



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

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C12Q 1/68, C12P 19/34</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 97/47768</b> <b>(43) International Publication Date:</b> 18 December 1997 (18.12.97)
<b>(21) International Application Number:</b> PCT/US97/09679 <b>(22) International Filing Date:</b> 4 June 1997 (04.06.97)  <b>(30) Priority Data:</b> 08/661,080                      10 June 1996 (10.06.96)              US  <b>(71)(72) Applicant and Inventor:</b> WANG, Fenggang [CN/US]; 1- k, 1749 Dobbins Drive, Chapel Hill, NC 27514 (US).  <b>(74) Agent:</b> BARBER, Lynn, E.; P.O. Box 6450, Raleigh, NC 27628 (US).	<b>(81) Designated States:</b> AU, CA, CN, JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>With international search report.</i>	
<b>(54) Title:</b> METHOD FOR RANDOMLY SYNTHESIZING BIOPOLYMERS  <b>(57) Abstract</b>  <p>A method for synthesis of random biopolymers, particularly oligonucleotides. The method which can randomly incorporate different residues at defined positions with a controllable rate while preserving certain positions as single type residues. The oligonucleotides can be manually or automatically synthesized. For synthesis of oligonucleotides, "site doping synthesis" (SDS) utilizes reservoirs of conventional phosphoramidites and other reservoirs with the same phosphoramidites which have been mixed ("doped") with a mixture of the other phosphoramidites or other nucleotide analogues. The method can be used for mutagenesis with mutation at specific positions. Synthesized oligonucleotides may be used for isolating agonists and antagonists of a protein, such as a protein inhibitor or other therapeutic drug, for producing specified random peptides or proteins, and for developing antisense, antigene and ribozyme therapeutic drugs.</p>		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

<b>AL</b>	Albania	<b>ES</b>	Spain	<b>LS</b>	Lesotho	<b>SI</b>	Slovenia
<b>AM</b>	Armenia	<b>FI</b>	Finland	<b>LT</b>	Lithuania	<b>SK</b>	Slovakia
<b>AT</b>	Austria	<b>FR</b>	France	<b>LU</b>	Luxembourg	<b>SN</b>	Senegal
<b>AU</b>	Australia	<b>GA</b>	Gabon	<b>LV</b>	Latvia	<b>SZ</b>	Swaziland
<b>AZ</b>	Azerbaijan	<b>GB</b>	United Kingdom	<b>MC</b>	Monaco	<b>TD</b>	Chad
<b>BA</b>	Bosnia and Herzegovina	<b>GE</b>	Georgia	<b>MD</b>	Republic of Moldova	<b>TG</b>	Togo
<b>BB</b>	Barbados	<b>GH</b>	Ghana	<b>MG</b>	Madagascar	<b>TJ</b>	Tajikistan
<b>BE</b>	Belgium	<b>GN</b>	Guinea	<b>MK</b>	The former Yugoslav Republic of Macedonia	<b>TM</b>	Turkmenistan
<b>BF</b>	Burkina Faso	<b>GR</b>	Greece	<b>ML</b>	Mali	<b>TR</b>	Turkey
<b>BG</b>	Bulgaria	<b>HU</b>	Hungary	<b>MN</b>	Mongolia	<b>TT</b>	Trinidad and Tobago
<b>BJ</b>	Benin	<b>IE</b>	Ireland	<b>MR</b>	Mauritania	<b>UA</b>	Ukraine
<b>BR</b>	Brazil	<b>IL</b>	Israel	<b>MW</b>	Malawi	<b>UG</b>	Uganda
<b>BY</b>	Belarus	<b>IS</b>	Iceland	<b>MX</b>	Mexico	<b>US</b>	United States of America
<b>CA</b>	Canada	<b>IT</b>	Italy	<b>NE</b>	Niger	<b>UZ</b>	Uzbekistan
<b>CF</b>	Central African Republic	<b>JP</b>	Japan	<b>NL</b>	Netherlands	<b>VN</b>	Viet Nam
<b>CG</b>	Congo	<b>KE</b>	Kenya	<b>NO</b>	Norway	<b>YU</b>	Yugoslavia
<b>CH</b>	Switzerland	<b>KG</b>	Kyrgyzstan	<b>NZ</b>	New Zealand	<b>ZW</b>	Zimbabwe
<b>CI</b>	Côte d'Ivoire	<b>KP</b>	Democratic People's Republic of Korea	<b>PL</b>	Poland		
<b>CN</b>	Cameroon	<b>KR</b>	Republic of Korea	<b>PT</b>	Portugal		
<b>CU</b>	Cuba	<b>KZ</b>	Kazakstan	<b>RO</b>	Romania		
<b>CZ</b>	Czech Republic	<b>LC</b>	Saint Lucia	<b>RU</b>	Russian Federation		
<b>DE</b>	Germany	<b>LI</b>	Liechtenstein	<b>SD</b>	Sudan		
<b>DK</b>	Denmark	<b>LK</b>	Sri Lanka	<b>SE</b>	Sweden		
<b>EE</b>	Estonia	<b>LR</b>	Liberia	<b>SG</b>	Singapore		

## METHOD FOR RANDOMLY SYNTHESIZING BIOPOLYMERS

BACKGROUND OF THE INVENTIONField of the Invention

This invention relates to the synthesis of biopolymers, and in particular,  
5 pertains to the synthesis of oligonucleotide sequences.

Description of the Related Art

Randomly synthesized biopolymers, which generate a combination pool of  
variable molecules of a known sequence, have had wide applications for many years.  
Various methods have been developed for synthesizing these type of biopolymers.

10 Oliphant et. al., Gene 44:177-193 (1986) previously described a method for  
cloning random oligonucleotide sequences. The disclosure of this publication and of  
all other patents and publications referred to herein is incorporated herein by  
reference. This method is used for many applications such as mutagenesis, aptamer  
technology and making random peptides in phage display technology. In this method,  
15 random oligonucleotides are synthesized on a synthesizer using an equal molar mixture  
of all four phosphoramidites at a continuous DNA sequence region. Recognition sites  
for certain restriction endonucleases are also synthesized at the 5' and 3' ends of the  
oligonucleotides. The single strand oligonucleotides synthesized by this method can  
be paired to double strands and incorporated into a vector sequence at a low  
20 efficiency. Although the method can generate highly degenerated sequences, one  
limitation of the method is that it is poorly suited for highly specific consensus  
sequences. For example, if a genetic element has a specificity that is equivalent to  
10 exact bases, the probability at which such a sequence occurs is only about  $10^{-6}$  by  
this method. Such a low frequency greatly limits the identification of a desired  
25 mutated sequence.

Another limitation of the prior methodology is when it is applied to produce  
mutant proteins. One of the major goals of mutagenesis is a complete understanding  
of the correlation between protein structure and function. A gene or a protein  
contains hundreds to thousands of residues. The sequence order of a gene has been  
30 selected by evolution to best fit its selected function. It is known that most of these  
residues can be changed without affect on the function of the protein. However, some  
parts of the sequence cannot be changed. Mutagenesis methods are frequently used  
to define sites that cannot be changed without affecting the function of the protein.

-2-

Since a protein is coded by a DNA or RNA sequence, the complexity of a DNA or RNA sequence also represents the complexity of a protein. For easy understanding and for calculating the complexity of a protein, we use DNA herein as an example to demonstrate how a conventional random library generates a complex protein. The number of mutants generated by the random library method is about  $4^n$  for a DNA or RNA sequence, where "n" is the number of base pairs in the mutated region in a DNA or RNA sequence. The number of mutants will be substantial for a reasonably long DNA sequence. For example, if the studied sequence is only 20 bases long, the number of mutants generated will be  $4^{20}=10^{12}$ , which is a thousand times larger than the human genome. It would not be possible to isolate all of these mutants from such a large number of sequences. Moreover, most mutants in a completely random library are not related to the sequence of interest. Even the isolation of single site or double site mutants is difficult because they represent only a small percentage of the random library. This method would not be useful in mutating a DNA or protein sequence that has hundreds or thousands of residues.

The method of saturation mutagenesis (Hutchison C. A. et. al., Proc. Natl. Acad. Sci. USA 83:710-714 (1986)) was invented in order to reduce the generation of unnecessary mutants in a target sequence. With this method, the oligonucleotides in the library are synthesized by using a substrate phosphoramidite that is mixed with a small amount of other phosphoramidites. Since this method is different from the conventional random synthesis method which uses equal molar concentration of all four of the different phosphoramidites, it reduces the chance of the incorporation of mutant residues into the synthesized oligonucleotides. Although this method advances the complete random library method in reducing the generation of unrelated mutants, it generates mutations at all positions in the synthesized oligonucleotides. In other words, it also generates unnecessary mutants and increases the complexity of mutagenic libraries. The minimum length of the oligonucleotides for this type of mutagenesis is about 25 to 30 base pairs long which is a limitation. Moreover, it is often not desired to mutate all the positions in the oligonucleotide. Furthermore, it is statistically difficult to isolate all of the mutants from all positions using this method. For example, isolation of 2000 mutants only gives 95% of all possible single mutants in a 30 base pair long sequence. In an actual experiment this number can be

even much larger due to other factors. For example, the Kunkle method (Sambrook et. al., Molecular Cloning- A Laboratory Manual, 2nd Edition, 1989) only gives 50-70% incorporating efficiency of mutagenic oligonucleotides into the vector templates and the cell transformation efficiency also limits the isolation of mutants. It is almost impossible to isolate all of the expected mutants from all sites for a sequence having hundreds of residues with these methods. This limits the application of the prior method for isolating a complete set of mutants from long sequences.

Random synthesis methods are also used for other technologies for isolating protein, DNA or RNA binding sequences. Isolation of these binding sequences can lead to drug development and the discovery of diagnostic reagents. One example is the application of random oligonucleotides for aptamer technology (Ellington, A. D. Nature 346:818-822 (1990)). Aptamer technology, also called SELEX technology (Tuerk, C. et. al., Science 249:505-510(1990); and Gold, L. et. al., U. S. Pat. No. 5,270,163), systematically selects DNA or RNA sequences from a random synthesized library that bind to proteins or other molecules. The selected DNA or RNA sequences are potential protein inhibitors. One requirement of this technology is that it requires completely random DNA or RNA libraries. However, the conventional random library, which is made from equal molar amounts of four nucleotides, is often incomplete for various reasons. One factor is that it is limited by transforming efficiency when the synthesized DNA is introduced to a host cell as discussed above. Another factor is that the quantity of oligonucleotides synthesized is often not suited to all compositions. Therefore, like many other applications using random synthesis libraries, the SELEX procedure often loses many random combinations of the design libraries. The lost sequences can be synthesized by using the present invention.

Peptides also are key mediators of biochemical information in all mammalian systems and are important therapeutic and diagnostic reagents. The isolation of peptides from a randomly synthesized peptide library can be very useful. For example, in examining structure-function relationships in peptides, one can first generate a mixture of peptides having different amino acid substitutions at one or more defined polypeptide residue positions and then isolate the desired peptide. Moreover, a polypeptide having a special function, such as a high binding affinity to a given receptor or antibody, can be identified from randomly synthesized libraries. The

process involved in the development of a novel compound for diagnostic or therapeutic use requires the synthesis and screening of hundreds to thousands of analogs of an original active sequence.

The random synthesis libraries are also used for generating random peptide  
5 libraries for the selection of peptides. Since the triplet codons are not evenly distributed to code the amino acids in a protein, one limitation of this method is that when it is used for random peptide synthesis, the randomly synthesized codons bias the amino acids which are incorporated during translation of the DNA by the cell into polypeptides, so there is not an even distribution of amino acids. In other words,  
10 using four different nucleotides (A, C, G, and T) to make random triple combinations yields  $4^3=64$  different triplets (called "codons"), which are unevenly assigned to 20 different amino acids or coding a protein or polypeptide. For example, one amino acid may only have one triplet from the 64 combinations to code it, while others may be coded by three or four triplets. Therefore, a complete random combination of  
15 nucleotides will not yield a simple random amino acid sequence.

In order to generate random peptides and to avoid this limitation, random codon oligonucleotide synthesis (Huse, W. D., U. S. Pat. No. 5,264,563) was invented. In this method, the coding triplets for all of the twenty amino acids are first synthesized from phosphoramidite monomers. The random codon oligonucleotides are  
20 then synthesized by equal proportions of the triplets which represent codons of amino acids. This method is more costly than the synthesis from individual monomers and the synthesis yield is too low for applications requiring large oligonucleotides, i.e., oligonucleotides incorporating multiple mutagenic codons (Ono A. et. al., Nucleic Acids Res. 23:4677-4682(1995)).

In order to generate wild type closed random mutant libraries, site specific  
25 random synthesis is sometimes required. Column-splitting synthesis technology was invented to achieve this. The technology relies upon standard synthetic DNA chemistry. The incorporation of entire codons at mutation sites is performed on two separate columns (Cormark, B. P. and Struhl, K. Nature 262: 244-248, (1993)). The  
30 mutation frequency is controlled by the ratio of solid support split between the mutants and the wild-type columns. The goal of this technique is to mutate a small portion of synthesized oligonucleotides at a specific codon. However, each time the positions

are mutated, the synthesis is interrupted and the column is dismantled, split, and reunited manually. This technique is not practical for the synthesis of oligonucleotides with three or more separated mutagenic positions because this results in loss of yield and interruption of the synthesis. This process is also not easily automated.

5           The methods mentioned above have several disadvantages. First, although theoretically the completely random method can generate all possible combinations, the method often cannot generate a really complete random library due to the various factors mentioned above. Second, a desired sequence cannot be isolated easily because the composition of a library is too large. Third, it is inconvenient that the  
10       positions of the mutations in the sequences cannot be separated using these methods. This factor severely limits the application of the saturation mutagenesis method. Fourth, although site directed mutagenesis can be used to produce all of the desired mutants, unfortunately, this technique cannot be used to produce a population of different mutants efficiently and it can be costly and time consuming. For example,  
15       30 or more oligonucleotides need to be synthesized in order to obtain mutations at ten positions which is too costly (each position has three possible substitutions). Last, although it is possible to make some improvements of the prior methods, these modifications are insufficient (Barbas, C. F. et. al., Proc. Natl. Acad. Sci. USA 91:3809-3813 (1994)) or too costly and time consuming (Cormack, B. P. and Struhl,  
20       K. Nature 262: 244-248, (1993), and Huse, W. D., U. S. Pat. No. 5,264,563) and they are therefore not practical to produce large quantities of site specific random libraries.

          It is therefore an object of this invention to overcome the disadvantages mentioned above. The invention allows control of position selection, and a random  
25       substitution rate. The method can allow free choice of certain sites to be efficiently and randomly mutated while preserving the other sites as wild type in a DNA, RNA or protein peptide sequence and is valuable for biotechnology research, drug discovery and diagnostic reagent development. This method is easy to use, automatic, cost effective, time saving, and increases efficiency. Here we disclose a technology that  
30       can be used to randomly synthesize biopolymers such as oligonucleotides and which can also have potential applications in the synthesis of peptides and other polymers. The proposed method has no restriction on site selection or the rate of mutation that

is obtained. The method can automatically generate mutants having random mutations at different positions with a controllable rate. Mutants can be made by using two sets of phosphoramidites for the mutagenic oligonucleotide synthesis. One set contains only pure phosphoramidites for the wild type position synthesis and another set  
5 contains mixtures, each of which contains mostly one of the phosphoramidites together with a small amount of the 'others'. The term 'others' as used herein with reference to the invention includes both other phosphoramidites and also could include totally different molecules that can be incorporated into the oligonucleotide sequences instead of or in addition to the other phosphoramidites.

10 Other objects and advantages will be more fully apparent from the following disclosure and appended claims.

#### SUMMARY OF THE INVENTION

The method of the invention allows synthesis of random biopolymers, particularly oligonucleotides. Use of the method allows random incorporation of  
15 different residues at defined positions with a controllable rate while preserving certain positions as single type residues. The oligonucleotides can be manually or automatically synthesized. For synthesis of oligonucleotides, the "site doping synthesis" (SDS) of the invention utilizes reservoirs of conventional phosphoramidites and other reservoirs preferably with the same phosphoramidites which have been  
20 mixed ("doped") with a mixture of the other phosphoramidites or other nucleotide analogues or other incorporable residues.

The invention be used for mutagenesis for mutation at specific positions. Oligonucleotides synthesized according to the invention may be used for isolating agonists and antagonists of a protein, such as a protein inhibitor or other therapeutic  
25 drug, for producing specified random peptides or proteins, and for developing antisense, antigene and ribozyme therapeutic drugs. The invention may also be used to develop diagnostic kits with higher efficiency in detection, for example, of hyper mutational viruses, such as HIV.

Other aspects and features of the invention will be more fully apparent from  
30 the following disclosure and appended claims.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a graph showing the automatic synthesis of an oligonucleotide by



an oligosynthesizer by the present invention. Two groups of phosphoramidites supply the monomers for the oligonucleotide DNA synthesis. One group, labeled in the Figure with A, C, and T, is for the synthesis of "wild type" positions, and the other group, labeled with A\*, C\*, G\* and T\*, is for the synthesis of the mutated positions.

5 The positions of random substitution can be defined and programmed. Substitution occurs only at specifically defined positions. The synthesized oligonucleotide can carry 0, 1, 2 or more substituted bases.

Figure 2 is a photograph showing *in vitro* synthesized DNAs for mutagenesis using the random synthesized oligonucleotide of the present invention as a primer.  
10 All synthesis reactions were set with 0.8 pmol of single strand DNA as a template. The added primer was 0, 0.2 and 0.8 pmol.

DETAILED DESCRIPTION OF THE INVENTION AND PREFERRED  
EMBODIMENTS THEREOF

The present invention provides a method for randomly synthesizing  
15 biopolymers at specific positions, which, as a result, yields random libraries. The method is named "site doping synthesis (SDS)". Biopolymers are biologically functional molecules, such as oligonucleotides and polypeptides, and include DNA, RNA, peptides, macromolecules, or any combination of these molecules as well as others, such as, for example, peptide nucleic acids (PNA) or modified DNA  
20 oligonucleotides.

In general, an oligonucleotide sequence can either comprise one or more of the deoxyribonucleotide DNA bases, A, T, C, or G, or one or more of the ribonucleoside RNA bases, A, U, C, or G. For the purposes of the present invention, the bases T and U are considered equivalents for DNA and RNA oligonucleotides. The bases  
25 used in a DNA or RNA sequence are often abbreviated as follows:

A = Adenine	G = Guanine
C = Cytosine	T = Thymine
	U = Uracil

The sequences may contain a non-standard base, i.e. a base that is not A, T, C, or G. These non-standard nucleosides, for example, may be deoxyinosine (dI), hypoxanthine, bromodeoxyuridine (BrdU), and other bases which can be incorporated. Residues of the synthesized biopolymer also can be substituted with any other

chemical or nucleotide analog that can be incorporated, for example, the methylphosphonate and phosphorothioate nucleotides (Cohen, J. S. *Tibtech* 10:87-91, (1992)), and/or a labeling reagent, such as, biotin, by known techniques (Cook, A. F. et. al., *Nuclei Acid Res.* 16: 4077(1988)). The invention also can be used for randomly incorporating dimers or trimers of such sequences. The synthesis of these small polymers has been described by Akira Ono et. al. (Ono A., et. al. *Nucleic Acids Res.* 23, No. 22, 4677-4682, (1995)).

The wild type sequence refers to the original sequences from which the degenerated sequences are derived, which are the base (source) sequences at the beginning of the synthesis according to the invention. The wild type sequences of these biopolymers are known prior to the synthesis process. These sequences may be isolated as DNA, RNA, or peptide sequences that are present in their natural form. Such isolation and sequencing techniques are known by one skilled in the art. The wild-type sequence also may created by hypothetical imagination, rational design or computer modeling (e.g., Zhou, G. et al., *Science* 256:1059-1064 (1994)). Rational design means constructing a DNA, a RNA or a protein using the information obtained from various sources, such as a homologue study and DNA, RNA or protein structure study. Wild type sequences may be from biochemical selection, such as the SELEX procedure (Gold, L. et. al., U. S. Pat. No. 5,270,163) or aptamer technology (Ellington, A. D. *Nature* 346:818-822 (1990)). They also can be generated from any combination of the methods mentioned above. The sequences can be from several different sequence sources.

A random library refers to a pool of different biopolymer species derived from the same wild type sequences. The differences between each of the biopolymers only occur at specified positions in the biopolymer sequence with the present invention.

Specified positions refer to any residues in a selected biopolymer sequence which are subjected to random synthesis by the present invention. The selection of the positions for random substitution may be derived from what has been learned in previous experiments. The selected positions can be from a rational design or can be accomplished by means of computer modeling and predictions. The positions also may be chosen in order to test a hypothesis or the function of a molecule. In addition, positions may be selected for the purpose of the creation of therapeutic drugs

or molecules used for diagnostic purposes.

Doping synthesis refers to the random replacement or substitution of a known residue or residues at a given probability with any other molecule that may be incorporated while preserving all other residues as wild type. The substitution residue  
5 may be any other deoxynucleotide, ribodeoxynucleotide, amino acid or the analog of these molecules or even may be any totally different chemical molecule. The specified position for random synthesis can be at one or more positions. The positions may be next to each other or may be dispersed in the sequence. In the synthesis a mixture of incorporable molecules is used for the random synthesis, and the  
10 probability of the random synthesis is determined by the mixture portions of the ingredients. In Fig. 1, four doping reservoirs are depicted. The actual number of doping reservoirs can vary from 1, 2, 3 or more. The number of mixtures (doping reservoirs) used for the random synthesis is unlimited by the site doping synthesis.

The substitution rate refers to a given probability of any one given substitution  
15 within the wild type sequence. The substitution rate can be from 0 to 100 percent and may vary from position to position. For example, one position may be randomly substituted at a 10 percent rate while the other position may be designed for 90 percent substitution by increasing the incorporation rate. Random substitution refers to an even chance of replacement of any individual synthetic molecule. The rate is  
20 also calculated as the average number of substitutions per synthesized molecule. The average number of substitutions can be zero or more up to the size of the synthesized sequence.

The relationship between mutation rate and the percentage of doping mixture can be expressed by the formula:

$$25 \quad C(\%) = (M \times 4 \times 100)/(N \times 3) \quad (1)$$

In the formula (1) (Hutchison, C. A. et al., *Methods in Enzymology* 202:356-390 (1991)), C(%) is the percentage by volume of equimolar phosphoramidite mix added to a pure phosphoramidite, M is the desired average number of substitution mutations per clone, and N is the size of the mutagenic target in nucleotides. The factor 4/3 is  
30 present because only three of the four phosphoramidites in the equimolar mixture will be different from the wild-type base and hence mutagenic. Once the average mutation rate is determined, one can calculate the amount of doping of a pure synthesis reagent.

The synthesized polymers can consist of many different types of molecules. They can be classified into zero site replacement (wild type), single site replacement, double site replacements, triple and quadruple or more sites of replacement with oligonucleotides. If the doping process is evenly random, each class of oligonucleotide within the synthesized library will obey the Poisson distribution which can be described using the formula:

$$f(i) = (e^{-m} m^i) / i! \quad (2)$$

where in the formula (2) (Suzuki, David T. et. al.: An Introduction to Genetic Analysis (3rd. edition) pp 102, (1986)),

- 10  $f(i)$  is probability distribution of one particular class;
- $e$  is the base of natural logarithms (it is approximately equal to 2.7);
- $!$  is the factorial symbol. For example,  $3! = 3 \times 2 \times 1 = 6$ , and by definition  $0! = 1$ ;
- $m$  represents the mean number of replacements in the library;
- $i$  represents the number of replacements for a particular class;

- 15 Using this formula, the percentage of one particular class in the library can be calculated. The distribution of each class of the replacement is determined by the average replacement "m". For example, when an average of one replacement per oligonucleotide is designed, the synthesized oligonucleotide will contain  $f(0) = 36.8\%$  which will contain no replacements within the oligonucleotide,  $f(1) = 36.8\%$  will contain one replacement and  $f(2) = 18.4\%$  will contain two replacements per oligonucleotide and so on. The library contains almost no oligonucleotide having more than 5 replacement sites. When the average replacement ( $m$ ) is increased, the proportion of the oligonucleotides having multiple replacement sites will also increase. The maximum number of replacements within an oligonucleotide in the library is
- 25 limited by the number of positions designed for the substitution synthesis in the present invention. For example, if an average substitution rate is designed as 1.5 replacements per oligonucleotide (the length of the oligonucleotide being 40 bases), and ten positions are designed using the invention, the resultant oligonucleotide molecules will contain either 0, 1, 2...10 positions randomly substituted by others.
- 30 Therefore, the maximum number of positions that can be substituted is 10 per molecule, with the rest of the positions (30, where there are 40 substitutions) being preserved as the wild type base.

The replacements in the oligonucleotides of a particular class may also vary from molecule to molecule. For example, one oligonucleotide might have a substitution at one position while the other oligonucleotides in the same class may have a substitution at another position. The number of different type of molecules in one particular class can be calculated by the combination rule. For example, when the designed substitution number for a library is n, the number of different molecule types in the i class is

$$\text{Combinations} = \frac{n!}{i!(n-i)!} * 3^i \quad (3)$$

This formula (Feller, William (1970). An Introduction to Probability Theory and Its Applications. 3rd Ed., John Wiley & Sons, New York) assumes that for the average number of three possible nucleotide replacements, i.e., among the four different nucleotides A, C, G, T, one should be the wild type and is not considered as a substitution, where:

n = total designed mutational positions

15 m = the number of mutations in a synthesized oligonucleotide

3 means every position having 3 substitutions evenly.

The presentation of each class can be calculated by the formula given above although in practicality the synthesis does not always result in a random oligonucleotide synthesis since other factors such as the reaction rate of the four phosphoramidites may also be involved. The individual molecules can be identified by any established techniques (Sambrook et. al., Molecular Cloning- A Laboratory Manual, 2nd Edition, (1989)). The synthesized biopolymer libraries also can be used before the purification and the identification of individual molecules.

25 The synthesized oligonucleotide may also contain reactive sites for enzymes such as cleavage sites for restriction endonucleases or promoter sites for RNA polymerases. Such sites would allow, for example, for cloning of amplification products or transcription of amplification products.

The synthesis of the oligonucleotides of the present invention can be performed using a special oligonucleotide synthesizer which contains two groups of reservoirs: one is for pure reagents and the other is for doped reagents. The contents of each reservoir in the two groups may be completely different. For example, one set of reservoirs may contain the conventional A, T, G or C phosphoramidites or any other

pure reagents. The second group of reservoirs may contain these same phosphoramidites with each one "doped" with a mixture of the other three phosphoramidites or any nucleotide analogues. By programming the synthesizer to use either the conventional or modified set of reagents, one can choose which sites in the oligonucleotide remain "wild type" and which will contain the mutation sites. The number of the "doped" reservoirs used in this example is four, but there can be more if a special doping rate or a special mixture is used. The resultant mixture of synthesized biopolymers, in this case, oligonucleotides, can be termed a biopolymer library.

In order to control quality in a production line or for other reasons, it may be necessary to synthesize several different types of oligonucleotides (several biopolymer libraries or pure synthesized oligonucleotides) and mix them to make a random oligonucleotide library. The mixed oligonucleotides will yield a random library with a precise composition. This is particularly useful for clinical drug production.

The invention herein may be used with a wide variety of previous technologies, some of which are discussed below.

#### Use of the invention in mapping

In random peptide and epitope mapping, the peptide of interest may be an antigenic peptide, a peptide hormone, or an antibiotic peptide. Typically, in structure-function studies, it is desired to determine the effect on activity of one of a variety of amino acid substitutions at one or more selected residue positions. Peptides that interact with proteins, for example, enzyme inhibitors or receptor agonists and antagonists, are used in clinical therapy and in drug development research. Discovery of peptides having affinity to enzymes, hormones, receptors or other molecules may lead to the development of a new therapeutic drug or diagnostic reagent. In general, these small peptides can be developed by rational design or isolated by screening large numbers of naturally occurring or synthetic compounds. Frequently, there is not enough information on the amino acid sequences available for design. Also, assembling and screening a large library of compounds can be time consuming and expensive.

Phage display is a technology using biological expression systems such as bacteriophages or bacteria to facilitate both the production of large libraries of random

peptide sequences and the screening of these libraries for peptide sequences that bind to particular proteins (Devlin, J. J. et al, Science, 249:404-406 (1990)). Filamentous bacteriophage, such as M13, display three to five copies of the gene III protein (g3p) at one end of the virion (Lin. T.-C., et. al., J. Biol. Chem. 255:10331 (1980)). This display is essential for proper phage assembly and infectivity by attachment to the pili protein of *E. coli*. Electron micrographs have shown that g3p appears as a nodule linked to the phage by a flexible linkage that contains a series of Gly-Gly-Gly-Ser repeats. The phage display technique inserts small randomly synthesized DNA fragments between the amino- and carboxy-terminal domains of gene III in the phage genome. The progeny phages hence display small random peptides. The phage library displaying random peptides are further selected by exposure to a target protein immobilized on a solid support. After washing away the non-binding phage, the bound phage are recovered and the identity of binding peptides easily determined from the DNA sequences of the isolated phage. Entire libraries can be quickly screened in one tube (Devlin, J.J. et. al, Science, 249:404-406 (1990)). However the technology is limited by the library size which is usually determined by the transformation efficiency of the DNA into the *E. coli* host. The transformation efficiency is often very low in comparison with the desired library. Thus, certain random sequences may be eliminated in the process and these sequences, which may contain the sequence having high affinity for the target protein, will never be present in the phage library. In theory these lost potential high affinity sequences usually have similar sequences that are already isolated by the phage display.

The present invention can help to recover these sequences having higher affinity in combination using the conventional phage display technology. After the sequence is isolated with the conventional phage display technology, the degenerated sequences can be made by site doping synthesis as described in present invention. The degenerated sequence may be designed to have one, two or more residues which are randomly replaced. The replacement can occur at specified positions within the sequence to reduce the complexity of the library. The synthesized library then is incorporated into its host phage genome for phage display selection.

Phage display technology is also used for antibody epitope mapping. Tens of millions of short peptides can be easily surveyed for tight binding to an antibody,

receptor or other binding protein using an "epitope library". As described above, the library is incorporated into a vast mixture of filamentous phage clones and the peptide sequences are displayed on the virion surface. The survey is accomplished by using the binding protein affinity to purify phage that display tight-binding peptides on the phage followed by sequencing the corresponding coding region in the viral DNA's. Although the library used in phage display can be very large, like other libraries, it is often incomplete. Using the technology of the present invention, a complementary library can be made. The library can be more sequence oriented and more specific. Potential applications of the epitope library include investigation of the specificity of antibodies and discovery of mimetic drug candidates.

#### Use of the invention in aptamer technology

As discussed in the background section herein, aptamer technology is used to identify a protein inhibitor from within a mixture of oligonucleotides or to isolate high-affinity nucleic acid ligands for a protein, traditionally done using the Oliphant random oligonucleotide library method. Aptamer technology (Ellington, A. D. Nature 346:818-822 (1990)) employs alternate cycles of ligand selection from pools of variant sequences. The procedure is also called SELEX technology (Gold, L. et. al., U. S. Pat. No. 5,270,163). The technology is used to determine the optimal binding sequences for any nucleic acid binding to protein. The isolated single- or double-stranded nucleic acids are named aptamers which are capable of binding proteins or other small molecules. Large randomly generated populations can be enriched in aptamers by *in vitro* selection and polymerase chain reaction (PCR). Aptamers used for the purpose of therapeutics would most likely bind to proteins involved in the regulation and expression of genes (i.e. transcription factors). The presence of the aptamer would act as a sink for the protein factors preventing the factors from carrying out their normal functions and presumably modulating the expression of the genes dependent upon the activity of the protein.

The technology involves nucleic acid ligands that inhibit replicative proteins of epidemiologically important infections. The technology has been used successfully for isolating the inhibitor for Human Immunodeficiency Virus type I (HIV-1) tat protein of HIV infections. Many other ligands and inhibitors, such as those for anti-coagulation of human thrombin, can also be isolated using this technology (Schneider,



D. J. et al., *Biochemistry* 34:9599-9610, (1995)).

In general, the RNA or DNA libraries for the SELEX technology are made using the random strategies. However the conventional random libraries used in the experiments do not always contain all possible combinations of the DNA or RNA molecules because the amount of synthesis is often not sufficient to suit all kinds of combinations and therefore the libraries are often not complete. The technology is also limited by other factors such as the transforming efficiency of DNA into the host cell. Like many other applications using the conventional random libraries, the SELEX procedure often involves loss of random combinations for the designed libraries. Those lost combinations could be the best inhibitor molecules for a protein. The sequences of these best inhibitors are often similar to the RNA or DNA molecules previously isolated using the technology (Tuerk, C. et. al., *Science* 249:505-510(1990)). Therefore, with the assistance of computer modeling, using known techniques and using the isolated DNA or RNA sequence as the base sequences, one can redesign the libraries using the present invention. The library can contain many variants of the isolated ligands. The best binding sequences can be identified. Using the present invention a large DNA sequence that can bind to protein or other molecules also can be obtained.

#### Use of the invention in PCR diagnosis

PCR amplification can exponentially amplify DNA or RNA from a very small amount of template sequence (Mullis K. et al., US Pat. No. 4,683,202). In this technique, a denatured DNA sample is incubated with two oligonucleotide primers that direct the DNA polymerase-dependent synthesis of new complementary strands. A cycle of the synthesis results in an approximate doubling of the amount of the target sequence. Each cycle is controlled by simply varying the temperature to permit denaturation of the DNA strands, annealing the primers, and synthesizing the new DNA strands. The temperatures used in the PCR process are usually 93-95°C for the denaturing, 30-55°C for the annealing and 72°C for the synthesis (Saiki, R. et al., U.S. Pat. No. 4,683,194). Multiple cycle amplifications are usually used in the PCR process. One advantage of PCR is that it can increase the signal intensity by several million fold in a very short time. For example, twenty-five amplification cycles increase the amount of target sequence by approximately  $10^6$  fold. The amplified

sequence can be further analyzed if desired. The PCR techniques can be performed on samples of cells without prior DNA purification. Therefore numerous PCR diagnosis kits have been developed and commercialized for clinical purposes.

5 The accuracy of the PCR method is mainly determined by the hybridization of the primers to the template. In general the hybridization temperature varies from reaction to reaction. As discussed above, it can range from 30 to 55°C for the PCR reactions. The lower temperature makes the primer-template hybridization easier but this can also lead to a greater probability of a mismatch of primers to the DNA template. Raising the temperature can increase the accuracy of the hybridization  
10 process. One can lower the annealing temperature for hybridizing all possible sequences with the primer.

The ability of human immunodeficiency virus (HIV) and other viruses to escape the host's immune response due in part to the capacity for extensive genetic variation (Nowak, M.A. et al., *Science* 254:963-969 (1991)) is one of the key features  
15 of this virus that complicates development of an effective vaccine for the prevention of AIDS. Like many other RNA retroviruses, HIV mutates its genome at a rate about a million times faster than in the eukaryotic genome (Nowak, M., *Nature* 347:522 (1990)). Therefore HIV is capable of considerable mutational drug resistance. It has been known that HIV can completely mutate itself to form a new strain for drug  
20 resistance in about two weeks. Using conventional PCR technique it is extremely hard to accurately diagnose such hypermutational viruses, especially for quantitative purposes. This is because lower temperatures of hybridization can cause nonspecific binding to the contaminated DNA sequences and result in false-positive results. Although raising the annealing temperature can eliminate this nonspecific binding to  
25 the contaminated DNA, it also reduces the probability of detecting the mutated forms of virus DNA and leads to a higher rate of false-negative results (Bockstahler, L.E., *PCR Methods and Applications*, 3: 263-267(1994)).

The oligonucleotides synthesized by the present invention are especially useful for the amplification of regions of a nucleic acid which are highly variable due to a  
30 high mutational frequency. The primers prepared by the present invention can better hybridize to the mutated HIV genome sequence since it can simulate the mutation of the viruses. The sequence variety of the present invention can be controlled by

random synthesis at mutational hot positions of the target sequences. With the oligonucleotides prepared by the present invention, the PCR annealing temperature can be set very high so that only perfect matches between the primers and the template can hybridize and amplify. With the invention, both wild type and mutated sequences of hypermutable viruses can be detected simultaneously. Nonspecific primer-template hybridization is eliminated which could otherwise also be amplified from the mutated form of the DNA sequences. Therefore the detection accuracy of the diagnosis is raised using this method. Using the oligonucleotides made by the invention as primers, viral sequences that are randomly integrated in a genome or otherwise present in a cell and which have a tendency to have a high mutational rate also can be detected and amplified for further study.

Kits for detecting viruses having such high mutation rates can be developed from the present invention, which also can be used for the quantitative determination of the viruses which have high mutation rates, a process that may be difficult using conventional PCR techniques. Such kits may contain a certain amount of the randomly synthesized oligonucleotides made by the present invention, thermostable DNA or RNA polymerase and cofactors, such as buffer, and a sufficient amount of reagents such as deoxyadenosine triphosphate (dATP), deoxythymine triphosphate (dTTP), deoxycytosine triphosphate (dCTP), deoxyguanosine triphosphate (dGTP), cytosine triphosphate (CTP), adenosine triphosphate (ATP), guanosine triphosphate (GTP), uridine triphosphate (UTP) or nucleoside triphosphates for oligonucleotide synthesis.

#### Use of the invention to change target recognition in ribozymes

Ribozymes, or catalytic RNAs, are RNAs having catalytic activity which are capable of cleaving covalent bonds in a target RNA sequence (Cech, T.R., Science, 236:1532-1539 (1987), and Zaug, A. J. et al., Nature 324:429-433 (1986)). In general, ribozymes have two parts to their sequences: the core catalytic sequence (target digestion) and target recognition sequences. Catalytic activity requires the "ribozyme core" sequence and a divalent cation. The catalytic site is the result of the conformation adopted by the RNA-RNA complex in the presence of divalent cations. For most ribozymes, specificity is conferred by the recognition sequence, which is a two part sequence flanking the core sequence. A recognition sequence is also called

an internal guide sequence (IGS). When ribozymes cleave the target RNA, an IGS hybridizes to the target RNA sequence, where the two halves of ribozyme hold the target sequence together by base-pairing between complementary nucleotides. Each half of the ribozyme can be redesigned and produced (Haseloff, J. P. et al., U.S. Pat. No. 5254678). The sequence in the target RNA to which the IGS is hybridized is also called an external guide sequence (EGS). There is no consensus sequence required for the internal and external guide sequence and the only requirement is that they can hybridize to each other. The ribozyme can be engineered to cleave any RNA whose sequence is known. Thus the target sequence is extremely easy to manipulate. It has been demonstrated that ribozymes have great potential for the development of drugs used for interruption of gene expression and silencing of potentially directed gene sequence products.

Ribozymes having extensive therapeutic and biological applications have been documented (Stull R. A. et al. *Pharmaceutical Res.* 12:465-483 (1995)). For example, disease-causing viruses in man and animals may be inactivated by administering a ribozyme, which has been adapted to hybridize to and cleave RNA transcripts of the virus in accordance with the present invention, to a subject infected with a virus (Heidenreich, O. et al., *J. Biol. Chem.*, 267:1904-1909(1992)). The ribozymes substituted using the invention herein have particular application to viral diseases caused, for example, by the herpes simplex virus (HSV) or the AIDS virus. It has been reported there are several studies developing ribozyme drugs for antiviral purpose such as ribozymes directed against the HIV virus.

Synthetic oligonucleotides containing complementary sequences to variable target sequences can be synthesized according to the present invention. The ribozyme libraries can be made by ligating the prepared oligonucleotides with ribozyme core sequences. Such DNA manipulation can be done by one skilled in the art. The ribozymes generated by this method can recognize the randomly mutated sequence and degrade the mutated sequences, for example, the sequences from HIV or HSV because of the replacement of the target recognition sequences by random nucleotide sequences.

#### Use of the invention to synthesize antisense and antigene nucleotides

Antigene nucleic acid compounds are designed to bind to single-stranded or

double-stranded DNA. They are targeted to genomic DNA to interfere with transcription or replication processes. Antisense sequences are DNA or RNA sequences targeted to specific 'sense' sequences in the mRNA. Antisense sequences inhibit a specific protein expression. Like ribozymes, both antigene and antisense sequences have potential in drug development. An antigene or antisense sequence synthesized by a conventional method has only one type sequence and it often binds poorly to the high variable DNA or RNA sequences, for example, HIV. An antigene or antisense sequence generated by the present invention contains random composition sequences and is particularly useful for binding to these highly mutated sequences such as RNA viruses and retroviruses, such as HIV, which have a high mutation rate with highly variable genome sequences. Using the present invention, a library of DNA or RNA sequences having different randomly incorporated substitutions can be made for drug development to inhibit the pathogenic protein expression.

#### Use of the invention in mutagenesis to design proteins

One of the major goals of genetic engineering is the rational design of proteins (e.g. Ohana, B., et. al., Biochemistry, 29:6409-6415 (1990)), such as known proteins, by the transfer of genes coding for the material into appropriate host cells and through molecular biology. In protein engineering it is necessary either to create new genes or to mutate wild-type genes which may be transferred into an organism to produce a novel material. A method is needed which can efficiently generate a population of differing novel sequences from which desired novel sequences are selected. Such sequences can be transferred into host microorganisms to produce novel protein-like materials and to impart various desirable properties to microorganisms which are useful in other biotechnological processes. Furthermore, in a wide variety of biological studies relating to protein structure, synthesis and function such as studies of enzyme function, antibody binding, viral pathogenesis, transcription and translation, mutants with a wide variety of possible substitution mutations in a sequence are often desired.

Using the technique of the present invention, desired mutating positions can be selected and randomly synthesized. Mutation rate at each position can be controlled from 0% (wild type synthesis) to a 100% mutation rate. The process is fully automated and the number of positions is unlimited. The randomly mutated

positions can be localized or dispersed within the sequence. The method of the invention and the resulting oligonucleotides can be used by anyone skilled in the art. The method also can be used in combination with other methods if this is desired.

The features and advantages of the present invention will be more clearly understood by reference to the following examples, which are not to be construed as limiting the invention.

Example 1: Mutagenesis of A2-6 monomer using the site doping synthesis method

Source sequence used

A2-6 is a PCR generated DNA sequence derived from A-type promoter of mouse long interspersed element type 1 (LINEs-1) (Hutchison, C. A., et al., In Berg, D. E. and Howe, M. A. (ed.), Mobil DNA: 593-617). The sequence is 208 base pairs long, and has the following sequence (Sequence ID 1):

```

1      GTGCCTGCCC  CAATCCAATC  GCACGGAACT  TGAGACTGCA
      GTACATAGGG
15     51      AAGCAGGCTA  CCCGGGCCTG  ATCTGGGGCA  CAAGTCCCTT
      CCGCTCGACT
      101     CGTGACTCGA  GCCCCGGGCT  ACCTTGCCAG  CAGAGTCTTG
      CCCAACACCC
      151     GCAAGGGTCC  ACACAGGACT  CCCC GCGGGA  CCCTAAGACC
20     TCTGGTGAGT
      201     GGATCACA
  
```

The sequence was detected having promoter activity (Severynse, D. M. et al., Mammalian Genome 2:41-50 (1990)). In this project, we generated base pair substitution mutants to study the effect on the promoter activity of A2-6.

25 Primers synthesized by the method of the invention

Primers Design List 1 (used for mutagenesis; mutants from the designed positions in the primers were partially identified):

Primer Muta25; length: 45; Seq ID: 2

No. of mutation positions = 13

30 Average Mutations per oligo designed = 1.6

-21-

5' CGTGCGATTG GATTGGGGCA GGCACAAGCT TGGCGTAATC ATCCT 3'

Primer Muta21; length: 45; Seq ID: 3

Number of mutation positions=20

Average Mutations per oligo designed=1.6

5' GGTAGCCTGC TTCCCTATGT ACTGCAGTCT CAAGTTCCGT GCGAT 3'

Primer Muta53; Length 45; Seq ID: 4

Number of mutation positions=13

Average Mutations per oligo designed=1.6

5' CGAGCGGAAAG GGACTTGTGC CCCAGATCAG GCCCGGGTAG CCTGC 3'

10 Primer Muta93; Length:45; Seq ID: 5

Number of mutation positions=11

Average Mutations per oligo designed=1.6

5' GACTCTGCTG GCAAGGTAGC CCGGGGCTCG AGTCACGAGT CGAGC 3'

Primer Muta132; Length:45; Seq ID: 6

15 Number of mutation positions=20

Average Mutations per oligo designed=1.6

5' CGCGGGGAGT CCTGTGTGGA CCCTTGCGGG TGTTGGGCAA GACTC  
3'

Primer Muta165; Length:44; Seq ID: 7

20 Number of mutation positions=10

Average Mutations per oligo designed=1.6

5' TGTGATCCAC TCACCAGAGG TCTTAGGGTC CCGCGGGGAG TCCT 3'

### Primer List 2

Primer Design List 2:

25 Primer Muta46; Length:40; Seq ID: 8

Number of mutation positions=7

Average Mutations per oligo designed=1.6

5' ACTTGTGCCC CAGATCAGGC CCGGGTAGCC TGCTTCCCTA 3'

Primer Muta92; Length: 33; Seq ID: 9

30 Number of mutation positions=4

Average Mutations per oligo designed=1.6

5' AGGTAGCCCG GGGCTCGAGTC ACGAGTCGA GCG 3'

Primer Muta131; Length:45; Seq ID: 10

Number of mutation positions=5

Average Mutations per oligo designed=1.6

5' CGGGGAGTCC TGTGTGGACC CTTGCGGGTG TTGGGCAAGA CTCTG 3'

Primer Muta165-2; Length:44; Seq ID: 11

Number of mutation positions=4

Average Mutations per oligo designed=1.6

5' TGTGATCCAC TCACCAGAGG TCTTAGGGTC CCGCGGGGAG TCCT 3'

#### 10 Synthesis and purification of oligonucleotides

Oligonucleotides were synthesized with an Applied Biosystems model 380A DNA oligosynthesizer. The synthesis program was supplied by Applied Biosystems Inc. Standard operating procedures were used except for the preparation of the mutagenic nucleotide phosphoramidite mixtures. The contents of 0.5-g bottles of each of the four phosphoramidites were dissolved in the following amounts of anhydrous acetonitrile injected through the septum to give 0.1 M solutions (A, 5.8 ml; G, 6.0 ml; C, 6.0 ml; T, 6.7 ml). After the phosphoramidites were completely in solution, all four were uncapped. A solution was made by simply mixing equal molar amounts of the four pure solutions. This mixture was then added to each of the pure phosphoramidites to be used in the synthesis of randomly mutagenized oligonucleotides. The amount added for doping is calculated as:  $C(\%) = (M \times 4 \times 100) / (N \times 3)$ , where, as discussed above, C(%) is the percentage by volume of equimolar phosphoramidite mix added to each pure phosphoramidite, M is the desired average number of substitution mutations per clone, and N is the size of the mutagenic target in nucleotides. In the case of Site Doping Synthesis (SDS), N is the number of the positions needed for doping synthesis. After doping the mixture, the bottles were covered temporarily with Parafilm, swirled briefly, and placed on the machine as quickly as possible by using the bottle changing routines.

Since the 380A DNA synthesizer only has seven reservoir positions, and the SDS method requires minimum eight reservoir positions (four for pure reagents for synthesizing wild type residues and another four for the synthesis of the doping



positions), the procedure was not fully automatic using this synthesizer. In many situations of site doping synthesis, an interruption is therefore required in order to replace one reservoir, which was subsequently programmed into the method. For example, the synthesis of the oligonucleotide SeqID 2 was programmed for four  
5 positions for the wild type synthesis which uses the pure phosphoramidites and the remaining three positions were first programmed for A, T, and G doping synthesis. Since there was only one T needed for doping synthesis in Oligo SeqID 2, the synthesis was interrupted after the T was synthesized and the doped T reservoir then was replaced by the doped C reservoir and the synthesis was continued. A DNA  
10 synthesizer having a capacity of eight or more reservoirs can be used to fully automate site doping synthesis without interruption. Such machines are commercially available. (for example, EXPEDITE™ 8909 nucleic acid synthesis system from PerSeptive Biosystems).

After the oligonucleotide was synthesized, the detritylated oligonucleotides  
15 were incubated with ammonium hydroxide at 55°C overnight to remove the protection groups used for synthesis. The tube was capped tightly since the ammonia develops considerable pressure on heating and can escape at this temperature. After heating at 55°C, the tube was placed at -20°C for one hour to prevent the liquid from "boiling" out of the tubes in the Speed-Vac. Several holes were pierced in the cap. The  
20 contents were then dried in the Speed-Vac to a white residue. The dried substance was resuspended in 200 µl of TE and stored at -20°C. The oligonucleotide was electrophoresed with Form-dye (XC-BPB) and Formamide on a 20% polyacrylamide, 45% urea gel in 1x TBE buffer (1x TBE: Tris base, 48.2 g; glacial acetic acid, 11.4 ml; EDTA, 20 ml of a 0.5 M solution (pH 8.0), add water to 1 liter). The gel was  
25 made using cast cassettes as is known in the art. A lower percent acrylamide may be desirable for longer oligonucleotides.

The gel was run until the desired separation was reached (BPB Runs with 8-mer, XC runs with 28-mer), and the gel was taken off the plate and wrapped in SARAN-WRAP™. The gel was placed on thin layer chromatography and examined  
30 with a short wave UV lamp. Oligonucleotides appeared as dark bands against a bright fluorescent background. The band containing the oligonucleotides was cut from the gel with a scalpel and stored in a screw-capped vial at -20°C.

-24-

The gel slices were crushed into fine particles in the vial, using a glass rod, in 1.0 ml of 0.5 M ammonium acetate. The oligonucleotide was eluted by agitating the vial overnight at 37°C. The gel was filtered off from the DNA solution by using a Pasteur pipet with glass wool. The volume of oligonucleotide solution was reduced  
5 to 200  $\mu$ l by 2-butanol extractions. The reduced oligonucleotide solution then was transferred to a new 1.5 ml Eppendorf tube. The oligonucleotide was precipitated by adding 1 ml cold 100% ethanol in a -70°C freezer for 1 hour. The Eppendorf was centrifuged for 10 minutes. The ethanol was removed and the pellets were dried in a Speed-Vac for 15 minutes. The pellet was resuspended in 180  $\mu$ l TE by vortexing  
10 and then centrifuged for 1 minute to remove anything not resuspended and transferred to a new Eppendorf tube. 20  $\mu$ l of 3 M sodium acetate (pH 5.2) were added along with 1 ml 100% ethanol to precipitate the DNA. The sample was then placed in a -70°C freezer for 1 hour. The sample was centrifuged for 10 minutes, the ethanol removed, and the pellet was dried. The pellet was resuspended in 200  $\mu$ l TE and  
15 stored at -20°C. The concentration was determined on a spectrophotometer at 260 nm wavelength.

#### In vitro synthesis and library production

Oligonucleotides synthesized on automated synthesizers have 5'-OH termini. These oligonucleotides therefore must be phosphorylated before they can be used in  
20 an *in vitro* mutagenesis experiment. This is because after primed-DNA synthesis has proceeded completely around the circular template, a phosphodiester bond must be formed between the 3' end of the last polymerized base and 5' end of the mutagenic oligonucleotide in order to form a covalently closed circle.

Using a maximum of 15 pmol of oligonucleotide in a volume of 15  $\mu$ l or less,  
25 2  $\mu$ l of 10 x oligonucleotide kinase buffer (0.5 M Tris-HCl, pH 7.5, 0.1 M MgCl<sub>2</sub>, 50 mM dithiothreitol), 2  $\mu$ l of 10 mM rATP, and 10 units of T7 polynucleotide kinase were added. The volume was brought to 20  $\mu$ l with TE buffer and incubated at 37°C for 1 hr. The kinase was heat-inactivated by incubating at 65°C for 10 min. The solution can be stored at -20°C.

30 Kunkel's "uracil template" method (Sambrook et. al., Molecular Cloning- A Laboratory Manual, 2nd Edition, 1989) was used to produce populations of clones containing sequences derived from the randomly mutagenized oligonucleotide.

-25-

preparations. In this procedure the DNA to be used as template is prepared from an *E. coli* strain which carries mutations in the *dut* and *ung* genes (e.g. CJ236). The mutations result in the synthesis of DNA in which uracil residues are substituted for a small fraction of the thymines. However the uracil incorporated into the DNA does not synthesize a correct second strand DNA from this template which is now free of uracil. After *in vitro* synthesis of a second mutagenized DNA strand (free of uracil) the uracil-containing template strand can be selected against by introduction of the DNA into *E. coli* which is wild-type at the *dut* and *ung* loci (e.g. JM101). This method allows production of mutant libraries in which 70-80% of the clones contain sequences derived from the oligonucleotide preparations. The Kunkle method is well known by one skilled in the art.

To hybridize the oligonucleotide primer to the uracil template, 0.8 pmol of uracil-containing template, 0.8 pmol of phosphorylated mutagenic oligonucleotide, and 2.4  $\mu$ l of 10 x SSC (1.5 M NaCl, 150 sodium citrate) were mixed in an Eppendorf tube. The mixture then was brought to a final volume of 24  $\mu$ l with water. Two control reactions were set up in the same format except the primer was 0.2 pmol and 0 pmol (substituted with water) respectively. The Eppendorf tube was placed in a 1-liter beaker containing 500 ml of water, in a 65°C incubator. The beaker and Eppendorf tube were left at 65°C for 15 min and then removed to room temperature for about 3 hours until ambient temperature was achieved.

Next, the template strand in the annealing mixture was converted to a covalently closed circular DNA molecule by DNA synthesis and ligation. The reaction was begun at a lower temperature such that annealed molecules with unstable ends (i.e., a mismatch near the ends) would have a better chance of priming DNA synthesis. The bulk of the reaction occurred at 37°C, the optimal temperature for T7 DNA polymerase activity. The reaction was then placed at a lower temperature so that molecules with mismatches close to the 5' end of the oligonucleotide would have a better chance of ligating.

For each *in vitro* synthesis reaction, a mixture of 20  $\mu$ l of 100 mM Hepes (pH 7.8), 2  $\mu$ l of 100 mM dithiothreitol, 1  $\mu$ l of 1 M MgCl<sub>2</sub>, 0.5  $\mu$ l of 100 mM each of the four dNTPs, 10  $\mu$ l of 10 mM rATP, 2 units of T4 DNA ligase, 2.5 units of T7 DNA polymerase was made. The mixture was brought to a final volume of 76  $\mu$ l

with water. The 24  $\mu$ l of annealed oligonucleotide and template DNA were then added to the reaction mixture. The final mixture was first placed on ice for 5 min, then at room temperature for 5 min, 37°C for 2 hr, room temperature for 15 min, and then on ice for 15 min. A volume of three  $\mu$ l of 0.5 M EDTA was added to stop the synthesis reaction.

Aliquots of the products of the synthesis reactions were analyzed on a 0.8% agarose gel in 1 X TAE buffer containing 0.5  $\mu$ g/ml ethidium bromide. For comparison, single-stranded circular DNA and double-stranded replicative form I and form II DNAs was loaded to the adjacent lanes of the reaction samples. 100 ng of each sample was loaded. The gel was examined under a UV light source. Two main bands were seen in the synthetic product which comigrates with the two bands, which comprise form I, or covalently closed circular DNA, whereas the slower band represents form II, or open circular DNA (Fig 2).

#### Transformation and library making

The *in vitro* synthesized DNA was introduced into *E. coli* strain DH5 $\alpha$ F which was wild type at the *dut* and *ung* loci. In general, the transformed cells gave between 10<sup>4</sup> and 10<sup>5</sup> clones of the mutant library. Since it is possible to get clones with a mixed genotype, it was necessary to "purify" the library. This was accomplished by transferring the library from cells into filamentous particles by washing the library (represented by the bacterial particles) off the agar plate, growing in liquid culture, and infecting with helper phage M13 K07. The filament DNA was individually packaged into filamentous particles. The particles were then used to reinfect a clean background of DH5 $\alpha$ F. On infection with a filament, the DNA replicates as a double-stranded molecule because of the lack of the helper phage gene products. Since the filament carries an ampicillin resistance gene, the infected *E. coli* should be grown into a colony on an ampicillin plate. The subsequent colonies were used for phenotype and genotype analyses.

Individual colonies were picked from the plates having well-isolated colonies. A liquid culture was grown and used to prepare single-stranded DNA for sequencing to determine the genotype of each clone. At the same time a separate culture was inoculated for phenotypic assay of each clone and to store a sample from each colony to determine phenotypes at a later time. The isolated DNA then was sequenced by

the dideoxy sequencing method (Sambrook et. al., Molecular Cloning- A Laboratory Manual, 2nd Edition, 1989). This method is known by one skilled in the art.

Example 2. Mutagenesis results from the primers synthesized by the site doping synthesis method

5           The ultimate goal of the project was to isolate single substitution mutants from all sites of the A2-6 promoter. In minimizing the repeat isolations we first randomly generated mutants of the A2-6 promoter sequence of mouse LINEs-1 using the saturation mutagenesis method (Hutchison C. A. et. al., Proc. Natl. Acad. Sci. USA 83:710-714 (1986)). Approximately two-thirds of the all positions of the A2-6  
10 promoter were found to be single substitution mutants. In order to avoid substantial repeat isolation of mutants, we redesigned the mutation library by the multiple site orientated mutagenesis (MSOM) strategy. With this strategy, mutagenic oligonucleotides were synthesized by using the site doping synthesis (SDS) method. The generated libraries with this method contain oligonucleotides randomly  
15 incorporated with non-wild type bases only at the specified positions (the bold type positions in oligonucleotide sequences in the primer lists 1 and 2). In the oligonucleotide sequences of list 1 and list 2, the bold type characters represent the positions for which no single substitution mutants were isolated. The oligonucleotides represented by plain text characters were synthesized as wild type and no mutation  
20 was expected from the SDS libraries. In this round we stopped isolating mutants when an average of about 2/3 of the redesigned positions were isolated single site mutation mutants (Table 1 and Table 2). We redesigned the mutagenesis libraries again. The average repeat isolations of single substitution mutants at this time was about 2.4 (Table 1).

25           The oligonucleotides were designed containing some part of overlap sequences. This is because the designed oligonucleotides will contain the non-wild type bases which are unstable compared to the wild type sequences and the exonuclease activity of the polymerase can remove these non-wild type substitutions near the 3' end of the oligonucleotides. Design of overlap sequences increases the chance of isolating  
30 mutants from these sites.

Table 1

Primers	SS	D	DS	$\geq 3$	DsgPsn	IsoPsn	Iso Rate	Rpt Iso	Total
Muta25	22	4	18	4	13	10/13	77%	2.2	48
Muta21	27	5	13	22	20	15/20	75%	1.8	67
Muta53	28	3	15	21	13	11/13	84%	2.5	67
Muta93	33	3	21	11	11	9/11	81%	3.6	68
Muta132	17	2	11	5	20	12/20	60%	1.4	35
Muta165	19	2	13	1	11	7/11	63%	2.7	35

Abbreviation legend: SS: single site mutants; D: deletion mutants; DS: double site mutants;  $\geq 3$ : mutants have 3 or more site substitutions; DsgPsn: number of designed mutating positions; IsoPsn: number of positions where single site mutants isolated; IsoRate: isolation rate of single site mutants in comparison with designed positions; RptIso: average repeated isolation.

After an appropriate portion of the designed positions were isolated as single substitution mutants from the first round of multiple site oriented mutagenesis (MSOM) library (example I), the second round of MSOM libraries was designed in completion of the isolation of single substitution mutants from all positions of the A2-6 sequence. Four oligonucleotides were designed for this round of SDS libraries. In completing this round of mutant generation and mutant isolation, one can fulfill the original goal, i.e., isolate single substitution mutants from all the positions in the A2-6 promoter which is 208 base pairs long. Mutants from the 5 positions left for isolating single substitution mutants should be easily isolated using the SDS method.

The strategy of multiple site orientated mutagenesis (MSOM) using the site doping synthesis method facilitates the positional precision of oligonucleotide mediated mutagenesis. It preserves certain sites as wild type while mutating other specific sites randomly. This is the key novelty and the utility of the site doping synthesis (SDS). So far we have not found any mutation from a site which is not specified by design. The random synthesis basically follows the Poisson distribution. We designed the library having an average substitution of 1.6. We observed the overall ratio of

mutants having single substitution, double substitutions and triple or more substitutions is in the ratio of 0.48:0.30:0.21 respectively. This is not significantly different from the Poisson distribution which is approximately 0.40:0.32:0.28. Here we could not use the isolated wild type clones to assess the results because they contain the clones from the other sources of the mutagenesis method. One observation is that the number of the isolated mutants having multiple site mutations is lower than expected. This is probably because the oligonucleotides containing more mutations are more difficult to hybridize to the template. The observation of multiple-site-mutation oligonucleotides difficult to hybridize to the wild type template strongly support the idea of using the SDS oligonucleotides to hybridize the mutated virus sequence as discussed in previous sections.

Deletion mutants which were not expected were also found from the mutagenesis library. This was also observed by other mutagenesis projects (Hutchison C. A. et. al., Proc. Natl. Acad. Sci. USA 83:710-714 (1986)). An explanation for this is that the detritylation is incomplete and the molecules bearing a 5' trityl group are unable to couple in one synthetic cycle, but they can participate in subsequent cycles. Or it may also be caused by the inefficiency of the capping reaction in the oligonucleotide synthesis process.

Except for a few unexpected observations, the MSOM strategy using the site doping synthesis method does not require numerous repeated isolations and is fast, simple to use, reliable and has many potential applications.

Table 2:

Primers	SS	D	DS	$\geq 3$	DsgPsn	IsoPsn	Iso Rate	Rpt Iso	Total
Muta46-2	9	ND	ND	ND	7	5/7	71%	1.8	9
Muta92	9	ND	ND	ND	4	4/4	100%	2.3	9
Muta131	11	ND	ND	ND	5	4/4	100%	2.2	11
Muta165-2	5	ND	ND	ND	4	3/4	75%	1.7	5

For abbreviations, see Table 1.

Although the method of the invention is preferred for making random

oligonucleotides for the variety of purposes discussed above, it is contemplated that for some new purposes discussed herein, where it is not critical to have completely accurate, random, site-specific substitution, oligonucleotides made according to prior, conventional methods may be adequate. There are a number of novel uses of random  
5 oligonucleotides made by any method, which are not evident from the prior work in this field. These include the use of random oligonucleotides to ribozymes, antisense and antigens, and PCR diagnosis kits. For more demanding methods such as aptamers, epitope mapping, random peptide synthesis and mutagenesis, however, use of the method of the invention is critical for success.

#### 10 Preferred Embodiment of the Invention

Preferably, the method of synthesizing biopolymers herein comprises obtaining a wild type biopolymer sequence with a plurality of residue positions; selecting one or more of the residue positions for random substitution; utilizing a first group of reservoirs containing at least one reservoir of unmixed residues; utilizing a second  
15 group of reservoirs, each of which reservoirs has been selected from the group consisting of reservoirs containing the same residues as the first group which have been doped with a selected mixture of other residues and reservoirs containing mixtures of residues different from the residues in the first group of reservoirs; and programming a synthesizer to use selected reservoirs from the two groups, in a  
20 predetermined sequence, resulting in synthesis of a first biopolymer library. Most preferably, the biopolymers are oligonucleotides and the residues are selected from the group consisting of phosphoramidites and any other incorporable residues.

#### Industrial Applicability

Synthetic oligonucleotides and other biopolymers are useful in the manufacture  
25 of therapeutic and diagnostic reagents, for development and clinical production of drugs, for kits, for example, for the detection of viruses, for the inactivation of viruses, for interfering with biological processes, for example, to inhibit specific protein expression, and to design proteins for specific purposes.

While the invention has been described with reference to specific embodiments  
30 thereof, it will be appreciated that numerous variations, modifications, and embodiments are possible, and accordingly, all such variations, modifications, and embodiments are to be regarded as being within the spirit and scope of the invention.



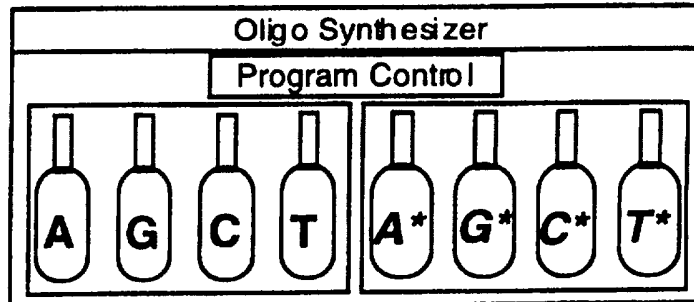
THE CLAIMSWhat Is Claimed Is:

1. A method of synthesizing biopolymers using a synthesizer to yield a library of biopolymers, comprising:
  - 5 (a) obtaining a wild type biopolymer sequence having a plurality of residue positions;
  - (b) selecting one or more of said residue positions for random substitution;
  - (c) utilizing a first group of reservoirs containing at least one reservoir of unmixed residues;
  - 10 (d) utilizing a second group of reservoirs, each of which reservoirs has been selected from the group consisting of reservoirs containing the same residues as the first group which have been doped with a selected mixture of other residues and reservoirs containing mixtures of residues different from the residues in the first group of reservoirs; and
  - 15 (e) programming said synthesizer to use selected reservoirs from said first and second groups, in a predetermined sequence, resulting in synthesis of a first biopolymer library containing biopolymers substituted at the residue positions.
2. The method of claim 1, wherein the biopolymers are oligonucleotides and the  
20 residues are selected from the group consisting of phosphoramidites and any other incorporable residues.
3. The method of claim 2, further comprising using the oligonucleotides for making random internal guide sequences of ribozymes.
4. The method of claim 2, further comprising using the oligonucleotides for  
25 making random antisense or antigene sequences for binding to a selected sequence of nucleic acid.

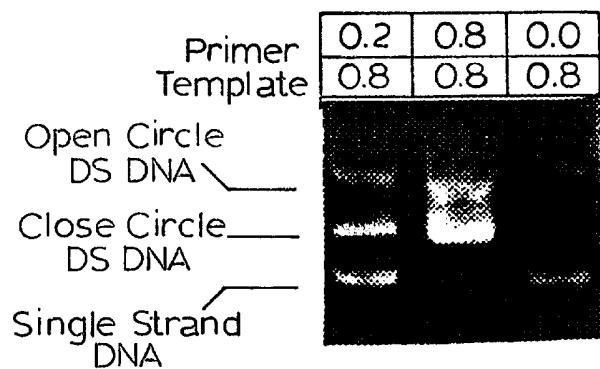
-32-

5. The method of claim 2, wherein the wild type biopolymer is isolated with conventional phage display technology, and the biopolymer synthesis is designed to make a high affinity target protein.
6. The method of claim 2, further comprising using the biopolymer synthesis to determine optimal binding sequences for binding of a nucleic acid to protein.
7. The method of claim 2, further comprising using the biopolymer synthesis to detect hypermutable base sequences.
8. The method of claim 2, further comprising using the oligonucleotides as primers in PCR amplification.
9. The method of claim 1, wherein the one or more residue positions are selected to increase mutation rate in the synthesized biopolymers.
10. The method of claim 1, further comprising mixing the first biopolymer library with one or more additional biopolymer libraries to make a random oligonucleotide library.
11. The method of claim 1, further comprising mixing the first biopolymer library with one or more pure synthesized biopolymers to make a random oligonucleotide library.
12. A method of making a random sequence selected from a group consisting of random antisense sequences, random antigene sequences and random internal guide sequences of a ribozyme, comprising:
  - (a) obtaining a wild type oligonucleotide sequence having a plurality of residue positions;
  - (b) selecting one or more of said residue positions for random substitution;
  - (c) utilizing a first group of reservoirs containing at least one reservoir of unmixed residues;

- (d) utilizing a second group of reservoirs, each of which reservoirs has been selected from the group consisting of reservoirs containing the same residues as the first group which have been doped with a selected mixture of other residues and reservoirs containing mixtures of residues different from the residues in the first group of reservoirs; and
- 5 (e) programming said synthesizer to use selected reservoirs from said first and second groups, in a predetermined sequence, wherein said residues are selected from the group consisting of phosphoramidites and any other incorporable residues.
- 10 13. A method of providing oligonucleotides useful for making random internal guide sequences of ribozymes; making random antisense or antigene sequences for binding to a selected sequence of nucleic acid; and as primers in PCR amplification, comprising synthesizing the oligonucleotides at one or more locations, said oligonucleotides having specificity for use in binding to a
- 15 selected target.
14. A kit for detecting biopolymers, comprising:
- (a) an oligonucleotide made by the method of claim 2;
- (b) a thermostable polymerase, selected from the group consisting of DNA polymerase and RNA polymerase; and
- 20 (c) reagents selected from the group consisting of deoxyadenosine triphosphate, deoxythymine triphosphate, deoxycytosine triphosphate, deoxyguanosine triphosphate, cytosine triphosphate, adenosine triphosphate, guanosine triphosphate, uridine triphosphate, and nucleoside triphosphates for oligonucleotide synthesis.



**FIG. 1**



**FIG. 2**

INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/09679

**A. CLASSIFICATION OF SUBJECT MATTER**  
 IPC(6) :C12Q 1/68; C12P 19/34  
 US CL :435/6, 91.2  
 According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**  
 Minimum documentation searched (classification system followed by classification symbols)  
 U.S. : 435/6, 91.1, 91.2, 183, 810; 536/23.1, 23.2, 24.33, 25.3; 514/44; 935/77, 78, 82

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
 Please See Extra Sheet.

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y,P -- X,P	US 5,589,332 A (SHIH et al.) 31 December 1996, see entire document.	1-12, 14 ----- 13
Y,P -- X,P	US 5,525,468 A (MCSWIGGEN) 11 June 1996, see entire document.	1-12, 14 ----- 13
Y -- X	US 5,468,613 A (ERLICH et al.) 21 November 1995, see entire document.	1-12, 14 ----- 13
Y -- X	US 4,965,188 A (MULLIS et al.) 23 October 1990, see entire document.	1-12, 14 ----- 13

Further documents are listed in the continuation of Box C.  See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 25 JULY 1997	Date of mailing of the international search report 29 AUG 1997
---------------------------------------------------------------------------	-------------------------------------------------------------------

Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer BRADLEY L. SISSON Telephone No. (703) 308-0196
-------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/09679

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y -- X	US 4,800,159 A (MULLIS et al.) 24 January 1989, see entire document.	1-12, 14 ----- 13
Y -- X	US 4,683,202 A (MULLIS) 28 July 1987, see entire document.	1-12, 14 ----- 13
Y -- X	US 4,683,195 A (MULLIS et al.) 28 July 1987, see entire document.	1-12, 14 ----- 13

**INTERNATIONAL SEARCH REPORT**

International application No.  
PCT/US97/09679

**B. FIELDS SEARCHED**

Electronic data bases consulted (Name of data base and where practicable terms used):

APS

Search Terms: Wang, Fenggang/in; random internal guide; ribozymes; synthesi?(P)oligonucleotides; hypermutable; hypermutable base; primers, per