Creating novel enzymes by applied molecular evolution

Adonis Skandalis, Lance P Encell and Lawrence A Loeb

Recent molecular techniques have made it feasible to simulate evolutionary processes and apply *in vitro* selection to evolve enzymes with novel properties that may have potential benefits for medical and industrial applications. The characterization of such mutants has also provided new insights into how molecular structure determines enzyme function.

Address: The Joseph Gottstein Memorial Cancer Research Laboratory, Departments of Pathology and Biochemistry, University of Washington School of Medicine, Box 357705, Seattle, WA 98195-7705, USA.

Correspondence: Lawrence A Loeb
E-mail: laloeb@u.washington.edu

Chemistry & Biology December 1997, **4**:889–898
http://biomednet.com/elecref/1074552100400889

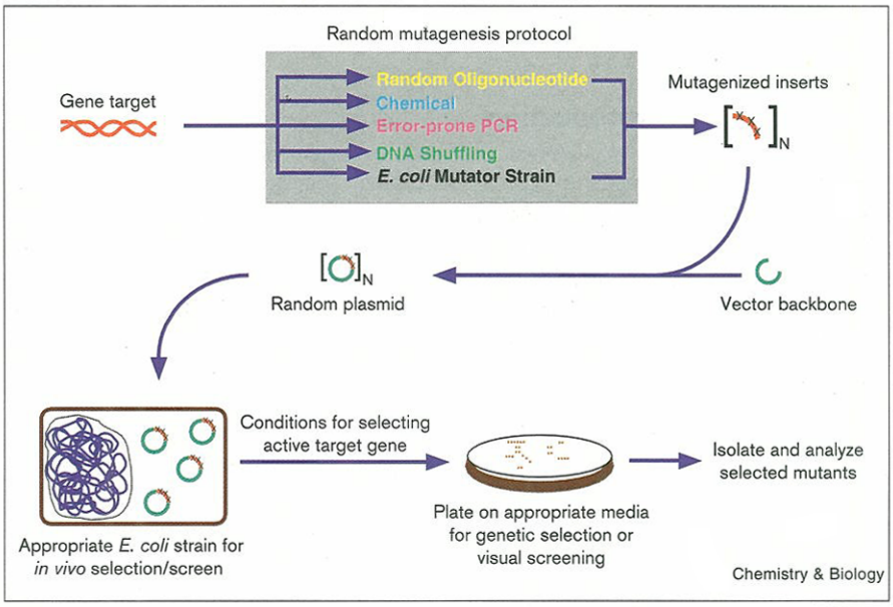
Introduction

The survival of a cell or organism is dependent on an intricate web of chemical reactions that are catalyzed, with remarkable efficiency, by enzymes. Clearly, the optimization of enzymatic processes is of paramount importance to the survival of a cell. In the context of evolution, however, where the unit of natural selection is the organism and not individual proteins, each enzyme is optimized for a variety of parameters in addition to catalytic efficiency. The 'perfect' enzyme in the mind of a biochemist might be a highly active one with a turnover equal to the diffusion rate of the substrate. Having a more active enzyme in the cell, however, may not be advantageous. Such an enzyme could generate the progressive accumulation of individual metabolites which probably would be disadvantageous in the cell, where there are complex and interrelated metabolic signal pathways for maintaining homeostasis Stability, substrate specificity and flexibility to respond to regulatory elements are all parameters likely to shape the properties of enzymes found in living organisms. Mutants of many different enzymes may therefore have enhanced catalytic activity or altered substrate specificities, but these mutants might not have been selected under natural conditions.

Another factor limiting the spectrum of mutant enzymes that can evolve naturally is historical contingency. It is thought that the majority of proteins evolved by gene duplication followed by sequential modification involving single amino acid substitutions [1]. The single amino acid replacements are then subjected to natural selection one at a time. For a mutation to be fixed, the resulting amino acid substitution must offer the cell a selective advantage over the wild-type protein, or at least be selectively neutral. Enzymatic enhancement or modulation by synergistic multiple mutations therefore cannot easily be explored in nature unless the mutations arise sequentially and each substitution results in a selective advantage. Often, however, an amino acid change that may be deleterious by itself, may enhance activity in the presence of other mutations (examples will be presented in this review), implying that a large number of sequences encoding active proteins have not been, nor will they probably ever be, naturally evolved.

The question of how many mutations a protein can tolerate has occupied scientists for a long time, and the answer has important implications for evolution as well as for protein engineering. Nearly 30 years ago, Maynard-Smith [1] proposed that the density of functional proteins differing from the wild-type by one amino acid (one

Figure 1



The potential pathways in the search for novel enzymes. The details of the methods for random mutagenesis are described in the text.

amino acid sequence space) must be high in order for evolution to be possible. In the last 30 years, it has become clear that indeed proteins can retain function even with several substitutions within their active site, and that some of these variants have different properties than the wild-type form.

Enzymes with enhanced or altered properties could benefit industry and medicine, and, at the same time, provide structure–function information about the protein. In industry, even marginal increases in enzyme stability or reaction rate could translate to significant cost savings. In medicine, gene therapy could benefit from the availability of certain mutant enzymes. Until recently, novel enzymes were obtained by searching for naturally occurring variants amongst exotic natural species that require a unique function in order to survive. One example is the group of thermostable DNA polymerases discovered in microorganisms living in hot springs that have revolutionized biological studies. These enzymes have made it feasible to dissect the human genome by amplifying single DNA molecules millions of times *in vitro* using the polymerase chain reaction (PCR).

In contrast to the millenniums required for natural evolution, it is now possible to generate mutant enzymes with altered properties rapidly *in vitro* using two general methodologies, summarized by the flow chart in Figure 1. One method involves the rational design of proteins by site-directed mutagenesis and the other utilizes random mutagenesis to mimic the evolutionary process. Random mutagenesis is also referred to as applied or directed evolution. This review will discuss these methods in protein engineering, with an emphasis on applied evolution of enzymes and the importance of genetic selection or high-throughput visual screening in obtaining mutants with the desired functions. In addition to reviewing current methods, we will discuss a few specific examples of proteins that have been successfully evolved to carry out novel or enhanced functions for potential use in gene therapy.

Rational protein design

The simplest approach to altering the structure of an enzyme involves the change of codons within a DNA sequence by site-directed mutagenesis. Amino acids are substituted, one residue at a time, at positions believed to be important on the basis of information about the protein's tertiary structure or previous mutagenesis experiments. The advantage of site-directed mutagenesis is that a site can be probed by specific substitutions. The results allow one to conclude whether a specific site can tolerate a specific change and to determine if changes alter the properties of a particular variant. No selection or screening is required because the mutant protein can be assayed directly. Bias towards certain substitutions when using site-directed mutagenesis can be overcome by making many changes at many positions. It should be noted, however, that interpretation of the results may be difficult. Loss of enzyme activity by certain substitutions does not indicate, by itself, that the replaced amino acid is directly involved in catalysis.

Site-directed mutagenesis is essentially a reductionist's approach which assumes that sufficient information can be obtained about amino acid interactions in a protein to

Table 1

The number of potential variants (sequence space) of polymerase β for various numbers of amino acid changes.

Number of amino acid changes	Sequence space
0	1*
1	6,061
2	18,310,281
3	36,760,940,821
319	19^{319}

*Theoretical sequence space of polymerase β (319 amino acids) based on the number of residues (*M*) being substituted by the 19 alternative amino acids. The number of sequences containing *M* amino acid substitutions is $19^M \times 319!/[(319-M)!M!]$.

make it possible to predict the effect of specific substitutions. Unfortunately, our current understanding of amino acid interactions within proteins and their effect on function is inadequate. Site-directed mutagenesis as a method to create new proteins with enhanced properties is therefore extremely limited. Moreover, it is not feasible to analyze multiple mutations. As an example, in order to examine all the possible single mutants (one amino acid sequence space) in rat polymerase β (pol β; 319 amino acids), 6061 site-specific mutations would be required. If one was interested in all the possible mutants, including those with multiple substitutions, 19^{319} variants would have to be constructed (Table 1).

Applied molecular evolution

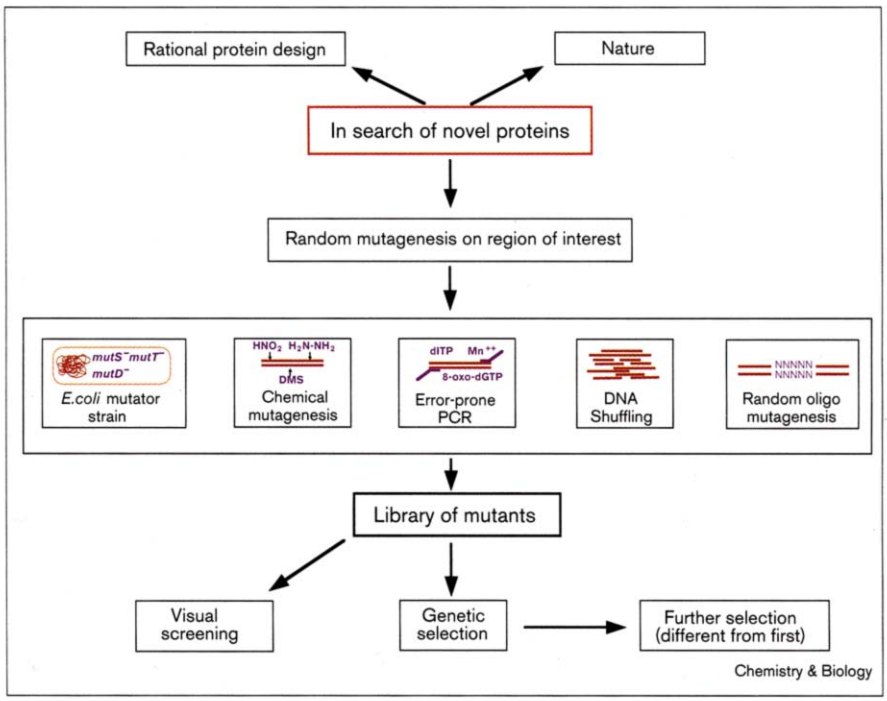
Applied molecular evolution circumvents the problems associated with rational design, as proteins are evolved for the characteristic(s) that we would like them to have. A detailed understanding of the changes necessary to reach such variants is not required. Instead of rationally designing proteins, powerful screening protocols can be used to sort out the enzyme with the desired properties from large libraries of mutant proteins. Spiegelman and Eigen [2,3] were among the early pioneers to use applied molecular evolution on Qb bacteriophage. The first *in vitro* evolution of DNA sequences involved insertion of random oligonucleotides followed by genetic selection [4]. We will discuss the specific approaches that have been used in applied evolution by reviewing first the methodologies used to generate molecular diversity (libraries of mutants) and then the methodologies used in the screening of these libraries for altered enzymes (Figure 2).

Methods for generating random diversity

The introduction of random mutations into a gene or gene segment is analogous to the natural processes that are operative during evolution, but the yield of mutations can be much higher. Mutations can be introduced throughout the coding region of a gene, or they can be confined to specific segments. In contrast to site-directed mutagenesis, a prior knowledge of the structure of the protein is not required. An overview of the methodologies to be presented in this review is shown in Figure 2.

Figure 2

The creation of random libraries by applied molecular evolution. A gene target can be randomized by one of several methods, followed by ligation of the library into a vector backbone and transformation into an appropriate strain of *Escherichia coli* for selection/screening.



Amino acid substitution by mutagenic processes

Chemical damage

Damage to nucleotide residues in DNA frequently results in their misincorporation during *in vitro* DNA synthesis or during DNA replication *in vivo*. This approach has been extensively utilized to produce a small number of mutations in a gene. An example is the use of nitrous acid (HNO_2) for inducing base deaminations. HNO_2-damaged DNA can be subjected to PCR amplification to introduce mutations. This method has been quantitated to consistently yield ~1 substitution per 500 bases [5,6]. Other chemical reactants used include: formic acid, sodium bisulfate, hydrazine and dimethylsulfate [7]. The region for mutagenesis can be limited by replacing a segment of the gene with a corresponding damaged sequence.

Mutagenic PCR

The intrinsic mutation frequency of PCR, in the absence of prior treatment with DNA damaging agents, may produce sufficient numbers of mutations, by itself, to establish a library. Standard PCR using *Taq* polymerase consistently yields ~1 substitution per 1000 bases and has been used with success to create mutant libraries [8]. Replacing magnesium with manganese as the divalent cation can lead to an increase in mutation frequencies of as much as 0.66% [9,10]. Dnase I digestion prior to sexual PCR methods, which will be discussed later, is also mutagenic. Mutation frequencies between 0.05 and 0.70% have been reported, depending on the exact conditions used [11,12]. Another approach that has been taken is to incorporate known mutagenic nucleoside analogs into the pool of normal deoxynucleoside triphosphates (dNTPs) during PCR. 8-oxo-substituted nucleosides including 8-oxodG have been used for this type of library construction, and have yielded libraries containing an array of transition mutations [13]. Additionally, biased pools of normal dNTPs can lead to the introduction of mutations [14]. In such protocols, incorrect bases are incorporated at a much slower rate than the normal base, but they can be ample for generating libraries of mutants. Deoxyinosine triphosphate (dITP) has been used to restore the efficiency of the PCR when a particular dNTP is absent [15]. Under these conditions, dITP gets incorporated in place of the missing normal dNTP. Interestingly, the recent production of error-prone mutant *Taq* DNA polymerases by random sequence mutagenesis [16] should provide a new set of mutant enzymes for increased mutagenesis in PCR reactions. Although these approaches have been successful, there are two major disadvantages. First, the frequency of mutations produced within a short gene segment is low and thus it is difficult to obtain multiple mutations within a catalytic site. Second, the introduction of mutations is not completely random. In other words, 'hot-spots' for damage/incorporation are likely to produce a non-random distribution of mutations.

DNA shuffling (sexual PCR)

DNA shuffling utilizes general homologous recombination *in vitro*. The approach involves the digestion of homologous yet diverse (mutagenized or from different species) target sequences by Dnase I and yields a pool of DNA fragments that can then be reassembled into a hybrid full-length sequence by mutagenic, primerless PCR [11]. The fragments prime each other on the basis of their homology and recombination occurs when fragments from one copy of a gene prime on another copy, causing a template switch. The method can be used in combination with other *in vitro* evolution techniques to generate hybrid molecules containing segments with different mutations. At the same time, irrelevant or deleterious mutations can be eliminated by back-crossing because the methodology can segregate mutations; mutations that do not contribute to the phenotype do not persist. DNA shuffling has been used successfully for the applied evolution of several proteins, including β-lactamase [11], green fluorescent protein (GFP) [17], β-fucosidase [18] and an arsenate detoxification operon [19].

Mutator *Escherichia coli* strains

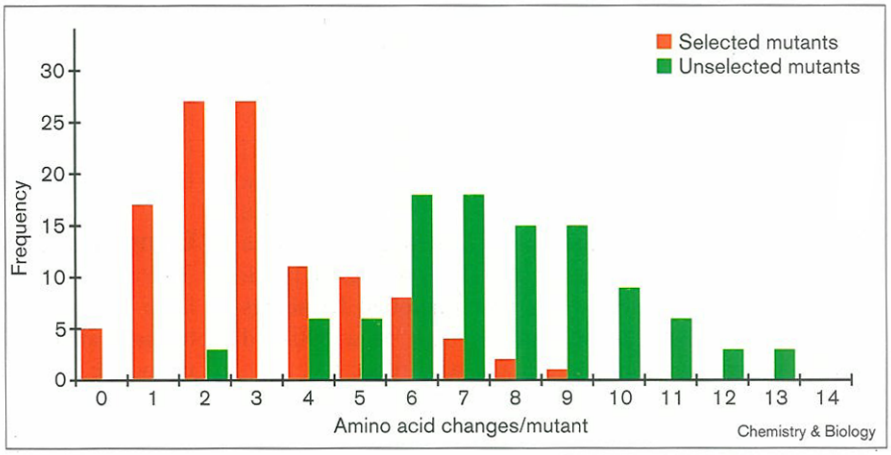
An *Escherichia coli* strain lacking mismatch repair (mutS), oxo-GTP repair (mutT) and 3′–5′ exonuclease activity of DNA pol III (mutD) has been shown to have a 5000-fold higher mutation rate than wild-type [20]. This strain, or a variety of other mutator strains, can be used to introduce mutations into plasmids and thus generate libraries of mutant genes. In principle, mutator strains can serve as a host to mediate the continuous selection for mutations with desired characteristics.

Random oligonucleotides

The use of degenerate DNA oligonucleotides provides a facile method of generating molecular diversity. The basic concept is to replace a portion of a plasmid-borne gene encoding an enzyme with a 'randomized' oligonucleotide containing all four nucleotides at specified positions. The entire population of recombinant plasmids contains random substitutions at the designated positions, whereas individual isolates contain unique sequences. The degree of randomization, and consequently the average number of amino acid substitutions per isolate, is an important parameter that can be controlled by varying the ratio of the wild-type base to the other three bases. Bias towards the wild-type nucleotide limits the number of substitutions produced within a given molecule. The methodology of generating the oligonucleotides for this technique has been reviewed extensively [21–23]. The approach was first utilized to introduce random substitutions into the promoter region of the tetracycline resistance gene [4] and then into the active site of β-lactamase [24]. Since these initial studies, random oligonucleotides have been used to study a variety of enzymes, some of which will be discussed in detail below.

Figure 3

The frequency (expressed as % total) of amino acid substitutions in 'unselected' and 'selected' AGT mutants (L.P.E., M.M. Coates and L.A.L., unpublished observations). 34 clones were sequenced from the unselected human AGT library containing random mutations amongst amino acids 150–172 and it was determined that on average there were 8 amino acid substitutions per mutant. 111 clones were sequenced from the *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine (MNNG)-selected library and it was found that there were on average 3 amino acid substitutions per mutant.



Genetic selection and visual screening

Genetic selection

Often the limiting factor in evolving a protein is not the method used to generate a library of mutants, but the approach used to identify the active variants of interest. In order to select the small subset of functional enzymes from the large pool of variants in a library, an efficient screening system is necessary. The most powerful technique for identifying functional viable mutants is positive genetic selection. An illustration of the power of this approach is the genetic complementation system for DNA polymerases [25]. The genetic selection is based on a bacterial strain that is temperature sensitive for DNA polymerase I (pol I). This strain is unable to grow at the non-permissive temperature of 37°C unless a plasmid encoding a functional exogenous polymerase is introduced. Several DNA polymerases have been shown to complement the pol I temperature sensitivity, including mammalian pol β, HIV reverse transcriptase (HIV RT) [26] and *Taq* polymerase [27].

Pol β is involved in DNA repair and its small size (39 kDa) and lack of exonucleolytic function have made it a model for studying polymerases [28]. As a result, a wealth of structural information exists that assists in establishing the functional domains of the polymerase and, in return, protein engineering studies help validate the structural studies. A pool of 10,000 pol β mutant clones generated by chemical mutagenesis (HNO_2) of 248 base pairs and a pool of 7000 mutants generated by mutagenic (Mn^{2+}) PCR of 89 base pairs were tested by the genetic complementation assay [5,25,29]. The survivors were subjected to a screen for identifying polymerases with lower replication fidelity to yield 13 low fidelity mutant polymerases. The mutants sequenced had only single amino acid changes. The same genetic selection was used to identify active mutant pol β from a library of 100,000 mutants generated by random mutagenesis of 15 nucleotides coding for the dNTP binding domain of the polymerase. Only 4% of the mutants were found to be active in complementing pol I function (A.S. and L.A.L., unpublished observations). Further analysis by direct assay of the individual mutants identified polymerases with both lower and higher fidelity than that of wild-type. Each of these mutants contained single amino acid changes.

Another example of a powerful selection scheme involves human O^6-alkylguanine-DNA alkyltransferase (AGT). Using a bacterial strain lacking alkyltransferase, large libraries of AGT mutants have been screened for those enabling the cells to survive treatment with alkylating agents. Figure 3 shows the effect that this type of genetic selection has on the extent of amino acid substitution compared to the unselected library. The figure shows that an AGT library contained prior to selection eight amino acid substitutions per randomized segment. Following genetic selection, the collection of mutants conferring resistance to alkylating agents had on average only three amino acid substitutions (L.P.E., M.M. Coates, and L.A.L, unpublished observations). This type of shift provides a valuable measure of the quality of the selection scheme. Most importantly, it also provides information about the amino acid substitutions that lead to loss of function. This is a unique property of the random oligonucleotide mutagenesis method and it assists greatly in identifying critical residues [27].

Positive selection for drug resistance was also used in the selection of β-lactamase mutants after extensive rounds of DNA shuffling [11]. The β-lactamase gene *TEM-1* provides weak resistance to the antibiotic cefotaxime. Previous attempts to generate and identify mutants that provided enhanced resistance to the antibiotic yielded a modest result of a 16-fold increase [24]. Three rounds of DNA shuffling and two rounds of back-crossing with an increasing concentration of the cefotaxime yielded mutants that were as much as 32,000-fold more resistant than the

wild-type. The highly resistant mutants identified had at least six amino acid changes. The authors note that high resistance can only be achieved by altering more than one gene region [11,30]. DNA shuffling can also be applied to gene families. By shuffling homologous genes from different species, one can take advantage of genetic diversity that has taken billions of years to evolve. Crameri *et al.* [31] found that an eightfold increase in moxalactam resistance compared to wild-type could be achieved from single libraries of shuffled microbial class C cephalosphorinase genes that are >50% identical at the DNA sequence level. The best clone obtained when four microbial homologs were shuffled, however, showed as high as a 540-fold increase in resistance compared to the wild-type.

Visual screening

Although genetic selection is most efficacious for sorting large libraries of mutants, high-throughput visual screening has also lead to the identification of mutants with significantly enhanced functions. This type of screening was used in bacteria to screen for increased whole cell fluorescence by bacteria expressing GFP [17]. GFP variants were generated by protein shuffling and approximately 10,000 bacterial colonies were screened per cycle, from which the 40 brightest colonies were picked. After three rounds of shuffling, a mutant with three amino acid substitutions that had a 45-fold increase in fluorescence was isolated. This increase was attributed to better solubility in bacterial cells rather than changes in emission and excitation maxima. Visual screening was also used in the applied evolution of a novel *para*-nitrobenzyl (pNB) esterase [8]. Such an enzymatic function is highly desirable for the industrial manufacture of antibiotics, as current methods generate significant solvent and zinc waste. The substrate of the parent enzyme was unknown but it exhibited weak pNB esterase activity in aqueous solutions. The industrial application frequently requires activity in an organic solvent environment, however. After each of four cycles of mutagenic PCR, 1000–7000 bacterial colonies were screened for esterase activity using an artificial substrate that generates colorimetric signals. One additional round of DNA shuffling produced a five amino acid change mutant that had a 30-fold increase in activity in organic solvent.

A remarkable screening scheme has recently been used to detect nonfunctional but full-length poly(ADPribose) polymerase (PARP) mutants in order to study inactivating mutations of the gene [32]. PARP is a very complex, multifunctional nuclear zinc-finger protein which is involved in the response of cells to DNA strand breaks. A library of PARP variants with mutations at the catalytic domain was generated by PCR and introduced into *E. coli*. The bacterial colonies were then replica plated onto nitrocellulose and screened for function using an activity assay. It was determined that 2% of the colonies carried inactive PARP. The same blot was immunostained in order to confirm full length product and exclude truncating mutations and then qualifying genes were sequenced. The data are very useful for structure–function analysis, but the authors point out that the same screen can be used for identifying rescue mutations, generating libraries of inactive PARP mutants, or for gain-of-function mutations. Indeed a PARP mutant with a ninefold increase in k_{cat} value has been identified using a similar methodology [33].

Generating unique enzymes for gene therapy

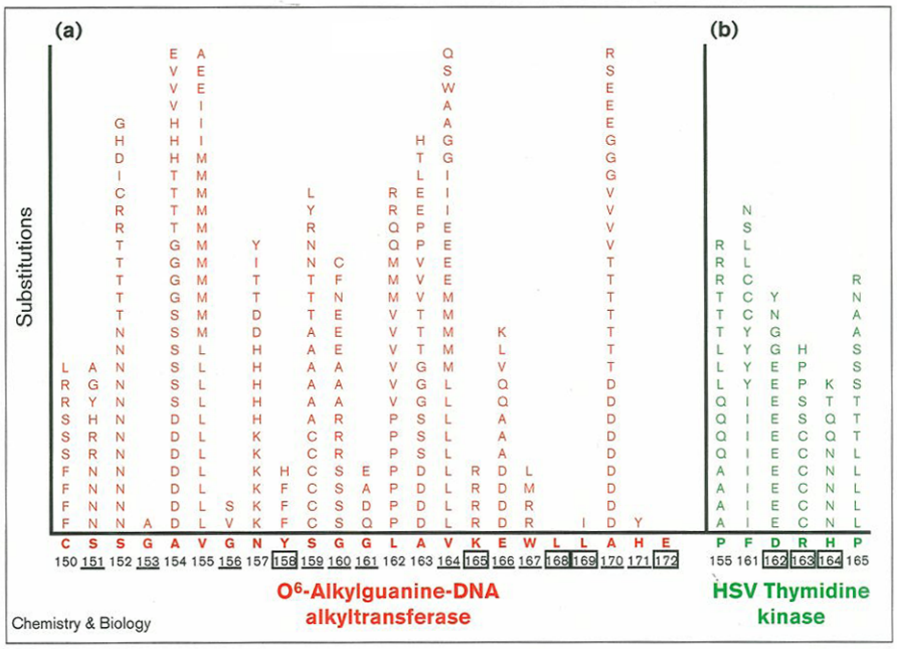
The *in vitro* evolution of new and unique enzymes could have direct implications in gene therapy for treating human cancers. A major goal of cancer gene therapy is to introduce lethal genes into tumor cells. Among the candidate genes would be those that render tumor cells uniquely sensitive to chemotherapeutic agents. The generation of mutant drug-activating enzymes with enhanced catalytic activity could accomplish this. The enzyme herpes simplex virus (HSV) thymidine kinase (TK), which is currently in clinical protocols for gene therapy, falls into this category and is discussed below. A more immediately achievable goal in cancer gene therapy, however, may be the introduction of genes expressing mutant proteins into normal tissues in order to protect them against the drugs currently being used in cancer therapy. Because myelosuppression often dictates the dosages used for many chemotherapeutic agents, enhanced protection of marrow cells could allow increased doses of many effective chemotherapeutic agents to be used. Bone-marrow stem cells expressing the protective enzyme should have a selective advantage and would be able to repopulate the marrow. AGT presents an example for which this type of approach could be used and the wild-type form has been proposed for clinical protocols. The production of mutant HSV TK and AGT with enhanced activities may offer unique advantages over wild-type forms.

HSV thymidine kinase

Human TK exhibits a high stringency in substrate selection, as it catalyzes the phosphorylation of thymidine and only a limited number of related analogs. In contrast, HSV TK is able to efficiently phosphorylate a variety of nucleotide substrates, including the guanine analogs gancyclovir and acyclovir. These analogs lack a 3′-hydroxyl terminus, terminate DNA synthesis, and prevent both viral and host cell replications [34]. One way to kill cancers selectively is to introduce HSV TK into cancer cells and then administer gancyclovir systemically. HSV TK has been introduced to tumors in animals by a variety of methods, and the tumors regress following treatment of the animals with acyclovir or gancyclovir [35–38]. The power of this method could be enhanced by identifying HSV TK mutants that phosphorylate analogs preferentially.

We have carried out remodeling of the nucleotide binding site of HSV TK by random sequence mutagenesis in

Figure 4



Amino acid substitutions in **(a)** human AGT mutants that were selected by resistance to three rounds of treatment with MNNG (L.P.E., M.M. Coates and L.A.L., unpublished observations); **(b)** TK mutants selected by the ability to complement TK deficient *E. coli* and allow growth on agar containing 5′-fluorodeoxyuridine [23]. For both (a) and (b), the one-letter code for amino acids is used. The wild-type sequence is shown below the solid line and the substitutions observed at each position and their frequency are indicated above each site. Phylogenetically conserved residues are boxed and, in the case of AGT, those that are invariant in mammals are underscored.

attempts to create such mutants. Active mutant HSV TKs were identified by genetic complementation of an *E. coli* strain lacking thymidine kinase [39]. The mutability of individual amino acids was analyzed by substituting a long contiguous nucleotide stretch with a random sequence biased towards wild-type sequence. Approximately 240 mutants were screened and 34% conferred TK activity. Certain residues appear to be highly conserved, while other positions allow a variety of changes (Figure 4b). We thereby identified amino acids that tolerate only a limited number of substitutions and completely randomized their codons. Thereafter active mutants were subjected to negative genetic selection to determine which members of the library conferred preferential killing by specific nucleoside analogs. Using this approach, a series of mutant HSV TKs that selectively phosphorylate gancyclovir and/or acyclovir were obtained [40]. Six codons were targeted with this technique and a library was obtained of more than 10^6 *E. coli* transformants; 426 of these were active as measured by genetic complementation. These mutants were analyzed for enhanced killing by gancyclovir and acyclovir. Of the active mutants, 26 conferred enhanced sensitivity to gancyclovir, 54 to acyclovir and 6 to both analogs. Ten of the mutants that were sequenced contain three to six amino acid replacements within the nucleoside binding site; notably, these combinations of substitutions would probably not have been chosen by rational remodeling of the active site. Hamster cells transfected with one of the mutants containing four substitutions, Ile160→Leu, Phe161→Leu, Ala168→Val, Leu169→Met, were more than 43-fold and 20-fold more sensitive to gancyclovir and acyclovir, respectively, than cells expressing equivalent amounts of the wild type HSV TK.

HSV TK mutants obtained by random mutagenesis have also been studied for altered thermal stability [41]. One mutant in particular has been found that is more stable at 42°C than wild-type when cell extracts were assayed for TK activity. Using purified proteins, it was found that this mutant lost only 10% of its activity over 45 min during a pre-incubation at 42°C. In contrast, wild-type HSV TK lost > 85% activity under the same conditions. This mutant contained three amino acid substitutions. Interestingly, none of the single or double mutants tested at these positions resulted in mutants with increased thermal stability.

O^6-Alkylguanine-DNA alkyltransferase

AGT is a suicide enzyme that repairs alkylation damage at the O^6 position of guanine in DNA. (The term O^6-alkylguanine-DNA alkyltransferase, shortened to alkyltransferase and abbreviated MGMT also refers to this protein.) O^6-alkylguanine is one of the major cytotoxic and mutagenic lesions produced in DNA by a variety of alkylating agents, including chemotherapeutic nitrosoureas. AGT excises alkyl groups from the O^6 position of guanine, and less efficiently from the O^4 position of thymine, to restore the nucleic acid's native structure [42,43]. Alkylating agents are used therapeutically for the treatment of many cancers and with these agents the limiting factor in dose escalation is myelosuppression. The sensitivity of the bone marrow to these treatments is, in part, due to inherently low levels of AGT in these cells. Overexpression of wild-type AGT in hematopoietic stem cells via retrovirus-mediated gene

transfer has been shown to protect both murine [44] and human [45] bone marrow cells from the clinically used alkylating agent 1,3-bis(2-chloroethyl)-1-nitrosourea (BCNU). These studies demonstrate the feasibility of expressing AGT in bone marrow for protection against myelosuppression induced by alkylating agents. Random mutagenesis was used to create libraries of AGT mutants with the goal of evolving new proteins with activities and characteristics that are enhanced over wild-type. Mutants were expressed in alkyltransferase-deficient *E. coli* and active clones were selected according to their ability to survive a stringent treatment with the toxic methylating agent *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG). Upon sequencing survivors, it was discovered that certain AGT residues were highly substitutable, while others were resistant to change (Figure 4a). One library contained a mutant that offers bacteria approximately tenfold increased protection against MNNG compared to wild-type [46]. O^6-Benzylguanine (BG) is a competitive inhibitor of AGT, and is currently in phase II clinical trials for sensitizing tumors to treatment with therapeutic alkylating agents [47]. The mutant with increased resistance to MNNG remains sensitive to BG. We have evolved new AGT proteins that are resistant to this inhibitor, however. By selecting mutants that are resistant to MNNG and BG, we have created a panel of BG-resistant alkyltranferases that are potential candidates for gene therapy to prevent myelosuppression in patients receiving the combination therapy of alkylating agents and BG [48] (L.P.E., M.M. Coates and L.A.L, unpublished observations). Interestingly, one of these mutants exhibits improved protection against MNNG. This mutant contains only a single amino acid change, yet nature has not selected the protein already. In contrast, the mutants that are the most resistant to BG all contain from three to eight amino acid substitutions. Clearly a mutant containing eight amino acid substitutions would have an extremely slim chance of being selected naturally.

Recently, the entire human AGT gene has been randomized by DNA shuffling (F.C. Christians, G. Dawes and W.P.C. Stemmer, unpublished observations). In these studies, the human gene was fragmented by Dnase I digestion (as described above) and oligonucleotides from four other mammalian alkyltransferases were doped into the reassembly reaction. Active mutants were selected from the resulting library as described above. The most active mutant selected showed nearly tenfold more resistance to MNNG than any other mutant found to date. This new mutant contains seven amino acids changes, all at positions not previously examined.

Glutathione S-transferase

Glutathione S-transferase (GST) detoxifies a broad range of electrophilic species capable of reacting with and damaging cellular macromolecules. Overexpression of this enzyme has been shown to protect cells from a variety of structurally different chemotherapeutic agents, including the alkylating agents used in treatment of brain tumors, and mechlorethamine used in the treatment of lymphomas. Nature evolved GST to metabolize structurally diverse compounds; however, random oligo mutagenesis techniques have been used to increase the specificity of the enzyme for a single agent, mechlorethamine [49]. In these studies, bacteria harboring a random library of GST active site mutants were subjected to eight rounds of treatment with increasing concentrations of mechlorethamine. Mutant enzymes were isolated from surviving bacteria and were shown to have up to a tenfold increase in resistance to the alkylating agent compared to wild-type GST. The most resistant mutants contained three amino acid substitutions. All three changes in the same mutant appeared to be required for optimal resistance, as none of the single mutants tested were found to offer resistance greater than that of wild-type GST. The mutants, which are candidates for gene therapy for protecting bone marrow, again demonstrate that multiple mutations provide optimum effect. This example also suggests that GST could be tailored to detoxify a variety of other drugs simply by using the drug of interest for the selection process. The approach could be applied to other drug metabolizing enzymes such as cytochromes P450 which have also evolved to display wide substrate specificities that could be focused and amplified by directed evolution.

Conclusions

The proteins described above represent only a few examples of applied molecular evolution. Many other proteins have been evolved *in vitro* by methods described here. It appears that within the eight amino acid sequence space of all the proteins tested to date, mutants exist that have remarkably diverse properties not present in the wild-type. A sequence space of this size contains over 10^{25} variants for an average protein. In Table 2 we list certain characteristics of some of the libraries that have been screened by our laboratory using genetic selection assays. The size of these libraries is nowhere near that of the sequence space. Success in identifying proteins with new properties may be indicative of the fact that the sequence space of proteins is densely packed with functional variants as Maynard-Smith predicted [1]. This conclusion is further supported by the results of experiments using visual screening, in which the library size is limited to 10^4. In our quest for creating proteins with more extraordinary enzymatic activities, it is our imagination for designing selection assays that may ultimately limit what can be achieved.

The results described in this review also demonstrate that the properties leading to the selection of an enzyme are not always apparent, and that very desirable properties for an enzyme may be relatively easy to obtain by applied evolution but are not adaptive and therefore not selected in an organism. Pol β mutants that had a single amino acid

Table 2

The size of libraries generated by random oligonucleotide mutagenesis and the number of active mutants recovered.

Target gene	Clones examined	Active mutants	Reference
HSV TK (sites 3 & 4)	1.1×10^6	426	[40]
HIV RT	1.2×10^4	1400	[26]
AGT	6.5×10^6	37	[48]
Taq DNA pol I	1.2×10^6	61	[50]

change and dramatically different fidelity were identified; the mutants might have served as a substrates for natural selection, but, for unknown reasons, they were not selected. It is intriguing that a higher fidelity mutant would not have been selected and fixed. On the other hand, multiple mutations were required for producing the enzymatic specificities required by the selections in the case of β-lactamase, *para*-nitrobenzyl esterase, AGT, HSV TK and GST. Such events are very rare in nature and therefore it would be hard to evolve such enzymatic activities naturally.

Increased catalytic activity is not the only goal that can be achieved by applied evolution. As was demonstrated by the HSV TK and pol β studies, it is possible to identify enzymes with altered fidelity. Such enzymes could have a role in gene therapy. Lower fidelity pol β could be used to sensitize tumors to DNA-damaging therapeutic agents as ablative gene therapy. Alternatively, a high fidelity pol β may have potential for protecting bone marrow as in the case of the previously mentioned gene therapy candidates. Protein stability has also been a target of applied molecular evolution as was the case with GFP and *para*-nitrobenzyl esterase [8,17]. Prior to protein engineering tools, we could only wonder how certain residues contribute to stability and function, and it is of particular interest that in certain cases the substitutions responsible for increased stability can be combined to show an additive effect [8].

Summary

The applied evolution of proteins, specifically enzymes, has become a basic tool for many different disciplines of science. With little or no knowledge of the three-dimensional structure of a given enzyme, the sequence space of entire genes or specific regions can be analyzed using the methods described in this review. The boundaries of this approach are defined by the imagination and the ability to devise a clever selection strategy to identify the functions of interest in a novel enzyme. So far, applied molecular evolution has produced new enzymes with enhanced properties or altered substrate specificities that make them desirable for both industrial and medical applications. The outcome of these studies can also provide structure–function information about the enzyme being studied.

Acknowledgements

We thank F.C. Christians, A. Crameri and W.P.C. Stemmer (Maxygen Inc.) for making available unpublished findings for this review. We also thank T.G. Newcomb and J.-C. Shen for useful editing comments on the manuscript. The laboratory of L.A.L. is supported in part by an OIG grant from NIH (R35-CA-39903).

References

1. Maynard-Smith, J. (1970). Natural selection and the concept of a protein space. *Nature*, **225**, 563-564.
2. Mills, D.R., Peterson R.L. & Spiegelman S. (1967). An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc. Natl Acad. Sci. USA*, **58**, 217-224.
3. Biebricher, C.K., Eigen M. & Gardiner W.C.J. (1983). Kinetics of RNA replication. *Biochemistry*, **22**, 2544-2559.
4. Horwitz, M.S. & Loeb L.A. (1986). Promoters selected from random DNA sequences. *Proc. Natl Acad. Sci. USA*, **83**, 7405-7409.
5. Sweasy, J.B. & Loeb L.A. (1993). Detection and characterization of mammalian DNA polymerase β mutants by functional complementation in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **90**, 4626-4630.
6. Diaz, J.-J., Rhoads D.D. & Roufa D.J. (1991). PCR-mediated chemical mutagenesis of cloned duplex DNAs. *BioTechniques*, **11**, 204-211.
7. Roufa, D.J. (1996). PCR-mediated chemical mutagenesis. *Meth. Mol. Biol.*, **57**, 357-367.
8. Moore, J.C. & Arnold F.H. (1996). Directed evolution of a *para*- nitrobenzyl esterase for aqueous-organic solvents. *Nat. Biotech.*, **14**, 458-467.
9. Beckman, R.A., Mildvan A.S. & Loeb L.A. (1985). On the fidelity of DNA replication: manganese mutagenesis *in vitro*. *Biochemistry*, **24**, 5810-5817.
10. Cadwell, R.C. & Joyce G.F. (1992). Randomization of genes by PCR mutagenesis. *Cold Spring Harbor Laboratory*, **2**, 28-33.
11. Stemmer, W.P.C. (1994). DNA shuffling by random fragmentation and reasembly: *In vitro* recombination for molecular evolution. *Proc. Natl Acad. Sci. USA*, **91**, 10747-10751.
12. Zhao, H. & Arnold F.H. (1997). Functional and nonfunctional mutations distinguished by random recombination of homologous genes. *Proc. Natl Acad. Sci. USA*, **94**, 7997-8000.
13. Zaccolo, M., Williams D.M., Brown D.M. & Gherardi E. (1996). An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *J. Mol. Biol.*, **255**, 589- 603.
14. Pezo, V. & WainHobson S. (1997). Hypermutagenic *in vitro* transcription employing biased NTP pools and manganese cations. *Gene*, **186**, 67-72.
15. Kuipers, O.P. (1996). Random mutagenesis by using mixtures of dNTP and dITP in PCR. *Meth. Mol. Biol.*, **57**, 351-356.
16. Suzuki, M., Avicola A.K., Hood L. & Loeb L.A. (1997). Low fidelity mutants in the O-helix of *Thermus aquaticus* DNA polymerase I. *J. Biol. Chem.*, **272**, 11228-11235.
17. Crameri, A., Whitehorn E.A., Tate E. & Stemmer W.P.C. (1996). Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotech.* **14**, 315-319.
18. Zhang, J.-H., Dawes G. & Stemmer W.P. (1997). Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening. *Proc. Natl Acad. Sci. USA*, **94**, 4504-4509.
19. Crameri, A., Dawes G., Rodriguez E., Silver S. & Stemmer W.P.C. (1997). Molecular evolution of an arsenate detoxification pathway by DNA shuffling. *Nat. Biotech.* **15**, 436-438.
20. Greener, A., Callahan M. & Jerpseth B. (1996). An efficient random mutagenesis technique using an *E. coli* mutator strain. *Meth. Mol. Biol.*, **57**, 375-385.
21. Black, M.E., Munir K.M. & Loeb L.A. (1993). Random sequence selection for generation of active recombinant molecules. *Meth. Molec. Genet.* **2**, 26-43.
22. Black, M.E. & Loeb L.A. (1996). Random sequence mutagenesis for generation of active enzymes. *Meth. Mol. Biol.* **57**, 335-349.
23. Loeb, L.A. (1996). Unnatural nucleotide sequences in biopharmaceutics. *Adv. Pharmacol.*, **35**, 321-347.
24. Dube, D.K. & Loeb L.A. (1989). Mutants generated by the insertion of random oligonucleotides into the active site of the β-lactamase gene. *Biochemistry* **28**, 5703-5707.
25. Sweasy, J.B. & Loeb L.A. (1992). Mammalian DNA polymerase β can substitute for DNA polymerase I during DNA replication in *Escherichia coli*. *J. Biol. Chem.* **267**, 1407-1410.
26. Kim, B. & Loeb L.A. (1995). Human immunodeficiency virus reverse transcriptase substitutes for DNA polymerase I in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **92**, 684-688.

27. Suzuki, M., Christians F.C., Kim B., Skandalis A., Black M.E. & Loeb L.A. (1996). Tolerance of different proteins for amino acid diversity. *Mol. Div.*, **212**, 111-118.
28. Singhal, R.K., Prasad R. & Wilson S.H. (1995). DNA polymerase β conducts the gap-filling step in uracil-initiated base excision repair in bovine testis nuclear extract. *J. Biol. Chem.* **270**, 949-957.
29. Washington, S.L., Yoon M.S., Chagovetz A.M., Li S.-X., Clairmont C.A., Preston B.D., Eckert K.A. & Sweasy J.B. (1997). A genetic system to identify DNA polymerase β mutator mutants. *Proc. Natl Acad. Sci. USA* **94**, 1321-1326.
30. Stemmer, W.P.C. (1994). Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* **370**, 389-391.
31. Crameri, A., Raillard S.-A., Bermudez E. & Stemmer W.P.C. (1997). DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature*, in press.
32. Rolli, V., O'Farrell M.O., de-Murcia J.M. & de-Murcia G. (1997). Random mutagenesis of the poly(ADP-ribose) polymerase catalytic domain reveals amino acids involved in polymer branching. *Biochemistry*, **36** 12147-12154.
33. Miranda, E.A., Dantzer F., O'Farrell M., de-Murcia G. & de-Murcia J.M. (1995). Characterisation of a gain-of-function mutant of poly(ADP-ribose) polymerase. *Biochem. Biophys. Res. Commun.* **212**, 317-325.
34. Elion, G.B. (1980). The chemotherapeutic exploitation of virus-specified enzymes. *Advan. Enzy. Regul.* **18**, 53-66.
35. Culver, K.W., Ram Z., Walbridge S., Ishii H., Oldfield E.H. & Blaese R.M. (1992). *In vivo* gene transfer with retroviral vector-producer cells for treatment of experimental brain tumors. *Science* **256**, 1550-1552.
36. Ram, Z., Culver K.W., Walbridge S., Frank J.A., Blaese R.M. & Oldfield E.H. (1993). Toxicity studies of retroviral-mediated gene transfer for the treatment of brain tumors. *J. Neurosurg.* **79**, 400-407.
37. Caruso, M., Panis Y., Gagandeep S., Houssin D., Salzmann J.-L. & Klatzmann D. (1993). Regression of established macroscopic liver metastases after *in situ* transduction of a suicide gene. *Proc. Natl Acad. Sci. USA* **90**, 7024-7028.
38. Vile, R.G. & Hart I.R. (1993). Use of tissue-specific expression of the herpes simplex virus thymidine kinase gene to inhibit growth of established murine melanomas following direct intratumoral injection of DNA. *Cancer Res.* **53**, 3860-3864.
39. Black, M.E. & Loeb, L.A. (1993). Identification of important residues within the putative nucleoside binding site of HSV-1 thymidine kinase by random sequence selection: analysis of selected mutant *in vitro*. *Biochemistry* **32**, 11618-11626.
40. Black, M.E., Newcomb T.G., Wilson H.-M.P. & Loeb L.A. (1996). Creation of drug-specific herpes simplex virus type 1 thymidine kinase mutants for gene therapy. *Proc. Natl Acad. Sci. USA* **93**, 3525-3529.
41. Munir, K.M., French D.C. & Loeb L.A. (1993). Thymidine kinase mutants obtained by random sequence selection. *Proc. Natl Acad. Sci. USA* **90**, 4012-4016.
42. Sassanfar, M., Dosanjh M.K., Essigmann J.M. & Samson L. (1991). Relative efficiencies of the bacterial, yeast and human DNA methyltransferases for the repair of O^6-methylguanine and O^4-methylthymine. *J. Biol. Chem.* **266**, 2767-2771.
43. Pegg, A.E., Dolan M.E. & Moschel R.C. (1995). Structure, function and inhibition of O^6-alkylguanine-DNA alkyltransferase. *Prog. Nucleic Acid Res.* **51**, 167-223.
44. Moritz, T., Mackay W., Glassner B.J., Williams D.A. & Samson L. (1995). Retrovirus-mediated expression of a DNA repair protein in bone marrow protects hematopoietic cells from nitrosourea-induced toxicity *in vitro* and *in vivo*. *Cancer Res.* **55**, 2608-2614.
45. Marathi, U.K., Harris L.C., Venable C.C. & Brent T.P. (1997). Retroviral transfer of a bacterial alkyltransferase gene (*ada*) into human bone marrow cells protects against O^6-benzylguanine plus 1,3-bis(2-chloro-ethyl)-l-nitrosourea cytotoxicity. *Clin. Cancer Res.* **3**, 301-307.
46. Christians, F.C. & Loeb L.A. (1996). Novel human DNA alkyltransferases obtained by random substitution and genetic selection in bacteria. *Proc. Natl Acad. Sci. USA* **93**, 6124-6128.
47. Dolan, M.E. & Pegg A.E. (1997). O^6-Benzylguanine and its role in chemotherapy. *Clin. Cancer Res.* **3**, 837-847.
48. Christians, F.C., Dawson B.J., Coates M.M. & Loeb L.A. (1997). Creation of human alkyltransferases resistant to O^6-benzylguanine. *Cancer Res.* **57**, 2007-2012.
49. Gulick, A.M. & Fahl W.E. (1995). Forced evolution of glutathione S-transferase to create a more efficient drug detoxification enzyme. *Proc. Natl Acad. Sci. USA* **92**, 8140-8144.
50. Suzuki, M., Baskin D., Hood L.E. & Loeb L.A. (1996). Random mutagenesis of *Thermus aquaticus* DNA polymerase I: concordance of immutable sites *in vivo* with the crystal structure. *Proc. Natl Acad. Sci. USA* **93**, 9670-9675.