

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
2 March 2006 (02.03.2006)

PCT

(10) International Publication Number  
WO 2006/023144 A2

(51) International Patent Classification:  
C12P 21/06 (2006.01) C12N 15/74 (2006.01)

(21) International Application Number:  
PCT/US2005/024140

(22) International Filing Date: 6 July 2005 (06.07.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/585,918 6 July 2004 (06.07.2004) US

(71) Applicant (for all designated States except US): **BIOREN INC.** [US/US]; 100 Glenn Way, Suite #1, San Carlos, CA 94070-6264 (US).

(72) Inventor; and

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

(74) Agents: **REMILLARD, Jane, E.** et al.; Lahive & Cockfield, LLP, 28 State Street, Boston, MA 02109 (US).

(81) 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, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, 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, IS, IT, LT, LU, LV, 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, TG).

**Published:**

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

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



WO 2006/023144 A2

(54) Title: LOOK-THROUGH MUTAGENESIS FOR DEVELOPING ALTERED POLYPEPTIDES WITH ENHANCED PROPERTIES

(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 is disclosed. The method is based on the premise that certain amino acids play a crucial role in the structure and function of proteins and thus is capable of identifying and distinguishing functional amino acid residues ("hot spots") from non-functional amino acid residues ("cold spots") within a polypeptide or portion thereof. 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.

## LOOK-THROUGH MUTAGENESIS FOR DEVELOPING ALTERED POLYPEPTIDES WITH ENHANCED PROPERTIES

### *Related Information*

5           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*

10           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  
15           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  
20           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-  
25           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  
30           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  
35           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: 5,830,650; 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.

20

### ***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). Alone, this 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.

However, in a preferred embodiment, the LTM method is improved by using it to identify functional amino acids (or so-called "hot spots") from non-functional amino acids (or so-called "cold spots") within a polypeptide, or portion thereof, to further

reduce the number of residues to be altered in order to screen and obtain a desired property in a polypeptide. Accordingly, the improved method of look-through mutagenesis (LTM) (hereafter the improved LTM being referred to as LTM2) allows for the identification and building of a subset of candidate molecules representing only the most relevant functional alterations in the polypeptide which can then be efficiently screened free of any "noise". Importantly, LTM2 also allows for the construction of an LTM2 library having superior advantages over traditional libraries because it has been designed to include only alterations in the amino acid residues of the polypeptide most likely to have an effect on the function of the polypeptide and therefore, upon screening, more likely to yield an altered polypeptide having an enhanced property. Thus, LTM2 allows one to "look-through" the structural and functional consequences of separately substituting a predetermined amino acid at each functional amino acid position within a defined region of the polypeptide, thereby segregating a specific protein chemistry to the defined region without any interference or "noise" from the generation of unwanted polypeptide analogs (*i.e.*, analogs containing amino acid substitutions other than those that follow the LTM2 scheme) (see, for example, Fig. 1).

Accordingly, the present invention allows for highly efficient and accurate systematic evaluation of the role of a specific amino acid change in one or more defined regions of a polypeptide. This becomes particularly important when evaluating (by mutating) two or more defined regions, such that the number of polypeptide analogs required greatly increases and, thus, the presence of undesired analogs also increases. The present invention obviates this problem by completely eliminating undesired analogs and, thus, the potential that any changes in protein structure or function observed are the result of anything but substitution of the predetermined amino acid. Thus, the effect of segregating a specific protein chemistry to even multiple regions with a protein can be studied with high accuracy and efficiency. Importantly, this includes studying how mutagenesis can effect the interaction of such regions, thereby improving the overall structure and function of the protein.

In a particular embodiment of the invention, the methods of the invention are suitable for identifying a particular chemical motif that maps to one or more functional amino acid residues or positions. The amino acid residue(s) that contribute to such a chemical motif can occur at one or more positions that are contiguous, non-contiguous, within one or more CDR regions, and/or within one or more polypeptides, for example, antibody heavy and light chains. The methods of the invention allow for the further exploration of a chemical motif in that they allow for the systematic testing (or chemical profiling) of related amino acid chemistries at selected amino acid position(s) or defined region(s). Accordingly, in one embodiment, the invention provides a method for identifying a desired chemistry and then exploring the consequences of incorporating

related or unrelated chemistries to achieve either an enhanced property or to remove a deleterious property. Typical amino acid side chain chemistries suitable for profiling by the methods of the invention are polar, positively charged, negatively charged, and hydrophobic amino acid side chain chemistries. In one embodiment, a charged  
5 chemistry is identified as resident at a selected amino acid residue(s), position, or defined region(s) and other charged amino acids are substituted for the parental amino acid such that an alteration in a measurable property is achieved. In a preferred embodiment, the alteration in a measurable property is an enhanced property in an antibody, for example, improved antigen-binding or effector function.

10 Accordingly, the invention also provides antibody libraries comprising related amino acid side group chemistries introduced at selected amino acid positions(s) / defined regions having, for example, related chemistry, for the efficient screening of antibodies with improved properties.

In another embodiment of the invention, the library of polypeptide analogs is  
15 generated and screened by first synthesizing individual polynucleotides encoding a defined region or regions of a polypeptide where, collectively, the polynucleotides represent all possible variant polynucleotides according to the look-through criteria described herein. The method is used to identify and distinguish functional amino acid residue(s) (positions) from non-functional amino acid residue(s) (positions). A subset of  
20 variant polynucleotides are expressed, for example, using *in vitro* transcription and translation and/or using a display technology, such as ribosome display, phage display, bacterial display, yeast display, arrayed display, or any other suitable display system known in the art.

The expressed polypeptides are then screened and selected using functional  
25 assays, such as binding assays or enzymatic/catalytic assays. In one embodiment, the polypeptides are expressed in association with the polynucleotide that encodes the polypeptide, thereby allowing for identification of the polynucleotide sequence that encodes the polypeptide. In yet another embodiment, the polypeptides are directly synthesized using protein chemistry.

30 In yet another embodiment of the invention, a combinatorial beneficial library of the  $V_L$  and  $V_H$  CDR amino-acid sequence variations is constructed. This second library is constructed by generating coding sequences having, at each amino acid variation position, codons for the wildtype amino acid and for each of the previously identified beneficial variant amino acids at that position.

35 Thus, the present invention provides a method of intelligent mutagenesis that can be used to generate libraries of polypeptide analogs that are of a practical size for screening, in part, because the libraries are devoid of any undesired analog polypeptides or so-called noise. The method 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, binding fragments or analogs thereof, single chain antibodies, catalytic antibodies, enzymes, and ligands. In addition, the method can be performed with the benefit of *a priori* information, *e.g.*, *via* computer modeling, that can be used to select an  
5 initial subset of polypeptide analogs to be produced and studied using LTM2.

Other advantages and aspects of the present invention will be readily apparent from the following description and Examples.

### ***Brief Description of the Figures***

10 **Figure 1** illustrates the advantages of improved LTM (LTM2) over LTM in that functional amino acids are distinguished from non-functional amino acids such that a more beneficial subset of candidate molecules is obtained and screened.

**Figure 2** illustrates a general approach for the use of polymerase chain reaction (PCR) to build defined regions of an antibody heavy and light chain for identifying  
15 functional amino acid residues, into a larger gene context.

**Figure 3** illustrates the arrangement of variable light-chain ( $V_L$ ) and variable heavy chain ( $V_H$ ) CDRs in a synthetic single chain antibody (scFv) anti-ovalbumin gene context. In the application of LTM, a leucine amino acid is introduced into each of the fourteen residues 56-69 in  $V_H$  CDR2 of the antibody. For the application of LTM2, only  
20 those residues identified as functional are further explored by mutagenesis.

**Figure 4** illustrates the by single overlap extension polymerase chain reaction (SOE-PCR) for the production of an LTM  $V_H$  CDR2 library; the production of multiple LTM  $V_H$  CDR libraries; and an array of LTM library combinations containing both  $V_H$  and  $V_L$  CDRs

25 **Figure 5** illustrates the diversity of the libraries of the invention with the x and y axes of the matrix representing the CDRs of each of the light and heavy chains wherein an "0" indicates a wild-type CDR and a "1" indicates a mutated CDR and the intersected number representing the complexity of the resultant subset library (*e.g.*, 4 means four CDRs are simultaneously mutated).

30 **Figure 6** shows a schematic of a yeast expression vector for displaying proteins of interest, *e.g.*, polypeptide analogs of the invention, on the surface of yeast for efficient identification of function (phenotype) and corresponding encoding sequence (genotype).

**Figure 7** represents a Fluorescence-Activated Cell Sorter (FACS<sup>TM</sup>) plot of the binding of biotinylated ovalbumin and streptavidin FITC to wild type anti-ovalbumin  
35 scFv (gray line); pYD1 vector alone (solid gray area); and control scFv (black line).

**Figure 8** represents Fluorescence-Activated Cell Sorter (FACS<sup>TM</sup>) plots showing a selection gate (the R1 trapezoid) for identifying only those LTM clones that expressed the scFv fusion with a higher binding affinity to ovalbumin than the anti-ovalbumin wild

type antibody (left panel), the distribution of binding affinities of the total LTM library (center panel), and a post sort FACS analysis (right panel) to confirm that >80% of the pre-screen anti-ovalbumin scFv clones were within the predetermined criteria.

*Figure 9* illustrates steps in the screening of scFv antibodies (*e.g.*, anti-ovalbumin) formed in accordance with the present invention for improved binding affinity based on equilibrium binding kinetics (*e.g.*, to ovalbumin).

*Figure 10* shows equilibrium binding curves for anti-ovalbumin scFv expressing cells prior to selection (circles), after one round of selection (light triangles), after two rounds of selection (dark triangles), and for the anti-ovalbumin wild-type reference antibody (black squares).

*Figure 11* illustrates typical steps for screening of antibodies formed in accordance with the present invention for high binding affinity based on particular binding kinetics, *e.g.*, antibody  $K_{off}$  constants, using the test antigen ovalbumin.

*Figure 12* shows the identification of enhanced properties in two clones (*i.e.*, higher relative  $K_{off}$  as compared to a reference antibody (square)) using the methods of the invention.

*Figure 13* represents the enhanced properties (see fold better than wild type) of a subset of improved clones having lower  $EC_{50}$  values with respect to an anti-ovalbumin wild-type reference antibody control (square).

*Figure 14* shows a matrix representing the functional (hot spots) and non-functional (cold spots) amino acid positions of an exemplary antibody. Mutations associated with enhanced affinity (relative to the reference wild type antibody) based on equilibrium binding ( $EC_{50}$ ) and/or kinetic binding experiments are shown below each  $V_H$  and  $V_L$  CDR position.

### ***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, and peptides 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. In another embodiment, the binding molecule, in particular, binding molecules comprising CDR region(s), can comprise nontraditional scaffolds or framework regions

derived from other antibodies, immunoglobulins, or immunoglobulin-like molecules (e.g., fibronectin), or be in part or in whole, of synthetic origin.

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  
5 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), e.g., a single domain binding region, or a complete heavy and/or light chain variable region (Fv) of an antibody. Thus, a  
10 functional site may include a single or multiple defined regions that contribute to the functional activity of the molecule.

The terms “functional amino acid(s)” and “non-functional amino acid(s)” refer to, respectively, the amino acid residues (or corresponding amino acid residue position) within a polypeptide (or portion thereof) that are determined (using, for example, the  
15 methods of the invention) to contribute to a measurable property or activity of the polypeptide. Accordingly, a functional amino acid residue(s) (or corresponding position(s)) is referred to as a “hot spot(s)” as it is a residue or residue position that influences the activity of the polypeptide as compared to a non-functional residue(s) or position(s) which does not influence the activity of the polypeptide and therefore  
20 referred to as a “cold spot(s)”. A functional amino acid residue (or position) is distinguished from a non-functional amino acid residue (or position) as being suitable for mutagenesis. Typically, when applying the methods of the invention to the investigation of an antibody molecule, amino acid residues that alter, for example, antigen binding, are considered functional residues/positions (*i.e.*, hot spots) whereas  
25 residues that do not alter such binding are referred to as non-functional residues/positions (*i.e.*, cold spots).

The term “measurable property” refers to a functional property or activity of a polypeptide (or portion thereof) that can be measured, determined, or assayed for, using standard techniques and include, binding activity, kinase activity, catalytic activity,  
30 thermal stability, or enzymatic activity. Measurable properties of polypeptides that are antigen-binding polypeptides, e.g., antibodies, typically include binding specificity, binding avidity, binding affinity, Fc receptor binding, glycosylation, complement binding, half-life stability, solubility, thermal stability, catalytic activity, and enzymatic activity.

35 The term “look-through mutagenesis” or “LTM” refers to a method for 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 term "improved look-through mutagenesis" or "LTM2" refers to LTM conducted so as to identify or distinguish functional amino acid residues (hot spots) from non-functional amino acid residues (cold spots). Accordingly, the LTM2 method allows for selectively introducing a predetermined amino acid into the functional amino acid residue positions within a polypeptide (or portion thereof). Corresponding LTM2 libraries are therefore enriched for polypeptides analogs having amino acid alterations most likely to confer an altered or enhanced property. LTM2 can be carried out subsequent to LTM or based on a priori information as to the functionality of a given amino acid residue or residue position.

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. Libraries of the invention can contain 2 or more molecules or polypeptide analogs, for example about 2 to 10, about 10 to 50, about 50 to  $10^2$ , about  $10^3$ , about  $10^4$ , about  $10^5$ , about  $10^6$ , about  $10^7$ , about  $10^8$ , about  $10^9$ , about  $10^{10}$ , about  $10^{11}$ , about  $10^{12}$ , about  $10^{13}$ , or more, or any interval or range of the foregoing.

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 "mutagenesis" refers to, unless otherwise specified, any art recognized technique for altering a polynucleotide or polypeptide sequence. Preferred types of mutagenesis include walk-through mutagenesis (WTM), beneficial walk-through mutagenesis, look-through mutagenesis (LTM), improved look-through mutagenesis (LTM2), or combinations thereof.

The term "combinatorial beneficial mutagenesis" refers to a combination library of coding sequences that encode degenerate mixtures of  $V_L$  and/or  $V_H$  CDR amino-acid sequence variations initially identified from the predetermined LTM amino acid mutagenesis screen as having an alteration on a measurable property. In the combinatorial beneficial mutation approach, oligonucleotide coding sequences are generated which represent combinations of these beneficial mutations identified by LTM. These combinations may be combinations of different beneficial mutations within a single CDR, mutations within two or more CDRs within a single antibody chain, or mutations within the CDRs of different antibody chains.

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  
5 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-  
10 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.

15 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  
20 polypeptide analogs to form libraries representing the look-through mutagenesis (LTM) or improved look-through mutagenesis (LTM2) 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  
25 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  
30 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, and in a preferred embodiment, at amino acid  
35 positions determined to be functional (hot spots). 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). Use of non-traditional amino acid residues (*e.g.*, homocysteine) are also within the scope of the invention and can be  
5 introduced using any art recognized techniques.

### *Detailed Description*

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  
10 amino acids participate in the binding of an antibody to an antigen or are involved in the catalytic event of an enzyme.

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,  
15 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 CDRs of the variable heavy and light chain regions of an antibody contribute to the antigen or ligand-binding pocket.

20 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 polypeptides. This is especially true given that desirable combinations are often accompanied by the presence of vast amounts of undesirable combinations or so-called  
25 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, within a defined region of a polypeptide, in the structure or function of the polypeptide and, thus, for producing improved polypeptides.

30

#### 1. Selecting a Defined Region

In accordance with the present invention, a defined region or regions within a protein are selected for mutagenesis. Typically, the regions are believed to be important to the protein's structure or function. This can be deduced, for example, from what  
35 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 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 (see, *e.g.*, Fig. 1). 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  
5 for mutagenesis.

## 2. Selecting a Predetermined Amino Acid Residue

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  
10 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 or all of the twenty naturally occurring amino acids generally can be selected, as well as non-traditional amino acid residues (*e.g.*, homocysteine). Thus, previous methods of  
15 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 create a practical number of analogs for each amino acid substitution and, thus, allows for the evaluation of a greater variety of protein chemistries within a segregated region or regions of a protein.

20 In contrast to protein binding, only a subset of amino acid residues typically participates in enzymatic or catalytic events. For example, from the chemical properties of the side chains, only a selected number of natural amino acids preferentially participate in catalytic events. Such groupings of amino acid side chain chemistries are useful for selecting an appropriate amino acid residue for use in the chemical profiling  
25 of a particular amino acid residue or position. 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  
30 role 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 CH<sub>2</sub> group than Asn which renders the polar group more flexible and reduces its interaction with the main chain.  
35 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 polypeptide backbone. Cys can form disulfide bridges.

Histidine (His) has a heterocyclic aromatic side chain with a pK value of 6.0. In 5 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 10 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 15 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 20 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.

### 25 3. Synthesizing Polypeptide Analog Libraries

In one embodiment, a library of polypeptide analogs is generated for screening 30 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 35 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.

- 5 Two or more regions can simultaneously be altered. In one embodiment, the amino acid residues (positions) within the defined region that are mutagenized are functional amino acid residues (positions). In another embodiment, the functional amino acid residues (positions) are exclusively mutagenized.

The mixture of oligonucleotides for generation of the library can be synthesized  
10 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  
15 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  
20 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  
25 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  
30 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. 8). Convenient restriction sites can be designed near the ends of the synthetic gene for  
35 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.

10 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.

25 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.

35 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 (see, *e.g.*, Fig. 2).

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.

5 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  
10 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  
15 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  
20 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).

25 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.*,  
30 Profusion™ (see, *e.g.*, U.S. Patent Nos. 6,348,315; 6,261,804; 6,258,558; and 6,214,553).

Alternatively, the polynucleotides of the invention can be expressed in a convenient prokaryotic expression systems such as the bacterial *E. coli* expression system, such as that described by Pluckthun and Skerra. (Pluckthun, A. and Skerra, A., *Meth. Enzymol.* 178: 476-515 (1989); Skerra, A. *et al.*, *Biotechnology* 9: 273-278 (1991)).  
35 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 chain antibodies (scFv) (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 Improved 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 a particular  
5 embodiment, functional amino acid residues (positions) or hot spots are identified as suitable for mutagenesis whereas nonfunctional amino acid residues (positions) or cold spots, are excluded. 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  
10 from previously tested analogs, so that the *in silico* database becomes more precise in its predictive ability.

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  
15 analogs that bind well or not so well or as being enzymatic/catalytic or not so enzymatic/catalytic. In this way, the 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  
20 particular function, such as binding affinity (*e.g.*, specificity), stability (*e.g.*, half life) and/or effector function (*e.g.*, complement activation and ADCC) by targeting hot spots for mutagenesis. 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  
25 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  
30 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  
35 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, the six CDRs of an antibody that make up the unique aspects of the antigen binding site (Fv 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. Functional amino acid residues are then distinguished from non-functional amino acid residues and identified as suitable for further mutagenesis.

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 (*e.g.*, functional amino acid residues (positions)) by the method of the invention.

### 7. Improved Look-Through Mutagenesis (LTM2) 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 (*e.g.*, scFv), 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  
5 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 targets  
10 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). In particular, the mutagenesis targets functional amino acid residues (positions) that have been determined to contribute to a measurable property of the antibody (*e.g.*, antigen-binding, Fc receptor binding, *etc.*).  
15 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. Improved Look-Through Mutagenesis (LTM2) for Making / Improving Catalytic / Enzymatic Polypeptides

20 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  
25 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.

30 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.  
35 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 a preferred embodiment, functional amino acid residues (positions) that have been determined to contribute to a measurable property of the antibody, *e.g.*, catalytic activity, are targeted.

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 identification of improvements to the foregoing is expedited by the methods of the invention by preferentially targeting functional amino acid residues (positions) that are more likely to contribute to the desired activity or property for which improvements are sought.

#### 9. Combinatorial Mutagenesis Methods

In the combinatorial beneficial mutation approach, coding sequences are generated which represent combinations of the beneficial mutations identified by LTM. These combinations may be combinations of different beneficial mutations within a single CDR, mutations within two or more CDRs within a single antibody chain, or mutations within the CDRs of different antibody chains.

One combinatorial approach resembles the WTM method except that the selected codon substitutions within the CDRs are the different beneficial amino-acid substitutions identified by LTM. Thus, not every residue position in an antibody CDR will contain a mutation, and some positions will have multiple different amino acids substituted at that

position. Overall, many if not all, combinations of beneficial mutations within a CDR or an antibody chain will be represented by at least one of the coding sequences in the library. As shown in Table 1, this coding-sequence library can be prepared by a modification of the WTM method, except that instead of placing codons for a single amino acid at each different position in the variable coding region, the codons that are introduced are those corresponding to all beneficial mutations detected in the LTM method. In order to keep the size of this library manageable, the mutations may be confined to one of the two heavy or light chains only.

In a second approach, individual gene fragments containing a single CDR region, and having a codon variation encoding all combinations of beneficial mutations within CDR reconstructed, e.g., by gene shuffling methods, to produce V<sub>L</sub> and V<sub>H</sub> chain coding sequences having combinations of beneficial mutations in all CDRs of a given chain or all CDRs in both chains.

A combinatorial library of mutations may also be generated by known gene shuffling methods, such as detailed in U.S. patent application 2003/005439A1, and U.S. Patent No. 6,368,861, and (Stemmer WP (1994) *Proc Natl Acad Sci* 91(22):10747-51), all of which are incorporated herein by reference. The method involves limited DNase I digestion of the collected mixed mutation clones to produce a set of random gene fragments of various pre-determined sizes (e.g. 50-250 base pairs). The fragments are then first denatured and the various separate fragments are then allowed to re-associate based on homologous complementary regions. In this manner, the re-natured fragments may incorporate differing mixed mutation CDRs in the re-assembled segments which are then extended by SOE-PCR as above, and a re-assembled chimera may then incorporate, at a minimum, at least two sets of beneficial CDR mixed mutations from each parental DNA source donor.

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

### *Exemplification*

Throughout the examples, the following materials and methods were used unless 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 technology, PCR technology, immunology (especially, e.g., antibody technology), expression systems (e.g., cell-free expression, phage display, ribosome display, and

Profusion™), 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, (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., 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., 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 Profusion™ as described by Szostak et al. in U.S. Patent Nos. 6,258,558; 6,261,804; and 6,214,553.

#### Construction of the Test Gene

The ovalbumin wild type Fab sequences were used as templates for the V<sub>L</sub> and V<sub>H</sub> portions (SEQ ID NOS:1 and 2 respectively) for scFv construction using PCR and appropriate primers (SEQ ID NOS: 11 and 12) and V<sub>H</sub> oligonucleotides (SEQ ID NOS: 13-14). PCR reactions consisted of 2 µl each of 10uM oligonucleotide stock, 0.5 µl Pfx DNA polymerase (2.5 U/µl), 5 µl Pfx buffer, 1 µl 10mM dNTP, 1µl 50 mM MgSO<sub>4</sub> and here conducted at 37.5 µl dH<sub>2</sub>O at 94C for 2 min, followed by 24 cycles of 30 sec at 94C, 30 sec at 50C, and 1 min at 68C followed by incubation at 68C for 5 min. Oligonucleotides were synthesized on the 3900 Oligosynthesizer by Syngen Inc. (San Carlos, CA).

The above V<sub>L</sub> and V<sub>H</sub> PCR reactions were then separately extracted and purified (Qiagen PCR purification Kit as per manufacturer instructions) and an aliquot (1 µl) of each reaction was combined into a new tube. By means of the linker sequence at the 3' end of the V<sub>H</sub> reverse oligonucleotide (SEQ ID NOS: 14) and the 5' end of the V<sub>L</sub> forward oligonucleotide (SEQ ID NOS: 11), complementary binding allowed for single

overlap extension PCR (SOE-PCR) assembly reaction to generate the full length ovalbumin scFv. The ovalbumin scFv PCR reaction was extracted and purified (Qiagen) for subsequent *EcoR* I and *Not* I endonuclease digestion (New England Biolabs as per manufacturer's directions). Full length ovalbumin scFv is subcloned into pYD1 vector and sequenced to confirm that no mutations, deletions or insertions were introduced (SEQ ID NO: 18) from the above PCRs. Once sequence verified, full length V<sub>H</sub> and V<sub>L</sub> ovalbumin serves as the wild type template for the subsequent strategies of building LTM libraries.

10 *Additional PCR Conditions and Primer Design*

Reaction conditions for T1 and T2 PCR are; 5 µl of 10 uM oligonucleotide mix, 0.5 µl Pfx DNA polymerase (2.5 U/µl), 5 µl Pfx buffer (Invitrogen), 1 µl 10mM dNTP, 1µl 50 mM MgSO<sub>4</sub> and 37.5 µl dH<sub>2</sub>O at 94C for 2 min, followed by 24 cycles of 30 sec at 94C, 30 sec at 50C, and 1 min at 68C and then incubated for a 68 C for 5 min. The reactions are performed using a programmable thermocycler (MJ Research). T1 and T2 PCR reactions were gel purified (Qiagen) and equimolar aliquots from both were combined for SOE-PCR.

SOE-PCR is a fast and simple method for combining DNA fragments that does not require restriction sites, restriction endonucleases, or DNA ligase. The T1 and T2 PCR products are designed to share end overlapping complementary sequences (Fig. 4) that would hybridize and allow PCR extension to produce a full length LTM ovalbumin scFv gene (Figs. 2 and 4). The scFv PCR extension reaction used T1 and T2 aliquots (approximately 2 ul each) with 0.5 µl Pfx DNA polymerase (2.5 U/µl), 5 µl Pfx buffer (Invitrogen), 1 µl 10mM dNTP, 1 µl 50 mM MgSO<sub>4</sub> and 37.5 µl dH<sub>2</sub>O at 94C for 2 min, followed by 20 cycles of 30 sec at 94C, 30 sec at 50C, and 1 min at 68C and then incubated for a 68 C for 5 min.

A set of ovalbumin end specific 5' *EcoR* I sense (SEQ ID NO: 18) and ovalbumin 3' *Not* I antisense primers (SEQ ID NO: 19) were added to facilitate LTM ovalbumin amplification and incorporation of the restriction enzyme sites in the PCR amplicons (Figure 4). The PCR extension reaction consisted of 4 µl of 10 uM oligonucleotide stock, 0.5 µl Pfx DNA polymerase (2.5 U/µl), 5 µl Pfx buffer (Invitrogen), 1 µl 10mM dNTP, 1µl 50 mM MgSO<sub>4</sub> and 37.5 µl dH<sub>2</sub>O at 94C for 2 min, followed by 24 cycles of 30 sec at 94C, 30 sec at 50C, and 1 min at 68C and then incubated for a 68 C for 5 min.

35

### PCR Product Cloning into a Yeast Cell Expression Vector pYD1

The recipient plasmid pYD1 (Figure 6) for anti-ovalbumin scFv expression, was prepared from an *E. coli* host by plasmid purification (Qiagen), digesting with the restriction enzymes, *EcoRI* and *NotI*, and terminally dephosphorylated with calf  
5 intestinal alkaline phosphatase. Ligation of the pYD1 vector and the above SOE-PCR products (digested by *EcoRI* and *NotI*), before subsequent *E. coli* (DH5 $\alpha$ ) transformation performed using standard techniques.

### Yeast Cell Expression System

10 The pYD1 (Fig. 6) is an expression vector designed to display proteins of interest on the extracellular surface of *Saccharomyces cerevisiae*. By the sub-cloning the scFv gene into pYD1, scFvs becomes a fusion proteins with the AGA2 agglutinin receptor allowing cell surface secretion and display.

### Transformation of Yeast Host Cells with pYD1 AGA2-scFv Constructs

15 Competent yeast host cells (500  $\mu$ l) were prepared as per the manufacturer's instructions (Zymo Research Frozen-EZ Yeast Kit). Briefly, 500  $\mu$ l of competent cells were mixed with 10-15  $\mu$ g pYD1 scFv library DNA after which 5 ml of EZ3 solution was added. The cell mixture was then incubated for 45 minutes at 30°C with occasional  
20 mixing (three times). The transformed cells were centrifuged and resuspended in glucose selection media (Invitrogen).

### Induction of AGA2-scFv

Cells were induced after growth in glucose selection media at 30°C under  
25 shaking and aeration conditions for 48 hours until the OD<sub>600</sub> = 7 (OD<sub>600</sub> = 1 represents 10<sup>7</sup> cells/ml). In particular, cells were collected and re-suspended in galactose selection / induction media (Invitrogen), until an OD<sub>600</sub> = 0.9 was reached after 48 hours at 20°C. Expression of the Aga2-scFv fusion protein from pYD1 was tightly regulated by the GAL1 promoter and depended on available galactose in the medium for GAL1 promoter  
30 induction.

### Biotinylated Ovalbumin Preparation

Biotinylation of the target antigen ovalbumin was carried out according to the manufacturer's instructions (Molecular Probes FluoReporter Biotin-XX Labeling Kit  
35 (cat# F-2610)). Briefly, ovalbumin 300 $\mu$ l of 1mg/ml stock (Sigma), was added to 30 $\mu$ l 1M Sodium Bicarbonate Buffer at pH 8.3 and 5.8  $\mu$ l of Biotin-XX solution (20mg/ml Biotin-XX solution in DMSO). The mixture was then incubated for 1 hour at 25°C

transferred to a micron filter tube, centrifuged, and the protein concentration was determined by light absorbance at OD<sub>280</sub>.

FACS monitoring of AGA2-scFv Expression and Ovalbumin Binding

5 To monitor induction and expression of the scFv construct, an aliquot of yeast cells ( $8 \times 10^5$  cells in 40  $\mu$ l) from the culture medium was collected by centrifugation for 5 minutes at 2300 rpm. The supernatant fluid was aspirated and then the cell pellet washed with 200  $\mu$ l of ice cold PBS/BSA buffer (PBS/BSA 0.5% w/v). The cells were re-pelleted and supernatant removed before re-suspending in 100  $\mu$ l of buffer containing  
10 the biotinylated ovalbumin (200 nM). The cells were left to bind the ovalbumin at 20°C for 45 minutes after which they were washed twice with PBS/BSA buffer before the addition and incubation with streptavidin-FITC (2 mg/L) for 30 minutes on ice. Another round of washing in buffer was performed before final re-suspension in a volume of 400  $\mu$ l in PBS/BSA. The cells were then analyzed on FACSscan (Becton Dickinson) using  
15 the CellQuest software package.

**EXAMPLE 1**  
**IMPROVED LOOK-THROUGH MUTAGENESIS FOR THE DEVELOPMENT OF HIGH AFFINITY ANTIBODIES**

20 In this example, the improved look-through mutagenesis of an exemplary scFv antibody, is described.

Briefly, improved look-through mutagenesis (LTM) was used to identify V<sub>H</sub> and V<sub>L</sub> CDR mutations in the V<sub>H</sub> and V<sub>L</sub> CDR regions of an antibody that enhance binding affinity to a chosen antigen, *i.e.*, to identify the functional amino acid residues  
25 (positions) or hot spots that impart binding activity. The purpose of improved look-through mutagenesis (LTM2) is to introduce a selected substitution at targeted positions in a region of a polypeptide, *e.g.*, the CDR regions of the variable antibody chain. Initially, coding libraries for both the V<sub>H</sub> and V<sub>L</sub> chains were constructed. Amino acid sequences of the reference anti-ovalbumin antibody V<sub>L</sub> and V<sub>H</sub> chains were identified  
30 (see SEQ ID NOS: 3-4) and a library of starting sequences for CDR1, CDR2, and CDR3 regions of the anti-ovalbumin V<sub>H</sub> chain identified by SEQ ID NOS: 5, 6 and 7, respectively, and those for CDR1, CDR2, and CDR3 regions of V<sub>L</sub> chain identified by SEQ ID NOS: 8, 9, and 10, were selected. Additionally, each V<sub>H</sub> or V<sub>L</sub> coding sequence library contains mutations for a selected representative functional amino acid in one,  
35 two, or all three CDRs in the selected chain.

Predetermined amino acids of V<sub>H</sub> CDR2 segment DINPSNGYTIYNQKFKG) (positions 56 to 69) from the ovalbumin wild type V<sub>H</sub> section SLEWIG-DINPSNGYTIYNQKFKG -FKGKATL, were selected for analysis. The polypeptide sequences SLEWIG and KATLTVD are portions of the V<sub>H</sub> frameworks 2 and 3

respectively flanking V<sub>H</sub> CDR2. In the design and synthesis of V<sub>H</sub> and V<sub>L</sub> CDR LTM oligonucleotides, the flanking framework sequence lengths of approximately 21 base pairs allow for complementary overlap and SOE-PCR. A reference oligonucleotide coding for the above CDR-H2 wild type sequence (in bold) (SEQ ID NO: 16) containing the flanking V<sub>H2</sub> and V<sub>H3</sub> portions (lowercase letters below) is shown below:

5' -agc ctt gag tgg att gga-GAT ATT AAT CCT AGC AAT GGT TAT ACT  
ATC TAC AAC CAG AAG TTC AAG GGC-ttc aag ggc aag gcc aca ttg -3'

10 The leucine LTM of V<sub>H</sub> CDR2 involves serially substituting only one leucine at a time, in every or at a plurality of CDR2 positions. Fig. 3 illustrates LTM application for introducing a leucine amino acid into each of the fourteen residues (positions 56-69) in the V<sub>H</sub> CDR2 region of anti-ovalbumin scFv. In performing leucine LTM, fourteen separate oligonucleotides encoding all possible V<sub>H</sub> CDR2 leucine positional variants were synthesized (SEQ ID NOS:17-33) with each having only one leucine replacement codon (in bold) bordered by anti-ovalbumin wild type sequence. A total of fourteen different peptides are then generated, and no "undesired" or multiple-substitution sequences are produced.

20 In an alternative to introducing leucine to a "plurality of positions", one may choose not to include one or more of the above CDR2 LTM oligonucleotides. Hence, for example, the LTM oligonucleotides (SEQ ID NOS: 20 and 21) for positions 59 and 60 respectively can be excluded and thus improved look-through mutagenesis (LTM2) would only be performed on positions 56-58 and 61 to 69. Based on the different combinations of using a "plurality of positions", permutations of LTM2 can be performed without necessarily restricting to each and every position within a CDR.

25 The approach in making the LTM CDR2 library is summarized in Figs. 2 and 4. Separate PCR reactions, T1 and T2, are carried out using primer pairs FR1 sense (SEQ ID NO: 34) and FR2 antisense (SEQ ID NO: 35) and the above pooled CDR-2 LTM leucine oligonucleotides (SEQ ID NO: 18 - 33) with FR4 anti-sense primer (SEQ ID NO: 36) respectively. Primer FR1 sense contains sequences from the 5' terminus of the ovalbumin gene and FR2 anti-sense contains the antisense sequence from the 3' terminus of ovalbumin framework 2 so that the ovalbumin CDR1, framework regions 1 and 2 was amplified in the T1 PCR reaction (Figs. 2A and 2B). The primer FR4 AS contains anti-sense sequence from the 3' terminus of the ovalbumin gene, CDR2 LTM oligonucleotides contain sequences from the 5' terminus of the ovalbumin CDR2 region with the incorporated CDR2 LTM codon mutations to amplify the remaining portion of ovalbumin (fragment CDR2, FR3, CDR3, FR4 and the entire V<sub>L</sub> segment) while concurrently incorporating the mutagenic codon(s).

Double and Triple CDR mutations (in different combinations of CDR1, 2, and 3) were created as above but instead of using the wild type ovalbumin gene as a PCR template, a previously generated LTM ovalbumin library was selected. For example, to create  $V_H$  chains in which both CDR1 and CDR2 are mutated with wild-type CDR3 and  $V_L$  segments, the previously constructed LTM CDR2 mutant genes were used as templates and then SOE-PCR was conducted to incorporate the CDR1 oligonucleotides to generate the Double LTM mutations (summarized in Figures 2 and 4).

In this example, the T3 PCR reaction used primer pairs FR1 sense (SEQ ID NO: 34) and FR5 antisense (SEQ ID NO: 37) to amplify the framework region 1 (FR 1). The T4 PCR reaction used the pooled CDR1 LTM histidine oligonucleotides (SEQ ID NO: 38-41) with FR4 anti-sense primer (SEQ ID NO: 36) to amplify the remaining FR2, CDR2 LTM, FR3, CDR3, FR4 and  $V_L$  portions of ovalbumin (Fig. 4). T3 and T4 PCR reactions were then purified and equimolar aliquots from both were combined for SOE-PCR (Fig. 4) to produce an ovalbumin scFv His CDR1 and Leu CDR2 double LTM library. A set of ovalbumin end specific 5' Eco RI sense (SEQ ID NO: 13) and ovalbumin 3' *NotI* antisense primers (SEQ ID NO: 12) were then added to facilitate LTM ovalbumin amplification and cloning into the pYPD1 expression vector.

The His CDR1 and Leu CDR2 double LTM library was then used as templates to further incorporate LTM CDR3 oligonucleotides to make the Triple CDR LTM libraries. By progressively utilizing the starting single and double LTM libraries, a more complex array of LTM library combinations in both the  $V_H$  and  $V_L$  CDR was developed. For example, once the  $V_H$  Triple LTM CDR1, CDR2, and CDR3 library was constructed, designated as the 111 library template in the top row of Fig. 5, introduction of LTM  $V_L$  CDR1 into the  $V_H$  111 templates produces a library of 4 LTM CDRs (Fig. 5).

The FACS plot shown in Fig. 7 illustrates that ovalbumin scFv binding molecules were generated in the construction of the foregoing library. In particular, molecules of biotinylated ovalbumin and streptavidin FITC (the "green" line) produced a peak signal response a magnitude higher compared to signal from the empty vector pYDI with biotinylated ovalbumin and streptavidin FITC (dark shaded area).

30

## EXAMPLE 2

### HIGH THROUGHPUT LIBRARY SCREENING FOR THE DEVELOPMENT OF ANTIBODIES WITH ENHANCED PROPERTIES

In this example, the high throughput screening of exemplary single chain antibodies with enhanced properties, is described.

Figure 9 depicts a generalized scheme for enriching the ovalbumin specific high affinity binding clones from the heterogeneous LTM2 scFv library. After induction in galactose media, the yeast cell library ( $10^7$ ) was resuspended in PBS/BSA buffer (total

volume of 500  $\mu$ l). Biotinylated ovalbumin was added to the yeast suspension for a final concentration 50nM and incubated at 25°C for 2-3 hours with shaking. The yeast cells were pelleted, washed 3 times, and resuspended in 300  $\mu$ l ice cold PBS/BSA buffer of buffer with  $1 \times 10^8$  streptavidin coated magnetic beads added. The bead cell mixture  
5 was incubated on ice for 2 minutes with gentle mixing by inversion to form a binding complex consisting of yeast high-affinity scFv expressing cells, biotinylated ovalbumin, and streptavidin coated magnetic beads. The column (tubes) containing bound complexes were then applied to the magnetic column holder for 2 minutes after which the supernatant was removed by aspiration. The column was removed from the magnet  
10 holder, and 300  $\mu$ l ice cold PBS/BSA was added to re-suspend the bound.. The bound complexes were washed again in order to remove those scFv clones of low-affinity and other non-specifically bound cells.

The column was then removed from the magnetic holder whereupon 1 ml of glucose selection media was added to the recovered yeast cells for a 4 hour incubation at  
15 30°C. The magnet holder was then re-applied to the culture tube to remove any remaining magnetic beads. The yeast culture was then grown in glucose selection media at 30°C for 48 hours before scFv induction in galactose selection media. In the second selection round, the ovalbumin concentration was then lowered from 10 nM to 0.5 nM. the ovalbumin binding, complex formation, and yeast cell enrichment and re-growth  
20 were performed as described above. For the final third selection round, the ovalbumin concentration was further reduced to 0.1nM.

Ovalbumin EC<sub>50</sub> binding, or “fitness” from each of the above rounds of progressive enrichment was evaluated by FACS (see Fig. 8). Results shown in Fig. 8 illustrates that the initially transformed V<sub>H</sub> LTM CDR2 yeast library, with no prior  
25 selection (closed circles), as well as the overall fitness in terms of percent binders (y-axis). Clones expressing functional anti-ovalbumin scFvs and their affinity, as measured by the ovalbumin EC<sub>50</sub> (x-axis), were inferior compared to the anti-ovalbumin wildtype antibody. However, after one round of selection (10nM), the “fitness” curve (light triangles) improved in percent binders and the EC<sub>50</sub> for ovalbumin binding was in the  
30 same nM range as the ovalbumin wild type. After the second selection round (0.1nM), the enriched population (dark triangles) exhibited an overall “fitness” that approached that of the ovalbumin wild type (solid squares) (Fig. 10). The recovered yeast cells from the second round of enrichment were then plated onto solid media in order to isolate single clones for individual binding analysis and sequence determination.

35 In an alternative methodology, the LTM2 yeast cell libraries were enriched for high affinity anti-ovalbumin scFv clones by FACS. Library construction, transformation, liquid media propagation and induction, were carried out as above. After scFv induction, the cells were incubated with biotinylated ovalbumin at saturating

concentrations (400 nM) for 3 hours at 25°C. After washing the cells, a 40 hour cold chase using unlabelled ovalbumin (1µM) at 25°C was performed. The cells were then washed twice with PBS/BSA buffer, labeled with Streptavidin PE (2 mg/ml) anti-HIS-FITC (25 nM) for 30 minutes on ice, washed and re-suspended as described above for  
5 FACS analysis.

Wild type anti-ovalbumin was initially FACS analyzed to provide a reference signal pattern for FACS sorting of the yeast LTM library (Figure 8, left panel). From the ovalbumin FACS plot, a selection gate (the R1 trapezoid) was drawn to obtain only those clones that expressed the scFv fusion (as detected by anti-HIS-FITC) and  
10 concomitantly would display a higher binding affinity to ovalbumin (a stronger PE signal) compared to wild type anti-ovalbumin. Figure 8 (middle panel) demonstrates that approximately 5% of the total LTM library screened was selected by the R1 gate. After FACS collection of these high anti-ovalbumin scFv clones, a post-sort FACS analysis (Figure 8, right panel) was performed to confirm that >80% of the pre-screen  
15 anti-ovalbumin scFv clones were within the predetermined criteria. The FACS sorted scFv clones were then grown in glucose selection media at 30 °C for 48 hours and plated on solid media to isolate individual clones. Clones were then grown in glucose selection media, re-induced in galactose selection media, and analyzed for their EC<sub>50</sub> and/or k<sub>off</sub> characteristics.

20

### EXAMPLE 3

#### HIGH THROUGHPUT LIBRARY SCREENING FOR THE DEVELOPMENT OF ANTIBODIES WITH IMPROVED K<sub>off</sub> RATES

In this example, the high throughput screening of exemplary single chain  
25 antibodies with enhanced K<sub>off</sub> rates, is described.

Briefly, pre-sorted clones from the above example were grown overnight in glucose selection media and then plated on solid media to isolate single colonies. From single colonies, liquid cultures of clones were grown in glucose selection media at 30°C with shaking for 48 hours before pelleting and re-suspending in galactose selection  
30 media for the appropriate OD time period. Because the FACS pre-sort enriches (by approximately 80%) but does not eliminate all undesirable clones, the EC<sub>50</sub> of the isolated clones was characterized to eliminate those that display binding values inferior to anti-ovalbumin wild type reference antibody (as detailed above). Only those isolates with comparable or superior EC<sub>50</sub> values were then selected for further K<sub>off</sub> analysis.

35 A scheme for analyzing the kinetics of candidate molecules is shown in Fig. 11. Specifically, yeast cells (approximately 5 x 10<sup>6</sup>) after induction in galactose selection media, were pelleted and re-suspended in PBS/BSA buffer (1 ml). Biotinylated ovalbumin (400nM final concentration) was then added to the re-suspended cells and

allowed to incubate for 2 hours at 25°C with continuous gentle mixing. The biotinylated-ovalbumin and yeast cell complex was then washed and re-suspended in PBS/BSA buffer and unlabelled ovalbumin was then added (to a final concentration of 1  $\mu$ M) to the yeast cell mixture which was further incubated for 24 hours at 25°C. Timed sample aliquots were then taken every two hours for the next 24 hours. The cell mixtures were then washed and re-suspended in chilled PBS/BSA buffer and staining antibody  $\alpha$ -SAPE (2  $\mu$ g/ml). After incubation for 30 minutes on ice with periodic mixing, the cell mixture was then twice washed and analyzed by FACS as above.

Results from the  $K_{off}$  assays (Fig. 12) demonstrated two clones (*i.e.*, 3ss-35; 3ss-30) as having a higher relative  $K_{off}$  as compared to the wild type anti-ovalbumin antibody. In other words, when exchanging the bound biotinylated ovalbumin for the unlabelled ovalbumin during the 24 hour sampling period, 3ss-35 and 3ss-30 released the previously bound biotinylated ovalbumin at a much slower rate (circles and triangles respectively in Fig. 12) compared to anti-ovalbumin wild type, (squares Fig. 12) which exhibited a much sharper decrease in MFI (mean fluorescence intensity) over the first 8 hours.

#### EXAMPLE 4

##### HIGH THROUGHPUT LIBRARY SCREENING FOR THE DEVELOPMENT OF ANTIBODIES WITH IMPROVED $EC_{50}$ BINDING

In this example, the high throughput screening of exemplary single chain antibodies with enhanced  $EC_{50}$  binding activity is described.

Briefly, a pre-determined amount of yeast cells ( $8 \times 10^5$  cells in 40  $\mu$ l) anti-ovalbumin scFvs (wild type and LTM library) was incubated with 1:4 serial dilutions of biotinylated ovalbumin (200 nM, 50 nM, 12.5 nM, 3.1 nM, 0.78 nM, and 0.19 nM final concentrations in a total volume of 80  $\mu$ l) at 20°C for 45 minutes followed by 5-10 minutes on ice. The yeast cells were then washed and re-suspended in 5 ml of PBS/BSA buffer after which streptavidin-PE (2 mg/ml) and  $\alpha$ HIS-FITC (25 nM) was added to label the cells during a 30 minute incubation on ice. Another round of washing was performed before re-suspending in 400  $\mu$ l of PBS/BSA buffer and analysis on FACSscan using CellQuest software.

Figure 13 exemplifies a subset of improved clones, relative to anti-ovalbumin, having lower  $EC_{50}$  values (the binding curves have shifted to the left with respect to the wild type). These relative  $EC_{50}$  values as compared to anti-ovalbumin wild type and the relative fold increase are shown. For example, the clone H3 S101Q exhibited a 2.1 fold improvement in ovalbumin binding. Nomenclature identification of this clone H3 S101Q, indicates that it was from a  $V_H$  CDR3 glutamine LTM single library.

The resulting enhanced-binding affinity antibodies produced provide a map of beneficial amino acid mutations in the V<sub>H</sub> and V<sub>L</sub> CDRs of the anti-ovalbumin antibody that were associated with enhanced binding activity. Putative individual amino acid V<sub>H</sub> and V<sub>L</sub> CDR mutations associated with the enhanced-affinity scFv antibodies are shown in Fig. 13. Mutations were found in each V<sub>H</sub> and V<sub>L</sub> CDR and single-, double-, and triple-CDR mutations, and include each of the nine different amino acids tested. Figure 13 exemplifies that four independent V<sub>H</sub> CDR-H3 H3 G102K clones identified from the above EC<sub>50</sub> and/or kinetics (K<sub>off</sub>) screen, can be recovered. The double LTM V<sub>L</sub> L1L2 S166H S193H mutant from the EC<sub>50</sub> screen (Fig. 14) also illustrates that enhanced ovalbumin binding occurs when there is a synergistic interaction between these two CDR1 S166H and CDR2 S193H substitutions.

Results also indicate that (see Fig. 14) unique sequences for scFv anti-ovalbumin antibodies selected in accordance with the above EC<sub>50</sub> and/or K<sub>off</sub> method using coding sequences containing single mutations at one, two or at all three CDRs in either the V<sub>H</sub> or V<sub>L</sub> chains, can be obtained. Some of the amino acid substitutions were found to be recovered in a higher preponderance in certain CDRs. For example in V<sub>H</sub> CDR1, there were seven independent single Lys substitutions in the positions of D30K, Y31K, M33K and W35K. The high preponderance of Lys in CDR1 indicates that enhanced ovalbumin binding occurs when there is a net increase of positive charges contributed by CDR1 during antigen contact. Examination of the recovered amino acid substitutions also revealed that a favored substitution occurs for V<sub>H</sub> CDR2 and CDR3. For example, it was observed that multiple scattered Lys replacements occur in CDR2 while CDR3 displays concentrated G102K and G104K replacements. Results also reveal that additionally, there were multiple Gln replacements, another polar amino acid, in CDR3 at positions 107, 108, 109, and 111. These results indicate that there is a unique “chemical motif” of corresponding negative and/or hydrophilic charges on ovalbumin that is contacted by these replaced CDR residues.

Accordingly, the recovered enhanced antigen binding mutations in each CDR after mutagenesis of all CDRs by each of the nine different amino acids tested was arrayed and results are shown in Fig. 14. Each CDR was determined to also display at least one position in which no mutations were found, *e.g.*, Ile57, Asn58, and Gly62 positions of V<sub>H</sub> CDR2, positions Tyr103, Ser105, Arg106 and Ala121 of V<sub>H</sub> CDR3 region. For V<sub>L</sub> positions; R164, A165, V169 and N174 of CDR1, N194 of CDR2, E208 and D209 of CDR3 were found not to be replaced after extensive screening. These results indicate that in addition to the “hot spot” replacement positions and any type of preferred amino acid recovered as described above, there are also “cold spot” CDR positions. The “cold spot” positions indicate that any of the substitutions either confer no antigen binding advantage or imparted inferior binding properties.

**EXAMPLE 5**  
**METHODS FOR CONDUCTING COMBINATORIAL BENEFICIAL**  
**MUTAGENESIS**

5 In this example, methods for conducting combinatorial beneficial mutagenesis, are described.

Briefly, the method incorporates the beneficial mutations found using LTM, and combines them together into a single library. Therefore, synergistic effects of multiple mutations can be explored in this process. This combinatorial approach resembles the  
 10 WTM method except that the selected codon substitutions within the CDRs are the different beneficial amino-acid substitutions identified by LTM. Thus, not every residue position in an antibody CDR will contain a mutation, and some positions will have multiple different amino acids substituted at that position. Overall, many if not all, combinations of beneficial mutations within a CDR or an antibody chain will be  
 15 represented by at least one of the coding sequences in the library. As will be seen below, this coding-sequence library can be prepared by a modification of the WTM method, except that instead placing codons for a single amino acid at each different position in the variable coding region, the codons that are introduced are those corresponding to all beneficial mutations detected in the LTM method.

20 In this example, a combinatorial beneficial heavy chain CDR1 library is constructed using the improved anti-ovalbumin variants described in Figure 14. The first amino acid position 30, encodes (at minimum) Asp (wild-type) and Lys; position 31, Tyr and Lys; position 32 Asn and Tyr; position 33 Met, His, and Lys; position 34 Asp and Pro; position 35 Trp and Lys. The DNA sequence is represented by a  
 25 degenerate CDR1 oligonucleotide that incorporates the combinatorial beneficial mutations identified from the LTM analysis (Table 1).

**Table 1 Nucleotide base possibilities from 5' to 3'**

5'	G	A	C	T	A	C	A	A	C	A	T	G	G	A	C	T	G	G	3'
	A		G	A		G	T			C	A	C	C	C		A	A	G	

30 A similar process can be performed with all CDR loops to produce a 6-CDR beneficial library, or sub-libraries can be produced and screened in combination. For example, if the library size is too large to be screened in a single library, then smaller sub libraries can be produced. For example, a 3-CDR heavy chain library and a 3-CDR light chain library can be produced for greater efficiencies. After screening, the  
 35 beneficial mutations of the heavy and light chain libraries can be further combined into a single library that incorporates mutations in the appropriate loops.

***Equivalents***

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  
5 described herein. Such equivalents are intended to be encompassed by the following claims.

What is claimed:

1. A method of distinguishing one or more functional amino acid residues from non-  
5 functional amino acid residues within a polypeptide comprising,  
selecting one or more amino acid residues within the amino acid sequence of the  
polypeptide;  
determining an amino acid residue to be substituted for the one or more amino  
acid residues selected;  
10 synthesizing polynucleotides encoding the polypeptide or portion thereof  
comprising the selected amino acid residues, the polynucleotides collectively  
representing possible variant amino acid substitutions according to the following  
criteria:  
15 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  
residues, and  
ii) each polynucleotide containing no more than one codon for the  
predetermined amino acid residue,  
20 thereby generating an expression library containing the polynucleotides;  
expressing the expression library to produce polypeptide analogs; and  
screening the polypeptide analogs for an alteration in a measurable property such  
that one or more amino acid residues(s) within the polypeptide, or portion thereof, is  
identified as contributing to the measurable property and therefore distinguished from a  
25 non-functional amino acid residue.
2. The method of claim 1, wherein the amino acid residue(s) identified as contributing  
to a measurable property is determined to be suitable for mutagenesis.
- 30 3. The method of claim 1, wherein the non-functional residue(s) is determined to be  
unsuitable for mutagenesis.
4. The method of claim 11, wherein one or more functional amino acid residue(s) are  
exclusively mutagenized.
- 35 5. The method of claim 2, wherein the mutagenesis is selected from the group  
consisting of look-through mutagenesis (LTM), walk-through mutagenesis (WTM), or a  
combination thereof.

6. The method of claim 1, wherein the polypeptide is an antibody or fragment thereof.
7. The method of claim 6, wherein the measurable property is selected from the group  
5 consisting of binding specificity, binding avidity, binding affinity, Fc receptor binding, glycosylation, complement binding, half-life stability, solubility, thermal stability, catalytic activity, and enzymatic activity.
8. The method of claim 1, wherein the method further comprises the step of identifying  
10 a polynucleotide that encodes a selected polypeptide analog having an alteration in a measurable property.
9. The method of claim 8, wherein the alteration in a measurable property is an  
enhanced property.
- 15 10. The method of claim 9, wherein the enhanced property is high affinity antigen binding.
11. The method of claim 9, wherein the enhanced property is improved effector  
20 function.
12. The method of claim 10, wherein the antigen is a therapeutic target in a human  
disease or disorder.
- 25 13. The method of claim 1, wherein the screening comprises,  
contacting a polypeptide with a target substrate, the polypeptide being associated  
with the polynucleotide encoding the polypeptide, the polynucleotide or polypeptide  
being associated with a detectable moiety, such that a variant polypeptide capable of  
binding a target substrate is detected and thereby identified as encoded by the  
30 polynucleotide.
14. The method of claim 13, wherein the detectable moiety is selected from the group  
consisting of a fluorescent moiety, a UV moiety, and a visible light absorbing moiety.
- 35 15. The method of claim 13, wherein the detectable moiety is selected from the group  
consisting of a biotin moiety, a GST moiety, a myc immunotag moiety, and a His tag  
moiety.

16. The method of claim 13, wherein the polynucleotide is associated with the polypeptide analog using expression display selected from the group consisting of ribosome display, polysome display, phage display, prokaryotic (bacterial) display, yeast display, eukaryotic cell display, and arrayed library display.

5

17. The method of claim 1, wherein the polypeptide is a single chain antibody (sFVs).

18. The method of claim 1, wherein the defined region comprises a functional domain of the polypeptide.

10

19. The method of claim 1, 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.

15 20. The method of claim 1, 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.

20 21. The method of claim 1, 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 receptor binding region.

22. 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,  
25 Gly, Ile, Leu, Met, Phe, Pro, Trp and Val.

23. A method of generating a library of polypeptide analogs in which a predetermined amino acid appears at one or more functional amino acid positions in a defined region of the polypeptide comprising:

30 selecting one or more functional amino acid positions in a defined region of the amino acid sequence of the polypeptide;

determining an amino acid residue to be substituted at each functional amino acid position within the defined region;

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

- i) each polynucleotide containing at each functional amino acid 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 amino acid residue,
- 5 thereby generating a library of polynucleotides in which the predetermined amino acid residue appears at each functional amino acid position within the defined region.
- 10 24. The method of claim 23, wherein the polynucleotides are pooled together.
25. The method of claim 23, wherein two or more functional amino acid positions within a defined region(s) within the polypeptide are mutagenized.
- 15 26. The method of claim 25, wherein the same predetermined amino acid is selected for substitution within each of the two or more functional amino acid positions.
27. The method of claim 23, wherein different predetermined amino acids are selected for substitution within each of the two or more functional amino acid positions,
- 20 respectively.
28. The method of claim 27, wherein the defined region or defined regions comprises a functional domain of the polypeptide.
- 25 29. The method of claim 28, wherein the functional domain is selected from the group consisting of an antibody binding site, an antibody framework region, an antibody effector region, a receptor binding site, and a catalytic site.
- 30 30. The method of claim 28, 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.
31. The method of claim 28, wherein the antibody framework region comprises a domain selected from the group consisting of FR1, FR2, FR3, FR4 and a combination
- 35 thereof.

32. The method of claim 28, wherein the antibody effector region comprises a domain selected from the group consisting of a complement binding site and an Fc binding region.
- 5 33. The method of claim 23, 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.
34. The method of claim 23, wherein the defined region comprises at least about 3 to 60  
10 amino acids.
35. The method of claim 23, wherein the polynucleotides are synthesized as expression library.
- 15 36. The method of claim 23, wherein the polynucleotides are synthesized using enzymatic means.
37. The method of claim 23, wherein the polynucleotides are synthesized using polymerase chain reaction.  
20
38. The method of claim 23, wherein the library is an expression library selected from the group consisting of a ribosome display library, a polysome display library, a prokaryotic (bacterial) display library, a yeast display library, a eukaryotic cell display library, and an arrayed display library.  
25
39. The method of claim 23, wherein one or more functional amino acid residue(s) are exclusively mutagenized.
40. A library of polypeptide analogs prepared by the method of claim 23.  
30
41. A library of polynucleotides encoding polypeptide analogs comprising one or more functional amino acid residues within a defined region wherein a predetermined amino acid residue is substituted at each functional amino acid position within the defined region, the polynucleotides collectively representing all possible variants according to  
35 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.

42. The library of claim 41, wherein one or more functional amino acid residues are  
5 exclusively mutagenized.

43. The library of claim 41, wherein the library is an expression library selected from the group consisting of a phage display library, a ribosome/polysome display library, a yeast display library, a prokaryotic (bacterial) display library and an arrayed library.  
10

44. The library of claim 41, wherein the polynucleotides further comprise one or more transcriptional regulatory elements.

45. The library of claim 44, wherein the polynucleotides, when transcribed and  
15 translated *in vitro*, are associated with the polypeptides encoded by the corresponding polynucleotides.

46. The library of claim 45, wherein the polynucleotides are associated with the polypeptide using a display library selected from the group consisting of a ribosome  
20 display library, a polysome display library, a prokaryotic (bacterial) display library, a yeast display library, a eukaryotic cell display library, and an arrayed display library.

47. The library of claim 46, wherein the polynucleotides comprise RNA.

48. The library of claim 47, wherein the polynucleotides further comprise a detectable  
25 moiety.

49. The library of claim 48, wherein the detectable moiety comprises a fluorescent  
30 moiety.

50. The library of claim 41, wherein the library comprises at least 45 to  $10^{12}$  different polynucleotides.

51. The library of claim 41, wherein the polypeptide encodes a binding polypeptide.  
35

52. The library of claim 41, 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 (scFv).

53. The library of claim 41, wherein the library is immobilized on a solid support.

54. The library of claim 41, wherein the solid support is a microchip.

5

55. The library of claim 41, wherein the library is an arrayed library.

56. A microchip comprising an array of immobilized polynucleotides according to the library of claim 41.

10

57. A polypeptide analog identified using the library of claim 41, 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 (scFv).

15

58. A method of identifying a subset of polypeptide analogs having a desired property comprising:

selecting one or more functional amino acids within a defined region of the amino acid sequence of the polypeptide;

20

determining an amino acid residue to be substituted at each functional amino acid position within the defined region;

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

25

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

30

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;

35

screening the expressed library to identify a polypeptide having a desired property;

comparing the property 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 as a nonresponder;

categorizing responders and nonresponders in a database; and

5 querying the database to determine the sequence of a subset polypeptides to be synthesized.

59. The method of claim 58, wherein one or more of the above steps is computer-assisted.

10 60. The method of claim 58, wherein the property is selected from the group consisting of an altered binding specificity, an altered binding on-rate ( $k_{on}$ ), an altered binding off-rate ( $k_{off}$ ), and combinations thereof.

15 61. The method of claim 58, wherein the control criterion is selected from the group consisting of binding specificity, binding avidity, binding affinity, effector function, Fc receptor binding, glycosylation, complement binding, half-life stability, solubility, thermal stability, catalytic activity, and enzymatic activity.

20 62. The method of claim 58, 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 (scFv).

63. A medium suitable for use in an electronic device having instructions for carrying out one or more steps of the method of claim 59.

25

64. A device for carrying out one or more steps of the method of claim 59.

65. A library of polynucleotides encoding antibodies or binding portions thereof comprising,

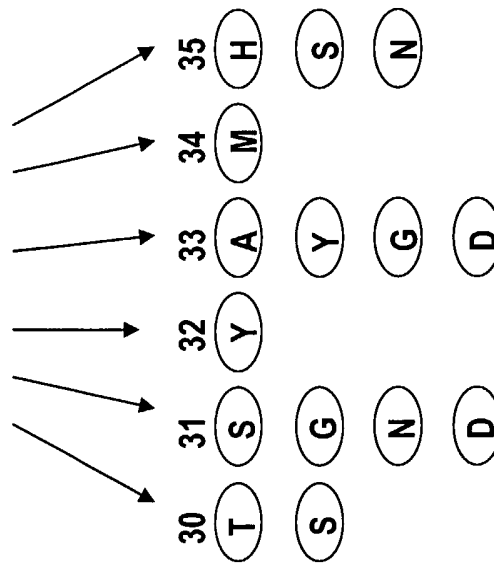
30 polynucleotides encoding a predetermined amino acid substitution at one or more functional amino acid residues, the polynucleotides collectively representing all possible variants suitable for mutagenesis for enhancing a property selected from the group consisting of an altered binding specificity, an altered binding avidity, an altered binding affinity, an altered Fc binding activity, an altered binding on-rate ( $k_{on}$ ), an altered  
35 binding off-rate ( $k_{off}$ ), and combinations thereof.

66. An antibody or binding portion thereof derived from the library of claim 65.

**LOOK-THROUGH MUTAGENESIS (LTM)**

Exemplary V<sub>H</sub> CDR1 region

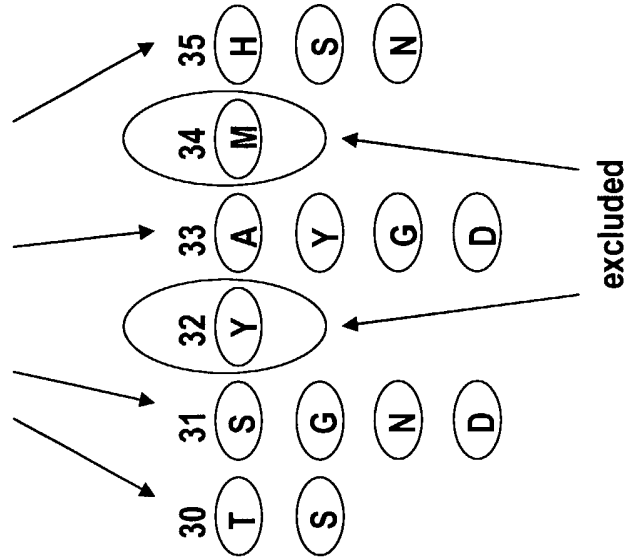
mutagenesis is performed on all residues within region



**IMPROVED LOOK-THROUGH MUTAGENESIS (ILTM)**

Exemplary V<sub>H</sub> CDR1 region

mutagenesis is performed on selected amino acids (e.g., functional residues that will change or improve binding) but NOT those residues that do not contribute function

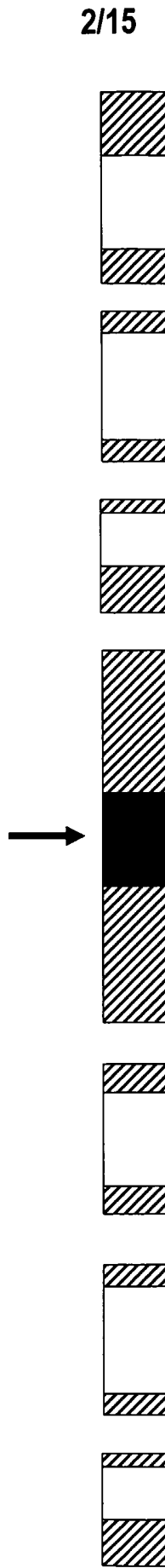


*Fig. 1*

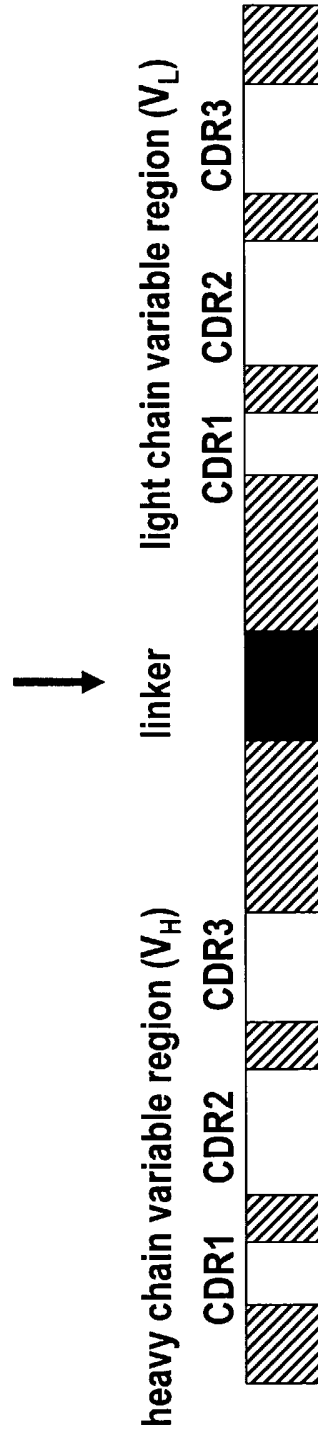
**Exemplary Assembly of Defined Regions  
into a Synthetic Gene (sFv)**



convert single-stranded nucleotides to double-stranded



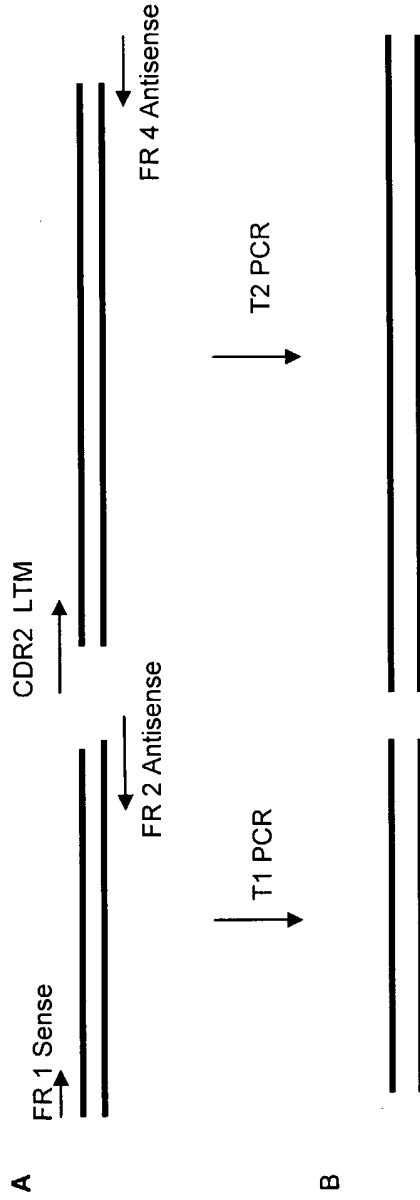
assemble w/ PCR / ligation



*Fig. 2*



SOE-PCR of CDR2 LTM of an anti-ovalbumin single chain antibody (scFv)



5'-----FR 2 -----cat ctc acc caa aga cgc tct----->3'  
 3-<-----'-gta gag tgg gtt tct gcg ata----- CDR2 LTM ----5'

Fig. 4A

SOE-PCR of CDR1 LTM into CDR2 LTM of an scFv template to make a double CDR1 and CDR 2 LTM library

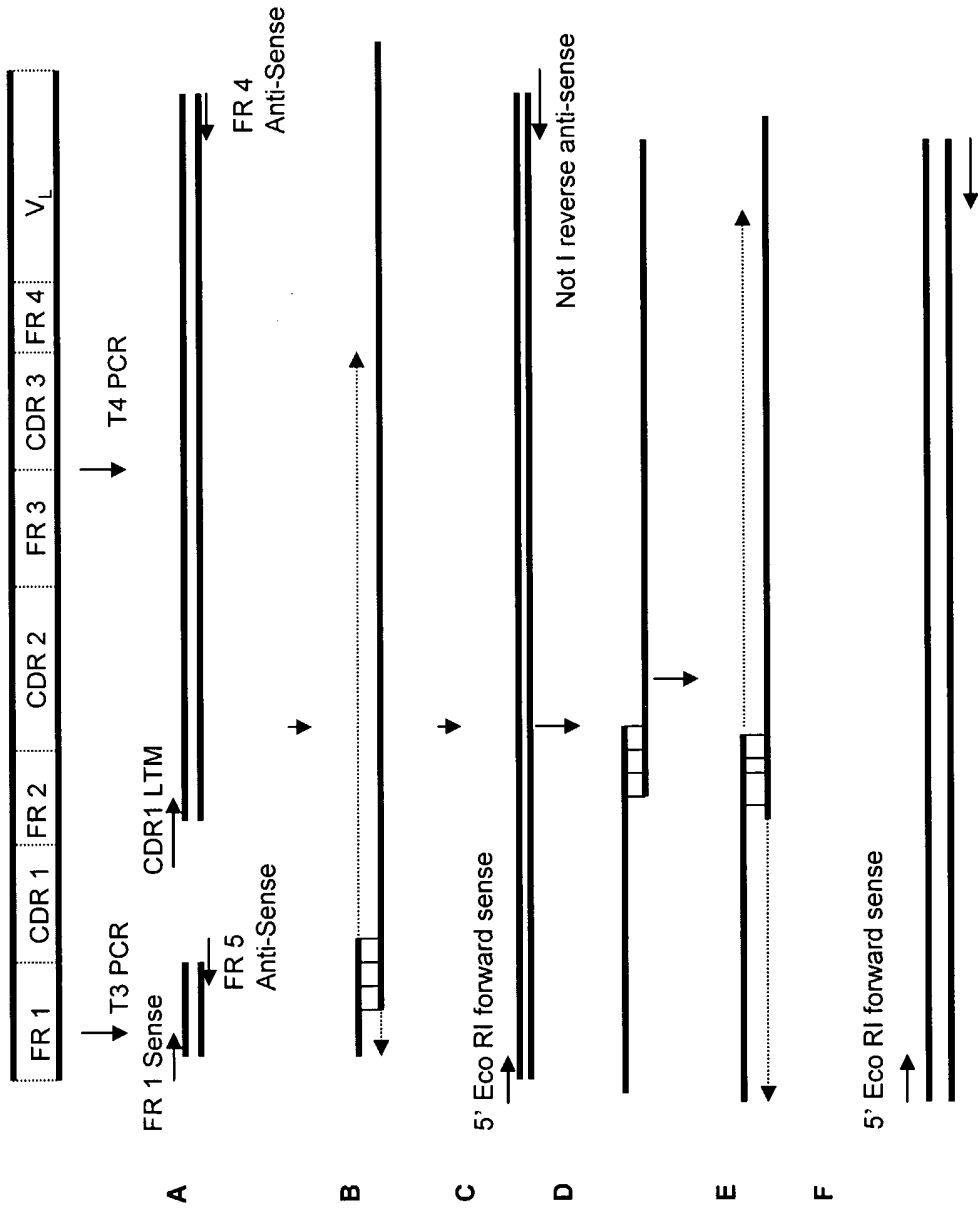


Fig. 4B

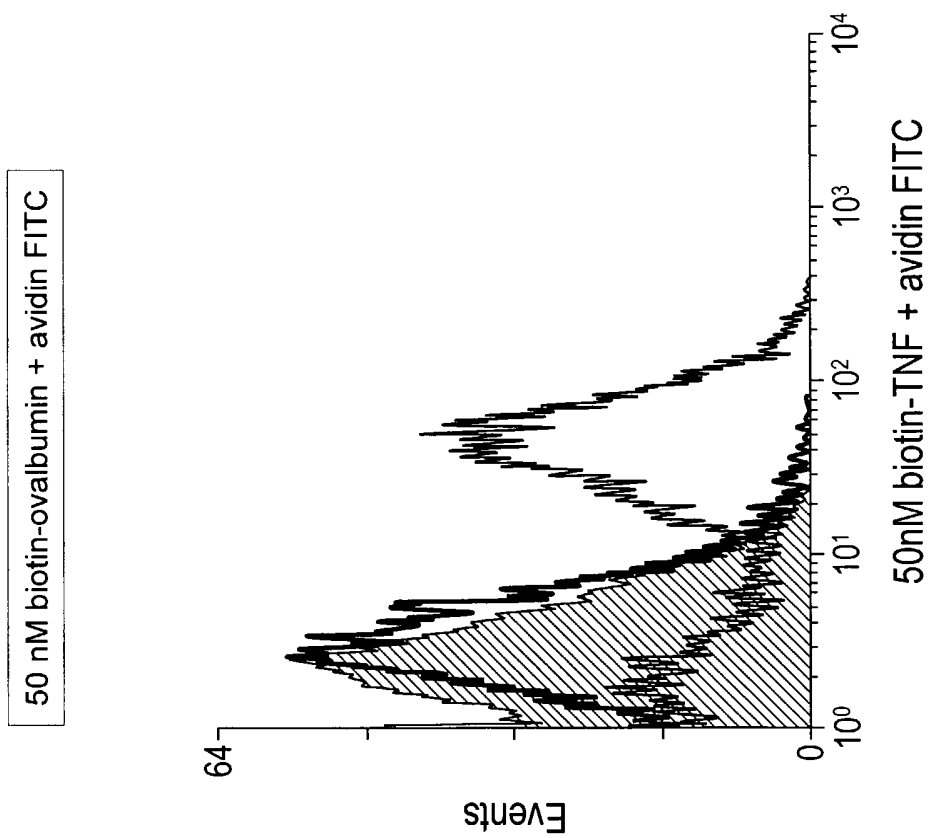
Matrix showing combined light and heavy chain CDR mutagenesis permutations

		heavy chain CDRs							
light chain CDRs	000	100	010	001	110	101	011	111	
000		1	1	1	2	2	2	3	
100	1	2	2	2	3	3	3	4	
010	1	2	2	2	3	3	3	4	
001	1	2	2	2	3	3	3	4	
110	2	3	3	3	4	4	4	5	
101	2	3	3	3	4	4	4	5	
011	2	3	3	3	4	4	4	5	
111	3	4	4	4	5	5	5	6	

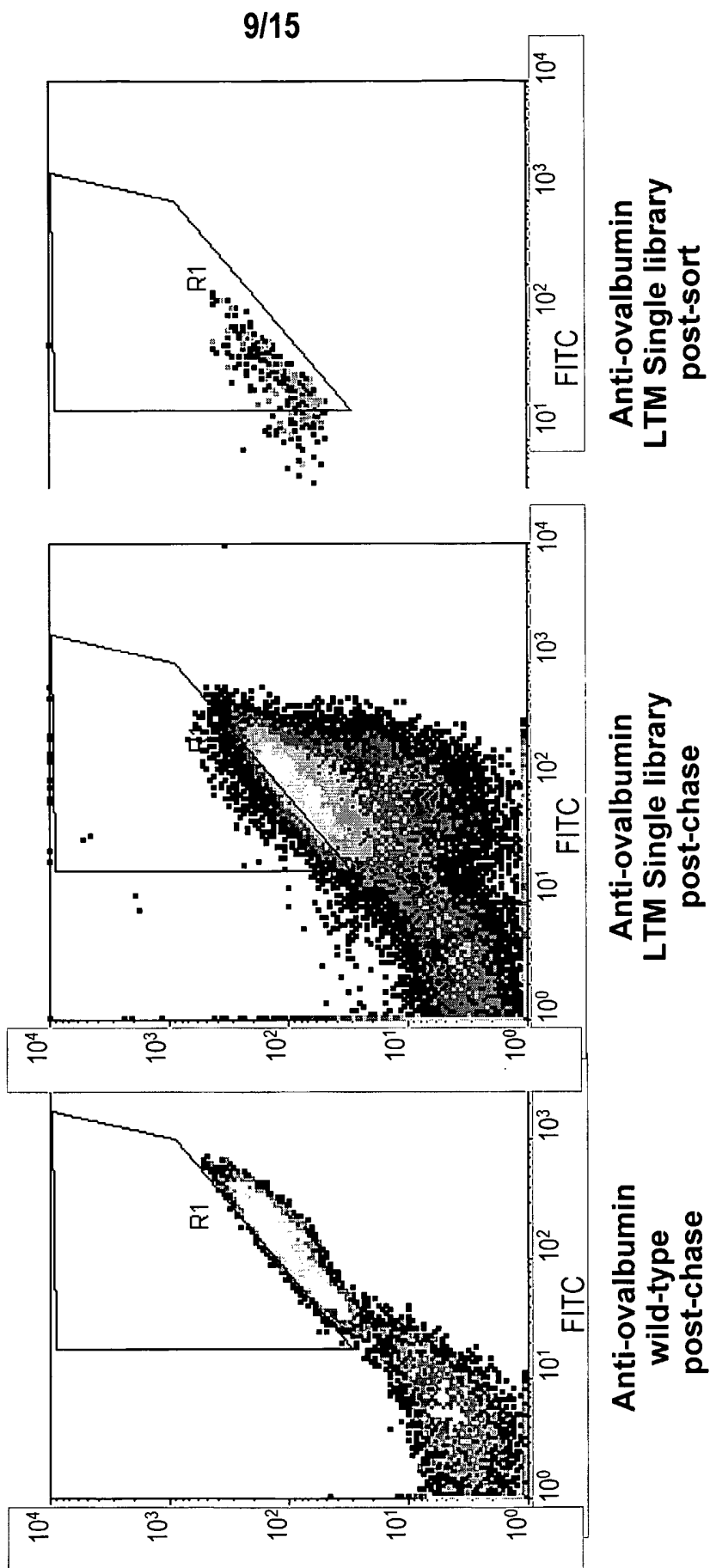
*Fig. 5*

**Schematic of single chain antibody (scFv) yeast expression, display, and screening cassette**

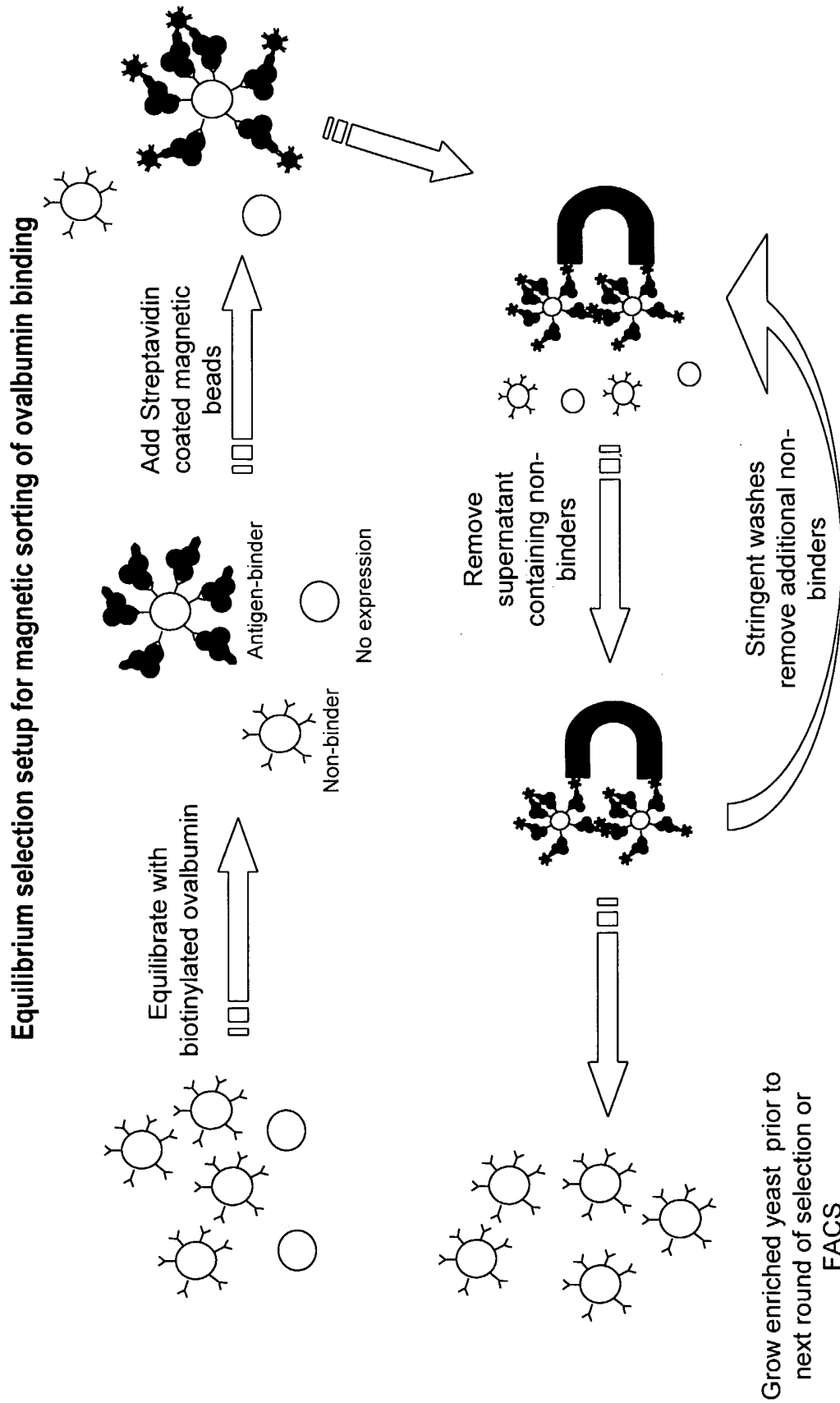




*Fig. 7*



*Fig. 8*



*Fig. 9*

VH CRD2 LTM anti-ovalbumin library selection scheme for increasing "fitness" of the library with each round of selection

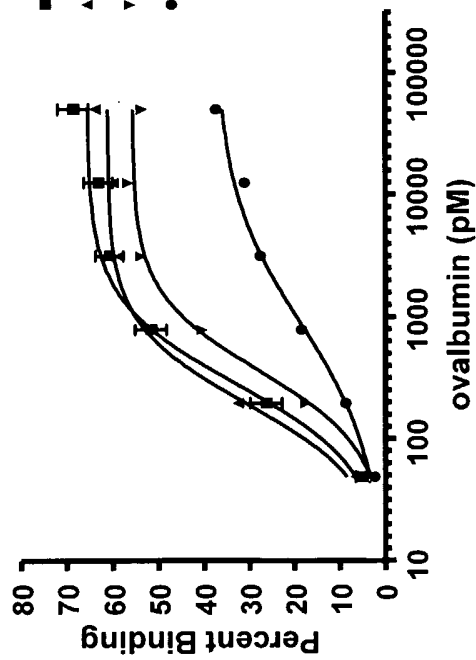
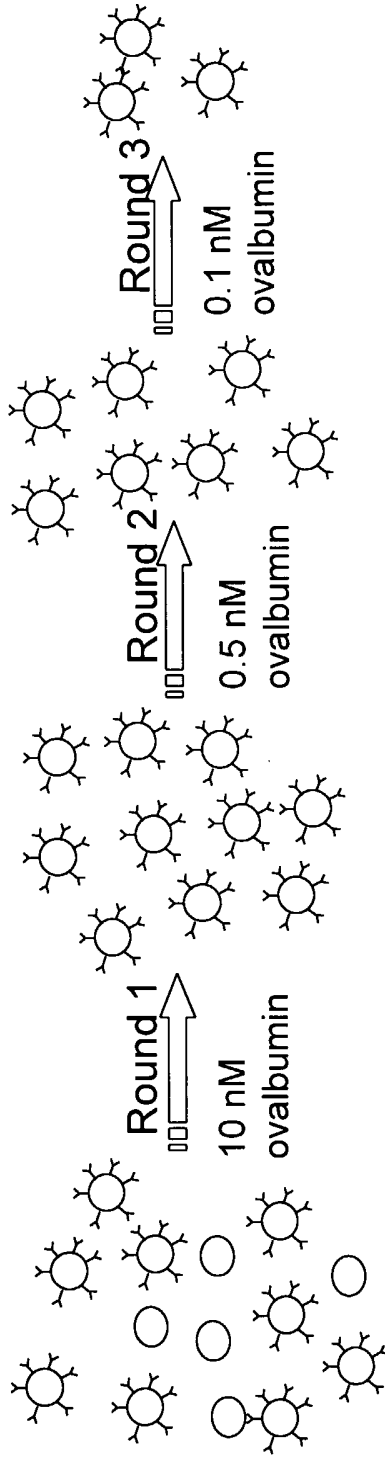
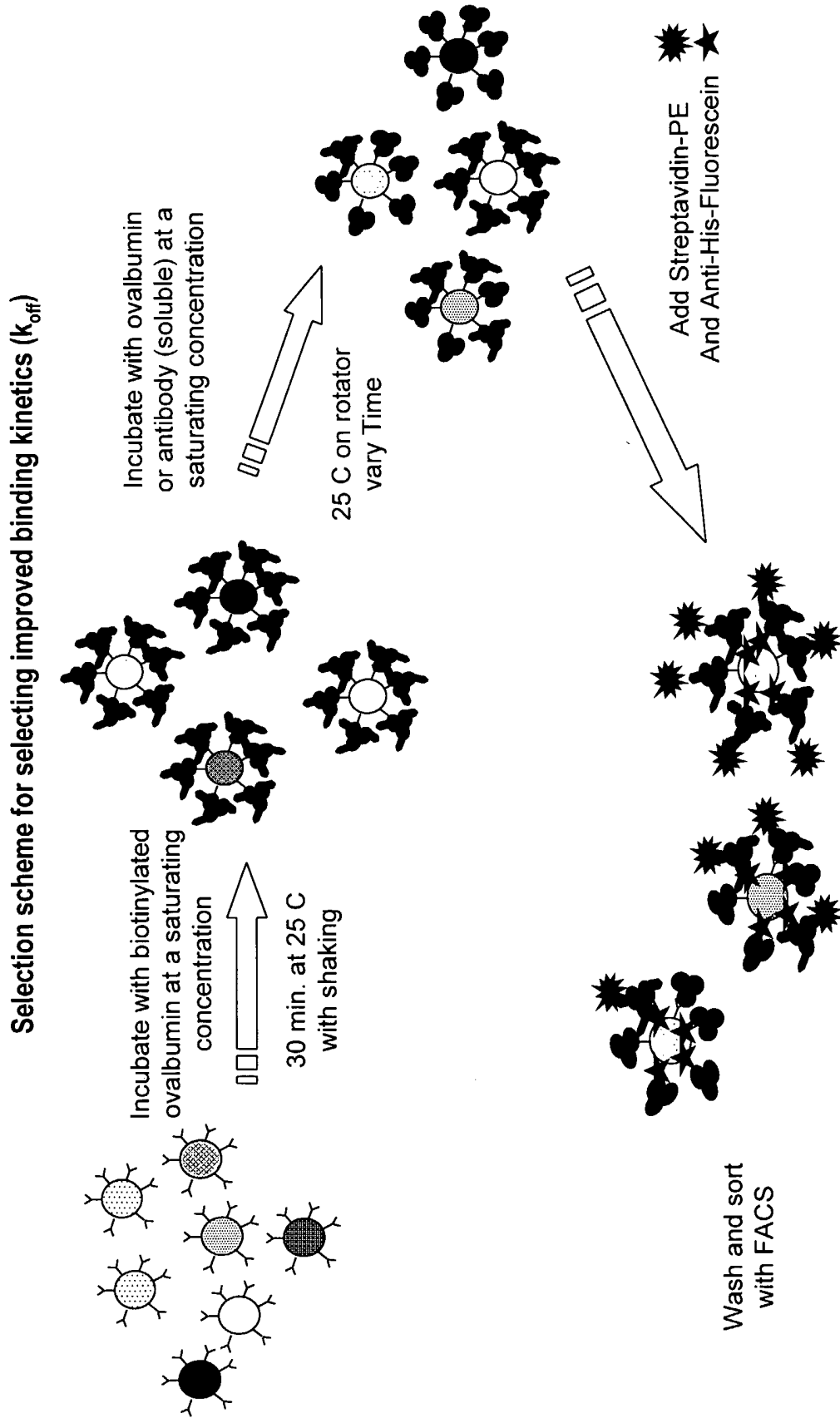


Fig. 10

12/15



*Fig. 11*

Kinetics ( $k_{off}$ ) assay results for improved antibody variants using yeast display

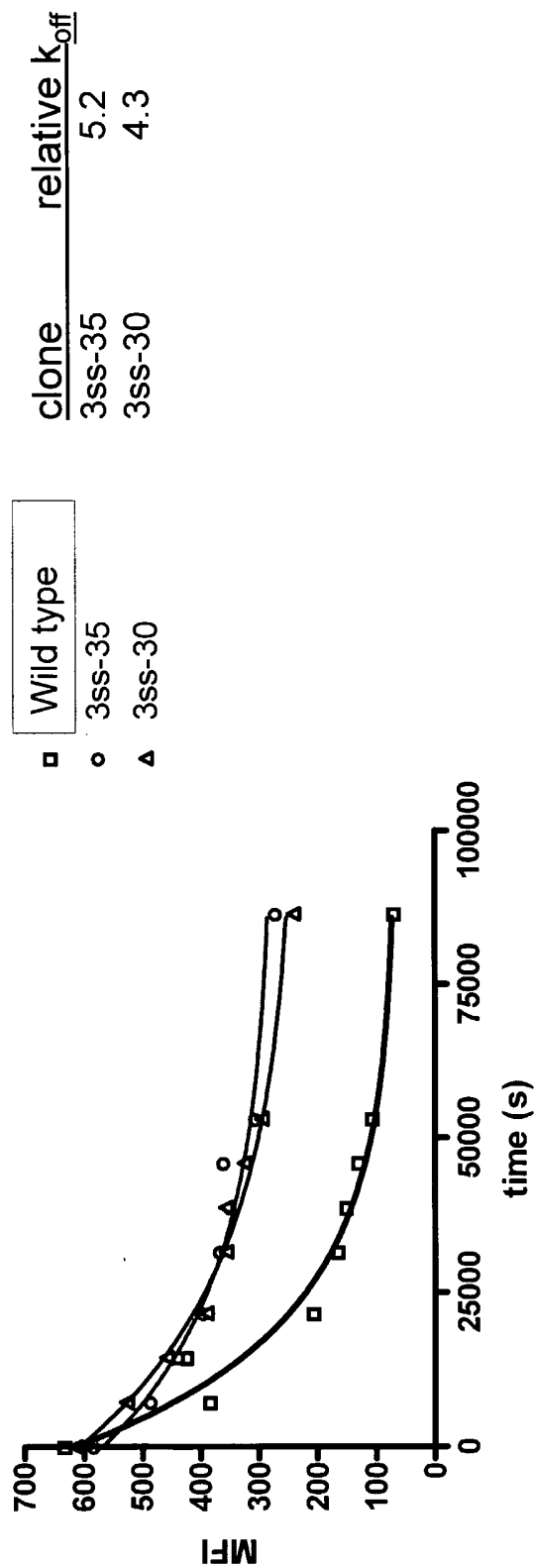


Fig. 12

A subset of improved anti-ovalbumin variants selected from LTM libraries

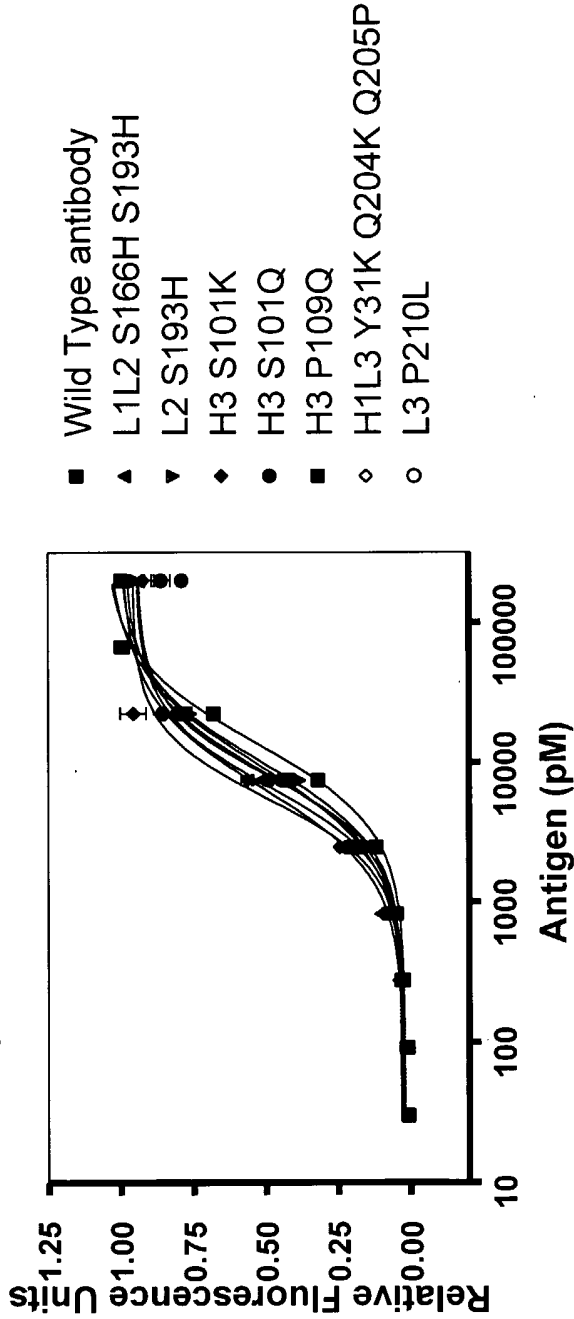
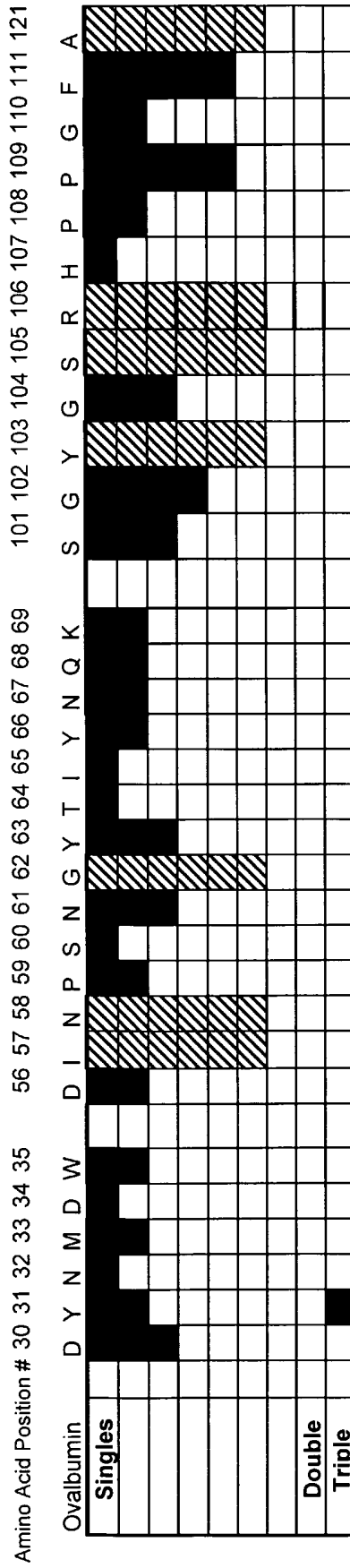


Table 1	WT Anti-ovalbumin	L1L2 S166H S193H	L2 S193H	H3 S101K	H3 S101Q	H3 P109Q	H1L3 Y31K Q204K Q205P	L3 P210L
Relative EC <sub>50</sub>	1.0	0.47	0.72	0.40	0.48	0.71	0.61	0.57
Fold better than WT		<b>4.0</b>	<b>1.4</b>	<b>2.5</b>	<b>2.1</b>	<b>1.4</b>	<b>2.9</b>	<b>1.7</b>

Fig. 13

Improved anti-ovalbumin variants selected from LTM libraries



15/15

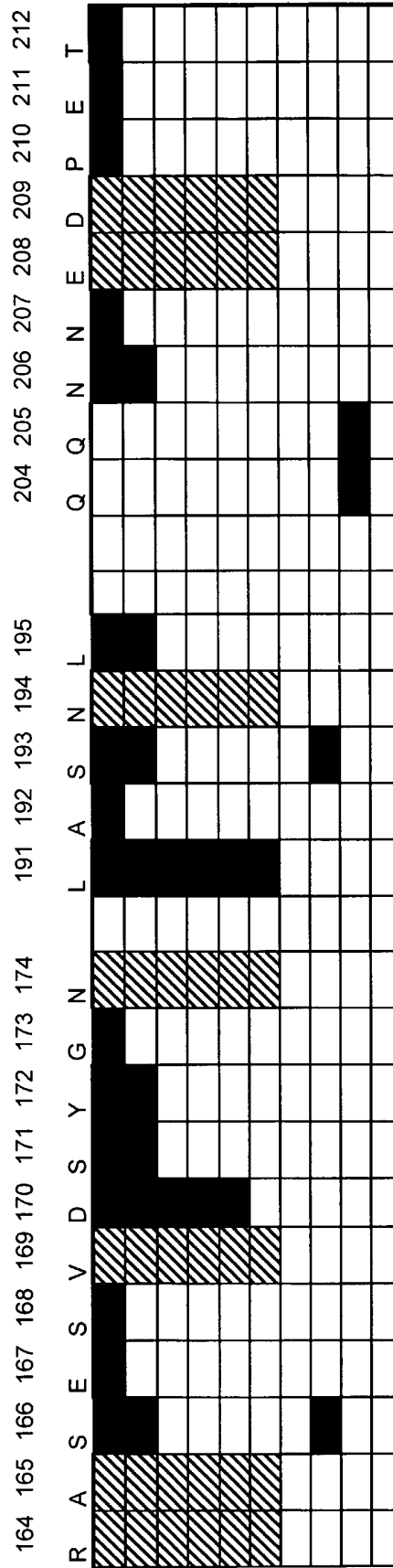


Fig. 14

*Sequence Listing*

SEQ ID NO: 1

nucleotide sequence of anti-ovalbumin antibody V<sub>L</sub>

5

1 atgaaatacctattgcctacggcggccgctggattggttactc  
 46 gctgccaaccagccatggccgacattgtgctgaccaatctcca  
 91 gcttctttggctgtgtctctagggcagagggccaccatatacctgc  
 136 agagccagtgaaagtgttgatagttatggcaatagttttatgcac  
 10 181 tggtagcagcagaaaccaggacagccacccaaactcctcatctat  
 226 cttgcatccaacctagaatctgggggtccctgccaggttcagtgcc  
 271 agtgggtctaggacagacttcaccctcaccattgatcctgtggag  
 316 gctgatgatgctacaacctattactgtcagcaaaataatgaggat  
 406 ccattcacgttcggctcggggacaaagttggaaataaaa

15

SEQ ID NO: 2

nucleotide sequences anti-ovalbumin antibody V<sub>H</sub>

1 gtgaaacaaagcactattgcactggcactcttaccgttactgttt  
 20 46 acccctgtgacaaaagcccaggtccaactgcagcagctctggacct  
 91 gagctggtgaagcctggggcttcagtgaagataccctgcaaggct  
 136 tctggatacacattcactgactacaacatggactgggtgaagcag  
 181 agccatggaaagagccttgagtggattggagatattaatcctagc  
 226 aatggttatactatctacaaccagaagtcaagggcaaggccaca  
 25 271 ttgactgtagacaagtcctccagcacagcctacatggagctccgc  
 316 agcctgacatctgaggacactgcagtctattactgtgcaagatca  
 361 gggtagcggctcacgacatccgcctgggtttgcttactggggccaa  
 406 gggactctgggtcactgtctctgca 429

30 SEQ ID NO: 3

amino acid sequences of the Ovalbumin antibody V<sub>L</sub> chain.  
 Ovalbumin V<sub>L</sub> CDR identified in bold.

Met LysTyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu  
 35 Ala Ala Gln Pro Ala Met Ala Asp Ile Val Leu Thr Gln Ser Pro  
 Ala Ser Leu Ala Val Ser Leu Gly Gln Arg Ala Thr Ile Ser Cys  
**Arg Ala Ser Glu Ser Val Asp Ser Tyr Gly Asn** Ser Phe Met His  
 Trp Tyr Gln Gln Lys Pro Gly Gln Pro Pro Lys Leu Leu Ile Tyr  
**Leu Ala Ser Asn Leu** Glu Ser Gly Val Pro Ala Arg Phe Ser Gly  
 40 Ser Gly Ser Arg Thr Asp Phe Thr Leu Thr Ile Asp Pro Val Glu  
 Ala Asp Asp Ala Thr Thr Tyr Tyr Cys **Gln Gln Asn Asn Glu Asp**  
**Pro Phe Thr** Phe G Ser Gly Thr Lys Leu Glu Ile Lys

SEQ ID NO: 4

45 amino acid sequences of the Ovalbumin antibody V<sub>H</sub> chain.  
 Ovalbumin V<sub>H</sub> CDR identified in bold.

Val Lys Gln Ser Thr Ile Ala Leu Ala Leu Leu Pro Leu Leu Phe  
 Thr Pro Val Thr Lys Ala Gln Val Gln Leu Gln Gln Ser Gly Pro  
 50 Glu Leu Val Lys Pro Gly Ala Ser Val Lys Ile Pro Cys Lys Ala  
 Ser Gly Tyr Thr Phe Thr **Asp Tyr Asn Met Asp** Trp Val Lys Gln  
 Ser His Gly Lys Ser Leu Glu Trp Ile Gly **Asp Ile Asn Pro Ser**  
**Asn Gly Tyr Thr Ile Tyr Asn Gln Lys Phe Lys Gly** Lys Ala Thr  
 Leu Thr Val Asp Lys Ser Ser Ser Thr Ala Tyr Met Glu Leu Arg  
 55 Ser Leu Thr Ser Glu Asp Thr Ala Val Tyr Tyr Cys Ala Arg **Ser**

**Gly Tyr Gly Ser Arg His Pro Pro Gly Phe Ala** Tyr Trp Gly Gln  
 Gly Thr Leu Val Thr Val Ser Ala

5 SEQ ID NO: 5  
 amino acid sequence of CDR1 region of the anti-ovalbumin V<sub>L</sub> chain  
 RASESVDSYGNSEFMH

10 SEQ ID NO: 6  
 amino acid sequence CDR2 region of the anti-ovalbumin V<sub>L</sub> chain  
 LASNLES

15 SEQ ID NO: 7  
 amino acid sequence CDR3 region of the anti-ovalbumin V<sub>L</sub> chain  
 QQNNEDPFT

20 SEQ ID NO: 8  
 amino acid sequence CDR1 region of the anti-ovalbumin V<sub>H</sub> chain  
 DYNMD

SEQ ID NO: 9  
 amino acid sequence CDR2 region of the anti-ovalbumin V<sub>H</sub> chain  
 DINPSNGYTIYNQKFKG

25 SEQ ID NO: 10  
 amino acid sequence CDR3 region of the anti-ovalbumin V<sub>H</sub> chain  
 SGYGSRRHPPGFAY

30 SEQ ID NO: 11  
 forward oligonucleotide for cloning anti-ovalbumin V<sub>L</sub> chain and  
 linker region into scFv

5'- GGT GGA GGC GGT TCT GGT GGA GGC GGT TCG GGT GGC GGA GGT TCA  
 AAA TAC CTA TTG CCT ACG GCA GCG GCT GG -3'

35

SEQ ID NO: 12  
 Not I Reverse end specific oligonucleotide for anti-ovalbumin  
 scFv cloning into pYD1

40

5'-TTT TCC TTT TGC GGC CGC TTT TAT TTC CAA CTT TGT CCC- 3'

SEQ ID NO: 13  
 Eco RI Forward oligonucleotides for cloning anti-ovalbumin V<sub>H</sub>  
 chain into scFv

45

5'- CCG GAA TTC ATG GTG AAA CAA AGC ACT ATT GCA CTG -3'

50 SEQ ID NO: 14  
 Reverse oligonucleotides for anti-ovalbumin V<sub>H</sub> chain including  
 linker for scFv

5'- TGA ACC TCC GCC ACC CGA ACC GCC TCC ACC AGA ACC GCC TCC ACC  
 TGC AGA GAC AGT GAC CAG-3'

55

SEQ ID NO: 15  
amino acid sequence for reference anti-ovalbumin wild type scFv  
construct (includes linker)

5  
QVQLQQSGPELVKPGASVKIPCKASGYTFT**DYNMD**WVKQSHGKSLEWIG**DINPSNGYTI**  
**YNQKFKG**KATLTVDKSSSTAYMELRSLTSEDVAVYYCAR**SGYGSRHPPGF**AYWGQGLVTV  
SSGGGGSGGGGSGGGGSDIVLTQSPASLAVSLGQRATISCRASE**SVDSYGNS**FMHWYQQKP  
GQPPKLLIY**LASNLES**GVPARFSGSGSRTDFTLTIDPVEADDATTYYC**QQNNE**DPFTFGSGT  
10 KLEIKAAAHHHHHHGEQKLISEEDL\*-(Histag)

SEQ ID NO: 16  
reference oligonucleotide coding for the above CDR2 wild type  
sequence (in bold) containing the flanking V<sub>H</sub>2 and V<sub>H</sub>3 portions  
15 (lowercase letters below)

5' -agc ctt gag tgg att gga-**GAT ATT AAT CCT AGC AAT GGT TAT ACT**  
**ATC TAC AAC CAG AAG TTC AAG GGC**-aag gcc aca ttg act gta gac -3'

20 SEQ ID NO: 17  
V<sub>H</sub> CDR2 LTM leucine oligonucleotide  
5' -agc ctt gag tgg att gga-**TTG** ATT AAT CCT AGC AAT GGT TAT ACT  
ATC TAC AAC CAG AAG TTC AAG GGC-aag gcc aca ttg act gta gac -3'

25 SEQ ID NO: 18  
V<sub>H</sub> CDR2 LTM leucine oligonucleotide  
5' -agc ctt gag tgg att gga-GAT **TTG** AAT CCT AGC AAT GGT TAT ACT  
ATC TAC AAC CAG AAG TTC AAG GGC-aag gcc aca ttg act gta gac -3'

30 SEQ ID NO: 19  
V<sub>H</sub> CDR2 LTM leucine oligonucleotide  
5' -agc ctt gag tgg att gga-GAT ATT **TTG** CCT AGC AAT GGT TAT ACT  
ATC TAC AAC CAG AAG TTC AAG GGC-aag gcc aca ttg act gta gac -3'

35 SEQ ID NO: 20  
V<sub>H</sub> CDR2 LTM leucine oligonucleotide  
5' -agc ctt gag tgg att gga-GAT ATT AAT **TTG** AGC AAT GGT TAT ACT  
ATC TAC AAC CAG AAG TTC AAG GGC-aag gcc aca ttg act gta gac -3'

40 SEQ ID NO: 21  
V<sub>H</sub> CDR2 LTM leucine oligonucleotide  
5' -agc ctt gag tgg att gga-GAT ATT AAT CCT **TTG** AAT GGT TAT ACT  
ATC TAC AAC CAG AAG TTC AAG GGC-aag gcc aca ttg act gta gac -3'

45 SEQ ID NO: 22  
V<sub>H</sub> CDR2 LTM leucine oligonucleotide  
5' -agc ctt gag tgg att gga-GAT ATT AAT CCT AGC **TTG** GGT TAT ACT  
ATC TAC AAC CAG AAG TTC AAG GGC-aag gcc aca ttg act gta gac -3'

50 SEQ ID NO: 23  
V<sub>H</sub> CDR2 LTM leucine oligonucleotide  
5' -agc ctt gag tgg att gga-GAT ATT AAT CCT AGC AAT **TTG** TAT ACT  
ATC TAC AAC CAG AAG TTC AAG GGC-aag gcc aca ttg act gta gac -3'

55

SEQ ID NO: 24  
V<sub>H</sub> CDR2 LTM leucine oligonucleotide  
5' -agc ctt gag tgg att gga-GAT ATT AAT CCT AGC AAT GGT **TTG** ACT  
ATC TAC AAC CAG AAG TTC AAG GGC-aag gcc aca ttg act gta gac -3'

5

SEQ ID NO: 25  
V<sub>H</sub> CDR2 LTM leucine oligonucleotide  
5' -agc ctt gag tgg att gga-GAT ATT AAT CCT AGC AAT GGT TAT **TTG**  
ATC TAC AAC CAG AAG TTC AAG GGC-aag gcc aca ttg act gta gac -3'

10

SEQ ID NO: 26  
V<sub>H</sub> CDR2 LTM leucine oligonucleotide  
5' -agc ctt gag tgg att gga-GAT ATT AAT CCT AGC AAT GGT TAT ACT  
**TTG** TAC AAC CAG AAG TTC AAG GGC-aag gcc aca ttg act gta gac -3'

15

SEQ ID NO: 27  
V<sub>H</sub> CDR2 LTM leucine oligonucleotide  
5' -agc ctt gag tgg att gga-GAT ATT AAT CCT AGC AAT GGT TAT ACT  
ATC **TTG** AAC CAG AAG TTC AAG GGC-aag gcc aca ttg act gta gac -3'

20

SEQ ID NO: 28  
V<sub>H</sub> CDR2 LTM leucine oligonucleotide  
5' -agc ctt gag tgg att gga-GAT ATT AAT CCT AGC AAT GGT TAT ACT  
ATC TAC **TTG** CAG AAG TTC AAG GGC-aag gcc aca ttg act gta gac -3'

25

SEQ ID NO: 29  
V<sub>H</sub> CDR2 LTM leucine oligonucleotide  
5' -agc ctt gag tgg att gga-GAT ATT AAT CCT AGC AAT GGT TAT ACT  
ATC TAC AAC **TTG** AAG TTC AAG GGC-aag gcc aca ttg act gta gac -3'

30

SEQ ID NO: 30  
V<sub>H</sub> CDR2 LTM leucine oligonucleotide  
5' -agc ctt gag tgg att gga-GAT ATT AAT CCT AGC AAT GGT TAT ACT  
ATC TAC AAC CAG **TTG** TTC AAG GGC-aag gcc aca ttg act gta gac -3'

35

SEQ ID NO: 31  
V<sub>H</sub> CDR2 LTM leucine oligonucleotide  
5' -agc ctt gag tgg att gga-GAT ATT AAT CCT AGC AAT GGT TAT ACT  
ATC TAC AAC CAG AAG **TTG** AAG GGC-aag gcc aca ttg act gta gac -3'

40

SEQ ID NO: 32  
V<sub>H</sub> CDR2 LTM leucine oligonucleotide  
5' -agc ctt gag tgg att gga-GAT ATT AAT CCT AGC AAT GGT TAT ACT  
ATC TAC AAC CAG AAG TTC **TTG** GGC-aag gcc aca ttg act gta gac -3'

45

SEQ ID NO: 33  
V<sub>H</sub> CDR2 LTM leucine oligonucleotide  
5' -agc ctt gag tgg att gga-GAT ATT AAT CCT AGC AAT GGT TAT ACT  
ATC TAC AAC CAG AAG TTC AAG **TTG** -aag gcc aca ttg act gta gac -3'

50

SEQ ID NO: 34  
FR1 sense oligonucleotide  
5' - ATG AAA CAA AGC ACT ATT GCA CTG GCA CTC TTA CCG TTA -3'

55

SEQ ID NO: 35

FR2 anti-sense oligonucleotide

5'- TCC ATT CCA CTC AAG GCT CTT TCC ATG GCT CTG CTT CAC CAA -3'

5 SEQ ID NO: 36

FR4 anti-sense oligonucleotide

5' - TTT TAT TTC CAA CTT TGT CCC CGC GCC- 3'

SEQ ID NO: 37

10 FR5 anti-sense oligonucleotide

5' - GTC AGT GAA TGT GTA TCC AGA AGC CTT GCA GGG TAT

SEQ ID NO: 38

V<sub>H</sub> CDR1 LTM histidine oligonucleotide

15 5'- gct tct gga tac aca ttc act - **CAC** TAC AAC ATG GAC - tgg gtg  
aag cag agc cat gga-3'

SEQ ID NO: 39

V<sub>H</sub> CDR1 LTM histidine oligonucleotide

20 5'- gct tct gga tac aca ttc act - GAC **CAC** AAC ATG GAC - tgg gtg  
aag cag agc cat gga-3'

SEQ ID NO: 40

V<sub>H</sub> CDR1 LTM histidine oligonucleotide

25 5'- gct tct gga tac aca ttc act - GAC TAC **CAC** ATG GAC - tgg gtg  
aag cag agc cat gga-3'

SEQ ID NO: 40

V<sub>H</sub> CDR1 LTM histidine oligonucleotide

30 5'- gct tct gga tac aca ttc act - GAC TAC AAC **CAC** GAC - tgg gtg  
aag cag agc cat gga-3'

SEQ ID NO: 41

V<sub>H</sub> CDR1 LTM histidine oligonucleotide

35 5'- gct tct gga tac aca ttc act - GAC TAC AAC ATG **CAC** - tgg gtg  
aag cag agc cat gga-3'