Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 13 (2005) 2701–2716
Bioorganic & Medicinal Chemistry

Review

Altering protein specificity: techniques and applications

Nina M. Antikainen and Stephen F. Martin*

Department of Chemistry and Biochemistry, The Institute of Cellular and Molecular Biology, The University of Texas at Austin, Austin, TX 78712, USA

Received 20 December 2004; accepted 26 January 2005

Abstract—Protein engineering constitutes a powerful tool for generating novel proteins that serve as catalysts to induce selective chemical and biological transformations that would not otherwise be possible. Protocols that are commonly employed for altering the substrate specificity and selectivity profiles by mutating known enzymes include rational and random methods as well as techniques that entail evolution, selection and screening. Proteins identified by these techniques play important roles in a variety of industrial and medicinal applications and in the study of protein structure–function relationships. Herein we present a critical overview of methods for creating new functional proteins having altered specificity profiles and some practical case studies in which these techniques have been applied to solving problems in synthetic and medicinal chemistry and to elucidating enzyme function and biological pathways.

Contents

1. Introduction

The discovery of new protocols whereby catalytically active proteins with novel specificity, function, catalytic activity, and/or stability are identified is an important area of contemporary bioorganic chemistry.[1,2] In addition to the obvious practical applications in medicine and industry, structural, and thermodynamic studies of engineered enzymes and their complexes with small molecules such as substrate analogues and inhibitors can lead to a better understanding of the structural basis for selectivity and function.

Keywords: Protein engineering; Enzyme specificity; Mutagenesis.
* Corresponding author. Tel.: +1 512 471 3915; fax: +1 512 471 4180; e-mail: sfmartin@mail.utexas.edu

0968-0896/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2005.01.059

A few examples of the computerized design of a protein fold have been reported.[3,4] However, our current level of understanding of protein folding and the relationships of protein structure with function is generally inadequate to enable the rational de novo design of proteins with specific biological and physical properties. Hence, contemporary protein engineering typically involves deliberate efforts to reprogram existing enzymes by mutagenesis to generate new phenotypes with the desired properties.

A number of different approaches have been developed for modifying proteins by mutagenesis of the parent gene, but these fall into the two main categories of rational and random methods as summarized in Figure 1. The category of random methods includes a variety of procedures for optimizing performance that involve evolutionary techniques. Irrespective of the precise details of the methods that are employed for randomly altering the amino acid sequence of the protein of interest, one of the major challenges that must be addressed in applying such approaches is developing efficient and rapid protocols for screening and/or selection of variants with the desired phenotype.

In this review, which is intended to be illustrative rather than comprehensive, some general approaches that have been utilized to alter the substrate specificity of known enzymes will first be presented. The practical utility of selected methods will then be highlighted by presenting examples of how enzyme variants with altered selectivities may be used in various applications, including those of synthetic, mechanistic, industrial, and medical importance.

2. Rational methods for altering protein substrate selectivity

Rational approaches directed toward altering enzyme selectivity involve replacing amino acids at specific sites of the protein with other amino acids that are selected based upon a combination of mechanistic and structural knowledge.[5] Rational methods depend upon a detailed understanding of the determinants of substrate selectivity and of the catalytic mechanism of the enzyme. Although this requirement limits the scope of the rational approach to well defined systems, an advantage of rational point- or site-directed mutagenesis relative to random and evolution-based approaches is that only a small number of variants must be examined. Hence, the specificities of the mutants can be quickly determined by comparing their respective specificity constants, which are defined by the ratio of the turnover

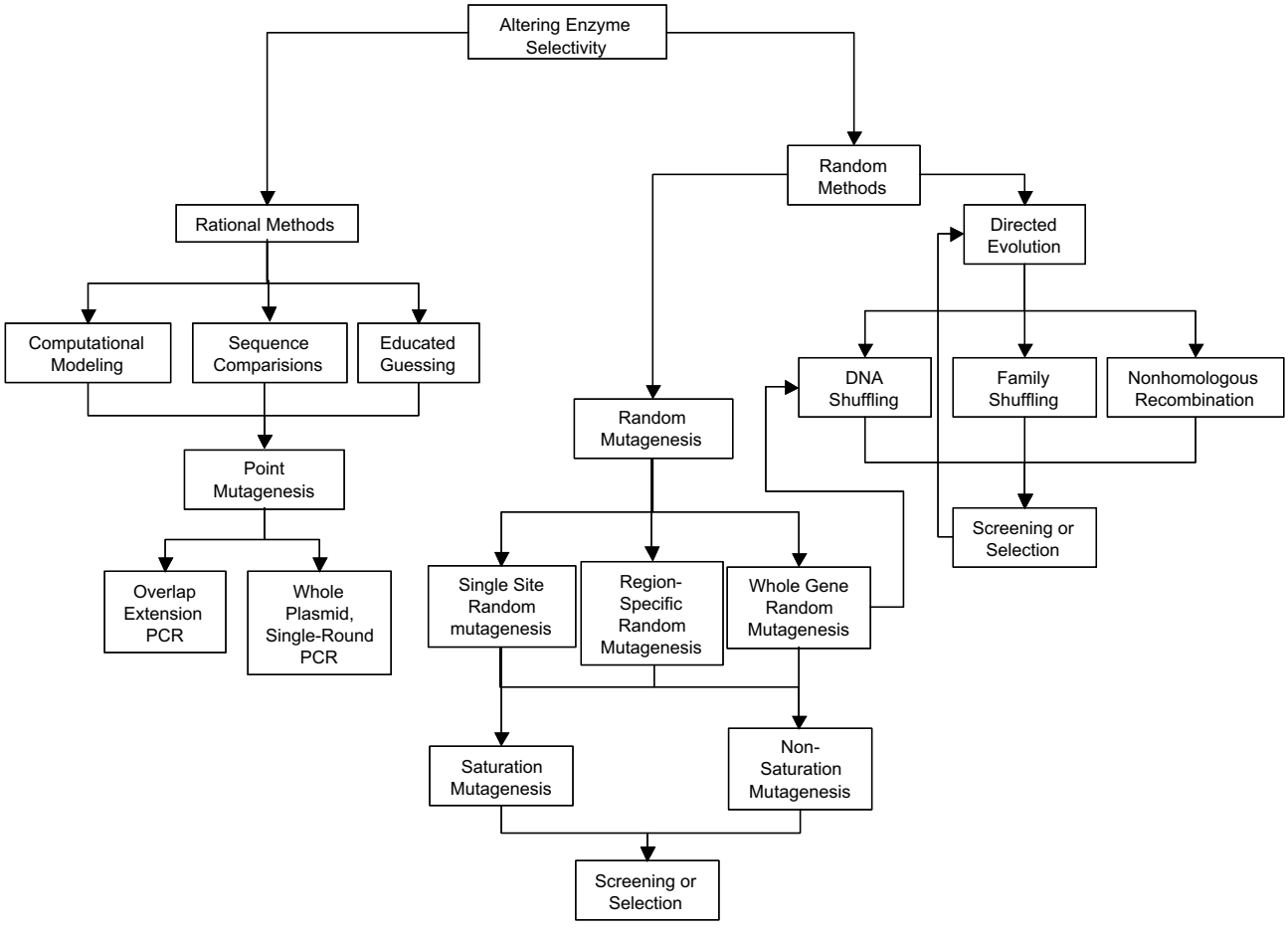

Figure 1. Methods for altering protein selectivity.

constant k_{cat} and the Michaelis constant K_M. Rational mutagenesis is also useful for testing hypotheses about the structural and functional roles of specific amino acid residues in a protein.

2.1. Representative techniques for site-directed mutagenesis

Two techniques are commonly used to introduce specific amino acid replacements into a target gene. The first of these is termed the overlap extension method and is summarized in Figure 2.[6,7] In this method, four primers, which are short sequences of synthetic DNA complementary to a section of the gene of interest, are used in the first polymerase chain reaction (PCR) step wherein two separate PCRs are performed. The primer pairs for these PCRs are 1/3 and 2/4, respectively, with primers 2 and 3 containing the mutant codon with a mismatched sequence. Two double-stranded DNA products are obtained. When these double stranded duplexes are denatured and then annealed, two heteroduplexes are produced (one shown) wherein each strand of the heteroduplex contains the desired mutagenic codon ×. The overlapping 3′ and 5′ ends of each heteroduplex are then filled in using DNA polymerase, and the second PCR step using primers 1 and 4 amplifies the mutagenic DNA.

A useful variant of the overlap extension method is the megaprimer method.[8] In this procedure, two rounds of PCR are performed employing two flanking primers and one internal mutagenic primer that contain the desired base substitutions. A benefit of this method is that mutations can be inserted into the flanking primers so multiple codons relatively far from each other can be replaced in one sequence of two PCRs.

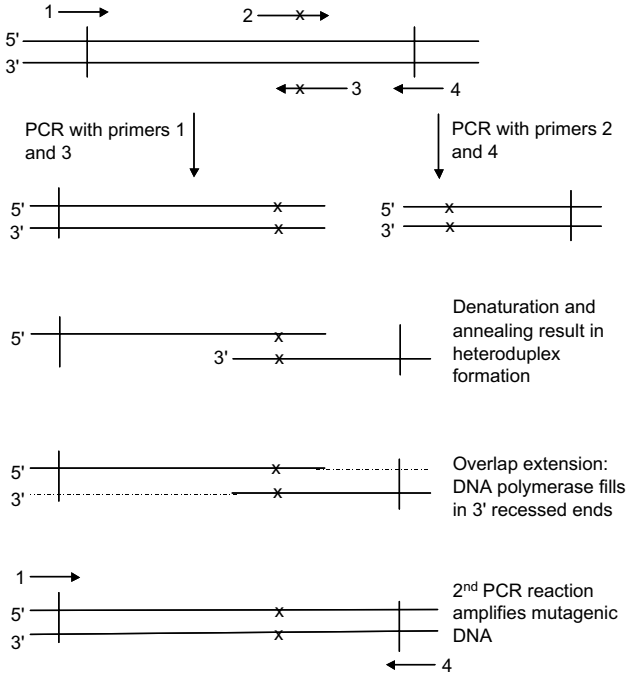

Figure 2. Overlap extension PCR method: → represents a primer, and × represents a mutagenic codon.

The second method for performing site-directed mutagenesis is referred to as whole plasmid, single-round PCR as illustrated in Figure 3.[9,10] In this protocol, two oligonucleotide primers containing the desired mutation(s), ×, and complementary to the opposite strands of a double-stranded DNA plasmid template are extended using DNA polymerase. In this PCR step, both strands of the template are replicated without displacing the primers to afford the mutated plasmid containing breaks that do not overlap. Because the original wild type plasmid originates from *E. coli* and is thus methylated on various A and C residues, it may then be selectively digested using *Dpn*I methylase endonuclease to cleanly give a circular, nicked vector containing the mutant gene. When this nicked vector is transformed into competent cells, the nick in the DNA is repaired by the cell machinery to give a mutated, circular plasmid. The popular 'QuikChange Site-Directed Mutagenesis Kit' from Stratagene (#200518) is useful for implementing this approach.

The advantages of whole plasmid, single-round PCR overlap extension PCR are that only one PCR needs to be performed and two, instead of four, primers are required. The disadvantages of this technique relative to overlap extension are that it does not work well with large plasmids (>10 kB) and typically only two nucleotides can be replaced at a time.

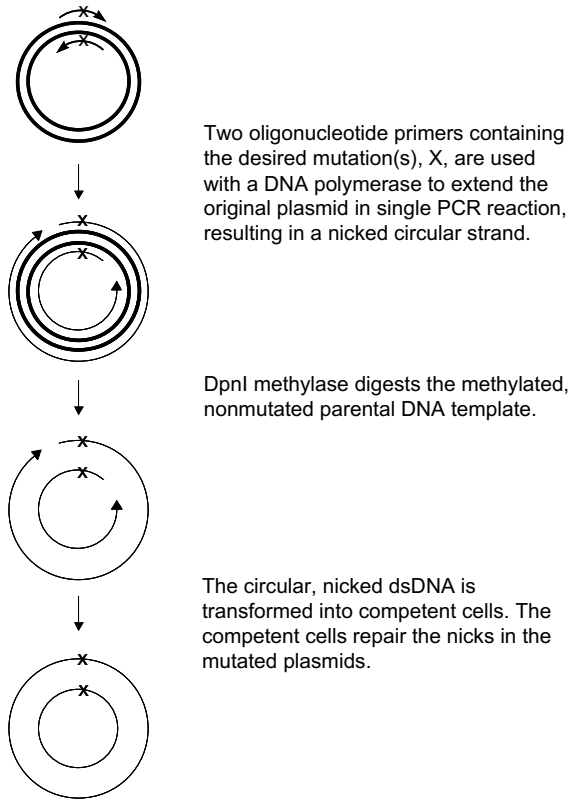

Figure 3. Whole plasmid, single-round PCR produces the desired mutation directly from a plasmid template: → represents a primer, × represents a mutagenic codon, a thick line represents the template plasmid DNA, and a thin line represents newly synthesized DNA.

2.2. Selected applications of site-directed mutagenesis

High resolution structural data of a protein and sometimes its complexes with biologically relevant small molecules have historically been used to provide a starting point for deciding what amino acid residues to vary. However, computational methods are being used with increasing frequency to provide insights that facilitate altering the properties of enzymes. There are numerous examples of the use of site-directed mutagenesis to create proteins that exhibit novel properties including specificity, function, and stability, but a few case studies will serve to illustrate the features of the method.

Owing to their exquisite chemo-, regio-, and stereospecificity, the use of enzymes as catalysts for effecting selective transformations in organic synthesis has become an important area in biocatalysis.[11,12] Early work in the field was focused upon natural enzymes that were readily available, but more recently there has been an increasing interest in redesigning existing enzymes to generate variants that exhibit not only improved efficiency and stability but also activity toward nonnatural substrates.[13,14] Indeed, tailoring the properties of enzymes for specific applications is critical to their development and use as mainstream catalysts in organic synthesis.

The value of engineering enzyme specificity to make biocatalysts for organic synthesis may be illustrated by considering the aldolases, which are enzymes that catalyze the highly stereoselective, reversible formation of carbon–carbon bonds via aldol reactions. For example, 2-deoxyribose-5-phosphate aldolase (DERA), which reversibly catalyzes the condensation of two aldehydes as shown in Figure 4a, has been used in the synthesis of the antitumor agent epothilone A.[15] However, DERA has a very strong preference for phosphorylated substrates, a requirement that severely limits its applicability in organic synthesis.[16]

In order to broaden the substrate range of DERA by mutagenesis, Wong and co-workers used the structure of *E. coli* DERA complexed with D-2-deoxyribose-5-phosphate (1.05 Å) to design five variants that were predicted to have improved activity for the unnatural nonphosphorylated substrate D-2-deoxyribose in the retroaldol reaction.[17] Thus, two basic active site residues were replaced with acidic residues to generate the K172E and R207E variants, and three neutral side chains in the phosphate binding pocket were converted to acidic ones, thereby creating the G205E, S238D, and S239E variants. The authors predicted that these replacements would change the substrate specificity of wild type DERA from preferring the negatively charged 2-deoxyribose-5-phosphate to the nonphosphorylated, neutral D-2-deoxyribose in the retroaldol reaction due to the electrostatic repulsion that would result between the new negatively charged amino acid side chains and the negatively charged phosphate moiety of the phosphorylated substrate.

Although activity toward the natural substrate D-2-deoxyribose-5-phosphate decreased for each of the five variants, S238D DERA was of particular interest as it was capable of catalyzing the retroaldol reaction of the unnatural substrate D-2-deoxyribose to yield D-glyceraldehyde and acetaldehyde (Fig. 4b) with a 2.5-fold improvement in activity (wild-type, $2\ s^{-1}\ M^{-1}$; S238D, $5\ s^{-1}\ M^{-1}$) compared to wild type DERA with the unnatural substrate. Although the increase in activity of S238D DERA toward D-2-deoxyribose relative to wild type was modest, it was approximately 7000-fold more selective than wild type toward this unnatural substrate compared to the natural substrate. The S238D variant also exhibited significant activity toward other unnatural substrates and was utilized to catalyze the aldol reaction of 3-azidopropinaldehyde and acetaldehyde, a reaction not catalyzed by wild type DERA, to afford a key intermediate in the synthesis of atorvastatin.[18] Given the broad importance of the aldol reaction, S238D DERA or other variants are likely to have considerable future utility as catalysts for organic synthesis.

Engineered enzymes may be employed to degrade toxic and environmentally hazardous chemicals such as organophosphates, polychlorinated biphenyls, and halogenated organic compounds.[19–21] For example, wild-type organophosphorus hydrolase (OPH) is a bacterial enzyme that hydrolyzes a variety of organophosphorus neurotoxins, including chemical warfare agents and popular pesticides.[22,23] Wild and co-workers used rational, site-directed mutagenesis to alter substrate

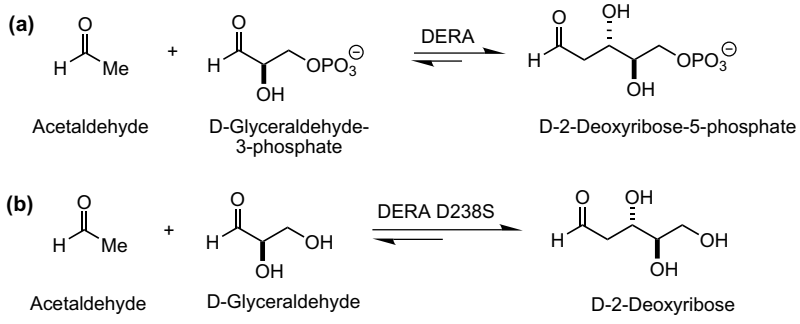

Figure 4. The reactions catalyzed by DERA on the natural substrate D-glyceraldehyde-3-phosphate (a) and by DERA D238S variant on the unnatural substrate glyceraldehyde (b).

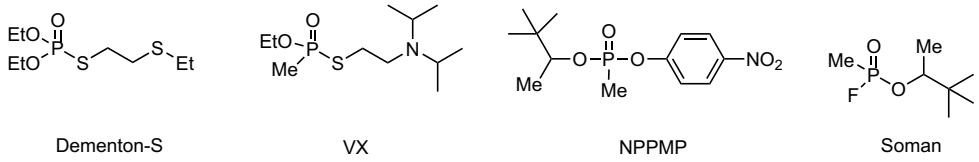

Figure 5. Structures of dementon S, VX, NPPMP, and soman.

selectivity of OPH to generate variants with higher activities and specificities toward such toxic compounds as dementon S, an analogue of *O*-ethyl *S*-(2-diisopropyl aminoethyl) methylphosphonothioate (VX) and *p*-nitrophenyl-*o*-pinacolyl methylphosphonate (NPPMP), an analogue of soman (Fig. 5).[24]

Because X-ray and NMR structures of OPH were not available, mechanistic considerations played a key role in developing a rationale for mutagenesis. For example, histidine residues bind divalent metal cations, which are known to play important roles in catalysis by a number of phosphoryl transfer enzymes. Histidine residues were thus presumed to be in the OPH active site and were chosen for replacement. The H254R and H257L variants of OPH exhibited corresponding four- and fivefold increases in specificity constants toward dementon S relative to wild type. Relative to wild type, H257L exhibited a fivefold improvement in specificity constant toward NPPMP. Combining these two replacements led to the double mutant H254R/H257L, which had a specificity constant that was 20-fold higher than wild type toward dementon S. It is noteworthy that each of these three mutants retained only one of the two native Co^{2+} ions in their active sites, although the precise importance of this observation is unclear. Whether these variants of OPH can be effectively used to hydrolyze unwanted organophosphorus compounds, including outdated pesticides and chemical warfare agents remains to be established.

Modifying the substrate specificity of enzymes can provide variants that can be used to study fundamental biological processes. For example, kinases play important roles in various signaling pathways, but identifying the natural substrates of a specific kinase is often difficult because several kinases may phosphorylate the same substrate. In order to identify the substrates of the tyrosine kinase of Rous sarcoma virus, Shokat and co-workers employed a rational strategy to alter the binding pocket by preparing the V323A/I338A variant.[25] Because there was no structural information for the tyrosine kinase of the Rous sarcoma virus, this variant was inspired by sequence alignments with the functionally homologous kinases *c*AMP-dependent kinase and cyclin-dependent kinase 2. Comparing the structures of these kinases complexed with ATP revealed that the amino acids corresponding to V323 and I338 of the Rous sarcoma virus tyrosine kinase were within 5 Å of *N*-6 of a bound ATP molecule. Replacing residues V323 and I338 of the Rous sarcoma virus tyrosine kinase with the smaller amino acid alanine was then envisioned to create space for an *N*-6 substituted ATP derivative such as *N*-6-cyclopentyl-ATP (Fig. 6).

The V323A/I338A variant of Rous tyrosine kinase was found to catalyze phosphoryl transfer from *N*-6-cyclopentyl-ATP (specificity constant, 3.3×10^3 min^{-1} M^{-1}), whereas the wild type kinase and other kinases did not. Moreover, this novel enzyme exhibited a 30-fold lower specificity constant toward ATP compared to wild type (V323A/I338A, 5.3×10^3; wild type, 1.6×10^5 min^{-1} M^{-1}). When radiolabeled [γ-^{32}P]*N*-6-cyclopentyl ATP was used in conjunction with the V323A/I338 variant of Rous tyrosine, it was possible to identify the natural substrates of the wild type kinase because they were labeled with ^{32}P. This approach appears rather general and may be extended to identify the substrates of other tyrosine kinases.[26–28]

Computational methods employing the three-dimensional coordinates of protein structures may be used to engineer proteins with designed substrate selectivity.[29–31] Computer modeling is particularly valuable for estimating short-range interactions involving steric contacts and hydrogen bonds between substrates and proteins.

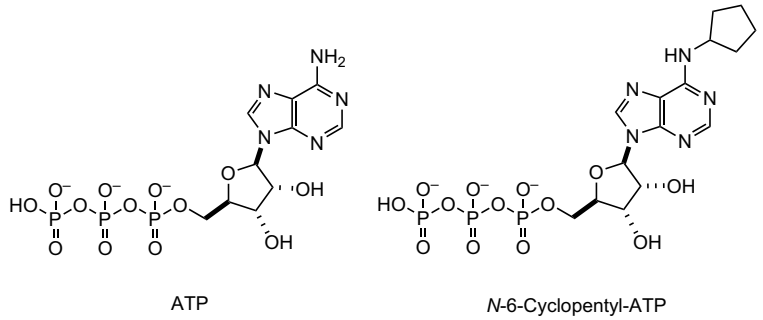

Figure 6. The structures of ATP and *N*-6-cyclopentyl–ATP analogue.

Hence, this technique can provide some clues regarding the molecular basis of substrate selectivity and how this specificity might be changed by mutagenesis. One major obstacle to using computational methods in the rational design of novel proteins is the difficulty of predicting long-range electrostatic interactions, but improved algorithms are being developed for this purpose.[32] Another challenge is predicting what effects individual mutations might have upon protein flexibility, which is now recognized as having a significant impact on catalysis.[33]

Hellinga and co-workers have developed the powerful computational tool DENZYMER to assist in reprogramming the specificities and properties of proteins.[34–36] A geometry dependent hydrogen-bonding term, a solvation term that represents the hydrophobic effect, and a term stipulating that potential hydrogen bond donors and acceptors in the ligand must be satisfied are critical for the success of the algorithm, although other factors including steric interactions and electrostatics are also important. This computational technique has been applied to the design of novel variants of *E. coli* periplasmic binding proteins to bind the nonnatural ligands trinitrotoluene, L-lactate, and serotonin with high affinities (Fig. 7).[37] The three-dimensional structures of the glucose, ribose, arabinose, glutamine, and histidine binding proteins were first modeled using target-ligand docking protocols in order to make predictions regarding what amino acid sequences formed a complementary surface for each of the ligands. The program then generated data for theoretical variants in which amino acids within contact distance of the ligand in the wild type binding pocket were replaced, and energy minimizations were performed for each variant.

Seventeen proteins, each of which contained 5–17 amino acid replacements, were thus identified that were predicted to have high affinities for trinitrotoluene, L-lactate, or serotonin. Each variant was characterized experimentally, and all exhibited significant affinities for their target ligands with some binding constants being in the nanomolar range. Several proteins exhibited excellent selectivity for their target ligand relative to

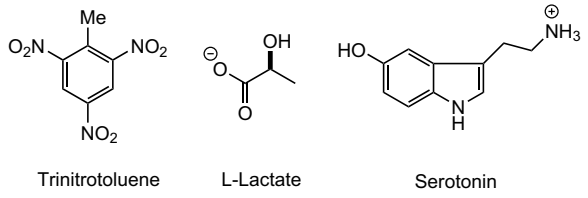

Figure 7. Structures of trinitrotoluene, L-lactate, and serotonin.

structurally related compounds. For example, an L-lactate selective receptor bound L-lactate more than 1000-fold better than the structurally related compounds D-lactate and pyruvate. Because the three target ligands (trinitrotoluene, L-lactate, and serotonin) are structurally very different, studying their interactions with their respective binding proteins may lead to a better understanding of the critical parameters (e.g., molecular shape, chirality, functional groups, molecular surface complementarity, internal flexibility, and water solubility) that are associated with molecular recognition. The success of the computational design process used in these studies strongly suggests that such techniques will play an increasingly important role in protein engineering.

When high resolution structural information for an enzyme of interest is not available, one strategy that may be used to select what amino acids might be replaced involves a comparative analysis of proteins with similar amino acid sequences but different substrate specificities. Such assessments can reveal key functional determinants of substrate selectivity as exemplified by reprogramming the substrate specificity of a choline acetyltransferase to give a dual specific enzyme that catalyzes acetyl transfer to both choline and carnitine. Choline acetyltransferase catalyzes the reversible transfer of an acetyl group between acetyl–CoA and choline (Fig. 8a), whereas carnitine acetyltransferase catalyzes a similar reaction between acetyl–CoA and carnitine (Fig. 8b).

The amino acid sequences of five choline acyltransferases and 13 carnitine acyltransferases from different

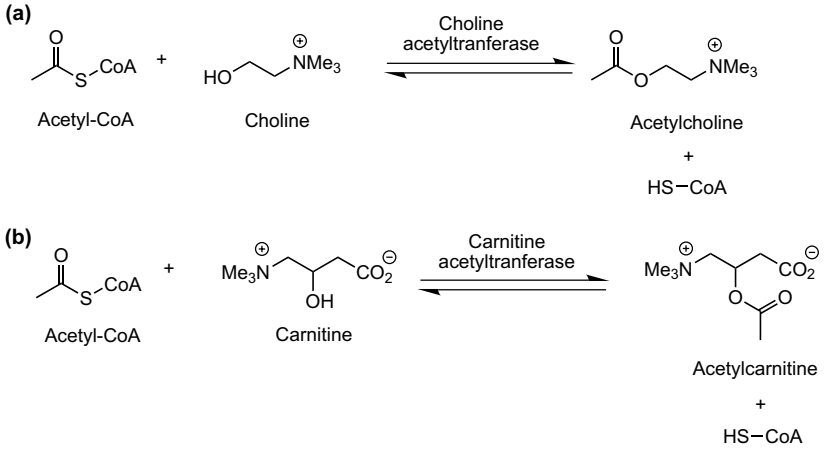

Figure 8. (a) Choline acetyltransferase catalyzes the reversible transfer of an acetyl group between acetyl–CoA and choline. (b) Carnitine acetyltransferase catalyzes the reversible transfer of an acetyl group between acetyl–CoA and carnitine.

organisms had been reported, but no structural information for any of these enzymes was available. The sequences of several choline acyltransferases and carnitine acyltransferases were thus compared, and four conserved amino acid residues in choline acyltransferases were identified as potentially essential for providing selectivity toward choline over carnitine. Based upon this analysis, Cronin made several replacements in the rat choline acyltransferase gene and determined the specificity constants of the respective protein products.[38] The quadruple mutant N514R/V459T/D460E/N461T was found to be a dual specific enzyme that was active toward both carnitine and choline, but the specificity constant for carnitine was more than three orders of magnitude greater than the wild type enzyme. In particular, the ratio of k_{cat}/K_M values toward carnitine versus choline of the N514R/V459T/D460E/N461T variant was 4.0×10^{-1}, whereas the corresponding ratio of k_{cat}/K_M for wild type was 1×10^{-6}: The catalytic discrimination between choline and carnitine was thus shifted by about 400,000-fold in favor of carnitine by changing only four amino acid residues. Unfortunately, because structural data are lacking, the altered selectivity profile of this variant of rat choline acyltransferase cannot be correlated with changes in structure.

Plant linoleate 9- or 13-lipoxygenases catalyze the regio- and stereoselective oxidation of linoleic acid to give 9- or 13-hydroperoxy derivatives, respectively. Feussner and co-workers have reported an interesting study in which site directed mutagenesis was used to explore the origin of the positional specificity of cucumber linoleate 13-lipoxygenase.[39] Modeling enzyme/substrate interactions suggested that linoleic acid bound to the active site cleft of this 13-lipoxygenase so the methyl, not the carboxyl, terminus was positioned in the hydrophobic portion of the binding site. When linoleic acid bound in this orientation, the methyl terminus was located near R758 of the lipoxygenase, and the positive charge on the arginine side chain appeared to be shielded from the substrate by the imidazole ring of H608. They reasoned that replacing H608 with an amino acid having a smaller side chain might allow formation of a salt bridge between the side chain of R758 and the carboxyl group of linoleic acid. Such an interaction would then favor binding of the substrate in the opposite orientation, thereby perhaps altering the specificity to create a 9-lipoxygenase. This hypothesis was supported by the observation that the positions of a number of 13-lipoxygenases that align with H608 in cucumber 13-lipoxygenase are occupied by histidine or phenylalanine, whereas in 9-lipoxygenases the corresponding residues are valine (Fig. 9).

In order to test this theory, the H608V variant of cucumber 13-lipoxygenase was prepared by site directed mutagenesis and incubated with linoleic acid to give a mixture of C-13/C-9 oxidation products in a ratio of 5:95. On the other hand, when the wild type was incubated with linoleic acid the ratio of products arising from C-13 and C-9 oxidation was 84:16. Thus, replacing H608 with valine resulted in a dramatic change of positional selectivity. It is tempting to conclude that this result supports the idea that altering substrate orientation in the active site of 13-lipoxygenase leads to modification of positional specificity. However, the authors noted that further experiments are needed because fatty acids are flexible, and the binding modes are not known. It is also possible that selectivity depends upon the orientations of the side chains of the amino acids lining the substrate binding cleft.

3. Random methods for altering protein substrate selectivity

Use of methods for the rational mutagenesis of proteins is often limited by the difficulty of predicting what specific amino acid replacements will lead to the preferred altered selectivities, even when high resolution structural data are available. Simple substitutions can lead to changes in protein tertiary structure and flexibility that affect catalytic activity and substrate specificity in unpredictable ways. Accordingly, methods have been designed and developed in which individual amino acids are randomly mutated to any of the other naturally occurring amino acids. Such methods, which require little understanding of structure and function, lead to the generation of a library of proteins that is then screened to identify novel enzyme variants that display the desired

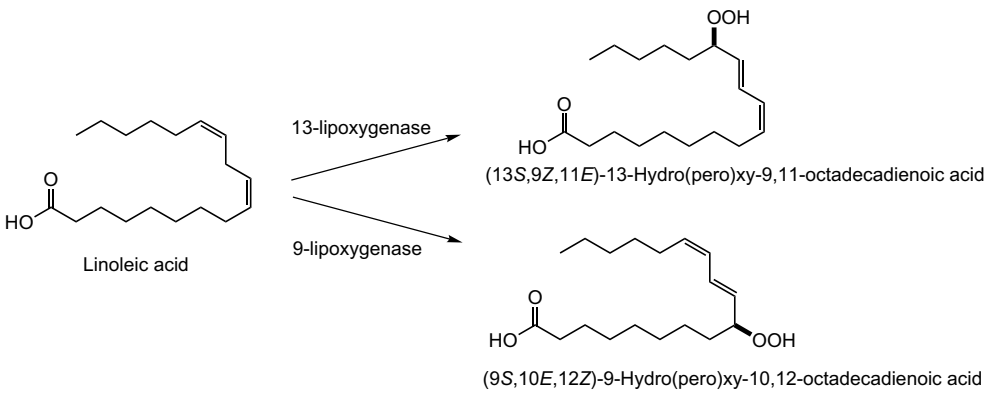

Figure 9. The oxidation of linoleic acid to 9- and 13-hydroperoxy linoleic acid derivatives.

phenotype. Critical to the success of these approaches to protein engineering is the availability of an efficient and reliable means of screening the library for activity. When the library size is modest, established assays for a particular enzymatic reaction can often be adapted in a relatively straightforward manner to determine the catalytic activity of the mutants. However, when large libraries are produced, other procedures must be developed and employed.

3.1. Representative techniques for random mutagenesis

The simplest technique for random mutagenesis is termed saturation mutagenesis and involves replacing a single amino acid in a protein with each of the remaining natural amino acids to give every possible variant at that site. This method, which is often used in combination with other techniques, is especially useful for gaining insights into the structure–function relationships of a biologically active peptide ligand[40] or an enzyme[41] because the effects of introducing each amino acid at a given position may be explicitly evaluated.

Localized or region-specific random mutagenesis is a hybrid of rational and completely random methods for protein engineering that involves the simultaneous replacement of a small number of amino acid residues, typically in a specific region of a protein, to generate proteins with novel specificities.[42–47] Generally, peptide domains containing several residues that are believed to be important for catalysis and/or binding are modified by random substitutions at each site, and the relatively small library thus produced is screened to identify variants that exhibit the desired phenotype.

Creation of protein variants by localized random mutagenesis utilizes procedures that are similar to those employed in the overlap extension and the whole plasmid, single-round PCR mutagenesis protocols discussed previously (Section 2.1, Figs. 2 and 3) with the important modification that the codons for the selected amino acids are randomized. The primers are designed so all three nucleotides in a randomized codon contain a statistical mixture of C, G, A, and T. Hence, for a process resulting in one randomized codon, a mixture of 64 different forward primers and 64 different reverse primers would be used. A modification of this technique restricts the third nucleotide of the codon to either C or G in order to remove unwanted stop codons and reduce the degeneracy of the genetic code by increasing the probability that rarely encoded amino acids, such as tryptophan, will be expressed. A mixture of 32 forward and 32 reverse primers containing a mutated codon are thus employed in a reaction resulting in one randomized codon. The term saturation mutagenesis is also sometimes applied to this technique, but it is important to recognize that there is no guarantee that all of the possible variants, the number of which is equal to 20^n wherein n is the number of amino acid residues being altered, will be generated. Statistical analysis is therefore used to determine the size of the library that must be screened to ensure that there is a high probability that each variant is encountered.

3.2. Selected applications of random mutagenesis

Three residues in the substrate binding pocket of the phosphatidylcholine-preferring phospholipase C from *Bacillus cereus* (PLC_{Bc}) were randomly varied via region-specific random mutagenesis in a study that led to a better understanding of the determinants for substrate specificity.[47] PLC_{Bc} catalyzes the hydrolysis of the phosphodiester bond of phosphatidylcholine (PC), phosphatidylethanolamine (PE), and phosphatidyl-L-serine (PS) with the ratios of the corresponding specificity constants, k_{cat}/K_M, being approximately 10:7:1 to provide a diacylglycerol (DAG) and the corresponding phosphorylated head group (Fig. 10).[48,49]

The crystal structure of a complex of PLC_{Bc} bound to a nonhydrolyzable analog of PC showed that Glu4, Tyr56, and Phe66 were proximal to the choline headgroup of the substrate derivative.[50] Some preliminary studies involving site-directed mutagenesis suggested that these residues might be good candidates for randomization,[51] and a library of about 6000 clones containing variants with one, two, or three amino acid replacements was generated in which Glu4, Tyr56, and Phe66 were randomly varied. The members of the resulting library were screened for their respective activities toward the three phospholipid substrates PC, PE, and PS using a novel colorimetric enzyme coupled assay for detecting inorganic phosphate that was adapted for use in a 96-well plate format (Fig. 11).[48] Ten enzymes with increased specificity constants for PS were initially identified. Of these, the C6PS selective variants E4G, E4K/Y56V, and E4Q/Y56T/F66Y exhibited four-, two-, and three times higher specificity constants (k_{cat}/K_M) toward C6PS than wild type, and their selectivities toward C6PC and C6PS were shifted 30-, 180-, and 60-fold, respectively.

A second round of studies was then performed in which the corresponding wild type residues were singly reinserted back into the E4K/Y56V and E4Q/Y56T/F66Y variants using site directed mutagenesis. It was found that the double and triple mutants uniformly had higher substrate preferences than the single mutants, thereby illustrating the advantage of randomly mutating all three residues in a combinatorial fashion rather than

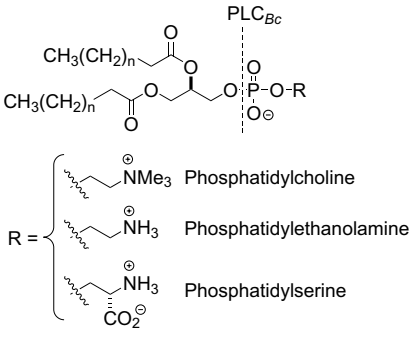

Figure 10. General structures of three families of phospholipids, PC, PE, and PS. The dotted line indicates the phosphodiester bond that is hydrolyzed by PLC_{Bc}.

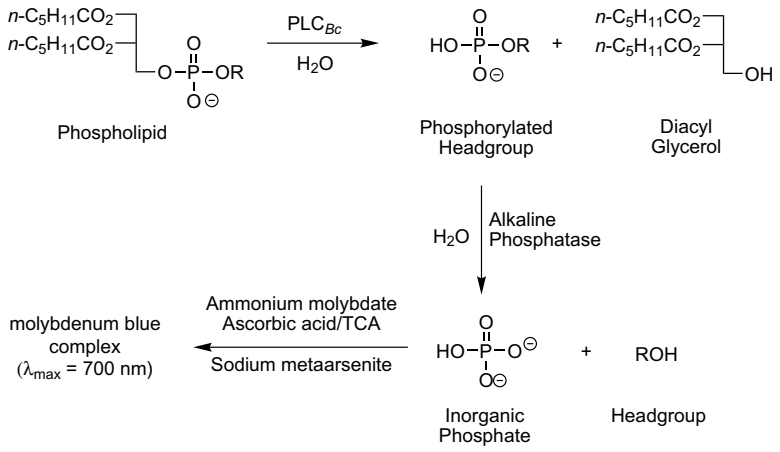

Figure 11. Reaction sequence for the chromogenic assay for detecting inorganic phosphate and screening a library of PLC$_{Bc}$ variants having altered substrate specificities.

singly mutating these residues by saturation mutagenesis. Some generalizations regarding the effect of replacements at each site upon substrate specificity were apparent. Namely, an aromatic residue at position 66 was important for significant catalytic activity toward all three substrates, especially PC and PE perhaps because of a cation interaction. Furthermore, a negatively charged residue at position 4 endowed the enzyme with a PC and PE preference, possibly due to a favorable electrostatic interaction, whereas a polar neutral or positively charged residue results in PS selectivity. Replacing Tyr56 with Val, Ala, Thr, or Ser greatly reduces activity toward PC.

DNA and RNA polymerases catalyze chain elongation reactions that are guided by single stranded DNA templates to generate polynucleotide chains containing 2′-deoxynucleotides (dNTPs) or ribonucleotides (rNTPs), respectively. Wild type *T. aquaticus* DNA polymerase I is unable to form products containing ribonucleotides, but Patel and Loeb utilized random mutagenesis to create variants that could incorporate rNTPs into a growing polynucleotide chain.[52] The section of 13 amino acids between L605-R617 within the dNTP binding site was randomized to create a library of 200,000 variants in which each residue could be any amino acid. The library was subjected to genetic selection in an *E. coli* strain that contained a temperature sensitive mutation in the *polA* gene, which encodes DNA polymerase I, so the strain can be propagated at 30 °C but not at 37 °C. Active variants are selected because they express an active DNA polymerase I that functions at 37 °C, thereby restoring the temperature-sensitive phenotype and enabling the mutant strain to survive at 37 °C. In this manner, 8000 independent clones were found to encode an active *T. aquaticus* DNA polymerase I. A group of 291 of these variants were then screened for their ability to incorporate rNTPs using a kinetic assay. Twenty three variants were thus identified that efficiently incorporated rNTPs into a growing polynucleotide chain, and in several cases the change in specificity constants was shifted by more than 1000-fold toward favoring rNTPs over dNTPs as substrates.

Sequencing these 23 variants in order to elucidate the structural origin of their enhanced specificities toward rNTPs revealed two protein subclasses, one of which incorporated replacements of I614 with a hydrophilic amino acid residue and the other in which E615 was replaced with aspartic acid. Those variants encoding the E615D mutation also contained one to three other substitutions. Modeling based upon X-ray structures suggested that replacing I614, which would be closely packed against the 2′-OH of a rNTP, with a hydrophilic residue might induce movement of the protein backbone at this site to create space for the 2′-OH group. Inspection of X-ray structures also suggested that the carboxyl group on E615 might interfere with binding of the 2′-OH of a rNTP, so shortening the side chain by replacing E615 with aspartic acid could create the requisite space for the 2′-OH group without disrupting the H-bond between the side chain carboxyl group and Y671.

Redesigning enzymes may also be applied to producing enzymes and other proteins that may serve as medicinal drugs that are known as macromolecular human therapeutics.[53,54] Although the use of enzymes as therapeutic agents currently constitutes only a minor portion of pharmaceutical production, such proteins have significant potential for treating a variety of disease states. Several examples will illustrate the essential concepts and some of the possibilities.

Thrombin is a serine protease that proteolytically cleaves fibrinogen and protein C, respectively, to stimulate coagulant and anticoagulant pathways. Toward the goal of creating a variant having enhanced anticoagulant activity, the crystal structure of thrombin was inspected to identify surface-exposed, polar or charged residues that would likely be involved in electrostatic and hydrogen bonding interactions with charged ligands, and alanine scanning was performed at these positions.[55] Because alanine replacements at W50, K52, E229, and R223 produced variants favoring the anticoagulant substrate protein C, these residues were individually substituted with all 19 naturally occurring amino acids by saturation mutagenesis. The resulting

library of thrombin variants was then screened for increased selectivity for protein C and clotting activity as measured by the release of fibrinopeptide A.[56] These studies led to the discovery of the E229K thrombin variant, which exhibited a substrate specificity that favored protein C activation over fibrinogen cleavage by 130-fold. Thrombin variants such as E229K having improved anticoagulant activity might be useful in treating strokes or myocardial infractions.

Engineered enzymes that metabolize prodrugs to their active forms may be used to treat various diseases, including cancer.[57] For example, cancer cells die when herpes simplex virus type 1 thymidine kinase (HSV-1 TK) is injected directly into a tumor that has been dosed with either ganciclovir or acyclovir (Fig. 12).[58,59] HSV-1 TK catalyzes the phosphorylation of thymidine analogues such as ganciclovir and acyclovir, and because these compounds lack a 3′-hydroxyl group, they act as chain terminators upon incorporation into DNA. Although this technique has shown promise in clinical trials, the large doses of HSV-1 TK required for tumor regression unfortunately have immunosuppressive effects, so more selective variants are needed.

In order to generate HSV-1 TK variants that might exhibit higher activity toward ganciclovir or acyclovir, six residues (L159, I160, F161, Ala168, L169, and L170) that are in the nucleoside-binding site were replaced by random mutagenesis, and the resulting library of 1.1×10^6 variants was screened for enhanced activity towards ganciclovir using a two step process.[59] The library was first plated onto TK selection plates to identify clones that exhibited TK activity. These variants were then subjected to ganciclovir and acyclovir, and the clones that did not grow were scored as positive because phosphorylation of these compounds led to cell death. Of these clones, 26 expressed HSV-1 TK variants with increased activity for ganciclovir, whereas 54 variants exhibited enhanced selectivity for acyclovir. Sequencing these HSV-1 TK variants revealed that three to six amino acid replacements had occurred in each one, indicating that it would have been difficult to design these enzymes rationally.

Pan and co-workers recently combined methods of site-directed and saturation mutagenesis to explore structure-activity relationships of vasoactive intestinal peptide (VIP), a 28-mer that binds to and activates VPAC2 and VPAC1 G-coupled protein receptors.[60] VPAC2 activation enhances glucose-induced insulin release, whereas VPAC1 activation elevates glucose output and induces diarrhea. Hence, VPAC2 specificity is crucial in developing VIP variants to treat type II diabetes.[40] In order to identify VPAC2-selective VIP analogues, Pan and co-workers selected a cyclic VPAC2-selective 31-mer analogue of VIP as a starting point. The structural elements of this peptide that precluded recombinant expression were removed, and the resultant 31-mer was found to retain VPAC2 agonist activity. An expression system was developed so VIP variants could be expressed as fusions that were *N*-linked to the C-terminus of glutathione *S*-transferase via a Factor Xa recognition site. The amino acids in the 31-mer that differed from VIP were replaced by site-directed mutagenesis with the corresponding residues in VIP, and replacements V19A, L27K, and N28K were identified as determinants of VPAC2-selectivity, although the later substitution was found to be context dependent. Replacing the *C*-terminal 29G and 30G residues of the 31-mer with the corresponding residues 29K and 30R from pituitary adenylate cyclase-activating peptide (PACAP), which is a 38-mer having a sequence similar to that of VIP and known to activate VPAC1 and VPAC2 receptors, led to potent VPAC2 agonists that exhibited selectivities for VPAC2 over VPAC1 several 100-fold better than VIP or PACAP. Saturation mutagenesis was then performed individually on positions 19, 27, 29, and 30 to reveal which amino acids at each site contributed maximally to VPAC2 selectivity. It was thus found that large hydrophobic residues were not well tolerated at position 19, whereas most amino acids were tolerated at positions 27, 29, and 30; however, VPAC2-selectivity was favored by a positively charged C-terminus.

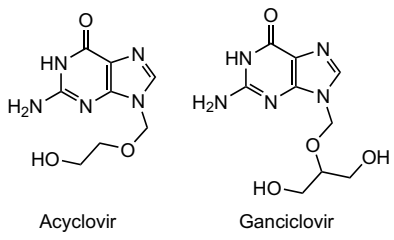

Figure 12. Structures of acyclovir and ganciclovir.

4. Evolutionary methods for altering enzyme substrate selectivity

Many efforts to engineer and reprogram enzyme selectivity via site-directed or region-specific random mutagenesis involving a few selected amino acids fall short of their objectives because of our limited knowledge of structure–function relationships and of protein dynamics and folding. Small changes in the geometry of an active site are often enough to induce large, unexpected consequences.[61] Moreover, multiple mutations, frequently in different domains of the protein, are often required to optimize substrate specificity and catalytic activity. Consequently, evolutionary methods are being developed and employed to alter a variety of properties of proteins, including substrate selectivity. For example, a process known as in vitro evolution has been developed that involves randomization of the entire gene of interest by error prone PCR followed by iterative rounds of screening and selection. This technique has been used to create new variants with remarkable changes in specificities.

4.1. DNA shuffling techniques for enzyme evolution

One useful protocol for evolving enzyme activity is a technique referred to as DNA shuffling, which was originally developed by Stemmer.[62,63] In this method, a group of genes, each consisting of a double stranded

DNA and having relatively similar sequences, is either obtained from different organisms or generated by error-prone PCR. These genes are then randomly cleaved into small fragments by digestion with the restriction enzyme *Dnase*I (Fig. 13). The fragments are then purified and reassembled in a PCR in the presence of a thermostable, error-prone DNA polymerase. This PCR uses the fragments as primers, and hence no other oligonucleotide primers are needed. The fragments align and cross-prime each other for replication to give a hybrid DNA strand with components from the different parent genes.

Several techniques related to DNA shuffling have been developed that offer advantages in certain situations. For example, in a simple modification of the original protocol, Kikuchi and co-workers employed a mixture of restriction endonucleases instead of *Dnase*I to cleave the parent genes in the first step of the process.[64] Use of such mixtures of restriction enzymes yielded a higher frequency of chimeras, which are DNA strands consisting of components from more than one parent gene, thereby resulting in a more diverse library. However, the method suffers from the disadvantage that the crossover sites were biased to coincide with existing restriction sites.

A variation of DNA shuffling that does not require fragmentation of the parent genes is the staggered extension process (StEP) that was developed by Arnold and coworkers.[65,66] In this procedure, terminal primers are employed to replicate the target DNA using PCR with very short extension times to produce short strands of replicated DNA (Fig. 14). After each round of replication, the product is separated from the parent strand by heat denaturation and allowed to anneal and prime the replication of a different template. The growing DNA strand changes templates multiple times and thus accumulates components from different parent genes to give variants with improved functions. Advantages of this procedure over the original DNA shuffling technique are that only small quantities of template are required and that the use of *Dnase*I, which hydrolyzes double-stranded DNA preferentially at sites adjacent to pyrimidine nucleotides and can consequently introduce sequence bias into the recombination, is avoided.

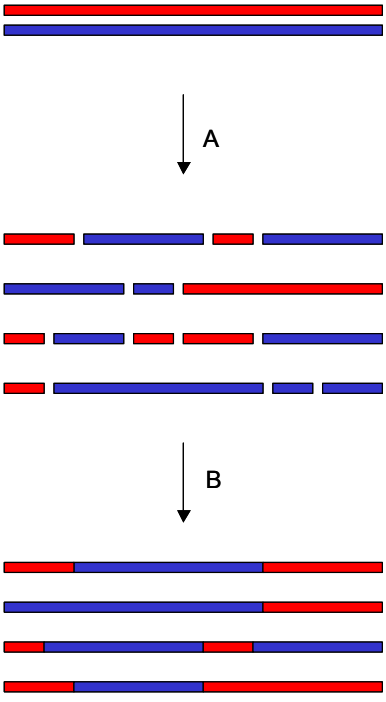

Figure 13. Essential elements of DNA shuffling. (A) Parent genes (only two are shown) are randomly fragmented with *Dnase*I or a mixture of other restriction enzymes. (B) The fragments are reassembled in a primerless PCR reaction in which the fragments serve as primers as well as the template. Replication results in hybrid DNA strands that have components from multiple parent genes.

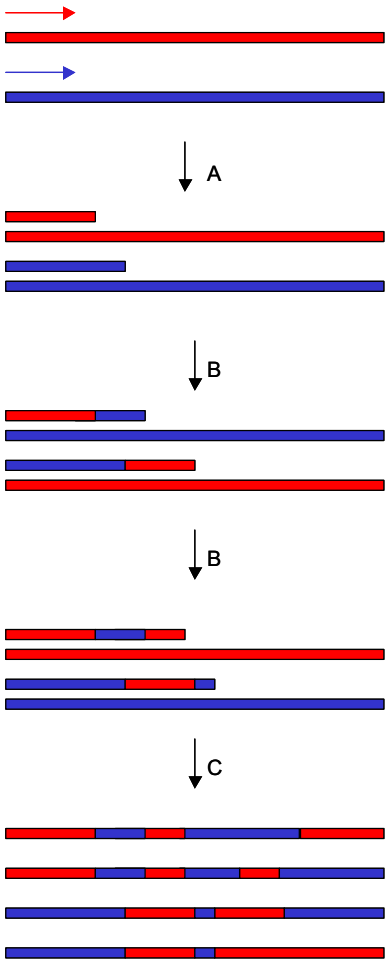

Figure 14. The staggered extension process (StEP). (A) Terminal primers are used to replicate the DNA strands with replication being stopped after a short time to produce small terminal segments of DNA. (B) The fragments are denatured from the template and allowed to anneal onto a different parent template, and replication is allowed to proceed for a short time increment; this step is repeated several times. (C) The procedure results in hybrid DNA strands that have components from multiple parent genes.

4.2. Selected applications of DNA shuffling

There are numerous cases where DNA shuffling and related techniques have been employed to evolve enzyme

variants that exhibit novel properties and/or function. The tremendous power of these methods resides in their ability to combine and recombine fragments from a group of related genes in an iterative fashion to give full length sequences. After each round, variants with the desired properties are selected and used as templates for subsequent gene recombination. A few representative examples will be presented to illustrate these methods.

Matsumura and Ellington used DNA shuffling and selection to evolve the *E. coli* β-glucuronidase to generate variants that would preferentially catalyze the hydrolysis of a β-galactoside substrate.[67] The libraries produced by DNA shuffling were screened for β-galactosidase activity by transforming into a β-galactosidase deficient *E. coli* strain and propagating the transformants on LB agar mother plates. The colonies were absorbed onto nitrocellulose filters and transferred to LB agar plates containing 4.5 mM X-gal. The colonies on the mother plate corresponding to those that turned blue on the plate containing X-gal were selected and subjected to a secondary screen to confirm the phenotypes. After three rounds of DNA shuffling and selection, a T509A/S557P/N566S/K568Q variant was identified that catalyzed the hydrolysis of a β-galactoside substrate 500 times more efficiently (k_{cat}/K_M) than wild type; this change in activity corresponded to a remarkable 52 million-fold inversion in specificity. The effects of the four different amino acid substitutions in the T509A/S557P/N566S/K568Q variant of *E. coli* β-glucuronidase were found to be synergistic as mutants containing only some of these replacements were less selective.

The family of cytochrome P450 monooxygenases catalyzes the oxidation of a wide range of organic substrates, and accordingly members of this family have considerable potential as biocatalysts. Camphor is the natural substrate of the cytochrome $P450_{cam}$ from *P. putida*, which requires nicotinamide adenine dinucleotide (NAD) as a co-factor; this P450 exhibits only weak activity towards naphthalene. In a significant achievement, Arnold and coworkers used random mutagenesis coupled with StEP to alter the substrate selectivity of cytochrome $P450_{cam}$ to hydroxylate naphthalene in the presence of hydrogen peroxide.[68] Indeed, the ability to effect such oxidations with hydrogen peroxide as the terminal oxidant greatly increases the potential of cytochrome P450s as biocatalysts.

In the first stage of evolving the substrate specificity of cytochrome $P450_{cam}$, approximately 200,000 random mutants were generated using mutagenic PCR. In order to screen these mutants, they were coexpressed with a variant of horseradish peroxidase in *E. coli*.[69] The horseradish peroxidase converted the phenolic products of the cytochrome P450 catalyzed oxidation of naphthalene into fluorescent compounds so that active colonies of *E. coli* could be detected directly on plates by digital fluorescence imaging (Fig. 15).[70] In this fashion, some 32,000 clones giving enhanced fluorescence were identified. For example, the R280L/E331K clone hydroxylated naphthalene with about 11-fold and 3-phenylpropionate with about 3-fold higher activities than the native enzyme. A second generation library of P450 variants was then prepared using StEP to recombine five improved $P450_{cam}$ clones from the first generation. Several variants with ~20-fold improvements in naphthalene hydroxylation activity over wild type were thus identified.

Reetz et al. have applied several tools for directed evolution to creating enzymes that may be used for the enantioselective synthesis of chiral organic compounds. For example, using a combination of error-prone PCR, saturation mutagenesis, and DNA shuffling, they were able to alter dramatically the enantioselectivity of a bacterial lipase from *Pseudomonas aeruginosa*, which catalyses the hydrolysis of a racemic *p*-nitrophenyl ester to the carboxylic acid with a selectivity factor $E = 1.1$ favoring hydrolysis of the (*S*)-ester (Fig. 16).[71]

Four cycles of error-prone PCR with an average mutation rate of one amino acid substitution per enzyme were first used to increase *E* from 1.1 to 11 in favor of forming the (*S*)-acid. Saturation mutagenesis of the 'hot spots', which were identified by DNA sequencing

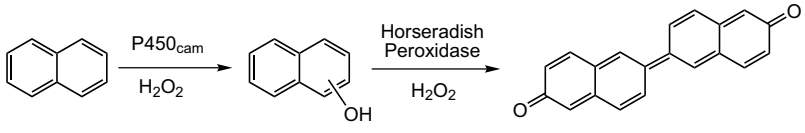

Figure 15. Cytochrome $P450_{cam}$ variants catalyze the hydroxylation of naphthalene, and horseradish peroxidase subsequently catalyzes the oxidative coupling of the product naphthols to form fluorescent dimer/polymer products that are detected by digital imaging.

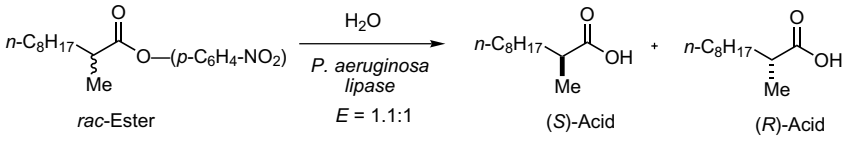

Figure 16. The bacterial lipase from *Pseudomonas aeruginosa* catalyses the hydrolysis of *rac*-ester with selectivity factor $E = 1.1$ favoring formation of (*S*)-acid.

of improved variants, then led to the identification of a variant that displayed higher enantioselectivity ($E = 20$). When one additional cycle of error prone mutagenesis was performed, a variant having five replacements (V47G, V55G, S149G, S155F, and S164G) and an E value of 25 was obtained. Applying DNA shuffling with the wild type gene and this variant led to no improvement in the hydrolytic kinetic resolution. Hence, the wild type gene was subjected to error prone PCR at a mutation rate of about three amino acid replacements per lipase in order to obtain higher diversity in the shuffling partner, and two new variants with enhanced enantioselectivity were identified. When DNA shuffling was performed with these two variants and the V47G/V55G/S149G/S155F/S164G variant, a new lipase (V47G, S149G, S155F, S199G, and T234S) with even higher enantioselectivity ($E = 32$) was created.

DNA shuffling and the derived methods discussed above only work if there is sufficient DNA sequence similarity between all the parent genes. If the parent genes are too diverse, their fragments cannot anneal and cross-prime each other. Recently progress has been made toward addressing this problem using nonhomologous recombination methods.[72,73] Although these methods are promising, they have yet to be used to alter enzyme substrate specificity.

5. Screening libraries generated by random mutagenesis

Developing effective protocols to screen large libraries of enzyme variants in protein engineering is a major challenge because each individual library member must be studied separately. If the direct products of the enzymatic reaction are chromogenic or can be observed directly on agarose plates, the screening process can be relatively simple. For example, when esters of *p*-nitrophenol are substrates for an enzymatic reaction, active clones can be easily identified by monitoring the increase in absorbance at 410 nm using UV/vis spectroscopy.[74]

In cases where product formation in the enzyme catalyzed reaction cannot be directly observed, it is sometimes possible to couple it with one or more subsequent reactions to produce a visible signal. For example, suspensions of cells of *E. coli* harboring variants of PLC_{Bc} were directly screened for activity on three phospholipid substrates in 96-well plates using an enzyme-coupled colorimetric assay for inorganic phosphate (see Fig. 11, Section 3.2).[47] Active colonies were characterized by a colorless-to-blue color change, and the relative intensities of the absorptions at 700 nm, which resulted from formation of a molybdate–inorganic phosphate complex, were roughly correlated with specificity.

The manipulations that are required to perform screens in 96-well plate format limit the applicability of such protocols to libraries of modest size, even when robots are used. Larger libraries are better screened by other methods. For example, a high-throughput screen was developed for identifying cytochrome $P450_{cam}$ variants that hydroxylated naphthalene (see Section 4.2).[68] A variant of horseradish peroxidase was coexpressed with the cytochrome $P450_{cam}$ variants so that the naphthols produced by oxidation would be converted into fluorescent compounds (see Fig. 15). Colonies expressing an active cytochrome $P450_{cam}$ variant were then screened by digital imaging because the wavelength of fluorescence generated by the coupled reaction varied with the site of hydroxylation.

Reetz et al. have developed a useful IR-thermographic detection method to screen the enantioselectivity of enzyme catalyzed reactions.[75] The feasibility of the method was established in a model study involving the enantioselective acylation of racemic 1-phenylethanol with vinyl acetate in the presence of a lipase from *Candida antarctica* (Fig. 17). This transformation is exothermic and proceeds with 99% enantioselectivity to favor formation of (*R*)-1-phenylethylacetate. Thus, separate solutions of (*R*)-1-phenylethanol, (*S*)-1-phenylethanol, and racemic 1-phenylethanol mixed with vinyl acetate were placed in the wells of a microtiter plate, and lipase from *C. antarctica* was added. The temperature changes in the reaction wells were then monitored with an infrared camera, and red spots corresponding to a significant rise in temperature were observed for the wells containing (*R*)-1-phenylethanol. No increase in temperature was detected in wells containing (*S*)-1-phenylethanol, whereas there was a moderate rise in temperature of wells containing racemic 1-phenylethanol. This technique has since been used by Reetz et al. in the context of generating more enantioselective variants of enzymes that catalyze useful reactions.[76]

An innovative screening approach that may be applicable to screening very large libraries involves coupling of cell surface display libraries with fluorescence activated cell sorting (FACS). In an illustrative example of this technique, Iverson, Georgiou and co-workers developed a strategy to identify active variants of the *E. coli* surface-displayed serine protease OmpT that exhibited altered substrate specificity.[77] Wild-type OmpT cleaves the peptide bond between two positively charged amino acids, such as Lys or Arg. In order to screen mutants of OmpT that would cleave a Arg-Val linkage, a substrate

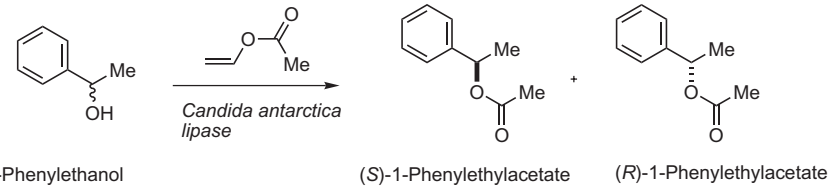

Figure 17. Lipase-catalyzed enantioselective synthesis of 1-phenylethylacetate from racemic 1-phenylethanol.

was synthesized containing the following components: a fluorophore, a positively charged moiety and the scissile bond (Arg-Val), and a quenching fluorophore that acts as an intramolecular fluorescence resonance electron transfer (FRET) partner (Fig. 18). The positively charged substrate first binds to the negatively charged surface of an *E. coli* cell by electrostatic interactions. If a mutant of OmpT hydrolyzes the scissile bond of the substrate, the fluorophore and the quenching fluorophore are separated, and fluorescence is observed. FACS may then be employed to isolate the clones in which cleavage of the scissile bond had ensued. In this fashion, a library of 6×10^5 OmpT variants was screened for cleaving the Arg-Val peptide linkage, and proteases with 60-fold improvement in specificity constants were obtained.

Another common technique for screening large protein libraries is phage display.[78] In this method, degenerate reverse primers are used in a PCR reaction to randomly mutate the starting cDNA throughout a target region, and the PCR products are subcloned into a M13 bacteriophage vector that encodes the coat protein III of the filamentous phage f1. Because each phage in the pool of mutants expresses a protein variant that is displayed on the coat protein of the phage surface, the library may be screened for variants that bind tightly to a given substrate by elution experiments. The phage–DNAs from the clones thus identified are then purified and sequenced.

The versatility of phage display to create proteins having novel binding properties is illustrated by its use to generate peptides that bind with high specificity to semiconductor surfaces.[79] Thus, Belcher and co-workers generated a combinatorial library containing 12 randomized amino acids that were fused to the pIII coat protein of M13 phage. Five copies of pIII were displayed on one end of the phage particle. The library members were screened against five different single crystal substrates, gallium arsenide(100), gallium arsenide(111)A, gallium arsenide(111)B, indium phosphide(100), and silicon(100). Proteins that bound to specific crystals were eluted from the surface, amplified, and re-reacted against the substrate under more stringent conditions. This protocol was repeated five times, and twenty peptides that selectively bound gallium arsenide were isolated and sequenced. These active peptides contained a large proportion of amino acids having Lewis-base functional groups (e.g., threonine, serine, asparagine, and glutamine) that could bind to Lewis-acid sites on the gallium arsenide surface. In this manner, they identified one peptide that was specific for gallium arsenide(100) and did not bind to either the gallium arsenide(111)A or the gallium arsenide(111)B lattices.

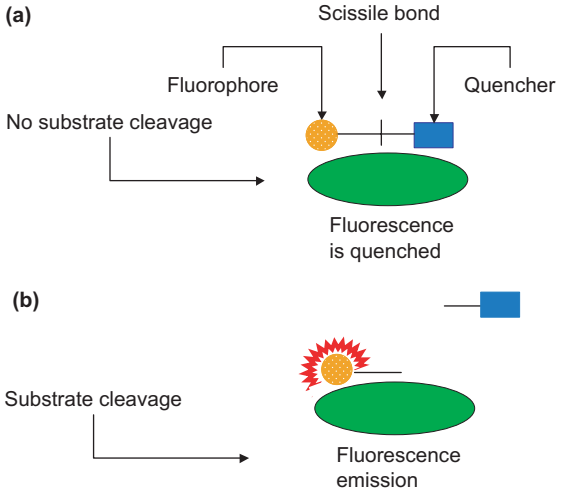

Figure 18. Coupling of cell surface display libraries with fluorescence activated cell sorting (FACS) represents an innovative screening approach that may be applicable to screening of very large libraries. (a) The scissile bond of the substrate is not cleaved by the enzyme, and the fluorophore emission is quenched by the quenching fluorophore so there is no fluorescence emission. (b) The scissile bond of the substrate is hydrolyzed, so the fluorophore and the quenching fluorophore are separated and fluorescence is observed.

6. Combining evolutionary methods with biological selection

It is apparent from the preceding discussions that random mutagenesis, which may then be coupled with evolution-based methods, generates large libraries of variants having altered substrate selectivity profiles. However, the technical difficulty of establishing a rapid and sensitive protocol to evaluate $>10^{10}$ mutants for the desired specificity often limits the practical utility of these methods. The identification of variants with chosen characteristics is greatly simplified when there is a direct connection between cell survival and enzyme activity. In such cases, genetic selection can be performed, wherein only those colonies containing variants with the desired phenotype are able to live. If a protocol involving genetic selection can be designed for the system of interest, large populations of biomolecules may be readily analyzed.[80–83] A specific example will serve to illustrate the power of this approach.

The conversion of aspartate aminotransferase from *E. coli* into a valine aminotransferase represents an excellent example of the use of genetic selection to generate enzymes with altered selectivities.[84] Wild type aspartate aminotransferase catalyzes the conversion of oxaloacetate to aspartate ($k_{cat}/K_M = 2.3 \times 10^7\ s^{-1}\ M^{-1}$), but it exhibits very low activity in transforming 2-oxovaline into valine ($k_{cat}/K_M = 0.057\ s^{-1}\ M^{-1}$) (Fig. 19a). Thus, the wild type aspartate aminotransferase gene was first subjected to error prone PCR to generate templates for DNA shuffling that led to mutants of aminotransferase. These mutants were then transformed into an auxotrophic strain of *E. coli* lacking the gene for branched-chain amino acid aminotransferase, which is required for the synthesis of valine. Because this strain cannot grow in the absence of added valine, only those colonies expressing a variant of aspartate aminotransferase capable of converting 2-oxovaline to valine survive (Fig. 19b). After five rounds of shuffling and selection, an aminotransferase variant having 13 amino

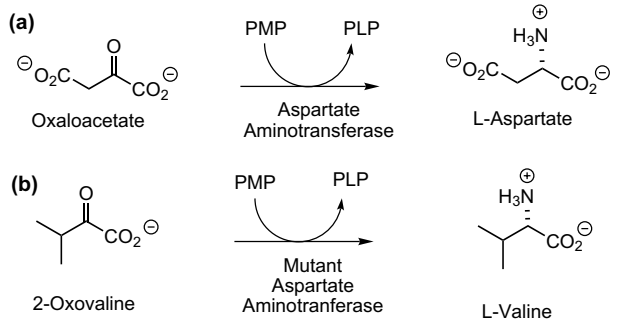

Figure 19. (a) Aspartate aminotransferase catalyzes the transamination of oxaloacetate to L-aspartate with concomitant conversion of the pyridoxamine phosphate cofactor (PMP) to pyridoxal phosphate (PLP). (b) Mutagenesis yielded an aspartate aminotransferase variant that catalyzes the conversion of 2-oxovaline to L-valine.

acid substitutions was identified whose activity for 2-oxovaline had increased by five orders of magnitude compared to wild type ($k_{cat}/K_M = 7.4 \times 10^3\ s^{-1}\ M^{-1}$). The specificity constant of this mutant toward oxaloacetate decreased approximately 10-fold, reflecting a large change in substrate selectivity. It is noteworthy that six (N34D, I37M, S139G, N142T, N297S, and V387L) of the 13 amino acid replacements accounted for 80–90% of the change in substrate specificity. Examination of the X-ray crystal structure of *E. coli* aspartate aminotransferase complexed with 2-methyl-L-aspartate suggested that only I37 was positioned so it could interact with the substrate. Thus, this study illustrates the significant effects residues distal to the active site of an enzyme can have upon catalysis and binding.

7. Conclusions and future directions

The advances that have been made in recent years toward reprogramming the substrate specificity of proteins are remarkable, but there remains considerable opportunity for new discoveries. Although rational methods for protein engineering are sometimes productive, our present knowledge of the relationships between protein structure and function does not yet generally enable the rational design of specificity in catalysis and binding. Indeed, altered substrate selectivity and activity profiles are frequently the result of multiple amino acid replacements, and variants with only some of these replacements are typically less selective as synergy plays an important role in specificity.[61] Moreover, amino acid replacements that alter the selectivity of a protein are often located too far from the binding pocket to interact directly with the substrate or ligand, indicating that subtle differences throughout the protein can have a large and unpredictable impact on selectivity. This observation is consistent with the induced fit model, which postulates that substrate binding is accompanied by changes in the three-dimensional structure of the protein, so flexibility and dynamics are important determinants of selectivity and catalytic activity.

The set of tools that is available for altering protein substrate selectivity is getting more diverse and effective. Rational, random, and evolution-based methods have all been successfully applied to altering substrate selectivity of proteins, but a combination of different approaches is often the best avenue for success. For example, altering the substrate selectivity by rational mutagenesis of an enzyme active site followed by employing various evolution-based techniques to increase the activity of the variant has been shown to be productive. In the future, improved computer algorithms that better predict the effect of amino acid replacements, especially those distal to the active site, on enzyme selectivity will likely emerge. Furthermore, as new screening and selection protocols are developed and automated, evolution-based approaches will become more useful and easier to apply. Methods that will enable the *facile* incorporation of unnatural amino acids into proteins via mutagenesis will greatly broaden the scope of what is possible. Ultimately, understanding the complex relationships between protein structure and function will enable us to design and synthesize proteins exhibiting desired activities and selectivities de novo. Rapid progress on numerous fronts is being made, but many challenges remain before we can achieve this goal–the holy grail of protein engineering.

Acknowledgements

We thank the National Institutes of Health, the Robert A. Welch Foundation, and the Texas Advanced Research Program for supporting our research.

References and notes

1. Penning, T. M.; Jez, J. M. *Chem. Rev.* **2001**, *101*, 3027.
2. Brannigan, J. A.; Wilkinson, A. J. *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 964.
3. Betz, S. F.; Liebman, P. A.; DeGrado, W. F. *Biochemistry* **1997**, *36*, 2450.
4. Severin, K.; Lee, D. H.; Kennan, A. J.; Ghadiri, M. R. *Nature* **1997**, *389*, 706.
5. Cedrone, F.; Menez, A.; Quemeneur, E. *Curr. Opin. Struct. Biol.* **2000**, *10*, 405.
6. Hedstrom, L.; Graf, L.; Stewart, C. B.; Rutter, W. J.; Phillips, M. A. *Methods Enzymol.* **1991**, *202*, 671.
7. Reikofski, J.; Tao, B. Y. *Biotechnol. Adv.* **1992**, *10*, 535.
8. Kammann, M.; Laufs, J.; Schell, J.; Gronenborn, B. *Nucleic Acids Res.* **1987**, 5404.
9. Nelson, M.; McClelland, M. *Methods Enzymol.* **1992**, *216*, 279.
10. Braman, J.; Papworth, C.; Greener, A. *Methods Mol. Biol.* **1996**, *57*, 31.
11. Davis, B. G.; Boyer, V. *Nat. Prod. Rep.* **2001**, *18*, 618.
12. Schoemaker, H. E.; Mink, D.; Wubbolts, M. G. *Science* **2003**, *299*, 1694.
13. Jaeger, K.-E.; Reetz, M. T. *Curr. Opin. Chem. Biol.* **2000**, *4*, 68.
14. Nixon, A. E.; Firestine, S. M. *IUBMB Life* **2000**, *49*, 181.
15. Liu, J.; Wong, C.-H. *Angew. Chem., Int. Ed.* **2002**, *41*, 1404.
16. DeSantis, G.; Liu, J.; Clark, D. P.; Heine, A.; Wilson, I. A.; Wong, C.-H. *Bioorg. Med. Chem.* **2003**, *11*, 43.
17. Heine, A.; DeSantis, G.; Luz, J. G.; Mitchell, M.; Wong, C. H.; Wilson, I. A. *Science* **2003**, *294*, 369.

18. Liu, J.; Hsu, C.-C.; Wong, C.-H. *Tetrahedron Lett.* **2004**, *45*, 2439.
19. Lockridge, O.; Blong, R. M.; Masson, P.; Froment, M.-T.; Millard, C. B.; Broomfield, C. A. *Biochemistry* **1997**, *36*, 786.
20. Crameri, A.; Dawes, G.; Rodriguez, E.; Silver, S.; Stemmer, W. P. C. *Nat. Biotechnol.* **1997**, *15*, 436.
21. Kumamaru, T.; Suenaga, H.; Mitsuoka, M.; Watanabe, T.; Furukawa, K. *Nat. Biotechnol.* **1998**, *16*, 663.
22. Lai, K.; Stolowich, N. J.; Wild, J. R. *Arch. Biochem. Biophys.* **1995**, *318*, 59.
23. Kolakowski, J. E.; Defrank, J. J.; Harvey, S. P.; Szafraniec, L. L.; Beaudry, W. T.; Lai, K.; Wild, J. R. *Biocatal. Biotransform.* **1997**, *15*, 297.
24. Di Sioudi, B. D.; Miller, C. E.; Lai, K.; Grimsley, J. K.; Wild, J. R. *Chem. Biol. Interact.* **1999**, *119–120*, 211.
25. Shah, K.; Liu, Y.; Deirmengian, C.; Shokat, K. M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 3565.
26. Liu, Y.; Shah, K.; Yang, F.; Witucki, L.; Shokat, K. M. *Chem. Biol.* **1998**, *5*, 91.
27. Eblen, S. T.; Kumar, N. V.; Shah, K.; Henderson, M. J.; Watts, C. K.; Shokat, K. M.; Weber, M. J. *J. Biol. Chem.* **2003**, *278*, 14926.
28. Ubersax, J. A.; Woodbury, E. L.; Quang, P. N.; Paraz, M.; Blethrow, J. D.; Shah, K.; Shokat, K. M.; Morgan, D. O. *Nature* **2003**, *425*, 859.
29. Kazlauskas, R. J. *Curr. Opin. Chem. Biol.* **2000**, *4*, 81.
30. Bornscheuer, U. T.; Pohl, M. *Curr. Opin. Chem. Biol.* **2001**, *5*, 137.
31. Bolon, D. N.; Voigt, C. A.; Mayo, S. L. *Curr. Opin. Chem. Biol.* **2002**, *6*, 125.
32. Wisz, M. S.; Hellinga, H. W. *Proteins* **2003**, *51*, 360.
33. Benkovic, S. J.; Hammes-Schiffer, S. *Science* **2003**, *301*, 1196.
34. Marvin, J. S.; Hellinga, H. W. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4955.
35. De Lorimier, R. M.; Smith, J. J.; Dwyer, M. A.; Looger, L. L.; Sali, K. M.; Paavola, C. D.; Rizk, S. S.; Sadigov, S.; Conrad, D. W.; Loew, L.; Hellinga, H. W. *Protein Sci.* **2002**, *11*, 2655.
36. Yang, W.; Jones, L. M.; Isley, L.; Ye, Y.; Lee, H.-W.; Wilkins, A.; Liu, Z.-R.; Hellinga, H. W.; Malchow, R.; Ghazi, M.; Yang, J. J. *J. Am. Chem. Soc.* **2003**, *125*, 6165.
37. Looger, L. L.; Dwyer, M. A.; Smith, J. J.; Hellinga, H. W. *Nature* **2003**, *423*, 185.
38. Cronin, C. N. *J. Biol. Chem.* **1998**, *273*, 24465.
39. Hornung, E.; Walther, M.; Kuhn, H.; Feussner, I. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 4192.
40. Tsutsumi, M.; Claus, T. H.; Liang, Y.; Li, Y.; Yang, L.; Zhu, J.; Dela Cruz, F.; Peng, X.; Chen, H.; Yung, S. L.; Hamren, S.; Livingston, J. N.; Pan, C. Q. *Diabetes* **2002**, *51*, 1453.
41. Miyazaki, K.; Wintrode, P. L.; Grayling, R. A.; Rubingh, D. N.; Arnold, F. H. *J. Mol. Biol.* **2000**, *297*, 1015.
42. Graham, L. D.; Hagget, K. D.; Jennings, P. A.; Le Brocque, D. S.; Whittaker, R. G. *Biochemistry* **1993**, *32*, 6250.
43. Choo, Y.; Klug, A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11163.
44. Viadiu, H.; Osuna, J.; Fink, A. L.; Soberon, X. *J. Biol. Chem.* **1995**, *270*, 781.
45. Cantu, C., III; Huang, W.; Palzkill, T. *J. Biol. Chem.* **1996**, *271*, 22538.
46. Hubner, B.; Haensler, M.; Hahn, U. *Biochemistry* **1999**, *38*, 1371.
47. Antikainen, N. M.; Hergenrother, P. J.; Harris, M. M.; Corbett, W.; Martin, S. F. *Biochemistry* **2003**, *42*, 1603.
48. Hergenrother, P. J.; Martin, S. F. *Anal. Biochem.* **1997**, *251*, 45.
49. Hergenrother, P. J.; Martin, S. F. *Top. Curr. Chem.* **2001**, *211*, 131.
50. Hansen, S.; Hough, E.; Svensson, L. A.; Wong, Y.-L.; Martin, S. F. *J. Mol. Biol.* **1993**, *234*, 179.
51. Martin, S. F.; Follows, B. C.; Hergenrother, P. J.; Trotter, B. K. *Biochemistry* **2000**, *39*, 3410.
52. Patel, P. H.; Loeb, L. A. *J. Biol. Chem.* **2000**, *275*, 40266.
53. Cho, M. J.; Juliano, R. *Trends Biotechnol.* **1996**, *14*, 153.
54. Weng, Z.; DeLisi, C. *Trends Biotechnol.* **2002**, *20*, 29.
55. Tsiang, M.; Jain, A. K.; Dunn, K. E.; Rojas, M. E.; Leung, L. L. K.; Gibbs, C. S. *J. Biol. Chem.* **1995**, *270*, 16854.
56. Tsiang, M.; Paborsky, L. R.; Li, W. X.; Jain, A. K.; Mao, C. T.; Dunn, K. E.; Lee, D. W.; Matsumura, S. Y.; Matteucci, M. D., et al. *Biochemistry* **1996**, *35*, 16449.
57. Encell, L. P.; Landis, D. M.; Loeb, L. A. *Nat. Biotechol.* **1999**, *17*, 143.
58. Black, M. E.; Loeb, L. A. *Biochemistry* **1993**, *32*, 11618.
59. Black, M. E.; Newcomb, T. G.; Wilson, H.-M. P.; Loeb, L. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 3525.
60. Yung, S. L.; Dela Cruz, F.; Hamren, S.; Zhu, J.; Tsutsumi, M.; Bloom, J. W.; Caudle, M.; Roczniak, S.; Todd, T.; Lemoine, L.; MacDougall, M.; Shanafelt, A. B.; Pan, C. Q. *J. Biol. Chem.* **2003**, *278*, 10273.
61. Koshland, D. E., Jr. *Nat. Medicine* **1998**, *4*, 1112.
62. Stemmer, W. P. C. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 10747.
63. Stemmer, W. P. C. *Nature* **1994**, *370*, 389.
64. Kikuchi, M.; Ohnishi, K.; Harayama, S. *Gene* **1999**, *236*, 159.
65. Shao, Z.; Zhao, H.; Giver, L.; Arnold, F. H. *Nucleic Acids Res.* **1998**, *26*, 681.
66. Zhao, H.; Giver, L.; Shao, Z.; Affholter, J. A.; Arnold, F. H. *Nat. Biotechnol.* **1998**, *16*, 258.
67. Matsumura, I.; Ellington, A. D. *J. Mol. Biol.* **2001**, *305*, 331.
68. Joo, H.; Lin, Z.; Arnold, F. H. *Nature* **1999**, *399*, 670.
69. Lin, Z.; Thorsen, T.; Arnold, F. H. *Biotechnol. Prog.* **1999**, *15*, 467.
70. Joo, H.; Arisawa, A.; Lin, Z.; Arnold, F. H. *Chem. Biol.* **1999**, *6*, 699.
71. Reetz, M. T.; Wilensek, S.; Zha, D.; Jaeger, K.-E. *Angew. Chem., Int. Ed.* **2001**, *40*, 3589.
72. Stevenson, J. D.; Benkovic, S. J. *J. Chem. Soc., Perkin Trans. 2* **2002**, 1483.
73. Lutz, S.; Ostermeier, M.; Benkovic, S. J. *Nucleic Acids Res.* **2001**, *29*, e16.
74. Reetz, M. T. *Tetrahedron* **2002**, *58*, 6595.
75. Reetz, M. T.; Becker, M. H.; Kuhling, K. M.; Holzwarth, A. *Angew. Chem., Int. Ed.* **1998**, *37*, 2647.
76. Reetz, M. T.; Becker, M. H.; Liebl, M.; Furstner, A. *Angew. Chem., Int. Ed.* **2000**, *39*, 1236.
77. Olsen, M. J.; Stephens, D.; Griffiths, D.; Daugherty, P.; Georgiou, G.; Iverson, B. L. *Nat. Biotechnol.* **2000**, *18*, 1071.
78. O'Neil, K.; Hoess, R. H. *Curr. Opin. Struct. Biol.* **1995**, *5*, 443.
79. Whaley, S. R.; English, D. S.; Hu, E. L.; Barbara, P. F.; Belcher, A. M. *Nature* **2000**, *405*, 665.
80. Hart, E. A.; Hua, L.; Darr, L. B.; Wilson, W. K.; Pang, J.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **1999**, *121*, 9887.
81. Liu, D. R.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 4780.
82. Jurgens, C.; Strom, A.; Wegener, D.; Hettwer, S.; Wilmanns, M.; Sterner, R. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9925.
83. Taylor, S. V.; Kast, P.; Hilvert, D. *Angew. Chem., Int. Ed.* **2001**, *40*, 3310.
84. Yano, T.; Oue, S.; Kagamiyama, H. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5511.