Description

Background of the Invention

Mutagenesis is a powerful tool in the study of protein structure and function. Mutations can be made in the nucleotide sequence of a cloned gene encoding a protein of interest and the modified gene can be expressed to produce mutants of the protein. By comparing the properties of a wild-type protein and the mutants generated, it is often possible to identify individual amino acids or domains of amino acids that are essential for the structural integrity and/or biochemical function of the protein, such as its binding and/or catalytic activity.

Mutagenesis, however, is beset by several limitations. Among these are the large number of mutants that can be generated and the practical inability to select from these, the mutants that will be informative or have a desired property. For instance, there is no reliable way to predict whether the substitution, deletion or insertion of a particular amino acid in a protein will have a local or global effect on the protein, and therefore, whether it will be likely to yield useful information or function.

Because of these limitations, attempts to improve properties of a protein by mutagenesis have relied mostly on the generation and analysis of mutations that are restricted to specific, putatively important regions of the protein, such as regions at or around the active site of the protein. But, even though mutations are restricted to certain regions of a protein, the number of potential mutations can be extremely large, making it difficult or impossible to identify and evaluate those produced. For example, substitution of a single amino acid position with all the other naturally occurring amino acids yields 19 different variants of a protein. If several positions are substituted at once, the number of variants increases exponentially. For substitution with all amino acids at seven amino acid positions of a protein, 19 x 19 x 19 x 19 x 19 x 19 x 19 or 8.9×10^8 variants of the protein are generated, from which useful mutants must be selected. It follows that, for an effective use of mutagenesis, the type and number of mutations must be subjected to some restrictive criteria which keep the number of mutant proteins generated to a number suitable for screening.

A method of mutagenesis that has been developed to produce very specific mutations in a protein is site-directed mutagenesis. The method is most useful for studying small sites known or suspected to be involved in a particular protein function. In this method, nucleotide substitutions (point mutations) are made at defined locations in a DNA sequence in order to bring about a desired substitution of one amino acid for another in the encoded amino acid sequence. The method is oligonucleotide-mediated. A synthetic oligonucleotide is constructed that is complementary to the DNA encoding the region of the protein where the mutation is to be made, but which bears an unmatched base(s) at the desired position(s) of the base substitution(s). The mutated oligonucleotide is used to prime the synthesis of a new DNA strand which incorporates the change(s) and, therefore, leads to the synthesis of the mutant gene. See Zoller, M. J. and Smith, M., Meth. Enzymol. 100, 468 (1983).

Variations of site-directed mutagenesis have been developed to optimize aspects of the procedure. For the most part, they are based on the original methods of Hutchinson, C.A. et al., J. Biol. Chem. 253:6551 (1978) and Razin, A. et al., Proc. Natl. Acad. Sci. USA 75:4268 (1978). For an extensive description of site-directed mutagensis, see Molecular Cloning, A Laboratory Manual, 1989, Sambrook, Fritsch and Maniatis, Cold Spring Harbor, New York, chapter 15.

A method of mutagenesis designed to produce a larger number of mutations is the "saturation" mutagenesis. This process is oligonucleotide-mediated also. In this method, all possible point mutations (nucleotide substitutions) are made at one or more positions within DNA encoding a given region of a protein. These mutations are made by synthesizing a single mixture of oligonucleotides which is inserted into the gene in place of the natural segment of DNA encoding the region. At each step in the synthesis, the three non-wild type nucleotides are incorporated into the oligonucleotides along with the wild type nucleotide. The non-wild type nucleotides are incorporated at a predetermined percentage, so that all possible variations of the sequence are produced with anticipated frequency. In this way, all possible nucleotide substitutions are made within a defined region of a gene, resulting in the production of many mutant proteins in which the amino acids of a defined region vary randomly (Oliphant, A.R. et al., Meth. Enzymol. 155:568 (1987)).

Methods of random mutagenesis, such as saturation mutagenesis, are designed to compensate for the inability to predict where mutations should be made to yield useful information or functional mutants. The methods are based on the principle that, by generating all or a large number of the possible variants of relevant protein domains, the proper arrangement of amino acids is likely to be produced as one of the randomly generated mutants. However, for completely random combinations of mutations, the numbers of mutants generated can overwhelm the capacity to select meaningfully. In practice, the number of random

mutations generated must be large enough to be likely to yield the desired mutations, but small enough so that the capacity of the selection system is not exceeded. This is not always possible given the size and complexity of most proteins.

Summary of the Invention

This invention pertains to a method of mutagenesis for the generation of novel or improved proteins (or polypeptides) and to libraries of mutant proteins and specific mutant proteins generated by the method. The protein, peptide or polypeptide targeted for mutagenesis can be a natural, synthetic or engineered protein, peptide or polypeptide or a variant (e.g., a mutant). In one embodiment, the method comprises introducing a predetermined amino acid into each and every position in a predefined region (or several different regions) of the amino acid sequence of a protein. A protein library is generated which contains mutant proteins having the predetermined amino acid in one or more positions in the region and, collectively, in every position in the region. The method can be referred to as "walk-through" mutagenesis because, in effect, a single, predetermined amino acid is substituted position-by-position throughout a defined region of a protein. This allows for a systematic evaluation of the role of a specific amino acid in the structure or function of a protein.

The library of mutant proteins can be generated by synthesizing a single mixture of oligonucleotides which encodes all of the designed variations of the amino acid sequence for the region containing the predetermined amino acid. This mixture of oligonucleotides is synthesized by incorporating in each condensation step of the synthesis both the nucleotide of the sequence to be mutagenized (for example, the wild type sequence) and the nucleotide required for the codon of the predetermined amino acid. Where a nucleotide of the sequence to be mutagenized is the same as a nucleotide for the predetermined amino acid, no additional nucleotide is added. In the resulting mixture, oligonucleotides which contain at least one codon for the predetermined amino acid make up from about 12.5% to 100% of the constituents. In addition, the mixture of oligonucleotides encodes a statistical (in some cases Gaussian) distribution of amino acid sequences containing the predetermined amino acid in a range of no positions to all positions in the sequence.

The mixture of oligonucleotides is inserted into a gene encoding the protein to be mutagenized (such as the wild type protein) in place of the DNA encoding the region. The recombinant mutant genes are cloned in a suitable expression vector to provide an expression library of mutant proteins that can be screened for proteins that have desired properties. The library of mutant proteins produced by this oligonucleotide-mediated procedure contains a larger ratio of informative mutants (those containing the predetermined amino acid in the defined region) relative to noninformative mutants than libraries produced by methods of saturation mutagenesis. For example, preferred libraries are made up of mutants which have the predetermined amino acid in essentially each and every position in the region at a frequency ranging from about 12.5% to 100%.

This method of mutagenesis can be used to generate libraries of mutant proteins which are of a practical size for screening. The method can be used to study the role of specific amino acids in protein structure and function and to develop new or improved proteins and polypeptides such as enzymes, antibodies, binding fragments or analogues thereof, single chain antibodies and catalytic antibodies.

Brief Description of the Figures

Figure 1 is a schematic depiction a "walk-through" mutagenesis of the Fv region of immunoglobulin MCPC 603, performed for the CDR1 (Asp) and CDR3 (Ser) of the heavy(H) chain and CDR2 (His) of the light chain (L).

Figure 2 is a schematic depiction of a "walk-through" mutagenesis of an enzyme active site; three amino acid regions of the active site are substituted in each and every position with amino acids of a serine-protease catalytic triad.

Figure 3 illustrates the design of "degenerate" oligonucleotides for walk-through mutagenesis of the CDR1 (Figure 3a) and CDR3 (Figure 3b) of the heavy chain, and CDR2 (Figure 3c) of the light chain of MCPC 603.

Figure 4 illustrates the design of a "window" of mutagenesis, and shows the sequences of degenerate oligonucleotides for mutation of CDR3 of the heavy chain (Figure 4a) and CDR2 of the light chain of MCPC 603 (Figure 4b).

Figures 5a and 5b illustrate the design of "windows" of mutagenesis and show the sequences of degenerate oligonucleotides for two different wlk-through mutagenesis procedures with His in CDR2 of the

heavy chain of MCPC 603.

Figure 6 illustrates the design and sequences of degenerate oligonucleotides for walk-through mutagenesis of CDR2 of the heavy chain of MCPC 603.

Figure 7 illustrates a "window" of mutagenesis in the HIV protease, consisting of three consecutive amino acid residues at the catalytic site. The design and sequences of degenerate oligonucleotides for three rounds of walk-through mutagenesis of the region with Asp, Ser and His is shown.

Figure 8 illustrates the design and sequence of degenerate oligonucleotides for walk-through mutagenesis of five CDRs of MCPC 603. The degenerate oligonucleotides for walk-through mutagenesis of the CDR1 (Figure 8a) and CDR3 (Figure 8b) of the light chain, and of CDR 1 (Figure 8c), CDR2 (Figure 8d), and CDR3 (Figure 8e) of the heavy chain are shown.

Detailed Description of the Invention

The study of proteins has revealed that certain amino acids play a crucial role in their structure and function. For example, it appears that only a discrete number of amino acids participate in the catalytic event of an enzyme. Serine proteases are a family of enzymes present in virtually all organisms, which have evolved a structurally similar catalytic site characterized by the combined presence of serine, histidine and aspartic acid. These amino acids form a catalytic triad which, possibly along with other determinants, stabilizes the transition state of the substrate. The functional role of this catalytic triad has been confirmed by individual and by multiple substitutions of serine, histidine and aspartic acid by site-directed mutagenesis of serine proteases and the importance of the interplay between these amino acid residues in catalysis is now well established. These same three amino acids are involved in the enzymatic mechanism of certain lipases as well. Similarly, a large number of other types of enzymes are characterized by the peculiar conformation of their catalytic site and the presence of certain kinds of amino acid residues in the site that are primarily responsible for the catalytic event. For an extensive review, see Enzyme Structure and Mechanism, 1985, by A. Fersht, Freeman Ed., New York.

Though it is clear that certain amino acids are critical to the mechanism of catalysis, it is difficult, if not impossible, to predict which position (or positions) an amino acid must occupy to produce a functional site such as a catalytic site. Unfortunately, the complex spatial configuration of amino acid side chains in proteins and the interrelationship of different side chains in the catalytic pocket of enzymes are insufficiently understood to allow for such predictions. As pointed out above, 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 proteins.

The method of this invention provides a systematic and practical approach for evaluating the importance of particular amino acids, and their position within a defined region of a protein, to the structure or function of a protein and for producing useful proteins. The method begins with the assumption that a certain, predetermined amino acid is important to a particular structure or function. The assumption can be based on a mere guess. More likely, the assumption is based upon what is known about the amino acid from the study of other proteins. For example, the amino acid can be one which has a role in catalysis, binding or another function.

With selection of the predetermined amino acid, a library of mutants of the protein to be studied is generated by incorporating the predetermined amino acid into each and every position of the region of the protein. As schematically depicted in Figures 1 and 2, the amino acid is substituted in or "walked-through" all (or essentially all) positions of the region.

The library of mutant proteins contains individual proteins which have the predetermined amino acid in each and every position in the region. The protein library will have a higher proportion of mutants that contain the predetermined amino acid in the region (relative to mutants that do not), as compared to libraries that would be generated by completely random mutation, such as saturation mutation. Thus, the desired types of mutants are concentrated in the library. This is important because it allows more and larger regions of proteins to be mutagenized by the walk-through process, while still yielding libraries of a size which can be screened. Further, if the initial assumption is correct and the amino acid is important to the structure or function of the protein, then the library will have a higher proportion of informative mutants than a library generated by random mutation.

In another embodiment, a predetermined amino acid is introduced into each of certain selected positions witin a predefined region or regions. Certain selected positions may be known or thought to be more promising due to structural constraints. Such considerations, based on structural information or modeling of the molecule mutagenized and/or the desired structure, can be used to select a subset of positions within a region or regions for mutagenesis. Thus, the amino acids mutagenized within a region

need not be contiguous. Walking an amino acid through certain selected positions in a region can minimize the number of variants produced.

The size of a 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^{10} mutants, and more preferably less than 10^9 mutants.

In a preferred embodiment, the library of mutant proteins is generated by synthesizing a mixture of oligonucleotides (a degenerate oligonucleotide) encoding selected permutations of amino acid sequences for the defined region of the protein. Conveniently, the mixture of oligonucleotides can be produced in a single synthesis. This is accomplished by incorporating, at each position within the oligonucleotide, both a nucleotide required for synthesis of the wild-type protein (or other protein to be mutagenized) and a single appropriate nucleotide required for a codon of the predetermined amino acid. (This differs from the oligonucleotides produced in saturation mutagenesis in that, for each DNA position mutagenized, only a single additional nucleotide, as opposed to three for "saturation", is added). The two nucleotides are typically, but not necessarily, used in approximately equal concentrations for the reaction so that there is an equal chance of incorporating either one into the sequence at the position. When the nucleotide of the wild type sequence and the nucleotide for the codon of the predetermined amino acid are the same, no additional nucleotide is incorporated.

Depending upon the number of nucleotides that are mutated to provide a codon for a predetermined amino acid, the mixture of oligonucleotides will generate a limited number of new codons. For example, if only one nucleotide is mutated, the resulting DNA mixture will encode either the original codon or the codon of the predetermined amino acid. In this case, 50% of all oligonucleotides in the resulting mixture will contain the codon for the predetermined amino acid at that position. If two nucleotides are mutated in any combination (first and second, first and third or second and third), four different codons are possible and at least one will encode the predetermined amino acid, a 25% frequency. If all three bases are mutated, then the mixture will produce eight distinct codons, one of which will encode the predetermined amino acid. Therefore the codon will appear in the position with a minimum frequency of 12.5%. However, it is likely that an additional one of the eight codons would code for the same amino acid and/or a stop codon and accordingly, the frequency of predetermined amino acid would be greater than 12.5%.

By this method, a mixture of oligonucleotides is produced having a high proportion of sequences containing a codon for the predetermined amino acid. Other restrictions in the synthesis can be imposed to increase this proportion (by reducing the number of oligonucleotides in the mixture that do not contain at least one codon for the predetermined amino acid). For example, when a complete codon (three nucleotides) must be substituted to arrive at the codon for the predetermined amino acid, the substitute nucleotides only may be introduced (so that the codon for the predetermined amino acid appears with 100% frequency at the position). The proportions of the wild type nucleotide and the nucleotide coding for the preselected amino acid may be adjusted at any or all positions to influence the proportions of the encoded amino acids.

In a protein library produced by this procedure, the proportion of mutants which have at least one residue of the predetermined amino acid in the defined region ranges from about 12.5% to 100% of all mutants in the library (assuming approximately equal proportions of wild type bases and preselected amino acid bases are used in the synthesis). Typically, the proportion ranges from about 25% to 50%.

The libraries of protein mutants will contain a number equal to or smaller than 2^n, where n represents the number of nucleotides mutated within the DNA encoding the protein region. Because there can be only a limited number of changes for each codon (one, two or three) the number of protein mutants will range from 2^m to 8^m, where m is the number of amino acids that are mutated within that region. This represents a dramatic reduction compared with the 19^m mutants generated by a saturation mutagenesis. For instance, for a protein region of seven amino acids, the number of mutants generated by a walk-through mutagenesis (of one amino acid) would result in a 0.000014% to 0.24% fraction of the number of mutants that would be generated by saturation mutagenesis of the region, a very significant reduction.

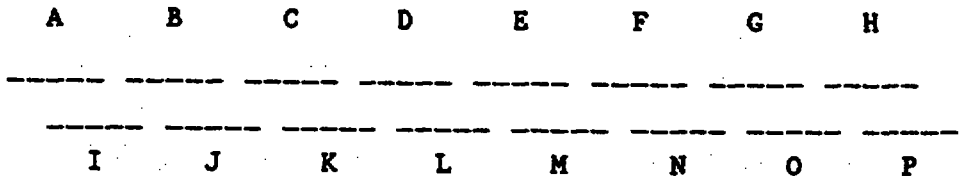
An additional, advantageous characteristic of the library generated by this method is that the proteins which contain the predetermined amino acid conform to a statistical distribution with respect to the number of residues of the amino acid in the amino acid sequence. Accordingly, the sequences range from those in which the predetermined amino acid does not appear at any position in the region to those in which the predetermined amino acid appears in every position in the region. Thus, in addition to providing a means for systematic insertion of an amino acid into a region of a protein, this method provides a way to enrich a region of a protein with a particular amino acid. This enrichment could lead to enhancement of an activity attributable to the amino acid or to entirely new activities.

The mixture of oligonucleotides for generation of the library can be synthesized readily by known methods for DNA synthesis. The preferred method involves use of solid phase beta-cyanoethyl phosphoramidite chemistry. See U.S. Patent No. 4,725,677. For convenience, an instrument for automated DNA synthesis can be used containing ten reagent vessels of nucleotide synthons (reagents for DNA synthesis), four vessels containing one of the four synthons (A, T, C and G)and six vessels containing mixtures of two synthons (A+T, A+C, A+G, T+C, T+G and C+G).

The wild type nucleotide sequence can be adjusted during synthesis to simplify the mixture of oligonucleotides and minimize the number of amino acids encoded. For example, if the wild type amino acid is threonine (ACT), and the preselected amino acid is arginine (AGA or AGG), two base changes are required to encode arginine, and three amino acids are produced (e.g., AGA, Arg; AGT, Ser; ACA, ACT Thr). By changing the wild type nucleotide sequence to ACA or ACG, only a single base change would be required to encode arginine. Thus, if ACG were chosen to encode the wild type threonine instead of ACT, only the central base would need to be changed to G to obtain arginine, and only arginine and threonine would be produced at that position. Depending on the particular codon and the identity of the preselected amino acid, similar adjustments at any position of the wild type codon may reduce the number of variants generated.

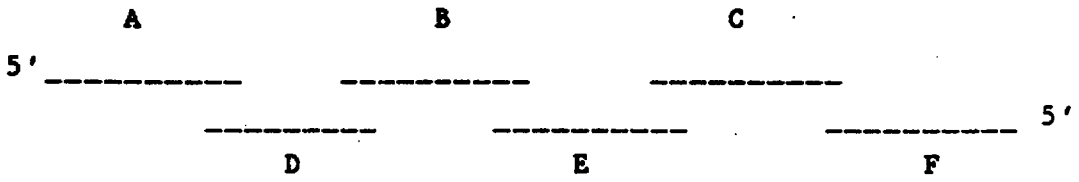
The mixture of oligonucleotides is inserted into a cloned gene of the protein being mutagenized in place of the nucleotide sequence encoding the amino acid sequence of the region to produce recombinant mutant genes encoding the mutant proteins. To facilitate this, the mixture of oligonucleotides can be made to contain flanking recognition sites for restriction enzymes. See Crea, R., U.S. Patent No. 4,888,286. The recognition sites are designed to correspond to recognition sites which either exist naturally or are introduced in the gene proximate to the DNA encoding the region. After conversion into double stranded form, the oligonucleotides are ligated into the gene by standard techniques. By means of an appropriate vector, the genes are introduced into a host cell suitable for expression of the mutant proteins. See e.g., Huse, W.D. et al., Science 246:1275 (1989); Viera, J. et al., Meth. Enzymol. 153:3 (1987).

In fact, the degenerate oligonucleotides can be introduced into the gene by any suitable method, using techniques well-known in the art. In cases where the amino acid sequence of the protein to be mutagenized is known or where the DNA sequence is known, gene synthesis is a possible approach (see e.g., Alvarado-Urbina, G. et al., Biochem. Cell. Biol. 64: 548-555 (1986); Jones et al., Nature 321: 522 (1986)). For example, partially overlapping oligonucleotides, typically about 20-60 nucleotides in length, can be designed. The internal oligonucleotides (B through G and I through O) are phosphorylated using T4 polynucleotide kinase to provide a 5' phosphate group. Each of the oligonucleotides can be 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 :



Convenient restriction sites can be designed near the ends of the synthetic gene for cloning into a suitable vector. The full length molecules can be cleaved with those restriction enzymes, gel purified, electroeluted 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 oligonucleotides representing the full-length double-stranded gene, oligonucleotides 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 the Klenow fragment of DNA polymerase and deoxynucleotide triphosphates to make a full length double-stranded gene. Typically, the overlapping oligonucleotides are from 40-90 nucleotides in length. The extended oligonucleotides are then ligated using T4 ligase. 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 gel-purified and ligated into a suitable vector. Alternatively, the gene fragment could be blunt end ligated into an appropriate vector.

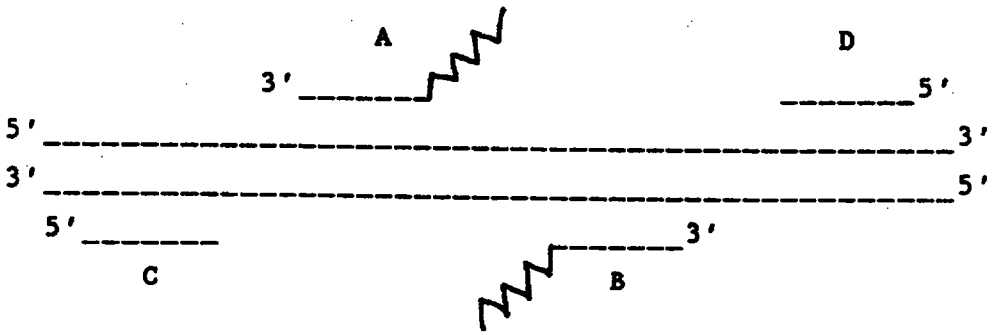


In these approaches, if convenient restriction sites are available (naturally or engineered) following gene assembly, the degenerate oligonucleotides can be introduced subsequently by cloning the cassette into an appropriate vector. Alternatively, the degenerate oligonucleotides can be incorporated at the stage of gene assembly. For example, when both strands of the gene are fully chemically synthesized, overlapping and complementary degenerate oligonucleotides can be produced. Complementary pairs will anneal with each other. An example of this approach is illustrated in Example 1.

When partially overlapping oligos are used in the gene assembly, a set of degenerate nucleotides can also be directly incorporated in place of one of the oligos. The appropriate complementary strand is synthesized during the extension reaction from a partially complementary oligo from the other strand by enzymatic extension with the Klenow fragment of DNA polymerase, for example. Incorporation of the degenerate oligonucleotides at the stage of synthesis also simplifies cloning where more than one domain of a gene is mutagenized.

In another approach, the gene of interest is present on a single stranded plasmid. For example, the gene can be cloned into an M13 phage vector or a vector with a filamentous phage origin of replication which 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 probes. The probes can be elongated and ligated, thus incorporating each variant strand into a population of molecules which can be introduced into an appropriate host (Sayers, J.R. et al., Nucleic Acids Res. 16: 791-802 (1988)). This approach can circumvent multiple cloning steps where multiple domains are selected for mutagenesis.

Polymerase chain reaction (PCR) methodology can also be used to incorporate degenerate oligonucleotides into a gene. For example, the degenerate oligonucleotides themselves can be used as primers for extension.



In this embodiment, A and B are populations of degenerate oligonucleotides encoding the mutagenic cassettes or "windows", and the windows are complementary to each other (the zig-zag portion of the oligos represents the degenerate portion). A and B also contain wild type sequences complementary to the template on the 3' end for amplification and are thus primers for amplification capable of generating fragments incorporating a window. C and D are oligonucleotides which can amplify the entire gene or region of interest, including those with mutagenic windows incorporated (Steffan, N.H. et al., Gene 77: 51-59 (1989)). The extension products primed from A and B can hybridize through their complementary windows and provide a template for production of full-length molecules using C and D as primers. C and D can be designed to contain convenient sites for cloning. The amplified fragments can then be cloned.

Libraries of mutants generated by any of the above techniques or other suitable techniques can be screened to identify mutants of desired structure or activity. The screening can be done by any appropriate means. For example, catalytic activity can be ascertained by suitable assays for substrate conversion and binding activity can be evaluated by standard immunoassay and/or affinity chromatography.

The method of this invention can be used to mutagenize any region of a protein, protein subunit or polypeptide. The description heretofore has centered around proteins, but it should be understood that the method applies to polypeptides and multi-subunit proteins as well. The regions mutagenized by the method of this invention can be continuous or discontinuous and will generally range in length from about 3 to about 30 amino acids, typically 5 to 20 amino acids.

Usually, the region studied will be a functional domain of the protein such as a binding or catalytic domain. For example, the region can be the hypervariable region (complementarity-determining region or CDR) of an immunoglobulin, the catalytic site of an enzyme, or a binding domain.

As mentioned, the amino acid chosen for the "walk through" mutagenesis is generally selected from those known or thought to be involved in the structure or function of interest. The twenty naturally occurring amino acids differ only with respect to their side chain. Each side chain is reponsible for chemical properties that make each amino acid unique. For review, see Principles of Protein Structure, 1988, by G.E. Schulz and R. M. Schirner, Springer-Verlag.

From the chemical properties of the side chains, it appears that only a selected number of natural amino acids preferentially participate in a catalytic event. 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.

Typical polar and neutral side chains are those of Cys, Ser, Thr, Asn, Gln and Tyr. Gly is also considered to be a borderline member of this group. Ser and Thr play an important role 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_2 group than Asn which renders the polar group more flexible and reduces its interaction with the main chain. 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 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.

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

In addition, Lys and Arg are found at the surface. They have long and flexible side chains. Wobbling in the surrounding solution, they increase the solubility of the protein globule. 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 proteins, Lys is a residue more frequently attacked by enzymes which either modify the side chain or cleave the peptide chain at the carbonyl end of Lys residues.

For the purpose of introducing catalytically important amino acids into a region, the invention preferentially relates to a mutagenesis in which the predetermined amino acid is one of the following group of amino acids: Ser, Thr, Asn, Gln, Tyr, Cys, His, Glu, Asp, Lys, and Arg. However, for the purpose of altering binding or creating new binding affinities, any of the twenty naturally occurring amino acids can be selected.

Importantly, several different regions or domains of a protein can be mutagenized simultaneously. The same or a different amino acid can be "walked-through" each region. This enables the evaluation of amino acid substitutions in conformationally related regions such as the regions which, upon folding of the protein, are associated to make up a functional site such as the catalytic site of an enzyme or the binding site of an antibody. This method provides a way to create modified or completely new catalytic sites. As depicted in Figure 1, the six hypervariable regions of an immunoglobulin, which make up the unique aspects of the antigen binding site (Fv region), can be mutagenized simultaneously, or separately within the V_H or V_L chains, to study the three dimensional interrelationship of selected amino acids in this site.

The method of this invention opens up new possibilities for the design of many different types of proteins. The method can be used to improve upon an existing structure or function of a protein. For example, the introduction of additional "catalytically important" amino acids into a catalytic domain of an enzyme may result in enhanced catalytic activity toward the same substrate. Alternatively, entirely new structures, specificities or activities may be introduced into a protein. De novo synthesis of enzymatic

activity can be achieved as well. The new structures can be built on the natural "scaffold" of an existing protein by mutating only relevant regions by the method of this invention.

The method of this invention is especially useful for modifying antibody molecules. As used herein, antibody molecules or antibodies refers to antibodies or portions thereof, such as full-length antibodies, Fv molecules, or other antibody fragments, individual chains or fragments thereof (e.g., a single chain of Fv), single chain antibodies, and chimeric antibodies. Alterations can be introduced into the variable region and/or into the framework (constant) region of an antibody. Modification of the variable region can produce antibodies with better antigen binding properties, and catalytic properties. Modification of the framework region could lead to the improvement of chemo-physical properties, such as solubility or stability, which would be useful, for example, in commercial production. Typically, the mutagenesis will target the Fv region of the immunoglobulin molecule - the structure responsible for antigen-binding activity which is made up of variable regions of two chains, one from the heavy chain (V_H) and one from the light chain (V_L).

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

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

The method of this 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 this 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, could be engineered. Such proteases could be useful anti-thrombotic agents or anti-viral agents against viruses such as AIDS, 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.

Assays for these activities can be designed in which a cell requires the desired activity for growth. For example, in screening for activites that degrade toxic compounds, the incorportation of lethal levels of the the toxic compound into nutrient plates would permit the growth only of cells expressing an activity which degrades the toxic compound (Wasserfallen, A., Rekik, M., and Harayama, S., Biotechnology 9: 296-298 (1991)). Alternatively, in screening for an enzyme that uses a non-toxic substrate, it is possible to use that substrate as the sole carbon source or sole source of another appropriate nutrient. In this case also, only cells expressing the enzyme activity will grow on the plates. In these methods, it is not necessary that the enzyme activity be secreted if the substrate or a product of the substrate (converted extracellularly by another activity) can be taken up by the cell. In addition, one can test directly for a novel function by incorporating a substrate into the medium which when acted upon leads to a visual indication of activity.

Illustrations of Walk-through Mutagenesis

Model I

To further illustrate the invention, a "walk-through" mutagenesis of three of the hypervariable regions or complemetarity determining regions (CDRs) of the monoclonal antibody MCPC 603 is described. CDR1 and CDR3 of the heavy chain (VH) and CDR2 of the light chain region (VL) were the domains selected for walk-through mutagenesis. For this embodiment, the amino acids selected are the three residues of the catalytic triad of serine proteases, Asp, His and Ser. Asp was selected for VH CDR1, Ser was selected for VH CDR3, and His was selected for VL CDR2.

MCPC 603 is a monoclonal antibody that binds phosphorylcholine. This immunoglobulin is recognized as a good model for investigating binding and catalysis because the protein and its binding region have been well characterized structurally. The CDRs for the MCPC 603 antibody have been identified. In the heavy chain, CDR1 spans amino acids 31-35, CDR2 spans 50-69, and CDR3 spans 101-111. In the light chain, the amino acids of CDR1 are 24-40, CDR2 spans amino acids 55-62, and CDR3 spans amino acids 95-103. The amino acid numbers in the Figures correspond to the numbers of the amino acids in the parent MCPC 603 molecule.

The cDNA corresponding to an immunoglobulin variable region can be directly cloned and sequenced without constructing cDNA libraries. Because immunoglobulin variable regions genes are flanked by conserved sequences, a polymerase chain reaction (PCR) can be used to amplify, clone and sequence both the light and heavy chain genes from a small number of hybridoma cells with the use of consensus 5' and 3' primers. See Chiang, Y.L. et al., BioTechniques 7:360 (1989). Furthermore, the DNA coding for the amino acids flanking the CDR regions can be mutagenized by site directed mutagenesis to generate restriction enzyme recognition sites useful for further "cassette" mutagenesis. See U.S. Patent No. 4,888,286, supra. To facilitate insertion of the degenerate oligonucleotides, the mixture is synthesized to contain flanking recognition sites for the same restriction enzymes. The degenerate mixture can be first converted into double stranded DNA by enzymatic methods (Oliphant, A.R. et al., Gene 44:177 (1986)) and then inserted into the gene of the region to be mutagenized in place of the CDR nucleotide sequence encoding the naturally-occurring (wild type) amino acid sequence.

Alternatively, one of the other approaches described above, such as a gene synthesis approach, could be used to make a library of plasmids encoding variants in the desired regions. The published amino acid sequence of the MCPC 603 VH and VL regions can be converted to a DNA sequence. (Rudikoff, S. and Potter, M., Biochemistry 13: 4033 (1974)). Note that the wild type DNA sequence of MCPC 603 has also been published (Pluckthun, A. et al., Cold Spring Harbor Symp. Quant. Biol., Vol. LII: 105-112 (1987)). Restriction sites can be incorporated into the sequence to facilitate introduction of degenerate oligonucleotides or the degenerate sequences may be introduced at the stage of gene assembly.

The design of the oligonucleotides for walk-through mutagenesis in the CDRs of MCPC 603 is shown in Figure 3. In each case, the positions or "windows" to be mutagenized are shown. It is understood that the oligonucleotide synthesized can be larger than the window shown to facilitate insertion into the target construct. The mixture of oligonucleotides corresponding to the VH CDR1 is designed in which each amino acid of the wild type sequence is substituted by Asp (Figure 3a). Two codons specify asp (GAC and GAT). The first codon of CDR1 does not require any substitution. The second codon (TTC, Phe) requires substitution at the first (T to G) and second position (T to A) in order to convert it into a codon for Asp. The third codon (TAC, Tyr) requires only one substitution at the first position (T to G). The fourth codon (ATG, Met) requires three substitutions, the first being A to G, the second T to A and the third G to T. The fifth codon (GAG, Glu) requires only one substitution at the third position (G to T). The resulting mixture of oligonucleotides is depicted below.

```
                T T   T     A T G     G
5'- G A C     C   A C         G A   - 3'
                G A   G     G A T     T
```

This represents a mixture of 2^7 = 128 different oligonucleotide sequences.

From the genetic code, it is possible to deduce all the amino acids that will substitute the original amino acid in each position. For this case, the first amino acid will always be Asp (100%), the second will be Phe

(25%), Asp (25%), Tyr (25%) or Val (25%), the third amino acid will be Tyr (50%) or Asp (50%); the fourth will be Met (12.5%), Asp (12.5%), Val (25%), Glu (12.5%), Asn (12.5%), Ile (12.5%) or Lys (12.5%); and the fifth codon will be either Glu (50%) or Asp (50%). In total, 128 oligonucleotides which will code for 112 different protein sequences (1 x 4 x 2 x 7 x 2 = 112) are generated. Among the 112 different amino acid sequences generated will be the wild type sequence (which has an Asp residue at position 31), and sequences differing from wild type in that they contain from one to four Asp residues at positions 32-35, in all possible permutations (see Figure 3a). In addition, some sequences, either with or without Asp substitutions, will contain an amino acid - neither wild type nor Asp - at positions 32, 34 or both. These amino acids are introduced by permutations of the nucleotides which encode the wild type amino acid and the preselected amino acid. For example, in Figure 3a, at position 32, tyrosine (Tyr) and valine (Val) are generated in addition to the wild type phenylalanine (Phe) residue and the preselected Asp residue.

The CDR3 of the VH region of MCPC603 is made up of 11 amino acids, as shown in Figure 3b. A mixture of oligonucleotides is designed in which each non-serine amino acid of the wild type sequence is replaced by serine (Ser), as described above for CDR1. Six codons (TCX and AGC, AGT) specify Ser. The substitutions required throughout the wild-type sequence amount to 12. As a result, the oligonucleotide mixture produced contains 2^{12} = 4096 different oligonucleotides which, in this case, will code for 4096 protein sequences. Among these sequences will be some containing a single serine residue (in addition to the serine 105) in any one of the other positions (101-104, 106-111), as well as variants with more than one serine, in any combination (see Figure 3b).

The CDR2 of the VL region of MCPC603 contains eight amino acids (56-63). Seven of these amino acids (56-62) were selected for walk-through mutagenesis as depicted in Figure 3c. The mixture of oligonucleotides is designed in which each amino acid of the wild type sequence will be replaced by histidine (His). Two codons (CAT and CAC) specify His. The substitutions required throughout the wild-type DNA sequence total 13. Thus, the oligonucleotide mixture produced contains 2^{13} = 8192 oligonucleotides which specify 8192 different peptide sequences (see Figure 3c).

As result of this mutagenesis method, by the synthesis and the use of three oligonucleotide mixtures, a library of Fv sequences can be produced which contains 112 x 4096 x 8192 = 3.76 x 10^9 different protein sequences. A significant proportion of these sequences will encode the amino acid triad His, Ser, Asp typical of serine proteases within the hypervariable regions.

The synthesis of the degenerate mixture of oligonucleotides can be conveniently obtained in an automated DNA synthesizer programmed to deliver either one nucleotide to the reaction chamber or a mixture of two nucleotides in equal ratio, mixed prior to the delivery to reaction chamber. An alternative synthetic procedure would involve premixing two different nucleotides in a reagent vessel. A total of 10 reagent vessels, four of which containing the individual bases and the remaining 6 containing all of the possible two base mixtures among the 4 bases, can be employed to synthesize any mixture of oligonucleotides for this mutagenesis process. For example, the DNA synthesizer can be designed to contain the following ten chambers:

Chamber	Synthon
1	A
2	T
3	C
4	G
5	(A+T)
6	(A+C)
7	(A+G)
8	(T+C)
9	(T+G)
10	(C+G)

With this arrangement, any nucleotide can be replaced by a combination of two nucleotides at any position of the sequence.

The following sequence of reactions is required to synthesize the desired mixture of degenerate oligonucleotides for:

VH CDR1: 4, 1, 3, 9, 5, 3, 9, 1, 3, 7, 5, 9, 4, 1, 9

VH CDR3: 1, 7, 3, 2, 6, 3, 2, 6, 3, 7, 4, 3, 1, 4, 3, 1, 10, 2, 2, 10, 4, 2, 6, 3, 2, 8, 3, 9, 6, 3, 9, 8, 2

VL CDR2: 10, 7, 2, 10, 6, 2, 6, 7, 3, 6, 6, 3, 3, 7, 2, 10, 1, 5, 8, 6, 2

As an alternative to this procedure, if mixing of individual bases in the lines of the oligonucleotide synthesizer is possible, the machine can be programmed to draw from two or more reservoirs of pure bases to generate the desired proportion of nucleotides.

Each mixture of synthetic oligonucleotides can be inserted into the gene for the respective MCPC 603 variable region. The oligonucleotides can be converted into double-stranded chains by enzymatic techniques (see e.g., Oliphant, A.R. et al., 1986, supra) and then ligated into a restricted plasmid containing the gene coding for the protein to be mutagenized. The restriction sites could be naturally occurring sites or engineered restriction sites.

The mutant MCPC 603 genes constructed by these or other suitable procedures described above can be expressed in a convenient 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)). 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).

These and other Fv variants, or antibody variants produced by the present method can also be produced in other microorganisms such as yeast, or in mammalian cells, such as myeloma or hybridoma cells. The Fv variants can be produced as individual VH and VL fragments, as single chains (see Huston, J.S. et al., Proc. Natl. Acad. Sci. USA 85: 5879-5883 (1988)), as parts of larger molecules such as Fab, or as entire antibody molecules.

In a preferred 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)). The mutant VH genes can be concurrently expressed with wild-type VL to produce Fv variants, or as described, with mutagenized VL to further increase the number and structural variety of the protein mutants.

Screening of these variants for acquisition of a proteolytic function can be accomplished in an as say as described below for the HIV protease variants (see also Example 4). Note also that since the catalytic triad of Asp-His-Ser has also been implicated in the mechanism of certain lipases, variants with lipase function may also be generated.

Model II

In a second model designed to generate a serine protease in the MCPC 603 Fv structure, Asp is selected for VH CDR1, His for VH CDR3, and Ser for VL CDR2. In this case, the degenerate oligonucleotides designed for the VH CDR1 Asp walk-through from model 1 can be reused, illustrating the interchangeable nature of the walk-through cassettes (Figure 3a).

For the His walk-through of VH CDR3, His the nucleotides required to specify histidine codons are introduced from positions 101-111 of the VH region. Figure 4a illustrates this walk-through procedure. Note that in this and other examples, the percentages of His produced are calculated for the case where approximately equal proportions of the wild-type or His nucleotide are introduced. These proportions can be adjusted to influence the frequency with which various amino acids are produced.

Figure 4b illustrates the Ser walk-through of VL CDR2 in each position (55-62). Here, the sequence at positions 58 and 62 is unchanged as serine is present in the wild type sequence. Note that at position 61, although four different nucleotide sequences are generated, only three different protein sequences would be produced. This outcome is due to the fact that TAA codes for a stop codon.

Application of the method in this case can produce a library of Fv sequences which contains 112 x 196,608 x 96 = 2.11×10^9 different protein sequences. Again, a significant proportion of these sequences will encode the catalytic Asp-His-Ser triad in the hypervariable regions.

Note that once a series of cassettes for a number of regions is designed, the series may be used in any permutation desired. For example, degenerate oligonucleotides may be designed for the CDRs, and these may be used together in any combination of regions and chains desired, as well as in different structures (e.g., single VL or VH chains, Fv molecules, single chain antibodies, full-size antibodies or chimeric antibodies).

Model III

In another approach to the design of a serine protease, only the heavy chain of the Fv molecule is used. Monomeric VH domains, known as single domain antibodies, with good antigen-binding affinities have

been prepared (Ward, E.S. et al., Nature 341: 544-546 (1989)). Thus, a single VH chain can provide a scaffold for walk-through mutagenesis. For this model, Asp was selected for VH CDR1 (Figure 3a), His for VH CDR2 and Ser for VH CDR3 (Figure 3b). Again, two of the degenerate nucleotide sequences described in Model I can be reused (Figures 3a and 3b). Figure 5a shows the His walk-through in a portion of VH CDR2.

Oligonucleotides comprising the windows shown in Figures 3a, 3b and Figure 5a and degenerate oligonucleotides complementary to these windows have been made. Furthermore, using complementary oligonucleotides, in addition to the degenerate oligonucleotides and their complements, a full length double-stranded VH gene variant was assembled. The assembled gene variants have been cloned into the vector pRB500 (Example 2), which contains the pelB leader sequence for secretion. These experiments are described in Example 1.

Synthesis of these oligonucleotides and incorporation into the VH gene as described, in all possible combinations, can theoretically generate $112 \times 2^{25} \times 4096 = 1.54 \times 10^{13}$ different peptide sequences. Due to the length of the region targeted in VH CDR2, a large number of variants are generated; however, a large proportion of the variants will have the preselected amino acids.

As an alternative to using the VH CDR2 window shown in Figure 5a, another window encompassing a different portion of VH CDR2 was designed (Figure 5b). In this window, certain positions in the region were selected (see Model VI below for further explanation) and subjected to walk-through mutagenesis using His as the preselected amino acid. If oligonucleotides designed as shown in Figure 5b are used instead of the oligonucleotides of Figure 5a, $112 \times 128 \times 4096 = 5.87 \times 10^{7}$ different peptide sequences can be generated.

Model IV

In another embodiment using the heavy chain of the Fv molecule, a different combination of windows is used. The Asp window previously described for CDR1 (Figure 3b; Models I, III) and the His window previously described for CDR3 (Figure 4a; Model II) are used with a new window in which Ser is walked through the amino-terminal portion of VH CDR2 from amino acids 50-60. This walk-through mutagenesis is illustrated in Figure 6.

Synthesis of these oligonucleotides and incorporation into the VH gene in all possible combinations can generate $112 \times 4096 \times 196{,}608 = 9.02 \times 10^{10}$ different peptide sequences.

Model V

In another embodiment, a protein with an existing catalytic activity is altered to generate a different mechanism of catalysis. In the process, the specificity and/or activity of the enzyme may also altered. The HIV protease was selected as an enzyme for mutagenesis. The HIV protease is an aspartic protease and has an Asp-Thr-Gly sequence typical of aspartic proteases which contain a conserved Asp-Thr(Ser)-Gly sequence at the active site (Toh et al., EMBO J. 4: 1267-1272 (1985)). For walk-through mutagenesis, the Asp-Thr-Gly sequence in the protease was selected as a target for mutagenesis. Walk-through mutagenesis was repeated three different times with three preselected amino acids, Asp, His and Ser. This approach is intended to result in the conversion of an aspartic protease into a serine protease and an alteration of the mechanism of catalysis. In addition, mutants of the HIV aspartic protease with altered activity, specificity, or an altered mechanism of catalysis are expected. altered

Figure 7 shows the three residues or window to be altered and illustrates three sequential walk-through procedures with Asp, His and Ser. At the first position, which is an Asp residue, only His and Ser are introduced. At the two remaining positions, Asp, His, and Ser are each introduced. Note that in the second position of the second codon and in the second position of the third codon, the A required in the His walk-through has already been introduced in the Asp walk-through (Figure 7). The sequence of the mixed probe which includes 324 different sequences and the encoded amino acids are also shown in Figure 7. This mutagenesis protocol will generate 324 different peptide sequences in the active site window.

For mutagenesis and expression of the HIV protease, plasmid pRB505 was constructed as described in Example 2. This plasmid will direct expression of the HIV protease from an inducible tac promoter (de Boer, H.A. et al., Proc. Natl. Acad. Sci. USA 80: 21 (1983)). In pRB505, the protease gene sequence is fused in frame to the 3' end of a sequence encoding the pelB leader sequence of pectate lyase, so that the protease can be secreted into the periplasmic space of E. coli. The construct is designed so that the leader sequence is cleaved and the naturally occurring N-terminal sequence of the protease is generated. Secretion of the HIV protease will facilitate assaying and purification of variants generated by mutagenesis.

The complement of the mixed probe shown in Figure 7 was synthesized, and a partially complementary oligonucleotide was also synthesized. These oligonucleotides are designed to allow production of a double-stranded sequence with convenient XhoI (CTCGAG) and BstEII (GGTNACC) restriction sites (underlined) flanking the active site window. (Note that the complement of the active site window's coding sequence was synthesized. Thus, the nucleotide sequence for the wild type for the active site window (5'-ACC AGT GTC-3') shown below is the complement of 5'- GAC ACT GGT -3', the latter which codes for Asp-Thr-Gly.)

```
                                                    G  TC   G
                                                   TT  CG  GA
5'- CAT TTC CTC GAG AAC GGT GTC ATC AGC ACC AGT GTC-
                                                  ---WINDOW--

CAG CAG AGC TTC CTT TAG TTG ACC ACC GAT TTT GAT GGT-
                                        3'-TA AAA CTA CCA

AAC CAG TGG - 3'
TTG GTC ACC TGC GAC GGT GTC TCA CTA AAC G- 5'
```

The oligonucleotides were annealed and extended in a reaction using the Klenow fragment of DNA polymerase. Extension of the short complementary oligonucleotide generates the complement of each of the variant oligonucleotides. The reaction mix was digested with BstEII and XhoI and the products were separated on an 8% polyacrylamide gel. A 106 bp band was recovered from the gel by electroelution. This band, containing the active site window fragments, was cloned between the BstEII and XhoI sites of pRB505, and the ligated plasmids were introduced into a TG1/pACYC177 $lacI^q$ strain. The resulting transformants were plated on LB amp plates, and yielded about 1000 colonies.

The colonies were screened using the protease screening assay described in Example 4. Ampicillin resistant colonies were screened for proteolytic activity by replica plating onto nutrient agar plates containing 2 mM IPTG for induction of expression, and either dry milk powder (3%) or hemoglobin as a protease substrate as described in Example 4. In this assay, if a colony secretes proteolytic activity leading to degradation of the substrate in the plate (e.g., dry milk), a zone of clearing appears against the opaque background of the plate. Because the wildtype HIV protease does not show activity in the assay (due to its substrate specificity), novel activities can be distinguished from the original activity. Preliminary data indicate that transformants with novel activity can be generated by the described procedure.

The novel variants generated can be screened further for acquisition of a different mechanism of action by differential inhibition with protease inhibitors. For example, serine proteases are inhibited by PMSF (phenylmethylsulfonyl fluoride), DFP (diisopropylphosphofluoridate), TLCK (L-1-chloro3-(-9-tosylamide)-7-amino-2-heptanone-hydrochloride).Transformants which generate a halo on plates can be grown in liquid media, and extracts from the cultures can be assayed in the presence of the appropriate inhibitors. Reduced activity in the presence of a serine protease inhibitor as compared to activity in the absence of such an inhibitor will be indicative that a variant functions with a serine protease catalytic mechanism. Among the variants generated by the walk-through mutagenesis procedure will be variants with altered activity, altered specificity, a serine protease mechanism or a combination of these features. These variants can be further characterized using known techniques.

Model VI

In this embodiment, walk-through mutagenesis of five out of six CDRs of the MCPC 603 Fv molecule is performed, and Asp, His and Ser are the preselected amino acids. In this model, "walk-through" mutagenesis is carried out from two to three times with a different amino acid in a given region or domain. For example, Ser and His are sequentially walked-through VL CDR1 (Figure 8a), and Asp and Ser are sequentially walked-through VL CDR3 (Figure 8b). VL CDR2 was not targeted for mutagenesis because structural studies indicated that this region contributes little to the binding site in MCPC 603.

In CDR1 of the VH chain of the Fv, Asp and His are walked through (Figure 8c). Ser can be introduced at two positions in CDR1 with a single base change (Figure 8c, positions 32 and 33). In VH CDR2, His and Ser are the preselected amino acids used (Figure 8d) and in VH CDR3, Asp, His and Ser are each walked through the amino terminal five positions of CDR3 (Figure 8e).

Furthermore, in this embodiment not all amino acids in a given region are mutagenized, although they do not contain the preselected amino acid as the wild type residue. For example, in Figure 8d, only positions 50, 52, 56, 58 and 60 are mutagenized. Similarly, in Figures 8a-d, it can be seen that one or more residues in the region are not mutagenized. Mutagenesis of noncontiguous residues within a region can be desirable if it is known, or if one can guess, that certain residues in the region will not participate in the desired function. In addition, the number of variants can be minimized.

For example, in the case of a serine protease, a design factor is the distance between the the preselected amino acids. In order to form a catalytic triad, the residues must be able to hydrogen bond with one another. This consideration can impose a proximity constraint on the variants generated. Thus, only certain positions within the CDRs may permit the amino acids of the catalytic triad to interact properly. Thus, molecular modeling or other structural information can be used to enrich for functional variants.

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

If the mutagensis is carried out as illustrated and the regions are randomly combined, then 17,280 x 27,648 x 432 x 2304 x 7776 = 5.2×10^{18} different peptide sequences can be generated.

Model VII

In each of the embodiments described above, mutagenesis is designed to create clusters of catalytically active residues. In the embodiment of Model VII, mutagenesis is designed to create a novel binding function. In this embodiment, residues implicated in the binding or chelating of a co-factor (e.g., Fe + + +) are introduced into regions of a molecule, in this case MCPC 603. Many enzymes use metal ions as cofactors, so it is desirable to generate such binding sites as a first step towards engineering such enzymes.

In this embodiment two histidine and two tyrosine residues are introduced into the CDRs of MCPC 603. Dioxygenases, which are members of the class of oxidoreductases, and which catalyze the oxidative cleavage of double bonds in catachols contain a bound iron at their active sites. Spectroscopic analysis and X-ray crystallography indicate that the ferric ion at the active site of the dioxygenases is bound by two tyrosine and two histidine residues.

The histidine windows designed for MCPC 603 (see e.g., Figure 3c, VL CDR2; Figure 4a, VH CDR3; and Figure 5a, VH CDR2) can be used to introduce histidine residues into one or more domains of MCPC 603 or additional windows can be designed. Similarly, the one or more CDRs of MCPC 603 can be targeted for walk-through mutagenesis with tyrosine. Using these cassettes, variants with 2 histidine and 2 tyrosine residues in a large variety of combinations and in different regions can be produced.

These variants can be screened for acquisition of metal binding. For example, pools of colonies can be grown and a periplasmic fraction can be prepared. The proteins in a the periplasmic fraction of a given pool can be labeled with an appropriate radioactive metal ion (e.g., ^{55}Fe) and the presence of a metal binding variant can be determined using high sensitivity gel filtration. The presence of radioactivity in the protein fraction from gel filtration is indicative of metal binding. Pools can be subdivided and the process repeated until a mutant is isolated.

Alternatively, a nitrocellulose filter assay can be used. Colonies of a strain which secretes the mutant proteins and which allows the proteins to leak into the medium can be grown on nitrocellulose filters. The mutant proteins leaking from the colonies can bind to the nitrocellulose and the presence of metal binding proteins can be ascertained by probing with radiolabeled metal ions.

Generation of a metal binding in the VL chain could provide a metal binding site for a catalytic VH chain. Production of Fv from these component chains could allow enhancement of catalysis mediated by one chain by co-factor binding in the other chain.

The present invention is further illustrated in the following examples.

Example 1

Construction of a VH Variant

Oligonucleotide Synthesis

β-cyanoethyl phosphoramidites and polymer support (CPG) columns were purchased form Applied Biosystems, Inc. (Foster City, CA). Anhydrous acetonitrile was purchased form Burdick and Jackson (Part no. 015-4). Oligonucleotides were synthesized on an Applied Biosystems Model 392 using programs provided by the manufacturer (Sinha, N.D., et al., Nucleic Acids Res., 12: 4539 (1984)). On completion of the synthesis, the oligonucleotide was freed from the support and the protecting cyanoethyl groups were removed by incubation in concentrated NH_4OH. Following electrophoresis on a 10% polyacrylamide gel, oligomers were excised from the gel, electroeluted, purified on C18 columns, freeze dried and dissolved in the appropriate buffer at a final concentration of 1 µg/ml.

Oligonucleotides

In order to construct the VH variant described in Model III, the following oligonucleotides and their complements (also shown), ranging in length from 30-54 bases were designed and synthesized as described. Codon utilization was adjusted to reflect the most frequently used E. coli codons.

A/a: 910372/910373

5'- AAG AAT TCC ATG GAA GTT AAA CTG GTA GAG -3'

5'- ACC ACC AGA CTC TAC CAG TTT AAC TTC CAT GGA ATT-
CTT- 3'

B/b: 910374/910375

5'- TCT GGT GGT GGT CTG GTA CAG CCG GGT GGA TCC-
CTG- 3'

5'- AGA CAG ACG CAG GGA TCC ACC CGG CTG TAC CAG-
ACC -3'

C/c: 910376/910377

5'- CGT CTG TCT TGC GCT ACC TCA GGT TTC -3'
5'- AGA GAA GGT GAA ACC TGA GGT AGC GCA -3'

D/d: 910378/910379

```
                                GA  G   GAT   T
5'-ACC TTC TCT GAC TTC TAC ATG GAG TGG GTA CGT-
CAG-3'

                                              A   ATC   C  TC
5'-ACC CGG GGG CTG ACG TAC CCA CTC CAT GTA GAA-
GTC -3'
```

E/e: 910380/910381

```
5'- CCC CCG GGT AAA CGT CTC GAG TGG ATC GCA GCT-
AGC- 3'
```

5'- GTT ACG GCT AGC TGC GAT CCA CTC GAG ACG TTT -3'

F/f: 910382/910383

```
                          CA  C   C T C   CA  CA  C T C   CA
5'-CGT AAC AAA GGT AAC AAG TAT ACT ACT GAA TAC AGC-

CA  CA  CA  C C CA
GCT TCT GTT AAA GGT CGT -3'

                           TG G G  TG  TG  TG  TG   G A G  TG
5'-GAT GAA ACG ACC TTT AAC AGA AGC GCT GTA TTC AGT-

TG   G A G   G  TG
AGT ATA CTT GTT ACC TTT -3'
```

G/g: 910384/910385

5'- TTC ATC GTT TCT CGT GAC ACT AGT CAA TCG ATC CTG-
TAC CTG- 3'

5'- ATT CAT CTG CAG GTA CAG GAT CGA TTG ACT AGT GTC-
ACG AGA AAC- 3'

H/h: 910386/910387

5'- CAG ATG AAT GCA TTG CGT GCT GAA GAC ACC GCT ATC-
TAC- 3'

5'-CGC GCA GTA GTA GAT AGC GGT GTC TTC AGC ACG CAA-
TGC- 3'

I/i: 910388/910389 OR 9104103/9104104

```
                          G   C   C  A        G   C   C
5'-TAC TGC GCG CGT AAC TAC TAT GGC AGC ACT TGG TAC-

  C  TC  TC
TTC GAC GTT TGG -3'

                        GA  GA  G   G   G   C        T
5'-ACC TGC ACC CCA AAC GTC GAA GTA CCA AGT GCT GCC

  G   G   C
ATA GTA GTT-3'
```

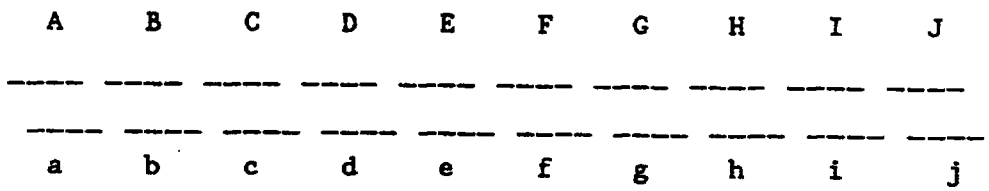
J/j: 910390/910391

5'- GGT GCA GGT ACC AAC GTT ACC GTT TCT TGA TAG CAG-
GTA AGC TTA A -3'

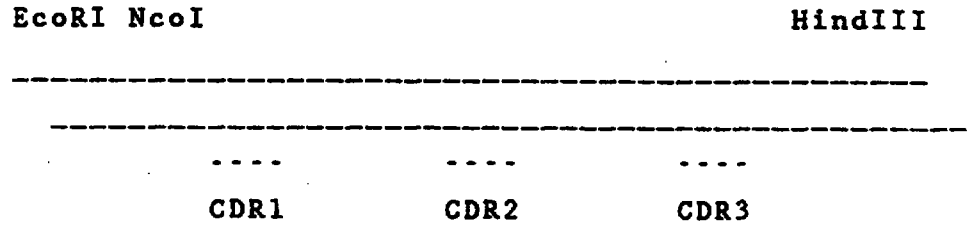
5'-TTA AGC TTA CCT GCT ATC AAG AAA CGG TAA CGG TGG
T -3'

Gene Assembly

These pairs of oligonucleotides can be assembled into a VH gene as depicted below:



Pairs D/d, F/f, and I/i are degenerate and complementary oligonucleotides encompassing the "windows" depicted in Figure 3a, Figure 5a, and Figure 3b, respectively. The design of the other oligonucleotides was similar to that described by Pluckthun et al., and included the introduction of a series of restriction sites (EcoRI, NcoI, BamHI, SauI, XmaI, XhoI, NheI, AccI, HaeII, SpeI, ClaI, PstI, NsiI, BssHII, KpnI, and HindIII useful for further manipulations (see Pluckthun, A. et al., Cold Spring Harbor Symp. Quant. Biol., Vol. LII, 105-112 (1987)). For gene assembly (Alvarado-Urbina, G. et al., Biochem. Cell. Biol. 64: 548-555 (1986)), eighteen of the oligonucleotides (B-I, b-i) were phosphorylated using T4 polynucleotide kinase. Each of ten complementary pairs was annealed separately. The annealed pairs were then mixed and ligated together using T4 DNA ligase. The product is shown schematically below:



The synthetic gene was designed to contain restriction sites for cloning. Following ligation, the fully assembled molecules were cleaved with NcoI and HindIII, gel purified, and inserted into vector pRB500 (see Example 2) at the NcoI and HindIII sites. About 1500 tranformants above the background were obtained on LB amp plates. The resulting constructs should contain the VH gene variants fused in frame to the pelB signal peptide.

Example 2

Construction of pRB505

Construction of pRB500

Two complementary oligonucleotides which code for the pelB leader sequence (Lei, S.P. et al.,J. Bacteriol. 169: 4379 (1987)) were chemically synthesized. The oligonucleotides, which were designed to have 5' and 3' overhangs complementary to NcoI and Pst I sites, were hybridized and cloned into the PstI and NcoI sites of vector pKK233.2 (Pharmacia). The oligonucleotides are shown below:

```
5'- C ATG AAA TAC CTA TTG CCT ACG GCA GCC GCT GCA-
      3'- TTT ATG GAT AAC GGA TGC CGT CGG CGA CGT

TTG TTA TTA GCT GCC CAA CCA GCC ATG GCG AAT TCC-
AAC AAT AAT CGA CGG GTT GGT CGG TAC CGC TTA AGG

CTG CA-3'
G -5'
```

The resulting plasmid, pRB500 has an inducible tac promoter upstream of the ATG start codon of the pelB sequence. There is a unique NcoI site (underlined) at the 3' end of the sequence coding for the pelB leader into which a gene encoding a product to be secreted, such as the HIV protease or the V_H or V_L regions of an antibody, may be inserted. (The NcoI site ligated to the 5' overhang of the fragment is not regenerated.)

Construction of pRB503

The HIV protease gene was obtained from pUC18.HIV (Beckman, catalog # 267438). The gene can be excised from this plasmid as a HindIII-EcoRI or HindIII-BamHI fragment. However, the HindIII site in the HIV protease cannot be directly cloned in frame to the pelB leader sequence present in plasmid pRB500. Therefore, a double-stranded oligonucleotide linker was designed so that the amino terminal methionine of the HIV protease coding sequence could be joined in frame to the coding sequence of the pelB leader peptide in pRB505. The following sequence was synthesized:

Met Ala Pro Gln Ile Thr ...

```
5'- AG CTT GCC ATG GCG CCG CAA ATC ACT CT- 3'
      3' -A CGG TAC CGC GGC GTT TAG TG -5'
             NcoI
```

This linker has a 5'- HindIII overhang and 3' DraIII overhang. The oligonucleotide was cloned into the unique HindIII and DraIII sites in pUC18.HIV. The resulting plasmid is called pRB503. The linker introduces an NcoI site into the vector at the initiator methionine of the HIV protease and reconstructs the sequence as found in pUC18.HIV.

Construction of pRB505

The HIV protease gene was isolated from pRB503 as an NcoI-EcoRI fragment and was cloned into the unique NcoI and EcoRI sites of pRB500. In the final construct, the HIV protease is fused in frame to the pelB leader sequence, and expression is driven by the inducible tac promoter. It is expected that the leader

peptidase will cleave the fusion protein between Ala and Pro (residues 2 and 3 above) of the HIV sequence, thereby generating an N-terminal proline just as in the wild type HIV protease.

Example 3

Walk-Through Mutagenesis of the HIV Protease Active Site

A degenerate oligonucleotide which spans the Asp-Thr-Gly active site residues of the HIV protease was designed and synthesized. This oligonucleotide has a sequence complementary to that shown in Figure 7.

```
                                                    G  TC   G
                                                   TT  CG  GA
5'- CAT TTC CTC GAG AAC GGT GTC ATC AGC ACC AGT GTC-

CAG CAG AGC TTC CTT TAG TTG ACC ACC GAT TTT GAT GGT-

AAC CAG TGG - 3'
```

A second oligonucleotide, partially complementary to the above sequence was synthesized to permit conversion of the above degenerate oligonucleotides to double-stranded form. The complementary oligonucleotide had the following sequence:

```
5'- GCA AAT CAC TCT GTG GCA GCG TCC ACT GGT TAC CAT-

CAA AAT -3'
```

The degenerate oligonucleotides and complementary oligonucleotides were annealed.

```
                                                    G  TC   G
                                                   TT  CG  GA
5'- CAT TTC CTC GAG AAC GGT GTC ATC AGC ACC AGT GTC-
             XhoI

CAG CAG AGC TTC CTT TAG TTG ACC ACC GAT TTT GAT-
                                          3'-TA AAA CTA-

GGT AAC CAG TGG - 3'
CCA TTG GTC ACC TGC GAC GGT GTC TCA CTA AAC G- 5'
```

The oligos were extended using the Klenow fragment of DNA polymerase. (Oliphant, A.R. and Struhl, K., Methods Enzymol., 155: 568-582 (1987)). The resulting mixture was cleaved with BstEII and XhoI, and separated on an 8% polyacrylamide gel. A 106 bp band containing the active site windows was isolated by electroelution from a gel slice, extracted with phenol:chloroform, and ethanol precipitated.

Vector pRB505 was cleaved with BstEII and XhoI and then treated with calf intestinal alkaline phosphatase to prevent religation. The vector band was purified from a low-melting agarose gel. The purified BstEII-XhoI active site windows (100 nanograms) were cloned into the BstEII and XhoI sites of

pRB505 (500 nanograms). The ligation mix was used to transform a TG1/pACYC177 $lacI^q$ strain and amplicillin resistant transformants were selected on LB amp plates (LB plus 50 µg/ml ampicillin; Miller, J.H., (1972), In: Experiments in Molecular Genetics, Cold Spring Harbor Laboratory (Cold Spring Harbor, NY), p. 433. Approximately 1000 transformants were obtained by this procedure. Several of these transformants were tested for novel activity using the protease plate assay described below in Example 4.

Example 4

Protease Activity Plate Assays

Sensitivity of the Plate Assay

In the case where the activity to be assayed is a proteolytic activity, substrate-containing nutrient plates can be used for screening for colonies which secrete a protease. Protease substrates such as denatured hemoglobin can be incorporated into nutrient plates (Schumacher, G.F.B. and Schill, W.B., Anal. Biochem., 48: 9-26 (1972); Benyon and Bond, Proteolytic Enzymes, 1989 (IRL Press, Oxford) p. 50). When bacterial colonies capable of secreting a protease are grown on these plates, the colonies are surrounded by a clear zone, indicative of digestion of the protein substrate present in the medium.

A protease must meet several criteria to be detected by this assay. First, the protease must be secreted into the medium where it can interact with the substrate. Second, the protease must cleave several peptide bonds in the substrate so that the resulting products are soluble, and a zone of clearing results. Third, the cells must secrete enough protease activity to be detectable above the threshold of the assay. As the specific activity of the protease decreases, the threshold amount required for detection in the assay will increase.

One or more protease substrates may be used. For example, hemoglobin (0.05 - 0.1%), casein (0.2%), or dry milk powder (3%) can be incorporated into appropriate nutrient plates. Colonies can be transferred from a master plate using and inoculating manifold, by replica-plating or other suitable method, onto one or more assay plates containing a protease substrate. Following growth at 37 °C (or the appropriate temperature), zones of clearing are observed around the colonies secreting a protease capable of digesting the substrate.

Four proteases of different specificities and reaction mechanisms were tested to determine the range of activities detectable in the plate assay. The enzymes included elastase, subtilisin, trypsin, and chymotrypsin. Specific activities (elastase, 81U/mg powder; subtilisin, 7.8 U/mg powder; trypsin, 8600 U/mg powder; chymotrypsin, 53 U/mg powder) were determined by the manufacturer. A dilution of each enzyme, elastase, subtilisin, trypsin, and chymotrypsin, was prepared and 5 µl aliquots were pipetted into separate wells on each of three different assay plates.

Plates containing casein, dry milk powder, or hemoglobin in a 1% Difco bacto agar matrix (10 ml per plate) in 50 mM Tris, pH 7.5, 10 mM $CaCl_2$ buffer were prepared. On casein plates (0.2%), at the lowest quantity tested (0.75 ng of protein), all four enzymes gave detectable clearing zones under the conditions used. On plates containing powdered milk (3%), elastase and trypsin were detectable down to 3 ng of protein, chymotrypsin was detectable to 1.5 ng, and subtilisin was detectable at a level of 25 ng of protein spotted. On hemoglobin plates, at concentrations of hemoglobin ranging from 0.05 and 0.1 percent, 1.5 ng of elastase, trypsin and chymotrypsin gave detectable clearing zones. On hemoglobin plates, under the conditions used, subtilisin did not yield a visible clearing zone below 6 ng of protein.

Assay of Variant of HIV Protease

Of the approximately 1000 ampicillin resistant transformants obtained by the procedure described in Example 3, 300 colonies were screened using the protease plate screening assay. The ampicillin resistant colonies were screened for proteolytic activity by replica plating onto nutrient agar plates (LB plus ampicillin) with a top layer containing IPTG (isopropylthiogalactopyranoside) for induction of expression, and either dry milk powder (3%) or hemoglobin as a protease substrate.

Protease substrate stock solutions were made by suspending 60 mg of hemoglobin or 1.8 g of powdered milk in 10 ml of deionized water and incubating at 60 °C for 20 minutes. The top layer was made by adding ampicillin and IPTG to 50 ml of melted LB agar (15 g/l) at 60 °C to final concentrations of 50 µg/ml and 2 mM, respectively, and 10 ml of protease substrate stock solution. 10 ml of the top layer was layered onto LB amp plates.

Colonies secreting sufficient proteolytic activity which degrades the particular substrate in the plate (e.g., dry milk) will have a zone of clearing around them which is distinguishable from the opaque background of the plate. Whereas none of the transformants gave a zone of clearing on hemoglobin plates, a large proportion of the transformants gave a zone of clearance on dry milk powder plates. Note that the dry milk powder plates had been incubated at 37 °C for about 1.5 days and then refrigerated. Although no halos appeared after the 1.5 day incubation at 37 °C, more than 90% of the colonies on the assay plates had halos after 3 days in the refrigerator. Three sample colonies which produced halos on the assay plate were streaked onto dry milk powder plates containing 2 mM IPTG. Two of the three streaks grew. Distinct zones of clearing were again observed for these two isolates under the same conditions (grown overnight at 37 °C, followed by refrigeration for three days). As a control, transformants of TG1/pACYC177 $lacI^q$ containing either pRB500, which encodes the pelB signal sequence, but no HIV protease, or containing pRB505, which encodes the pelB signal sequence fused to the "wild type" HIV protease, were also streaked onto dry milk powder plates with 2 mM IPTG. In contrast to the transformants obtained from the mutagenesis, these control transformants did not give a zone of clearance on dry milk powder plates. This observation is consistent with previous results indicating that retroviral proteases are selective for viral target proteins (Skalka, A.M., Cell 56: 911-913 (1984)). Using this assay novel protease activities generated by the walk-through mutagenesis procedure can be differentiated from the wild type HIV protease by altered substrate specificities.

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

Claims

1. A method of generating a protein library comprising a heterogeneous mixture of mutant proteins, said method comprising the step of specifically substituting each amino acid in one or more predefined regions of a protein with any one of one or more predetermined amino acids without saturation, the substitution of each predetermined amino acid being effected at essentially each and every sequence position of each predefined region of the protein in turn.

2. A method of generating a protein library comprising a heterogeneous mixture of mutant proteins, said method comprising the step of specifically substituting each amino acid in one or more predefined regions of a protein with any one of two predetermined amino acids, the substitution of each predetermined amino acid being effected at essentially each and every sequence position of each predefined region of the protein in turn.

3. A method of generating a protein library comprising a heterogeneous mixture of mutant proteins, said method comprising the step of specifically substituting each amino acid in a predefined region of the protein with a single predetermined amino acid, the substitution being effected at essentially each and every sequence position of the predefined region of the protein in turn.

4. The method of any one of Claims 1-3, wherein (a) a predetermined amino acid is introduced into two or more predefined regions of the protein; or (b) the predetermined amino acid is Ser, Thr, Asn, Gln, Tyr, Cys, His, Glu, Asp, Lys or Arg; or (c) the proportion of mutant proteins containing at least one residue of the predetermined amino acid in the predefined region ranges from about 12.5% to 100% of all mutant proteins in the library, and for example wherein the library comprises mutant proteins containing the predetermined amino acid in from one to all positions in the predefined region.

5. The method of any one of Claims 1-3, wherein substitution of the predetermined amino acid is effected at each and every sequence position of the predefined region or at least one predefined region of the protein.

6. The method of any one of Claims 1-3, wherein the predefined region or at least one predefined region comprises a functional domain of the protein.

Colonies secreting sufficient proteolytic activity which degrades the particular substrate in the plate (e.g., dry milk) will have a zone of clearing around them which is distinguishable from the opaque background of the plate. Whereas none of the transformants gave a zone of clearing on hemoglobin plates, a large proportion of the transformants gave a zone of clearance on dry milk powder plates. Note that the dry milk powder plates had been incubated at 37 °C for about 1.5 days and then refrigerated. Although no halos appeared after the 1.5 day incubation at 37 °C, more than 90% of the colonies on the assay plates had halos after 3 days in the refrigerator. Three sample colonies which produced halos on the assay plate were streaked onto dry milk powder plates containing 2 mM IPTG. Two of the three streaks grew. Distinct zones of clearing were again observed for these two isolates under the same conditions (grown overnight at 37 °C, followed by refrigeration for three days). As a control, transformants of TG1/pACYC177 $lacI^q$ containing either pRB500, which encodes the pelB signal sequence, but no HIV protease, or containing pRB505, which encodes the pelB signal sequence fused to the "wild type" HIV protease, were also streaked onto dry milk powder plates with 2 mM IPTG. In contrast to the transformants obtained from the mutagenesis, these control transformants did not give a zone of clearance on dry milk powder plates. This observation is consistent with previous results indicating that retroviral proteases are selective for viral target proteins (Skalka, A.M., Cell 56: 911-913 (1984)). Using this assay novel protease activities generated by the walk-through mutagenesis procedure can be differentiated from the wild type HIV protease by altered substrate specificities.

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

Claims

1. A method of generating a protein library comprising a heterogeneous mixture of mutant proteins, said method comprising the step of specifically substituting each amino acid in one or more predefined regions of a protein with any one of one or more predetermined amino acids without saturation, the substitution of each predetermined amino acid being effected at essentially each and every sequence position of each predefined region of the protein in turn.

2. A method of generating a protein library comprising a heterogeneous mixture of mutant proteins, said method comprising the step of specifically substituting each amino acid in one or more predefined regions of a protein with any one of two predetermined amino acids, the substitution of each predetermined amino acid being effected at essentially each and every sequence position of each predefined region of the protein in turn.

3. A method of generating a protein library comprising a heterogeneous mixture of mutant proteins, said method comprising the step of specifically substituting each amino acid in a predefined region of the protein with a single predetermined amino acid, the substitution being effected at essentially each and every sequence position of the predefined region of the protein in turn.

4. The method of any one of Claims 1-3, wherein (a) a predetermined amino acid is introduced into two or more predefined regions of the protein; or (b) the predetermined amino acid is Ser, Thr, Asn, Gln, Tyr, Cys, His, Glu, Asp, Lys or Arg; or (c) the proportion of mutant proteins containing at least one residue of the predetermined amino acid in the predefined region ranges from about 12.5% to 100% of all mutant proteins in the library, and for example wherein the library comprises mutant proteins containing the predetermined amino acid in from one to all positions in the predefined region.

5. The method of any one of Claims 1-3, wherein substitution of the predetermined amino acid is effected at each and every sequence position of the predefined region or at least one predefined region of the protein.

6. The method of any one of Claims 1-3, wherein the predefined region or at least one predefined region comprises a functional domain of the protein.

7. The method of Claim 6, wherein (a) the predefined region comprises a domain at or around the catalytic site of an enzyme or a binding domain; or (b) the predefined region comprises a hypervariable region of an antibody.

8. The method of any one of Claims 1-3, further comprising screening the library of mutant proteins to select mutant proteins having a desired structure or function.

9. The method of any one of Claims 1-3, wherein one or more of the predetermined amino acids is selected from the group consisting of: Asp, His, Tyr and Ser.

10. A library of mutant proteins prepared according to the method of any one of Claims 1-9.

11. A method of mutagenesis of a gene encoding a protein, comprising:
a) selecting a defined region of the amino acid sequence of the protein encoded by the gene to be mutagenized;
b) determining a single amino acid residue to be inserted into amino acid positions in the defined region;
c) synthesizing a mixture of oligonucleotides, comprising a nucleotide sequence for the defined region, wherein each oligonucleotide contains, at each sequence position in the defined region, either a nucleotide required for synthesis of the protein to be mutagenized or a nucleotide required for a codon of the predetermined amino acid, the mixture containing all possible variant oligonucleotides according to this criterion; and
d) generating an expression library of cloned genes containing said oligonucleotides.

12. A method of Claim 11, wherein (a) the defined region comprises a functional domain of the protein; or (b) the defined region comprises a domain at or around the catalytic site of an antibody; or (c) the defined region comprises a hypervariable region of an antibody; or (d) the predetermined amino acid is Ser, Thr, Asn, Gln, Tyr, Cys, His, Glu, Asp, Lys or Arg.

13. The method of Claim 11, further comprising one or more of the following steps:
a) expressing said library of cloned genes to produce mutant proteins; and/or
b) screening said library of cloned genes or encoded proteins to select for a desired structure or function.

14. A library of cloned genes prepared according to the method of Claim 11.

15. A method of producing a mutant protein having a desired structure or function by walk-through mutagenesis, comprising the steps of:
a) selecting one or more defined region(s) of the amino acid sequence of the protein to be mutagenized without saturation;
b) determining one or more amino acid residue(s) to be inserted into amino acid positions in said defined region(s);
c) synthesizing a mixture of oligonucleotides, comprising a nucleotide sequence for said defined region(s), wherein each oligonucleotide contains, at each sequence position in said defined region-(s), either a nucleotide required for synthesis of the protein to be mutagenized or a nucleotide required for a codon of said predetermined amino acid(s), the mixture containing all possible variant oligonucleotides according to this criterion;
d) generating an expression library of clones containing said oligonucleotides;
e) screening the library to detect a clone encoding a mutant protein having the desired structure or function; and
f) expressing a mutant protein having the desired structure or function by virtue of the presence of the oligonucleotide present in the clone detected in step (e).

16. The method of Claim 15, wherein (a) for at least one defined region, two predetermined amino acids are selected to be inserted into amino acid positions in said defined region(s); or (b) for at least one defined region, a single predetermined amino acid is selected to be inserted into amino acid positions in said defined region.

17. A library of mutants of a protein, comprising mutant proteins in which, collectively, a predetermined amino acid appears at least once in essentially every position in a region of the protein, wherein mutants containing at least one residue of the predetermined amino acid in a region of the protein comprise a proportion ranging from about 12.5% to 100% of the total number of different mutants in the library.

18. A library of Claim 17, wherein (a) the mutant proteins contain the predetermined amino acid in from one to all positions at once in the region, according to a statistical distribution; or (b) the protein is an enzyme and the region is at or around the catalytic site; or (c) the protein is an antibody or portion thereof and the region is a hypervariable region of the antigen-binding site; or (d) the predetermined amino acid is selected from the group consisting of: Ser, Thr, Asn, Gln, Tyr, Cys, His, Glu, Asp, Lys and Arg.

19. A library of HIV protease mutants, comprising mutant proteins in which, collectively, three predetermined amino acids appear at least once in all positions of the active site region of the protease.

20. The library of Claim 19, wherein the three predetermined amino acids are Asp, His and Ser.

21. A mutant protein of the library of Claim 20 wherein Asp, His and Ser appear in the active site region.

22. A method of producing a mixture of oligonucleotides for mutagenesis without saturation of a nucleotide sequence encoding a selected region of a protein to introduce a predetermined amino acid at each position in the region, comprising synthesizing a mixture of oligonucleotides comprising a nucleotide sequence for the preselected region, wherein each oligonucleotide contains, at each sequence position in the selected region, either a nucleotide required for synthesis of the amino acid of the region or a nucleotide required for a codon of the predetermined amino acid, the resulting mixture containing all possible variant oligonucleotides containing either of the two nucleotides at each position.

23. A mixture of oligonucleotides produced by the method of Claim 22, wherein about 12.5% to 100% of the oligonucleotides contain at least one codon for a single, predetermined amino acid.

24. An instrument for DNA synthesis having ten reagent vessels, each of four vessels containing a different one of the four nucleotide synthons corresponding to the four nucleotides of DNA and each of six containing vessels containing one of the six different mixtures of two synthons.

Patentansprüche

1. Ein Verfahren zur Bildung einer Protein-Bibliothek, die eine heterogene Mischung mutierter Proteine umfaßt, wobei besagtes Verfahren den Schritt der spezifischen Substitution jeder Aminosäure in einer oder mehreren zuvor bestimmten Regionen eines Proteins mit einer oder mehreren beliebigen, zuvor bestimmten Aminosäuren ohne Sättigung aufweist, wobei die Substitution jeder zuvor bestimmten Aminosäure im wesentlichen in allen und jeder Sequenzposition von jeder zuvor bestimmten Region des Proteins der Reihe nach bewerkstelligt wird.

2. Ein Verfahren zur Bildung einer Protein-Bibliothek, die eine heterogene Mischung mutierter Proteine umfaßt, wobei besagtes Verfahren den Schritt der spezifischen Substitution jeder Aminosäure in einer oder mehreren zuvor bestimmten Regionen eines Proteins mit einer beliebigen von zwei zuvor bestimmten Aminosäuren aufweist, wobei die Substitution jeder zuvor bestimmten Aminosäure im wesentlichen in allen und jeder Sequenzposition von jeder zuvor bestimmten Region des Proteins der Reihe nach bewerkstelligt wird.

3. Ein Verfahren zur Bildung einer Protein-Bibliothek, die eine heterogene Mischung mutierter Proteine umfaßt, wobei besagtes Verfahren den Schritt der spezifischen Substitution jeder Aminosäure in einer zuvor bestimmten Proteinregion mit einer einzigen, zuvor bestimmten Aminosäuren aufweist, wobei die Substitution im wesentlichen in allen und jeder Sequenzposition der zuvor bestimmten Proteinregion der Reihe nach bewerkstelligt wird.

* C L M S *

4. Das Verfahren gemäß einem der Ansprüche 1 bis 3, worin (a) eine zuvor bestimmte Aminosäure in zwei oder mehrere zuvor bestimmte Proteinregionen eingefügt wird; oder (b) die zuvor bestimmte Aminosäure Ser, Thr, Asn, Gln, Tyr, Cys, His, Glu, Asp, Lys, oder Arg ist; oder (c) der Anteil der mutierten Proteine, die mindestens einen Rest der zuvor bestimmten Aminosäure in der zuvor bestimmten Region enthalten, im Bereich von etwa 12,5% bis 100% aller mutierten Proteine der Bibliothek liegt, und worin zum Beispiel die Bibliothek mutierte Proteine umfaßt, welche die zuvor bestimmte Aminosäure in einer bis zu allen Positionen in der zuvor bestimmten Region enthalten.

5. Das Verfahren nach einem der Ansprüche 1 bis 3, worin die Substitution der zuvor bestimmten Aminosäure in allen und jeder Sequenzposition der zuvor bestimmten Region oder wenigstens einer zuvor bestimmten Region des Proteins bewerkstelligt wird.

6. Das Verfahren nach einem der Ansprüche 1 bis 3, worin die zuvor bestimmte Region oder wenigstens eine zuvor bestimmte Region eine funktionelle Domäne des Proteins aufweist.

7. Das Verfahren gemäß Anspruch 6, worin (a) die zuvor bestimmte Region eine Domäne an oder nahe des katalytischen Zentrums eines Enzyms oder eine Bindungsdomäne aufweist; oder (b) die zuvor bestimmte Region eine hypervariable Region eines Antikörpers aufweist.

8. Das Verfahren nach einem der Ansprüche 1 bis 3, das des weiteren ein Screening der Bibliothek mutierter Proteine zur Selektion von mutierten Proteinen mit einer gewünschten Struktur oder Funktion umfaßt.

9. Das Verfahren nach einem der Ansprüche 1 bis 3, worin eine oder mehrere der zuvor bestimmten Aminosäuren Asp, His, Tyr, oder Ser ist.

10. Eine Bibliothek mutierter Proteine, die nach dem Verfahren gemäß einem der Ansprüche 1 bis 9 gebildet ist.

11. Ein Verfahren zur Mutagenese eines Gens, das ein Protein codiert, wobei das Verfahren umfaßt:
    a) Auswählen einer bestimmten Region der Aminosäuresequenz des Proteins, das von dem zu mutagenisierenden Gen codiert wird;
    b) Bestimmen eines einzigen Aminosäurerests zur Insertion in Aminosäurepositionen in der bestimmten Region;
    c) Herstellen einer Oligonucleotidmischung, die eine Nucleotidsequenz für die bestimmte Region umfaßt, worin jedes Oligonucleotid in jeder Sequenzposition in der bestimmten Region entweder ein Nucleotid, das für die Synthese des zu mutagenisierenden Proteins erforderlich ist, oder ein Nucleotid, das für ein Codon der zuvor bestimmten Aminosäure erforderlich ist, enthält, wobei die Mischung alle möglichen verschiedenen Oligonucleotide diesem Kriterium gemäß enthält; und
    d) Bilden einer Expressionsbibliothek geklonter Gene, die besagte Oligonucleotide enthalten.

12. Ein Verfahren gemäß Anspruch 11, worin (a) die bestimmte Region eine funktionelle Domäne des Proteins aufweist; oder (b) die bestimmte Region eine Domäne an oder nahe des katalytischen Zentrums eines Antikörpers aufweist; oder (c) die bestimmte Region eine hypervariable Region eines Antikörpers aufweist; oder (d) die zuvor bestimmte Aminosäure Ser, Thr, Asn, Gln, Tyr, Cys, His, Glu, Asp, Lys oder Arg ist.

13. Das Verfahren gemäß Anspruch 11, das des weiteren einen oder mehrere der folgenden Schritte umfaßt:
    a) Expression besagter Bibliothek geklonter Gene, um mutierte Proteine herzustellen; und/oder
    b) Screening besagter Bibliothek geklonter Gene oder codierter Proteine, um eine gewünschte Struktur oder Funktion zu selektieren.

14. Eine Bibliothek geklonter Gene, die nach dem Verfahren gemäß Anspruch 11 hergestellt werden.

15. Ein Verfahren zur Herstellung eines mutierten Proteins mit einer gewünschten Struktur oder Funktion mittels walkthrough Mutagenese, wobei das Verfahren die Schritte umfaßt:

* C L M S *

a) Auswählen einer oder mehrerer bestimmten Regionen der Aminosäuresequenz des Proteins, das ohne Sättigung mutagenisiert werden soll;
b) Bestimmen eines oder mehrerer Aminosäurereste, die in Aminosäurepositionen in besagten bestimmten Regionen inseriert werden sollen;
c) Herstellen einer Oligonucleotidmischung, die eine Nucleotidsequenz für besagte bestimmte Regionen umfaßt, worin jedes Oligonucleotid in jeder Sequenzposition in besagten bestimmten Regionen entweder ein Nucleotid, das zur Synthese des zu mutagenisierenden Proteins erforderlich ist, oder ein Nucleotid, das für ein Codon besagter zuvor bestimmten Aminosäuren erforderlich ist, enthält, wobei die Mischung alle möglichen verschiedenen Oligonucleotide diesem Kriterium gemäß enthält;
d) Bilden einer Expressionsbibliothek von Klonen, die besagte Oligonucleotide enthalten;
e) Screening der Bibliothek zur Detektion eines Klons, der ein mutiertes Protein mit der gewünschten Struktur oder Funktion codiert; und
f) Expression eines mutierten Proteins, das aufgrund der Anwesenheit des Oligonucleotids, das in dem in Schritt (e) detektierten Klon vorhanden ist, die gewünschte Struktur oder Funktion besitzt.

16. Ein Verfahren gemäß Anspruch 15, worin (a) für wenigstens eine bestimmte Region zwei zuvor bestimmte Aminosäuren ausgewählt werden, die in Aminosäurepositionen in besagter oder in besagten bestimmten Region(en) inseriert werden sollen; oder (b) für wenigstens eine bestimmte Region, eine einzige zuvor bestimmte Aminosäure ausgewählt wird, die in eine Aminosäureposition in besagter bestimmten Region inseriert werden soll.

17. Eine Bibliothek von Mutanten eines Proteins, die mutierte Proteine umfaßt, in denen insgesamt eine zuvor bestimmte Aminosäure mindestens einmal im wesentlichen in jeder Position in einer Region des Proteins vorkommt, worin Mutanten, die mindestens einen Rest der zuvor bestimmten Aminosäure in einer Region des Proteins enthalten, einen Anteil im Bereich von etwa 12,5% bis 100% der Gesamtzahl der verschiedenen Mutanten in der Bibliothek aufweisen.

18. Eine Bibliothek gemäß Anspruch 17, worin (a) die mutierten Proteine die zuvor bestimmte Aminosäure in einer bis zu allen Positionen der Region gemäß einer statistischen Verteilung gleichzeitig enthalten; oder (b) das Protein ein Enzym ist, und die Region am oder nahe des katalytischen Zentrums ist; oder (c) das Protein ein Antikörper oder ein Teil dessen ist, und die Region eine hypervariable Region der Antigen-Bindungsstelle ist; oder (d) die zuvor bestimmte Aminosäure Ser, Thr, Asn, Gln, Tyr, Cys, His, Glu, Asp, Lys oder Arg ist.

19. Eine Bibliothek von HIV-Protease-Mutanten, die mutierte Proteine umfassen, in denen gemeinsam drei zuvor bestimmte Aminosäuren wenigstens einmal in allen Positionen im Bereich des aktiven Zentrums der Protease vorkommen.

20. Die Bibliothek gemäß Anspruch 19, worin die drei zuvor bestimmten Aminosäuren Asp, His und Ser sind.

21. Ein mutiertes Protein der Bibliothek gemäß Anspruch 20, worin Asp, His und Ser im Bereich des aktiven Zentrums vorkommen.

22. Ein Verfahren zur Herstellung einer Oligonucleotidmischung zur Mutagenese ohne Sättigung einer Nucleotidsequenz, die eine ausgewählte Region eines Proteins zum Einfügen einer zuvor bestimmten Aminosäure in jeder Position in der Region codiert, und die Herstellung der Oligonucleotidmischung umfaßt, die eine Nucleotidsequenz für die zuvor ausgewählte Region umfaßt, worin jedes Oligonucleotid an jeder Sequenzposition in der ausgewählten Region entweder ein für die Synthese der Aminosäure der Region erforderliches Nucleotid oder ein für ein Codon der zuvor bestimmten Aminosäure erforderliches Nucleotid enthält, und die resultierende Mischung alle möglichen verschiedenen Oligonucleotide enthält, die jedes der beiden Nucleotide an jeder Position enthalten.

23. Eine Oligonucleotidmischung, die nach dem Verfahren gemäß Anspruch 22 hergestellt wird, worin etwa 12,5% bis 100% der Oligonucleotide wenigstens ein Codon für eine einzige, zuvor bestimmte Aminosäure enthalten.

* C L M S *

24. Eine Vorrichtung zur DNA-Synthese mit zehn Reagenzgefäßen, wobei jedes von vier Gefäßen ein unterschiedliches der vier Nucleotidsynthone enthält, die mit den vier Nucleotiden der DNA übereinstimmen, und jedes von sechs Gefäßen eine der sechs unterschiedlichen Mischungen zweier Synthone enthält.

Revendications

1. Procédé pour la production d'une banque de protéines comprenant un mélange hétérogène de protéines mutantes, ledit procédé comprenant l'étape consistant à substituer de façon spécifique chaque amino-acide dans une ou plusieurs régions prédéfinies d'une protéine par un amino acide quelconque d'une série d'un ou plusieurs amino-acides prédéterminés, sans saturation, la substitution de chaque amino-acide prédéterminé étant réalisée tour à tour pratiquement en chacune des positions de la sequence de chacune des régions prédéfinies de la protéine.

2. Procédé de production d'une banque de protéines comprenant un mélange hétérogène de protéines mutantes, ledit procédé comprenant l'étape consistant à substituer de façon spécifique chaque amino-acide dans une ou plusieurs régions prédéfines d'une protéine par un amino-acide quelconque d'une série de deux amino-acides prédéterminés, la substitution de chaque amino-acide prédéterminé étant réalisée tour à tour pratique ment en chacune des positions de la séquence de chacune des régions prédéfinies de la protéine.

3. Procédé de production d'une banque de protéines comprenant un mélange hétérogène de protéines mutantes, ledit procédé comprenant l'étape consistant à substituer de façon spécifique chaque amino acide d'une région prédéfinie de la protéine par un amino-acide prédéterminé unique, la substitution étant réalisée tour à tour pratiquement en chacune des positions de la séquence de la région prédéfinie de la protéine.

4. Procédé selon l'une quelconque des revendications 1 à 3, dans lequel (a) un amino-acide prédéterminé est introduit dans deux ou plusieurs régions prédéfinies de la protéine; ou (b) l'amino acide prédé terminé est Ser, Thr, Asn, Gln, Tyr, Cys, His, Glu, Asp, Lys ou Arg; ou (c) la proportion de protéines mutantes contenant au moins un résidu de l'amino-acide prédéterminé dans la région prédéfinie correspond à environ 12,5 % jusqu'à 100% de toutes les protéines mutantes de la banque, la banque comprenant par exemple des protéines mutantes contenant l'amino-acide prédéterminé au niveau d'une ou plusieurs voire toutes les positions de la région prédéfinie.

5. Procédé selon l'une quelconque des revendications 1 à 3, dans lequel la substitution de l'amino-acide prédéterminé est réalisée à toutes les positions de la séquence, sans exception, de la région prédéfinie ou au moins dans une région prédéfinie de la protéine.

6. Procédé selon l'une quelconque des revendications 1 à 3, dans lequel la région prédéfinie ou au moins une région prédéfinie comprend un domaine fonctionnel de la protéine.

7. Procédé selon la revendication 6, dans lequel (a) la région prédéfinie comprend un domaine au niveau ou aux environs du site catalytique d'une enzyme ou un domaine de liaison; ou (b) la région prédéfinie comprend une région hypervariable d'une anticorps.

8. Procédé selon l'une quelconque des revendications 1 à 3, comprenant en outre l'étape de criblage de la banque de protéines mutantes afin de sélectionner les protéines mutantes présentant la structure ou la fonction souhaitée.

9. Procédé selon l'une quelconque des revendications 1 à 3, dans lequel un ou plusieurs des amino acides prédéterminés est choisi dans le groupe consituté de : Asp, His, Tyr, et Ser.

10. Banque de protéines mutantes préparée selon la méthode de l'une quelconque des revendications 1 à 9.

11. Procédé de mutagénèse d'une gène codant pour une protéines comprenant :

(a) la sélection d'une région définie de la séquence d'amino-acide de la protéine codée par le gène objet de la mutagénèse;
(b) la détermination du seul résidu amino-acide devant être inséré au niveau des positions des amino-acides dans la région définie;
(c) la synthèse d'un mélange d'oligonucléotides comprenant un séquence nucléotidique pour la région définie, dans lequel chaque oligonucléotide contient, au niveau de chaque position de la séquence de la région définie soit un nucléotide nécessaire à la synthèse de la protéine objet de la mutagénèse ou un nucléotide nécessaire pour un codon de l'amino-acide prédéterminé, le mélange contenant toutes les variantes d'oligonucléotides possibles répondant à ce critère; et
(d) la production d'une banque d'expression de gènes clonés contenant lesdits oligonucléotides.

12. Procédé selon la revendication 11, dans lequel (a) la région définie comprend un domaine fonc tionnel de la protéine; ou (b) la région définie comprend un domaine au niveau ou aux environs du site catalytique d'une anticorps; ou (c) la région définie comprend une région hypervariable d'un anticorps; ou (d) l'amino acide prédéterminé est Ser, Thr, Asn, Gln, Tyr, Cys, His, Glu, Asp, Lys ou Arg.

13. Procédé selon la revendication 11 comprenant en outre une ou plusieurs des étapes suivantes:
(a) exprimer ladite banque de gènes clonés de façon à produire les protéines mutantes; et/ou
(b) cribler ladite banque de gènes clonés ou de protéines codés de façon à sélectionner la structure ou fonction souhaitée.

14. Banque de gènes clonés préparés selon le procédé de la revendication 11.

15. Procédé pour la production d'une proteine mutante présentant la structure ou la fonction souhaitée par mutagénèse par cheminement, comprenant les étapes consistant à :
(a) sélectionner une ou plusieurs des régions définies de la séquence d'amino-acide de la protéine objet de la mutagénèse, sans saturation;
(b) déterminer un ou plusieurs des résidus amino-acides devant être insérés au niveau des positions d' amino acides dans ladite ou lesdites régions définies;
(c) synthétiser un mélange d'oligonucléotides comprenant une séquence nucléotidique pour ladite ou lesdites régions définies, dans lequel chaque oligonucléotide contient à chaque position de séquence dans chaque région définie, soit un nucléotide nécessaire pour la synthèse de la protéine objet de la mutagénèse ou un nucléotide nécessaire pour un codon dudit ou desdits amino-acides prédéterminés, le mélange contenant toutes les variantes d'oligonucléotides possibles répondant à ce critère; et
(d) générer une banque d'expression de clones contenant lesdits oligonucléotides;
(e) cribler la banque de façon à détecter un clone codant pour une protéine mutante présentant la structure ou la fonction souhaitée; et
(f) exprimer une protéine mutante présentant la fonction ou structure souhaitée grâce à l'oligonucléotide présent dans le clone détecté à l'étape (e).

16. Procédé selon la revendication 15, dans lequel (a) pour au moins une région définie, deux amino acides prédéterminés sont sélectionnés et insérés au niveau des positions d'amino-acides dans ladite ou lesdits régions définies; ou (b) pour au moins une des régions définies, un unique amino-acide prédéterminé est sélectionné de façon à être inséré au niveau des positions d'amino-acide dans ladite région définie.

17. Banque de mutants d'une protéine, comprenant des protéines mutantes dans lesquelles, dans leur ensemble, un amino acide prédéterminé apparaît au moins une fois au niveau de presque toutes les positions d'une région de la protéine, la proportion de protéines mutantes contenant au moins un résidu de l'amino-acide prédéterminé dans une région de la protéine correspondant à environ 12,5% jusqu'à 100% de la totalité des différentes protéines mutantes de la banque.

18. Banque selon la revendication 17, dans laquelle (a) les protéines mutantes contiennent l'amino-acide prédéterminé au niveau d'au moins une jusqu'à toutes les positions de la région, selon une distribution statistique, ou (b) la protéine est une enzyme et la région est celle du site catalytique ou une région voisine de celle-ci; ou (c) la protéine est un anti-corps ou une portion d'anticorps et la région est une région hypervariable du site de liaison de l'antigène; ou (d) l'amino-acide prédéterminé est choisi dans

* C L M S *

le groupe constitué de Ser, Thr, Asn, Gln, Tyr, Cys, His, Glu, Asp, Lys et Arg.

19. Banque de protéases mutantes de l'HIV comprenant des protéines mutantes dans lesquelles, dans leur ensemble, trois amino-acides prédéterminés appraissent au moins une fois dans toutes les positions de la région du site actif de la protéase.

20. Banque selon la revendication 19, dans laquelle les trois amino-acides prédéterminés sont Asp, His et Ser.

21. Protéine mutante de la banque selon la revendication 20, dans laquelle Asp, His et Ser appraissent dans la région du site actif.

22. Procédé pour la production d'un mélange d'oligonucléotides permettant la mutagénèse sans saturation d'une séquence nucléotidique codant pour une région sélectionnée d'une protéine de façon à introduire un amino acide prédéterminé à chaque position de la région, comprenant la synthèse d'un mélange d'oligonucléotides comprenant une séquence nucléotidique pour la région préalablement sélectionnée, dans lequel chaque oligonucléotide contient à chaque position de la séquence dans la région sélectionnée, soit un nucléotide nécessaire à la synthèse de l'amino-acide de la région ou un nucléotide nécessaire pour un codon de l'amino-acide prédéterminé, le mélange résultant comprenant toutes les variantes possibles d'oligonucléotides contenant l'un ou l'autre des deux nucléotides à chaque position.

23. Mélange d'oligonucléotides produits par le procédé de la revendication 22, dans lequel environ 12,5 à 100% des oligonucléotides contiennent au moins un codon pour un unique aminoacide prédéterminé.

24. Appareil pour la synthèse d'ADN comportant dix cuves de réactifs, une cuve sur quatre contenant un synthon nucléotidique différent des quatre synthons nucléotidiques correspondant aux quatre nucléotides de l'ADN et une cuve sur six contenant l'un des six mélanges différents de paires de synthons.