muchos equivalentes a las
5 realizaciones específicas de la invención descritas en esta memoria descriptiva. Tales equivalentes de pretende que están abarcadas por las siguientes reivindicaciones.

Habiendo descrito al invención como antecede, se declara como propiedad lo contenido en las siguientes

REIVINDICACIONES

5

1. Un procedimiento de generación de una librería de análogos d polipéptidos en los que un aminoácido predeterminado aparece en cada posición en una región definida del polipéptido que comprende: seleccionar una región definida de la secuencia de aminoácidos del polipéptido; determinar un resto aminoácido a sustituir en cada posición de aminoácido dentro de la región definida; sintetizar polinucleótidos individuales que codifican la región definida, representando los polinucleótidos colectivamente posibles polinucleótidos variantes de acuerdo con los siguientes criterios:

10

15

i) contener cada polinucleótido en cada posición de codón en la región definida, o bien un codón requerido para el resto aminoácido del polipéptido o un codón para el resto aminoácido predeterminado, y

20

ii) contener cada polinucleótido no más de un codón para el resto aminoácido predeterminado

por lo tanto generando una librería de polinucleótidos en la que el resto aminoácido predeterminado aparece en cada posición aminoácido dentro de la región definida.

25

2. El procedimiento de la reivindicación 1, en el que los polinucleótidos se combinan juntos.
3. El procedimiento de la reivindicación 1, en el que dos o más regiones definidas dentro del polipéptido se mutagenizan.
4. El procedimiento de la reivindicación 31, en el que el mismo aminoácido predeterminado se selecciona para sustitución dentro de cada una de las dos o más regiones definidas.
5. El procedimiento de la reivindicación 3, en el que diferentes aminoácidos predeterminados se seleccionan para sustitución dentro de cada una de las dos o más regiones definidas, respectivamente.
6. El procedimiento de una cualquiera de las reivindicaciones anteriores, en el que la región definida o regiones definidas comprende un dominio funcional del polipéptido.
7. El procedimiento de la reivindicación 6, en el que el dominio funcional se selecciona entre el grupo constituido por un sitio de unión de anticuerpo, una región de marco conservada de anticuerpo, una región efectora de anticuerpo, un sitio de unión de receptor, y un sitio catalítico.
8. El procedimiento de la reivindicación 7, en el que el sitio de unión de

anticuerpo o porción del mismo comprende un dominio CDR seleccionado entre el grupo constituido por CDR1, CDR2, CDR3, CDR4, CDR5, CDR6 y una combinación de las mismas.

- 5 9. El procedimiento de la reivindicación 7, en el que la región marco conservada de anticuerpo comprende un dominio seleccionado entre el grupo constituido por FR1, FR2, FR3, FR4 y una combinación de los mismos.
- 10 10. El procedimiento de la reivindicación 7, en el que la región efectora de anticuerpo comprende un dominio seleccionado entre el grupo constituido por un sitio de unión de complemento y una región de unión Fc.
- 15 11. El procedimiento de la reivindicación 1, en el que el resto aminoácido predeterminado se selecciona entre el grupo constituido por Ser, Thr, Asn, Gln, Cys, His, Asp, Lys, Arg, Ala, Gly, Ile, Leu, Met, Phe, Pro, Trp, y Val.
- 20 12. El procedimiento de una cualquiera de las reivindicaciones anteriores, en el que la región definida comprende al menos 3 a 40 aminoácidos.
13. El procedimiento de una cualquiera de las reivindicaciones anteriores, en el que los polinucleótidos se sintetizan como librería de expresión.

14. El procedimiento de una cualquiera de las reivindicaciones anteriores,
en el que los polinucleótidos se sintetizan usando medios enzimáticos.
- 5 15. El procedimiento de una cualquiera de las reivindicaciones anteriores,
en el que los polinucleótidos se sintetizan usando reacción en cadena
de la polimerasa.
- 10 16. El procedimiento de una cualquiera de las reivindicaciones anteriores,
en el que la librería de expresión se selecciona entre el grupo
constituido por una librería de representación de fagos, una librería de
representación de ribosomas/polisomas, una librería de
representación de levaduras, una librería de representación de
bacterias y una librería de representación ordenada.
- 15 17. Una librería de análogos de polipéptidos mediante el procedimiento
de la reivindicación 1.
- 20 18. Un procedimiento de identificación de un polipéptido que tiene una
estructura deseada o función que comprende:
seleccionar una región definida de la secuencia de aminoácidos del
polipéptido;
determinar un resto aminoácido a sustituir en cada posición de
aminoácido dentro de la región definida;
25 sintetizar polinucleótidos individuales que codifican la región definida,
-

representando los polinucleótidos colectivamente posibles polinucleótidos variantes de acuerdo con los siguientes criterios:

- i) contener cada polinucleótido en cada posición de codón en la región definida, o bien un codón requerido para el resto aminoácido del polipéptido o un codón para uno del resto aminoácido predeterminado, y
- ii) contener cada polinucleótido no más de un codón para el resto aminoácido predeterminado,

por lo tanto generando una librería de polinucleótidos que contiene los polinucleótidos;

expresar la librería de expresión para producir análogos de polipéptidos; y

seleccionar los análogos de polipéptidos para seleccionar un polipéptido que tenga una estructura o función deseada.

15

19. El procedimiento de la reivindicación 18, en el que el procedimiento comprende adicionalmente la etapa de identificar el polinucleótido que codifica el análogo de polipéptido seleccionado.

20 . 20. El procedimiento de la reivindicación 18, en el que la selección comprende,

poner en contacto un polipéptido con un sustrato diana, estando el polipéptido asociado al polinucleótido que codifica el polipéptido, comprendiendo el polinucleótido adicionalmente un resto detectable, de manera que se detecta un polipéptido variante capaz de unirse a un sustrato

25

diana y por lo tanto se identifica como codificado por el polinucleótido.

21. El procedimiento de la reivindicación 20, en el que el resto detectable se selecciona entre el grupo constituido por un resto fluorescente, un resto UV,
5 y un resto que absorbe la luz visible.

22. El procedimiento de la reivindicación 20, en el que el resto detectable se selecciona entre el grupo constituido por un resto de biotina, un resto GST, y un resto His tag.
10

23. El procedimiento de la reivindicación 20, en el que el polinucleótido está asociado con el análogo de polipéptido usando representación de ribosoma.

24. El procedimiento de la reivindicación 18, en el que dos o más regiones
15 definidas dentro del polipéptido se mutagenizan.

25. El procedimiento de la reivindicación 24, en el que el mismo aminoácido predeterminado se selecciona para sustitución dentro de cada una de las dos o más regiones definidas.
20

26. El procedimiento de la reivindicación 24, en el que los aminoácidos predeterminados diferentes se seleccionan para sustituciones dentro de cada una de las os o más regiones definidas.

25 27. El procedimiento de la reivindicación 18, en el que el polipéptido es un

anticuerpo de una sola cadena (sFVs)

28. El procedimiento de la reivindicación 18, en el que la región definida comprende un dominio funcional del polipéptido.

5

29. El procedimiento de la reivindicación 18, en el que la región definida comprende una CDR o porción de la misma seleccionada entre el grupo constituido por CDR1, CDR2, CDR3, CDR4, CDR5, CDR6 y una combinación de las mismas.

10

30. El procedimiento de la reivindicación 18, en el que la región definida es una región marco conservada de anticuerpo que comprende un dominio seleccionado entre el grupo constituido por FR1, FR2, FR3, FR4 y una combinación de los mismos.

15

31. El procedimiento de la reivindicación 18, en el que la región efectora de anticuerpo comprende un dominio seleccionado entre el grupo constituido por un sitio de unión de complemento y una región de unión Fc.

20

32. El procedimiento de la reivindicación 18, en el que el resto aminoácido predeterminado se selecciona entre el grupo constituido por Ser, Thr, Asn, Gln, Cys, His, Asp, Lys, Arg, Ala, Gly, Ile, Leu, Met, Phe, Pro, Trp, y Val.

33. Una librería de polinucleótidos que codifica análogos de polipéptidos que comprende una o más regiones definidas en el que un resto aminoácido

25

predeterminado está sustituido en cada posición aminoácido dentro de la región definida, representando los polinucleótidos colectivamente todas las posibles variantes de acuerdo con los siguientes criterios:

- 5 i) cada polinucleótido contiene en cada posición de codón en la región definida, o bien un codón requerido para el resto aminoácido del polipéptido o un codón para uno del resto aminoácido predeterminado, y
- ii) cada polinucleótido contiene no más de un codón para el resto aminoácido predeterminado.

10

34. La librería de la reivindicación 33, en el que dos o más regiones definidas dentro del polipéptido están mutagenizadas.

15 35. La librería de la reivindicación 33, en el que el mismo aminoácido predeterminado se selecciona para sustitución dentro de cada una de las dos regiones definidas.

20 36. La librería de la reivindicación 33, en el que los aminoácidos predeterminados diferentes se seleccionan para sustitución dentro de cada una de las dos regiones definidas.

37. La librería de la reivindicación 33, en el que la librería es una librería de expresión.

25 38. La librería de la reivindicación 33, en el que la librería de expresión se

selecciona entre el grupo constituido por una librería de representación de fagos, una librería de representación de ribosomas/polisomas, una librería de representación de levaduras, una librería de representación de bacterias y una librería de representación ordenada.

5

39. La librería de la reivindicación 33, en la que los polinucleótidos comprenden adicionalmente uno o más elementos reguladores transcripcionales.

10

40. La librería de la reivindicación 39, en la que los polinucleótidos, cuando se transcriben y traducen *in vitro*, están asociados con los polipéptidos codificados por los polinucleótidos correspondientes.

15

41. La librería de la reivindicación 40, en la que los polinucleótidos, están asociados con el polipéptido usando representación de ribosomas/polisomas.

42. La librería de la reivindicación 41, en la que los polinucleótidos comprenden ARN.

20

43. La librería de la reivindicación 42, en la que los polinucleótidos comprenden adicionalmente un resto detectable.

44. La librería de la reivindicación 43, en la el resto detectable comprende un resto fluorescente.

25

45. La librería de la reivindicación 33, en la que la librería comprende al menos 10^8 polinucleótidos diferentes.

46. La librería de la reivindicación 33, en la que la librería comprende al menos 45 - 10^{12} polinucleótidos diferentes.

47. La librería de la reivindicación 33, en la que el polipéptido codifica un polipéptido de unión.

48. La librería de la reivindicación 47, en la que el polipéptido de unión se selecciona entre el grupo constituido por una región variable de cadena pesada (V_H), una región variable de cadena ligera (V_L), y un anticuerpo de una sola cadena (sFv).

49. La librería de la reivindicación 33, en la que el polipéptido codifica una enzima.

50. La librería de la reivindicación 33, en la que el polipéptido codifica un inhibidor de enzima.

51. La librería de la reivindicación 33, en la que el polipéptido codifica un polipéptido catalítico.

52. La librería de la reivindicación 33, en la que la librería está inmovilizada sobre un soporte sólido.

53. La librería de la reivindicación 33, en la que el soporte sólido es un microchip.

54. La librería de la reivindicación 33, en la que la librería es una librería
5 ordenada.

55. Un microchip que comprende una serie de polinucleótidos inmovilizados de acuerdo con la librería de la reivindicación 33.

10 56. Un análogo de polipéptido identificado de acuerdo con la librería de la reivindicación 33, e el que el polipéptido se une a una molécula diana y comprende una región de unión seleccionada entre el grupo constituido por una región variable de cadena pesada (V_H), una región variable de cadena ligera (V_L), y un anticuerpo de una sola cadena (sFv).

15

57. La molécula de la reivindicación 56, en la que la molécula diana especificada es TNF α .

58. La molécula de la reivindicación 56, en la que el polipéptido tiene al
20 menos un 70% de identidad con la secuencia de aminoácidos de la SEQ ID N° 1, 2, 3, 4, 5 ó 6.

59. Un procedimiento de generación de un subconjunto de análogos de polipéptidos que tienen una estructura o función deseada que comprende:
25 seleccionar una región definida de la secuencia de aminoácidos del

polipéptido;

determinar un resto aminoácido a sustituir en cada posición de aminoácido dentro de la región definida;

sintetizar polinucleótidos individuales que codifican la región definida,

5 representando los polinucleótidos colectivamente posibles polinucleótidos variantes de acuerdo con los siguientes criterios:

i) contener cada polinucleótido en cada posición de codón en la región definida, o bien un codón requerido para la síntesis del resto aminoácido del polipéptido o un codón para el resto aminoácido predeterminado, y

10

ii) contener cada polinucleótido no más de un codón para el resto aminoácido predeterminado

por lo tanto generando una librería de expresión que contiene los polinucleótido;

15

exponer la librería de expresión a condiciones bajo las que la librería se expresa;

seleccionar la librería expresada para identificar un polipéptido que tiene una estructura o función deseada;

comparar la estructura o función del polipéptido cuando se compara con un criterio de control, en el que un polipéptido que corresponde o excede el criterio de control se clasifica como un efector y un polipéptido que incumple el criterio de control se clasifica como un no efector;

20

categorizar los efectores y no efectores en una base de datos; y

consultar la base de datos para determinar la secuencia de un subconjunto de polipéptidos a sintetizar.

25

60. El procedimiento de la reivindicación 59, en el que una o más de las etapas anteriores está computarizada.

5 61. El procedimiento de la reivindicación 59, en el que el criterio de control se selecciona entre el grupo constituido por afinidad de unión, estabilidad y función efectora.

62. El procedimiento de la reivindicación 59, en el que el criterio es
10 actividad catalítica sobre un sustrato especificado.

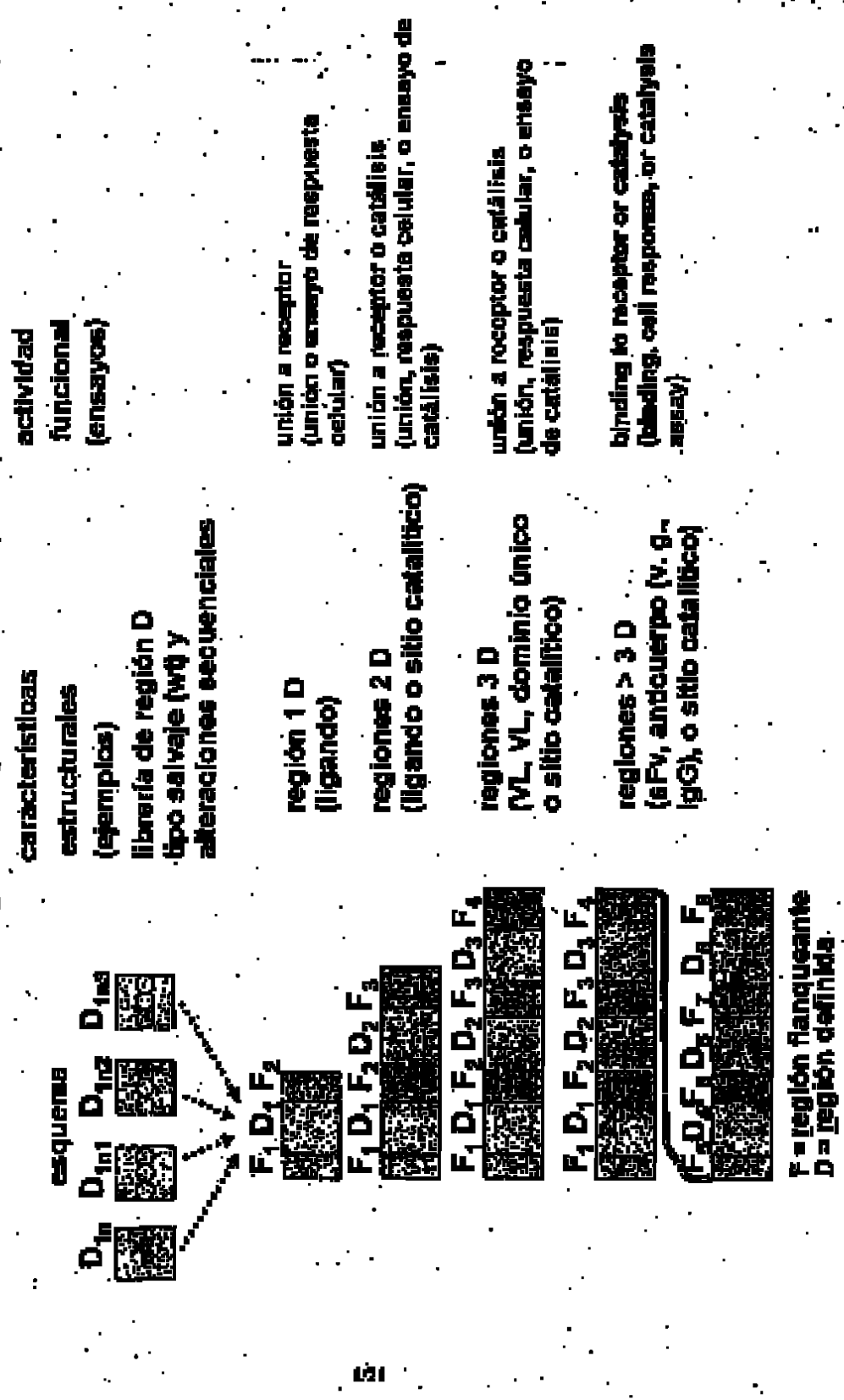
63. El procedimiento de la reivindicación 59, en el que el polipéptido se selecciona entre el grupo constituido por una región variable de cadena pesada (V_H), una región variable de cadena ligera (V_L), y un anticuerpo de una
15 sola cadena (scFv).

64. Un medio adecuado para uso en un dispositivo electrónico que tiene instrucciones para llevar a cabo una o más etapas del procedimiento de la reivindicación 59.

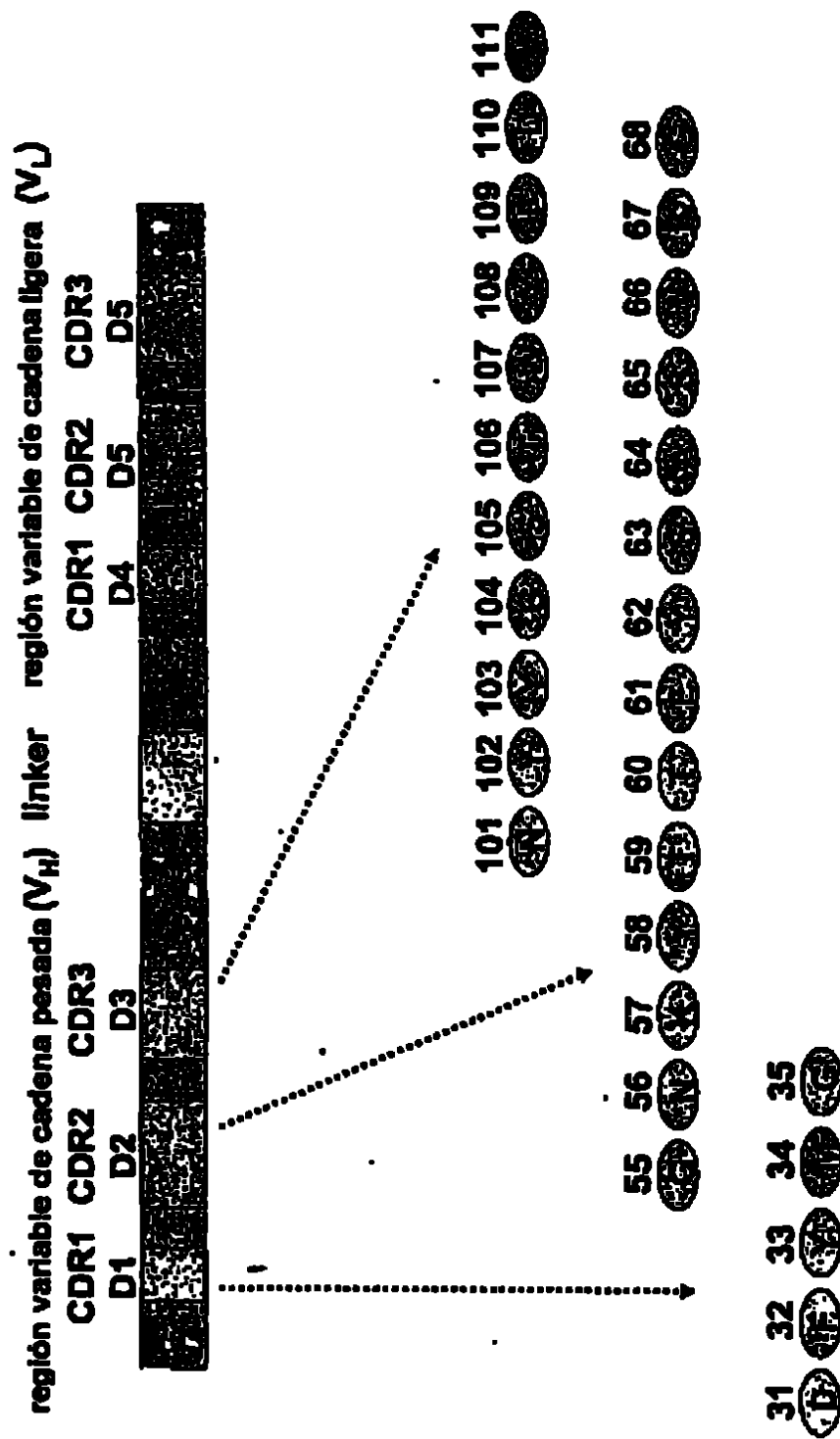
20

65. Un dispositivo para llevar a cabo una o más etapas del procedimiento de la reivindicación 59.

Formatos de librería de mutagénesis por revisión (LTM)



Identificación de 3 regiones definidas (D1 - D3) en un gen sintético (sFv) por mutagénesis de revisión



**LTM secuencial de una primera región definida (D1)
posiciones de aminoácidos 31-35 de V_HCDR1**

31	32	33	34	35

**LTM secuencial de una tercera región definida (D3)
(posiciones de aminoácidos 101-111 de V_H CDR3)**

[illegible]

Mutagénesis de espera (WTM)

V_H CDR1 Asp  **espera**

31 32 33 34 35

GAC	TTC	TAC	ATG	GAG	1 oligo
					rendimento de
GA	C	GAT	T	T ensayo	
1	4	2	7	2	2V = 128
(C)	(P)	(A)	(T)	(G)	oligo comb.

GA D CAT

1	4	7	$2^7 = 128$
1	4	7	

















9 22 2 11

112



secuencias

op

% de aminoácidos deseados (D)

% de aminoácidos no desejados

desodor

Mutagénesis de revisión (LTM)

V_H CDR1 Asp **revisión**

31 32 33 34 35

GAC TTC TAC ATG GAG

5 c/100

Ensayos

GAC GAC TAC ATG GAG GAG rendimento

Logo



combinaciones

5 secuencias

GAC TTC GAC ATG GAG

**% de aminoácidos
deseados** 100 100 100 100 100 100

% de aminoácidos	no	0	0	0	0
deseados					

प्लेसबो

Integración de LTM de regiones definidas D1 - D3 en un gen sintético (sFv)

región variable de cadena pesada (V_H) engarce región variable de cadena ligera (V_L)
CDR1 CDR2 CDR3 CDR1 CDR2 CDR3
D1 D2 D3 D4 D5 D6



Diagram showing a 4x4 grid of circles representing a sequence or structure.

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

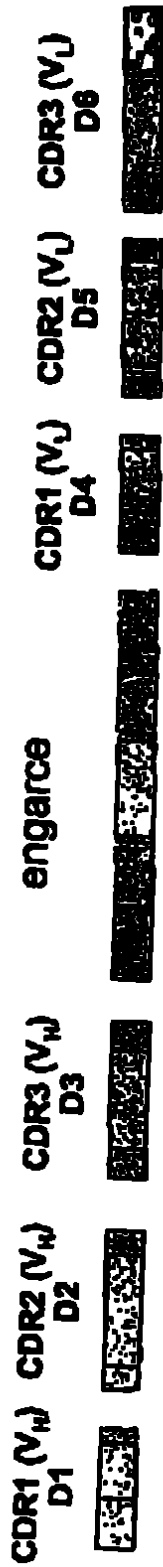
Diagram showing a 10x10 grid of circles representing a sequence or structure.

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

Diagram showing a 10x10 grid of circles representing a sequence or structure.

101	102	103	104	105	106	107	108	109	110	111
112	113	114	115	116	117	118	119	120	121	122
123	124	125	126	127	128	129	130	131	132	133
134	135	136	137	138	139	140	141	142	143	144
145	146	147	148	149	150	151	152	153	154	155
156	157	158	159	160	161	162	163	164	165	166
167	168	169	170	171	172	173	174	175	176	177
178	179	180	181	182	183	184	185	186	187	188
189	190	191	192	193	194	195	196	197	198	199
200	201	202	203	204	205	206	207	208	209	210

Ensamblaje general de regiones definidas
en un gen sintético (sFv)



conversión de nucleótidos de cadena sencilla a doble cadena



521



56

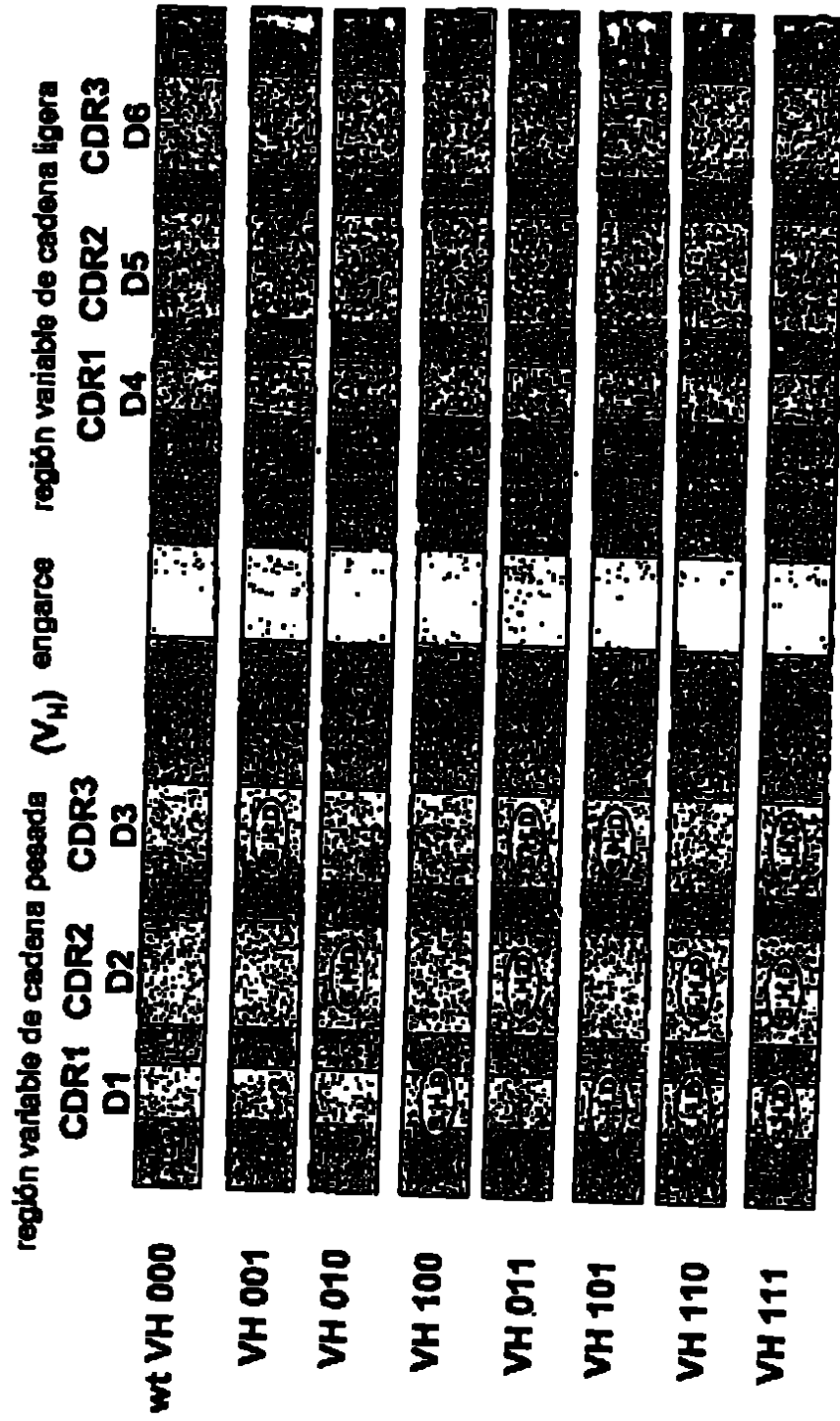
montaje w/ PCR / ligadura



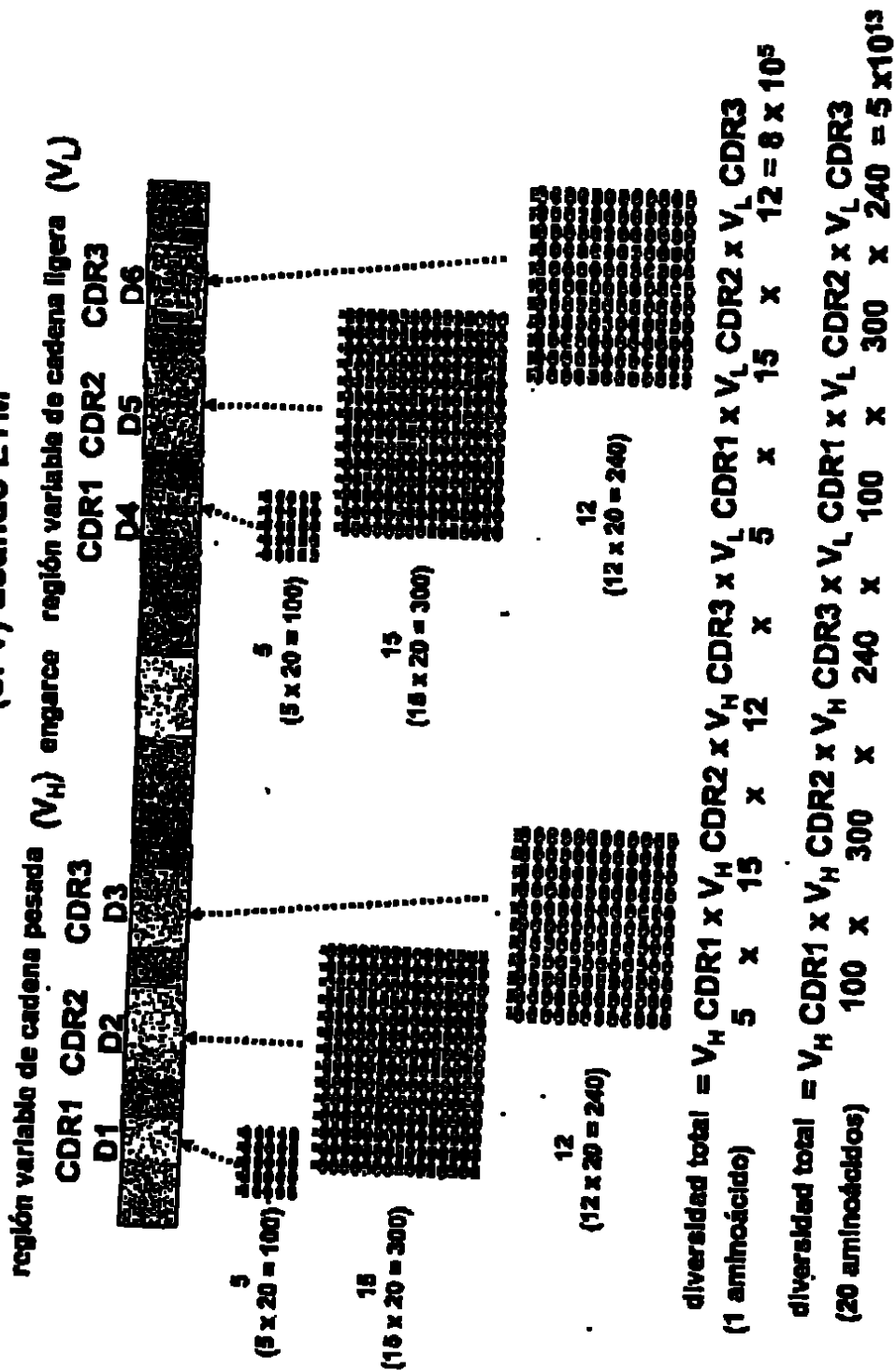
región variable de cadena pesada (V_H) engarce región variable de cadena ligera (V_L)



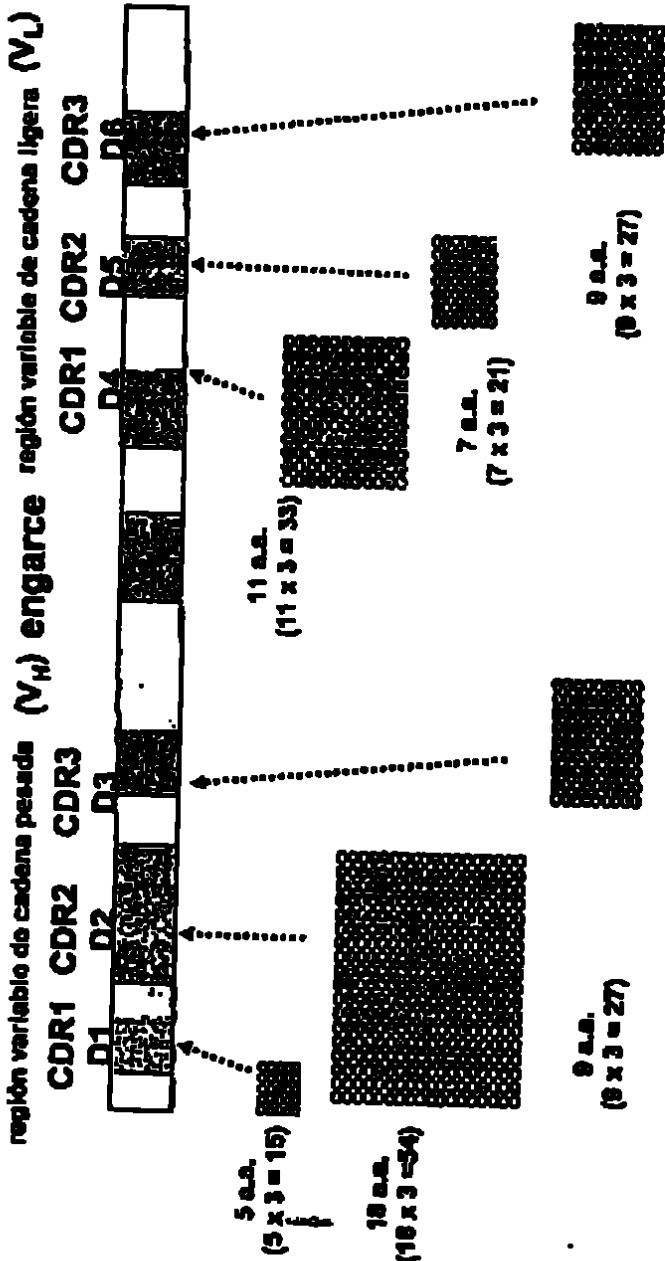
Formatos de diversidad ejemplar de regiones definidas D1-D3 en un gen sintético



Diversidad ejemplar de regiones definidas D1-D6 en un gen sintético (sFv) usando LTM



Diversidad ejemplar de regiones definidas D1-D6 en un gen sintético (a-TNF sFv) usando LTM



diversidad total = $V_H \text{ CDR1} \times V_H \text{ CDR2} \times V_H \text{ CDR3} \times V_L \text{ CDR1} \times V_L \text{ CDR2} \times V_L \text{ CDR3}$
 (1 aminoácido) $5 \times 15 \times 12 \times 5 \times 15 \times 12 = 8 \times 10^5$

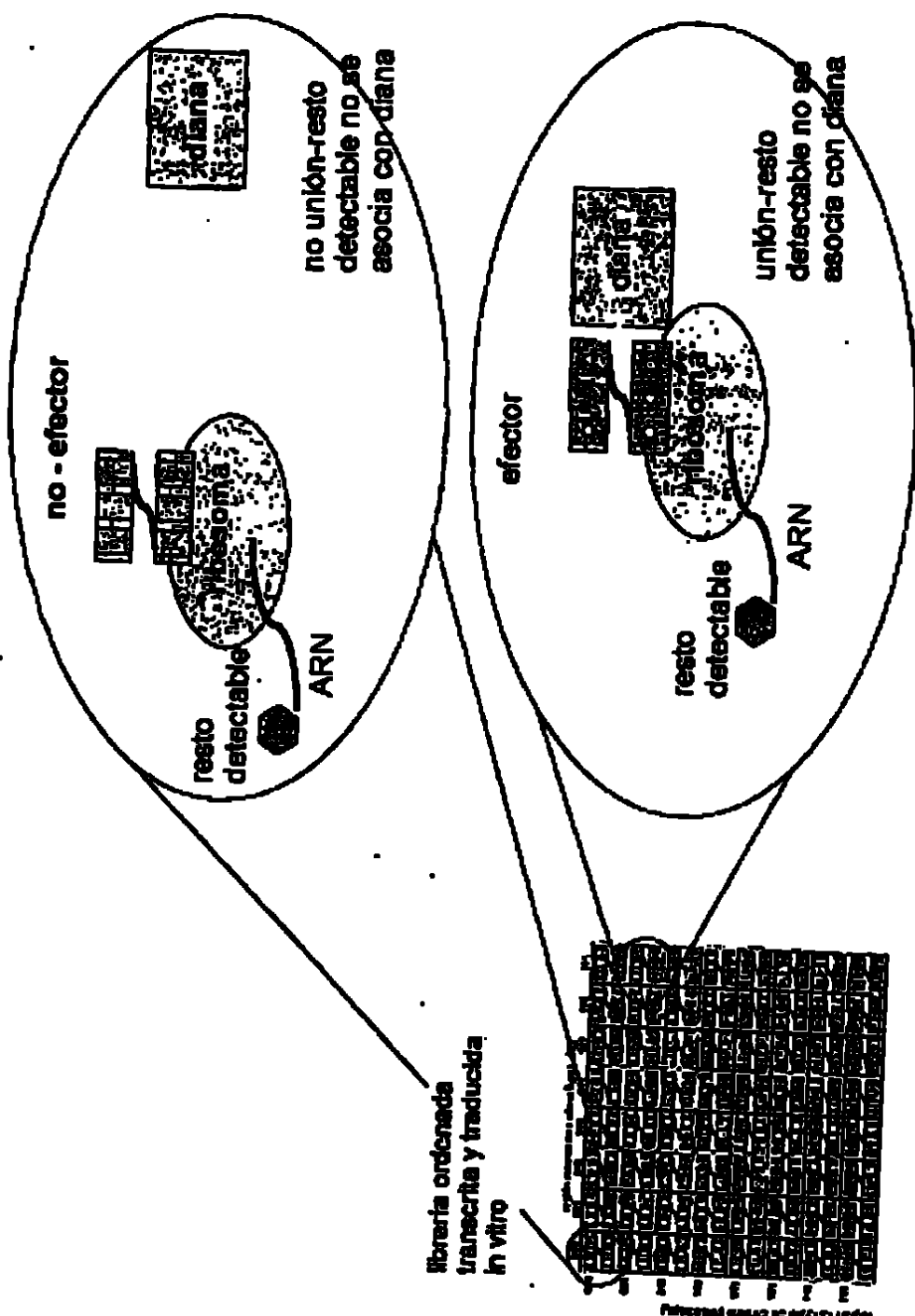
diversidad total = $V_H \text{ CDR1} \times V_H \text{ CDR2} \times V_H \text{ CDR3} \times V_L \text{ CDR1} \times V_L \text{ CDR2} \times V_L \text{ CDR3}$
 (3 aminoácidos) $15 \times 54 \times 27 \times 33 \times 21 \times 27 = 4.1 \times 10^8$

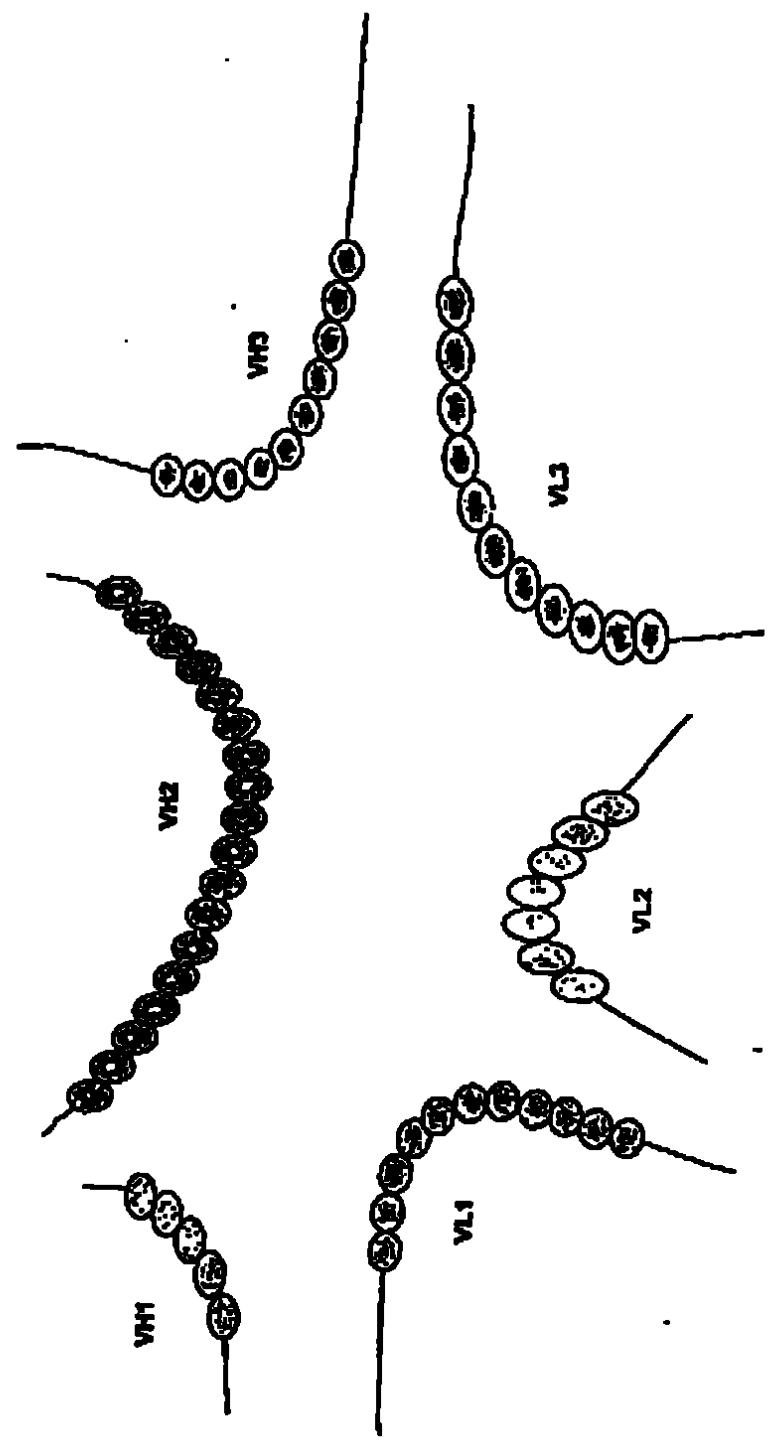
Asp, Ser, His

Formatos de diversidad ordenados de un gen sintético (sFv) usando LTM

región variable de cadena pesada (V _H)		región variable de cadena ligera (V _L)							
		000	001	010	100	011	101	110	111
000									
001									
010									
100									
011									
101									
110									
111									

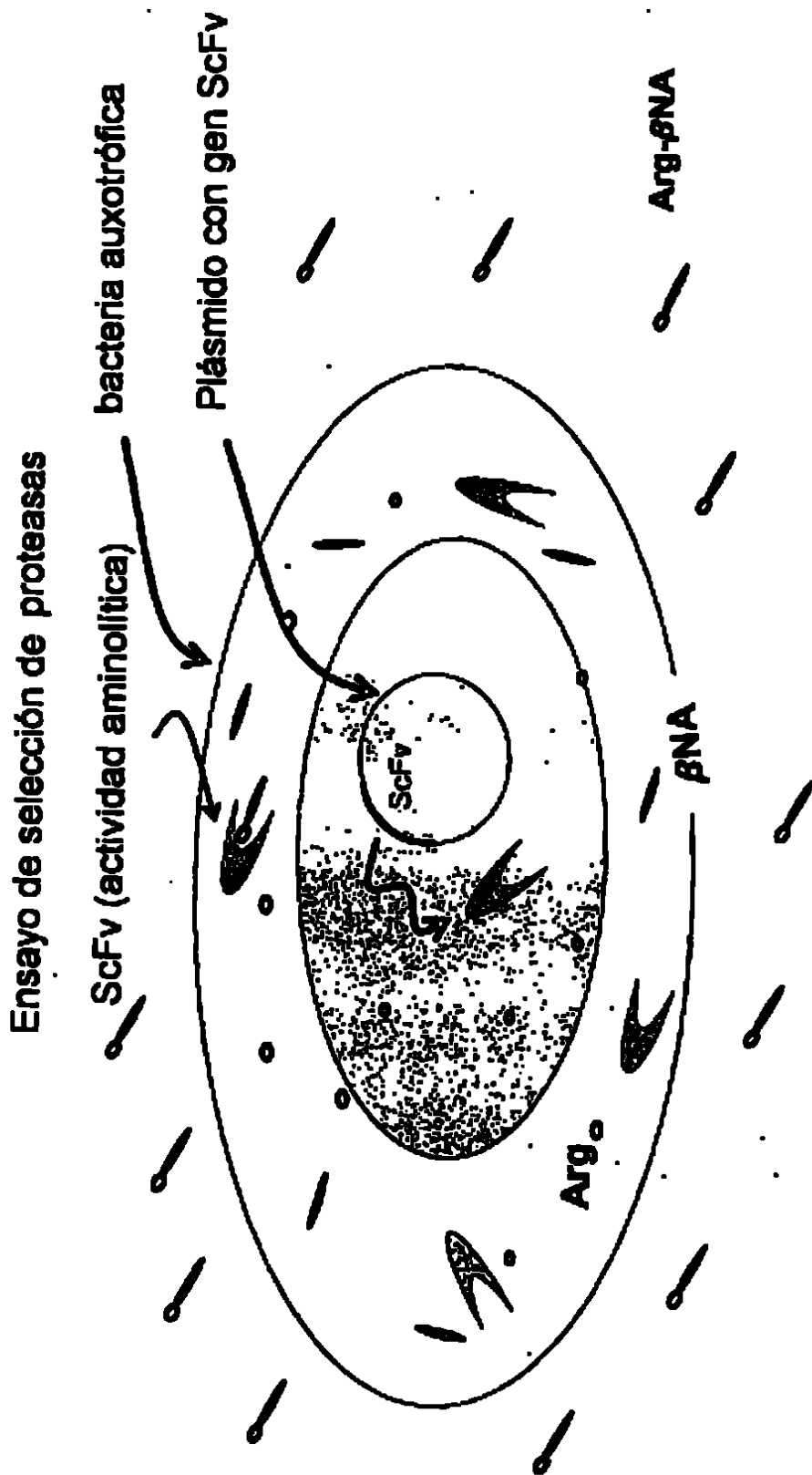
Sistema de detección universal para para seleccionar una librería ordenada



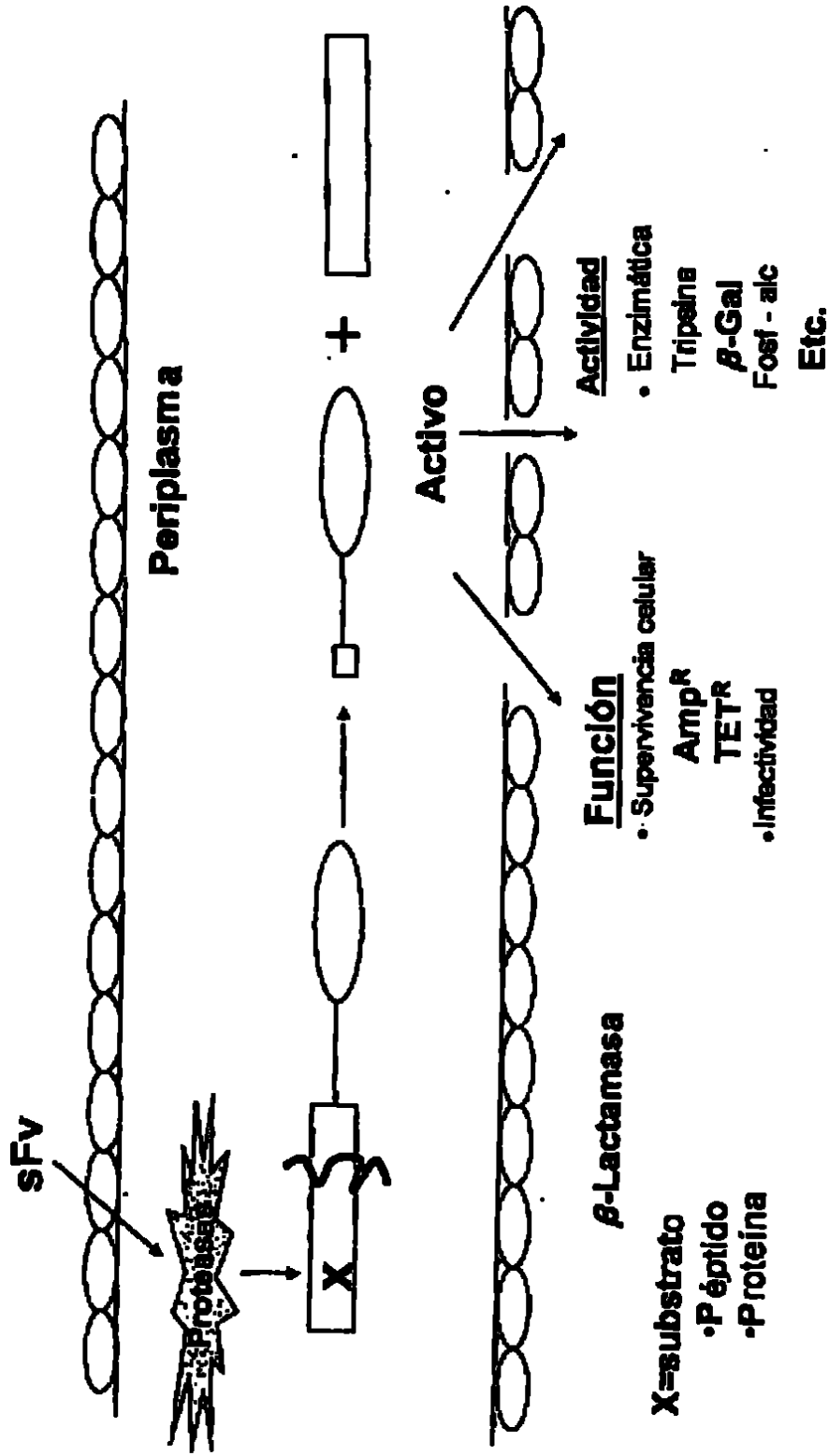


CA2

VH1: 5 A.A.	5 oligos	5 análogos de gen	VL1: 11 A.A.	11 oligos
VH2: 18 A.A.	18 oligos	18 análogos de gen	VL2: 7 A.A.	7 oligos
VH3: 9 A.A.	9 oligos	9 análogos de gen	VL3: 9 A.A.	9 oligos

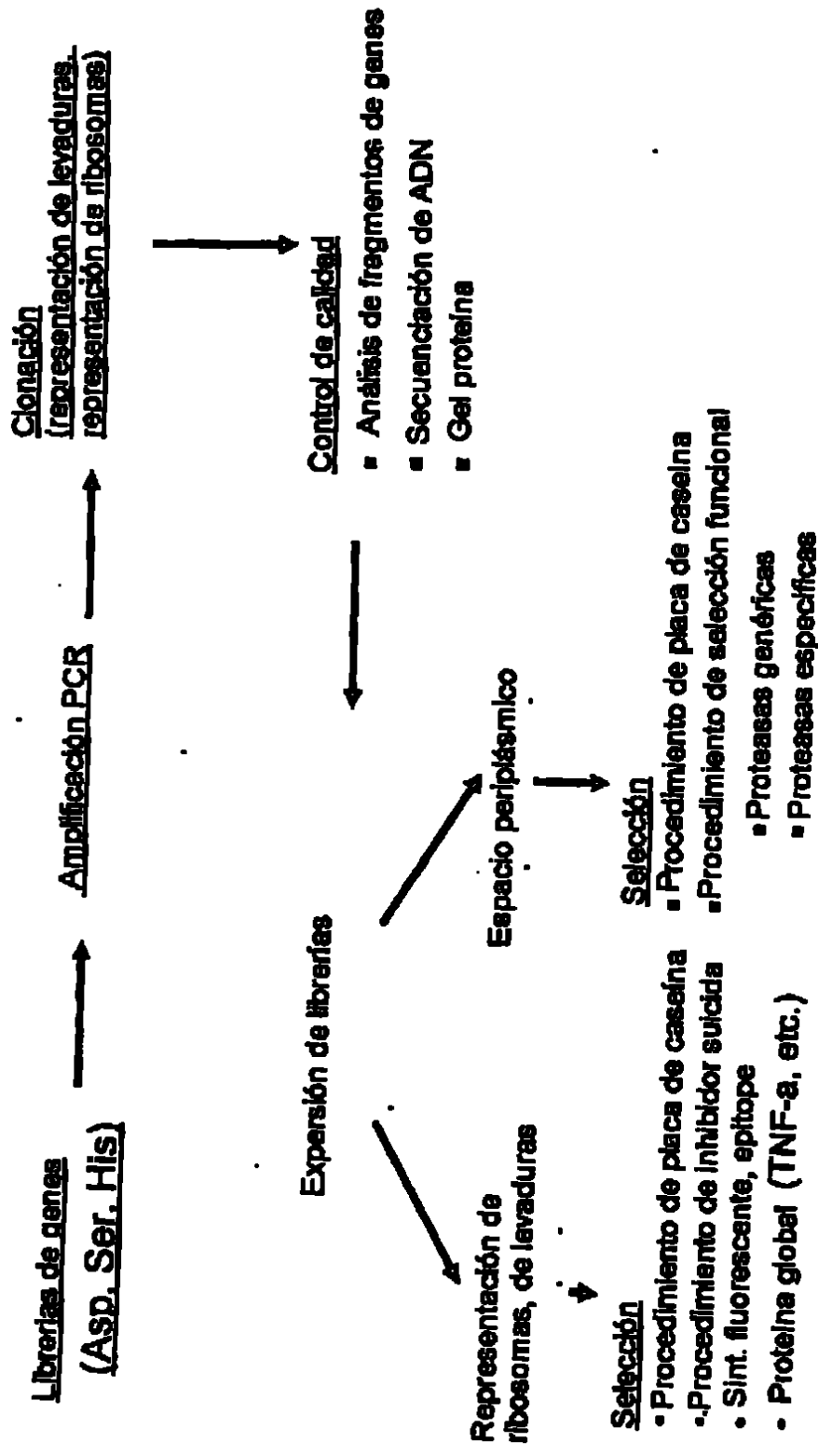


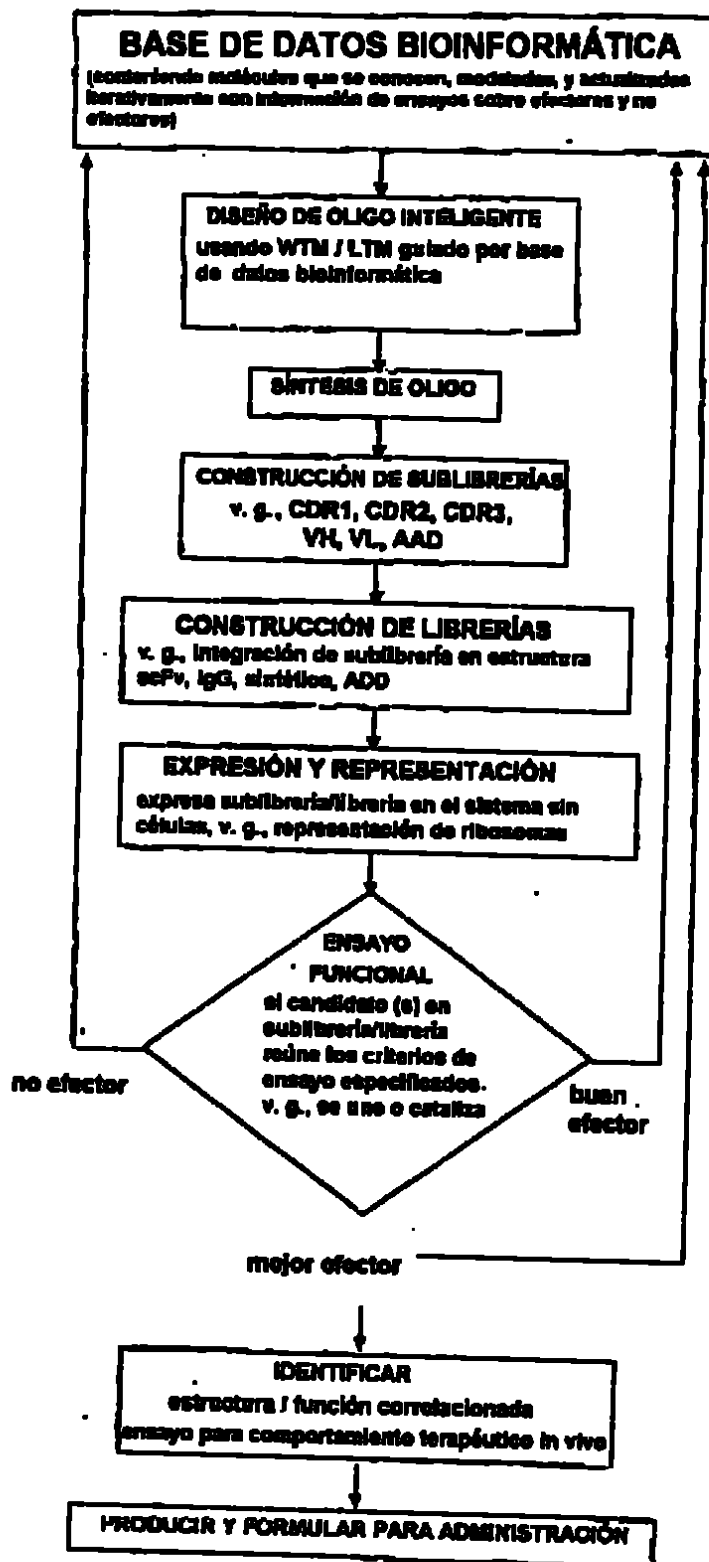
Espacio periplásmico bacteriano

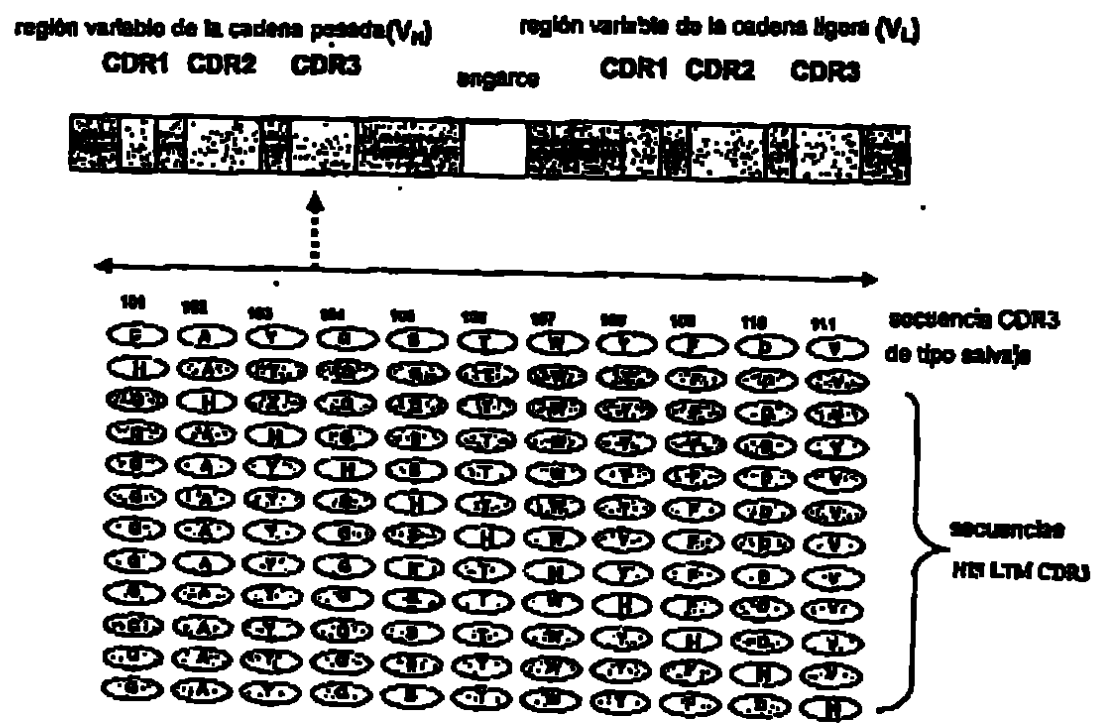


Librerías de genes
(Asp. Ser. His) → Amplificación PCR → Clonación
(representación de levaduras.
representación de ribosomas)

Sumario de procedimiento LTM

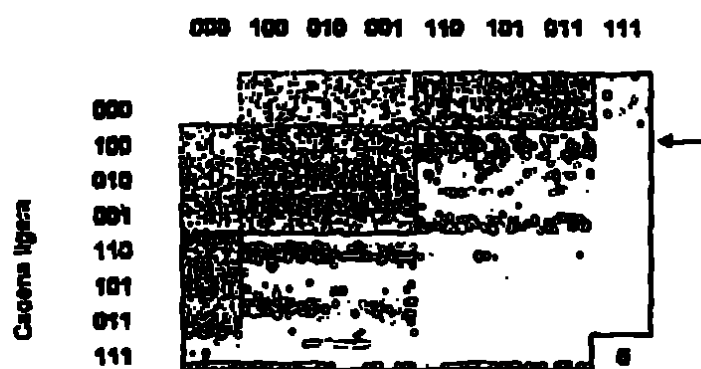


LSH
151



153
153

Cadena posada



LISTADO DE SECUENCIAS

<118> R. Cross & Co., et al.

<120> MUTAGÉNESIS POR REVISIÓN

<130> MRI-00100

<150> 60/463,362

<151> 2003-06-27

<160> 13

<170> FastSEQ para Windows Versión 4.0

<210> 1

<211> 131

<212> PRT

<213> Secuencia artificial

<220>

<221> Construcción artificial

<400> 1

```

Gln Val Gln Leu Val Gln Ser Gly Ala Gln Val Val Lys Pro Gly Ser
 1           5           10           15
Ser Val Lys Val Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Asp Tyr
 20           25           30
Asn Val Asp Trp Val Lys Gln Ala Pro Gly Gln Gly Leu Gln Trp Ile
 35           40           45
Gly Asn Ile Asn Pro Asn Asn Gly Gly Thr Ile Tyr Asn Gln Lys Phe
 50           55           60
Lys Gly Lys Gly Thr Leu Thr Val Asp Lys Ser Thr Ser Thr Ala Tyr
 65           70           75           80
Met Gln Leu Ser Ser Leu Thr Ser Gln Asp Thr Ala Val Tyr Tyr Cys
 85           90           95
Ala Arg Ser Ala Phe Tyr Asn Asn Tyr Gln Tyr Phe Asp Val Trp Gly
100           105           110
Gln Gly Thr Thr Val Thr Val Ser Ser
115           120

```

<210> 2

<211> 119

<212> PRT

<213> Secuencia artificial

<220>

<221> Construcción artificial

<400> 2

```

Gln Val Lys Leu Gln Gln Ser Gly Gly Gly Leu Val Gln Pro Gly Gly
 1           5           10           15
Ser Met Lys Leu Ser Cys Val Ala Ser Gly Phe Ile Phe Ser Asn His
 20           25           30
Trp Met Asn Trp Val Arg Gln Ser Pro Gln Lys Gly Leu Gln Trp Val
 35           40           45
Ala Gln Ile Arg Ser Lys Ser Ile Asn Ser Ala Thr His Tyr Ala Gln
 50           55           60

```

Ser Val Lys Gly Arg Phe Thr Ile Ser Arg Asp Asp Ser Lys Ser Ala
55 70 75 80
Val Tyr Leu Gln Met Thr Asp Leu Arg Thr Glu Asp Thr Gly Val Tyr
85 90 95
Tyr Cys Ser Arg Asn Tyr Tyr Gly Ser Thr Tyr Asp Tyr Trp Gly Gln
100 105 110
Gly Thr Thr Leu Thr Val Ser
115

<210> 3
<211> 114
<212> PRT
<213> Secuencia artificial

<220>
<223> Construcción sintética

<400> 3
Asp Ile Met Met Thr Gln Ser Pro Ser Thr Leu Ser Ala Ser Val Gly
1 5 10 15
Asp Arg Val Thr Ile Thr Cys Lys Ser Ser Gln Ser Leu Leu Tyr Ser
20 25 30
Asn Asn Gln Lys Asn Tyr Leu Ala Trp Tyr Gln Gln Lys Pro Gly Gln
35 40 45
Ala Pro Lys Leu Leu Ile Ser Trp Ala Ser Thr Arg Glu Ser Gly Val
50 55 60
Pro Ser Arg Phe Ile Gly Ser Gly Ser Gly Thr Glu Phe Thr Leu Thr
65 70 75 80
Ile Ser Ser Leu Gln Pro Asp Asp Val Ala Thr Tyr Tyr Cys Gln Gln
85 90 95
Tyr Tyr Asp Tyr Pro Trp Thr Phe Gly Gln Gly Thr Lys Val Gln Ile
100 105 110
Lys Arg

<210> 4
<211> 107
<212> PRT
<213> Secuencia artificial

<220>
<223> Construcción sintética

<400> 4
Asp Ile Leu Leu Thr Gln Ser Pro Ala Ile Leu Ser Val Ser Pro Gly
1 5 10 15
Glu Arg Val Ser Phe Ser Cys Arg Ala Ser Gln Phe Val Gly Ser Ser
20 25 30
Ile His Trp Tyr Gln Gln Arg Thr Asn Gly Ser Pro Arg Leu Leu Ile
35 40 45
Lys Tyr Ala Ser Gln Ser Met Ser Gly Ile Pro Ser Arg Phe Ser Gly
50 55 60
Ser Gly Ser Gly Thr Asp Phe Thr Leu Ser Ile Asn Thr Val Gln Ser
65 70 75 80
Glu Asp Ile Ala Asp Tyr Tyr Cys Gln Gln Ser His Ser Trp Pro Phe
85 90 95
Thr Phe Gly Ser Gly Thr Asn Leu Gln Val Lys
100 105

<210> 8
<211> 16
<212> PRT
<213> Secuencia artificial

<220>
<223> Construcción sintética

<400> 5
Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Ser
1 5 10 15

<210> 8
<211> 8
<212> PRT
<213> Secuencia artificial

<220>
<223> Construcción sintética

<400> 6
His His His His His His
1 5

<210> 7
<211> 10
<212> PRT
<213> Secuencia artificial

<220>
<223> Construcción sintética

<400> 7
Glu Glu Lys Leu Ile Ser Glu Glu Asp Leu
1 5 10

<210> 8
<211> 236
<212> PRT
<213> Secuencia artificial

<220>
<223> Construcción sintética

<400> 8
Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala
1 5 10 15
Ala Glu Pro Ala Met Ala Asp Ile Glu Met Thr Gln Ser Pro Ala Ser
20 25 30
Leu Ser Ala Ser Val Gly Glu Thr Val Thr Ile Thr Cys Arg Ala Ser
35 40 45
Gly Asn Ile His Asn Tyr Leu Ala Trp Tyr Glu Glu Lys Glu Gly Lys
50 55 60
Ser Pro Glu Leu Leu Val Tyr Asn Ala Lys Thr Leu Ala Asp Gly Val
65 70 75 80
Pro Ser Arg Phe Ser Gly Ser Gly Ser Gly Thr Glu Tyr Ser Leu Lys
85 90 95

Ile Asn Ser Leu Gln Pro Glu Asp Phe Gly Ser Tyr Tyr Cys Gln His
100 105 110
Phe Trp Ser Thr Pro Trp Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile
115 120 125
Lys Arg Ala Asp Ala Ala Pro Thr Val Ser Ile Phe Pro Pro Ser Ser
130 135 140
Glu Gln Leu Thr Ser Gly Gly Ala Ser Val Val Cys Phe Leu Asn Asn
145 150 155 160
Phe Tyr Pro Lys Asp Ile Asn Val Lys Trp Lys Ile Asp Gly Ser Glu
165 170 175
Arg Gln Asn Gly Val Leu Asn Ser Trp Thr Asp Gln Asp Ser Lys Asp
180 185 190
Ser Thr Tyr Ser Met Ser Ser Thr Leu Thr Leu Thr Lys Asp Glu Tyr
195 200 205
Glu Arg His Asn Ser Tyr Thr Cys Glu Ala Thr His Lys Thr Ser Thr
210 215 220
Ser Pro Ile Val Lys Ser Phe Asn Arg Asn Glu Cys
225 230 235

<210> 9
<211> 254
<212> PRT
<213> Secuencia artificial

<220>
<223> Construcción sintética

<400> 9
Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Ala
1 5 10 15
Ala Gln Pro Ala Ser Ala Gln Val Gln Leu Gln Gln Ser Gly Ala Glu
20 25 30
Leu Val Lys Pro Gly Ala Ser Val Lys Leu Ser Cys Thr Ala Ser Gly
35 40 45
Phe Asn Ile Lys Asp Thr Phe Met His Trp Val Lys Gln Arg Pro Glu
50 55 60
Gln Gly Leu Glu Trp Ile Gly Arg Ile Asp Pro Ala Asn Gly Asn Thr
65 70 75 80
Glu Tyr Asp Pro Lys Phe Gln Gly Lys Ala Thr Ile Thr Ala Asp Thr
85 90 95
Ser Ser Asn Thr Val Asn Leu Gln Leu Ser Ser Leu Thr Ser Glu Asp
100 105 110
Thr Ala Val Tyr Tyr Cys Ala Ser Gly Gly Glu Leu Gly Phe Pro Tyr
115 120 125
Trp Gly Gln Gly Thr Leu Val Thr Val Ser Ala Ala Lys Thr Thr Pro
130 135 140
Pro Ser Val Tyr Pro Leu Ala Pro Gly Ser Ala Ala Gln Thr Asn Ser
145 150 155 160
Met Val Thr Leu Gly Cys Leu Val Lys Gly Tyr Phe Pro Glu Pro Val
165 170 175
Thr Val Thr Trp Asn Ser Gly Ser Leu Ser Ser Gly Val His Thr Phe
180 185 190
Pro Ala Val Leu Gln Phe Asp Leu Tyr Thr Leu Ser Ser Ser Val Thr
195 200 205
Val Pro Ser Ser Thr Trp Pro Ser Glu Thr Val Thr Cys Asn Val Ala
210 215 220
His Pro Ala Ser Ser Thr Lys Val Asp Lys Lys Ile Val Pro Arg Asp
225 230 235 240
Cys Thr Ser Gly Gly Gly Gly Ser His His His His His
245 250

<210> 10
 <211> 236
 <212> PRT
 <213> Secuencia artificial

<220>
 <223> Construcción sintética

<400> 10
 Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala
 1 5 10 15
 Ala Gln Pro Ala Met Ala Asp Ile Gln Met Thr Gln Ser Pro Ala Ser
 20 25 30
 Leu Ser Ala Ser Val Gly Glu Thr Val Thr Ile Thr Cys Arg Ala Ser
 35 40 45
 Gly Asn Ile His Asn Tyr Leu Ala Trp Tyr Gln Gln Lys Gln Gly Lys
 50 55 60
 Ser Pro Gln Leu Leu Val Tyr Asn Ala Lys Thr Leu Ala Asp Gly Val
 65 70 75 80
 Pro Ser Arg Phe Ser Gly Ser Gly Ser Gly Thr Gln Tyr Ser Leu Lys
 85 90 95
 Ile Asn Ser Leu Gln Pro Glu Asp Phe Gly Ser Tyr Tyr Cys Gln His
 100 105 110
 Phe Trp Ser Thr Pro Trp Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile
 115 120 125
 Lys Arg Ala Asp Ala Ala Pro Thr Val Ser Ile Phe Pro Pro Ser Ser
 130 135 140
 Glu Gln Leu Thr Ser Gly Gly Ala Ser Val Val Cys Phe Leu Asn Asn
 145 150 155 160
 Phe Tyr Pro Lys Asp Ile Asn Val Lys Trp Lys Ile Asp Gly Ser Glu
 165 170 175
 Arg Gln Asn Gly Val Leu Asn Ser Trp Thr Asp Gln Asp Ser Lys Asp
 180 185 190
 Ser Thr Tyr Ser Met Ser Ser Thr Leu Thr Leu Thr Lys Asp Glu Tyr
 195 200 205
 Glu Arg His Asn Ser Tyr Thr Cys Glu Ala Thr His Lys Thr Ser Thr
 210 215 220
 Ser Pro Ile Val Lys Ser Phe Asn Arg Asn Gln Cys
 225 230 235

<210> 11
 <211> 254
 <212> PRT
 <213> Secuencia artificial

<220>
 <223> Construcción sintética

<400> 11
 Met Lys Tyr Leu Leu Pro Thr Ala Ala Val Gly Leu Leu Leu Leu Ala
 1 5 10 15
 Ala Gln Pro Ala Met Ala Glu Val Gln Leu Gln Gln Ser Gly Ala Glu
 20 25 30
 Leu Val Lys Pro Gly Ala Ser Val Lys Leu Ser Cys Thr Ala Ser Gly
 35 40 45
 Phe Asn Ile Lys Asp Thr Phe Met His Trp Val Lys Gln Arg Pro Glu
 50 55 60

Gln Gly Leu Glu Trp Ile Gly Arg Ile Asp Pro Ala Asn Gly Asn Thr
 68 70 75 80
 Glu Tyr Asp Pro Lys Phe Gln Gly Lys Ala Thr Ile Thr Ala Asp Thr
 85 90 95
 Ser Ser Asn Thr Val Asn Leu Gln Leu Ser Ser Leu Thr Ser Glu Asp
 100 105 110
 Thr Ala Val Tyr Tyr Cys Ala Ser Gly Gly Glu Leu Gly Phe Pro Tyr
 115 120 125
 Trp Gly Gln Gly Thr Leu Val Thr Val Ser Ala Ala Lys Thr Thr Pro
 130 135 140
 Pro Ser Val Tyr Pro Leu Ala Pro Gly Ser Ala Ala Gln Thr Asn Ser
 145 150 155 160
 Met Val Thr Leu Gly Cys Leu Val Lys Gly Tyr Phe Pro Glu Pro Val
 165 170 175
 Thr Val Thr Trp Asn Ser Gly Ser Leu Ser Ser Gly Val His Thr Phe
 180 185 190
 Pro Ala Val Leu Glu Ser Asp Leu Tyr Thr Leu Ser Ser Ser Val Thr
 195 200 205
 Val Pro Ser Ser Thr Trp Pro Ser Glu Thr Val Thr Cys Asn Val Ala
 210 215 220
 His Pro Ala Ser Ser Thr Lys Val Asp Lys Lys Ile Val Pro Arg Asp
 225 230 235 240
 Cys Thr Ser Gly Gly Gly Gly Ser His His His His His
 245 250

<210> 12

<211> 236

<212> PRT

<213> Secuencia artificial

<220>

<221> Construcción sintética

<400> 12

Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala
 1 5 10 15
 Ala Gln Pro Ala Met Ala Asp Ile Gln Met Thr Gln Ser Pro Ala Ser
 20 25 30
 Leu Ser Ala Ser Val Gly Glu Thr Val Thr Ile Thr Cys Arg Ala Ser
 35 40 45
 Gly Asn Ile His Asn Tyr Leu Ala Trp Tyr Gln Gln Lys Gln Gly Lys
 50 55 60
 Ser Pro Gln Leu Leu Val Tyr Asn Ala Lys Thr Leu Ala Asp Gly Val
 65 70 75 80
 Pro Ser Arg Phe Ser Gly Ser Gly Ser Gly Thr Gln Tyr Ser Leu Lys
 85 90 95
 Ile Asn Ser Leu Gln Pro Glu Asp Phe Gly Ser Tyr Tyr Cys Gln His
 100 105 110
 Phe Trp Ser Thr Pro Trp Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile
 115 120 125
 Lys Arg Ala Asp Ala Ala Pro Thr Val Ser Ile Phe Pro Pro Ser Ser
 130 135 140
 Glu Gln Leu Thr Ser Gly Gly Ala Ser Val Val Cys Phe Leu Asn Asn
 145 150 155 160
 Phe Tyr Pro Lys Asp Ile Asn Val Lys Trp Lys Ile Asp Gly Ser Glu
 165 170 175
 Arg Gln Asn Gly Val Leu Asn Ser Trp Thr Asp Gln Asp Ser Lys Asp
 180 185 190
 Ser Thr Tyr Ser Met Ser Ser Thr Leu Thr Leu Thr Lys Asp Glu Tyr
 195 200 205

Glu Arg His Asn Ser Tyr Thr Cys Glu Ala Thr His Lys Thr Ser Thr
210 215 220
Ser Pro Ile Val Lys Ser Phe Asn Arg Asn Glu Cys
225 230 235

<210> 13
<211> 254
<212> PRT
<213> Secuencia artificial

<400> 13
Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala
1 5 10 15
Ala Gln Pro Ala Met Ala Glu Val Gln Leu Gln Gln Ser Gly Ala Glu
20 25 30
Leu Val Lys Pro Gly Ala Ser Val Lys Leu Ser Cys Thr Ala Ser Gly
35 40 45
Phe Asn Ile Lys Asp Thr Phe Met His Trp Val Lys Gln Arg Pro Glu
50 55 60
Gln Gly Leu Glu Trp Ile Gly Arg Ile Asp Pro Ala Asn Gly Asn Thr
65 70 75 80
Glu Tyr Asp Pro Lys Phe Gln Gly Lys Ala Thr Ile Thr Ala Asp Thr
85 90 95
Ser Ser Asn Thr Val Asn Leu Gln Leu Ser Ser Leu Thr Ser Glu Asp
100 105 110
Thr Ala Val Tyr Tyr Cys Ala Ser Gly Gly Glu Leu Gly Phe Pro Tyr
115 120 125
Trp Gly Gln Gly Thr Leu Val Thr Val Ser Ala Ala Lys Thr Thr Pro
130 135 140
Pro Ser Val Tyr Pro Leu Ala Pro Gly Ser Ala Ala Gln Thr Asn Ser
145 150 155 160
Met Val Thr Leu Gly Cys Leu Val Lys Gly Tyr Phe Pro Glu Pro Val
165 170 175
Thr Val Thr Trp Asn Ser Gly Ser Leu Ser Ser Gly Val His Thr Phe
180 185 190
Pro Ala Val Leu Gln Ser Asp Leu Tyr Thr Leu Ser Ser Ser Val Thr
195 200 205
Val Pro Ser Ser Thr Trp Pro Ser Glu Thr Val Thr Cys Asn Val Ala
210 215 220
His Pro Ala Ser Ser Thr Lys Val Asp Lys Lys Ile Val Pro Arg Asp
225 230 235 240
Cys Thr Ser Gly Gly Gly Gly Ser His His His His His His
245 250

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau



INTERNATIONAL PATENT COOPERATION TREATY (PCT)

(43) International Publication Date
13 January 2005 (13.01.2005)

PCT

(10) International Publication Number
WO 2005/003345 A3

(51) International Patent Classification: C12N 15/10,
C07K 16/24

(31) International Application Number:
PCT/US2004/028906

(22) International Filing Date: 25 June 2004 (25.06.2004)

(23) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60483282 27 June 2003 (27.06.2003) US

(71) Applicant (for all designated States except US): R. CREA
& CO. [US/US]; 700 Occidental Avenue, San Mateo, CA
94402 (US).

(72) Inventor, and

(73) Inventor/Applicant (for US only): CREA, Roberto
[IT/US]; 700 Occidental Avenue, San Mateo, CA 94402
(US).

(74) Agents: RIEMLJARD, June, R. et al.; Lubrizo & Cock-
field, LLP, 28 State Street, Boston, MA 02109 (US).

(53) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,

AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN,
CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,
GB, GD, GE, GH, GM, GN, GU, HK, IL, IN, JP, KE,
KG, KP, KR, KZ, LC, LK, LS, LT, LU, LV, MA, MD,
MG, MK, MN, MW, MX, MY, NA, NI, NO, NZ, OM, PG,
PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TH,
TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM,
ZW.

(54) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,
ZW), European (AM, AZ, BY, KG, KZ, MD, RU, TL, TM),
European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,
FR, GB, GR, HU, IE, IT, LI, LU, MC, NL, PL, PT, RO, SE, SI,
SK, TR), OAPI (BF, BJ, CH, CI, CM, GA, GN, GQ,
GW, ML, MR, NI, SN, TD, TG).

Published:

- with international search report
- before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments

(55) Date of publication of the international search report:
24 March 2005

For two-letter codes and other abbreviations, refer to the "Guidance
Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

(54) Title: LOOK-THROUGH MUTAGENESIS

(57) Abstract: A method of mutagenesis by which a predetermined amino acid is introduced into each and every position of a
selected set of positions in a polypeptide region (or several different regions) of a polypeptide to produce a library of polypeptide
analogs. The method is based on the premise that certain amino acids play crucial role in the structure and function of proteins.
Libraries can be generated which contain only desired polypeptide analogs and are of reasonable size for screening. The libraries can
be used to study the role of specific amino acids in polypeptide structure and function and to develop new or improved polypeptides
such as antibodies, antibody fragments, single chain antibodies, enzymes, and ligands.

WO 2005/003345 A3

INTERNATIONAL SEARCH REPORT

Application No
Pv1/us2004/020306

A. CLASSIFICATION OF SUBJECT MATTER IPC 7 C12N15/10 C07K16/24		
According to International Patent Classification (IPC) or to both referred classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 7 C12N C07K		
Documentation searched other than minimum documentation to the extent and for documents are excluded in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used) EPO-Internal		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 01/44463 A (GENENTECH INC) 21 June 2001 (2001-06-21)	1-19, 24-39, 45-57, 59-65
Y	the whole document	20-23, 40-44
Y	HANES J & PLÜCKTHUN A: "In vitro selection and evolution of functional proteins by using ribosome display" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, NATIONAL ACADEMY OF SCIENCE, WASHINGTON, US, no. 94, 13 May 1997 (1997-05-13), pages 4937-4942, XP002077149 ISSN: 0027-8424 the whole document	20-23, 40-44
☐ Further documents are listed in the continuation of box C. ☒ Patent family members are listed in annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may raise doubts on priority claims or which is cited to establish the publication date of another citation for other special reason (as specified) "O" document relating to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is considered with one or more other documents, each document being obvious to a person skilled in the art. "Z" document member of the same patent family		
Date of the actual completion of the international search 21 January 2005		Date of mailing of the international search report 04/02/2005
Name and mailing address of the ISA European Patent Office, P.O. Box 2911, Patentstrasse 2 8001 Zürich, Switzerland Tel. (+41-76) 310 61 00, Fax (+41-76) 310 61 01		Authorized officer Schwachtgen, J-L

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2004/020305

Box II Observations where certain claims were found unsearchable (Continuation of Item 2 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(b) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☒ Claims Nos.: 58
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box III Observations where unity of invention is lacking (Continuation of Item 3 of first sheet)

This International Searching Authority found multiple inventions in this International application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without error justifying an additional fee, the Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

166

SUPERINDUSTRIA Y COMERCIO
Radicación : 05129963 00000000 Folios: 165
Fecha (AMD): 2005-12-25 16:40:34
Trámite : 011 PET-PATENT 379 FASE AECI 411 PATENTAC
Dependencia: 2020 DIVISION DE NUEVAS CREACIONES

REQUISITOS NECESARIOS PARA PRESENTACION Y AGILIZACION DEL TRAMITE

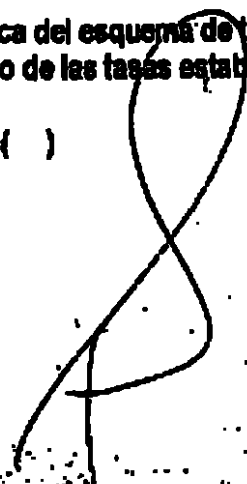
	USUARIO	RADICADOR
PATENTE DE INVENCION PCT MODELO DE UTILIDAD	()	()
- Indicación de que se solicita la concesión de una patente.	()	(x)
- Datos de identificación del solicitante o de la persona que se presenta la solicitud.	()	(x)
- Descripción de la invención.	()	(x)
- Dibujos de ser estos pertinentes.	()	(x)
- Comprobante de Pago de las tasas establecidas.		
COMPLETA (x)	INCOMPLETA ()	()

DISEÑO INDUSTRIAL ()		
- Indicación de que se solicita el registro de Diseño Industrial	()	()
- Datos de identificación del solicitante o de la persona que presenta la solicitud.	()	()
- Representación gráfica o fotográfica del Diseño Industrial o muestra del Material que incorpora el diseño.	()	()
- Comprobante de pago de las tasas establecidas	()	()
COMPLETA ()	INCOMPLETA ()	()

ESQUEMA DE TRAZADO ()		
- Indicación de que solicita el registro de un Esquema de trazado.	()	()
- Datos de identificación del solicitante o de la persona que presenta la solicitud.	()	()
- Representación gráfica del esquema de trazado	()	()
- Comprobante de pago de las tasas establecidas.	()	()
COMPLETA ()	INCOMPLETA ()	()

Firma Responsable:

Fecha:



167

REPUBLICA DE COLOMBIA
SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO

Bogotá,
2020

Doctor(a)
CLARTE GARCIA CARLOS REINALDO

REFERENCIA	Radicación No.	5 129963
	Solicitud internacional	US2004/020306
	Fecha inicio fase nacional	27/01/2006
	Trámite	11
	Evento	379
	Actuación	430
	Oficio No.	3819

Como resultado del examen de forma, realizado con fundamento en el artículo 38 de la Decisión 486 de la Comisión de la Comunidad Andina, artículo 27 del Tratado de Cooperación en Materia de Patentes (PCT), regla 51 bis del Reglamento y Circular Única de la Superintendencia de Industria y Comercio, se le comunica que:

1. Si fuere el caso, el solicitante debe adjuntar la traducción de los anexos del reporte de examen preliminar internacional (Art. 36.2.b) del tratado PCT.
2. La solicitud no contiene la copia del documento en que consta la cesión del derecho a la patente del inventor al solicitante o a su causante. (Art. 26-k) Decisión 486.
3. La solicitud no contiene el poder debidamente otorgado, con el lleno de los requisitos exigidos por los artículos 63 y subsiguientes del Código de Procedimiento Civil. Adicionalmente, deberá presentar el documento que acredita la existencia y representación legal de la sociedad solicitante, cuando ésta fuere persona jurídica.
4. Debe aclarar el nombre del solicitante, por cuanto la publicación internacional difiere del petitorio presentado en fase nacional.

En cumplimiento de lo dispuesto por el artículo 39 de la Decisión 486, se le requiero para que dentro del término de dos meses siguientes a la fecha de notificación del presente oficio, complete los requisitos anteriormente enunciados. En caso de no dar cumplimiento al presente requerimiento, la solicitud se considerará abandonada y perderá su prelación.

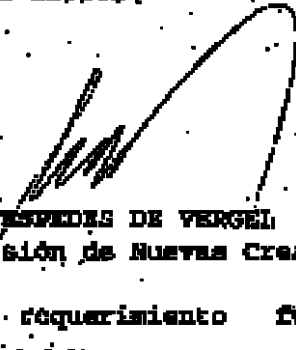
Notifíquese el presente oficio conforme a lo dispuesto en el numeral 5.2, literal e), del capítulo quinto del título primero de la Circular Única No.10 de 2001.

Continúa en la página siguiente...

168

REPÚBLICA DE COLOMBIA
SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO

Radiación No. 5 129963


ALIX CARMENZA CARREROS DE VERGEL
Jefe de la División de Nuevas Creaciones

El anterior requerimiento fue notificado por fijación en lista de
fecha 31 MAR. 2003
adpal