ENCYCLOPEDIA OF

BIOPROCESS TECHNOLOGY: FERMENTATION, BIOCATALYSIS, AND BIOSEPARATION

Michael C. Flickinger
University of Minnesota
St. Paul, Minnesota

Stephen W. Drew
Merck and Co., Inc.
Rahway, New Jersey

John Wiley & Sons, Inc.

ISBN 0-471-13822-3

ENZYMES, DIRECTED EVOLUTION

FRANCES H. ARNOLD
PATRICK L. WINTRODE
California Institute of Technology
Pasadena, California

KEY WORDS

Biocatalysis
Catalyst improvement
DNA shuffling
High-throughput screening
Protein engineering
Random mutagenesis
Recombination
Selection

OUTLINE

INTRODUCTION

Staunch Darwinists attribute all the complexity of living things to an algorithm of mutation and natural selection. The exquisite products of this evolution algorithm are apparent at all levels, from the amazing diversity of life all the way down to individual protein molecules. Scientists and engineers who wish to redesign these same molecules are now implementing their own versions of the algorithm. Directed evolution allows us to explore enzyme functions never required in the natural environment and for which the molecular basis is poorly understood. This bottom-up design approach contrasts with the more conventional, top-down one in which proteins are tamed "rationally" using computers and site-directed mutagenesis. We will describe how molecular evolution can be directed in the test tube in order to produce useful biocatalysts. It is not possible to provide a complete account of all the methods proposed for in vitro evolution; this article therefore introduces methods and strategies used successfully in our laboratory for directing the evolution of enzymes. Some alternative methods for biocatalyst evolution are also be discussed. The reader should also be aware that there is a rather substantial and largely separate literature on combinatorial approaches to engineering binding molecules. Recent advances in the ability to create genetic diversity and to screen or select for improved functions in large libraries of enzyme variants are being combined in a robust approach to solving difficult molecular design problems. With directed evolution we now have the ability to tailor individual proteins as well as whole biosynthetic and biodegradation pathways for biotechnology applications.

WHY DIRECTED EVOLUTION?

The Limitations of Nature's Biocatalysts Present Insurmountable Challenges to Rational Design

When natural enzymes are recruited for industrial applications—from serving as catalysts in chemicals synthesis to additives for laundry detergents—we discover that they are often not well suited to these tasks. Due to poor substrate solubility, breakdown of unstable products, or competing chemical reactions, the conditions for an enzyme reaction may be unsuitable for large-scale applications. Reflecting their participation in complex biochemical networks inside living cells, enzymes are often inhibited by their own substrates or products, either of which may severely limit the productivity of a biocatalytic process. Evolution is usually the culprit: enzymes are optimized and often highly specialized for specific biological functions within the context of a living organism. Biotechnology, in contrast, needs enzymes that are stable and active over long periods of time (a feature that might clash with the need for rapid protein turnover inside a cell), are active in nonaqueous solvents (a feature probably not required in most biological milieu), and can accept different substrates (substrates not present in nature).

It is possible to produce new enzymes in recombinant organisms, altering the amino acid sequence and therefore the properties through appropriate modifications at the DNA level. We are hobbled, however, by near complete ignorance of how the amino acid sequence affects every aspect of enzyme performance, from its ability to be expressed in a heterologous host to its catalytic activity in nonnatural environments. Numerous protein engineering experiments have demonstrated that changes in protein properties are brought about by the cumulative effects of many small adjustments, many of which are distributed or propagated over significant distances. Furthermore, proteins are usually teetering on the brink of instability, with folded structures that are more stable than unfolded—and therefore inactive—ones by the equivalent of a few hydrogen bonds out of the hundreds that form. Superimposing on the need to retain this relatively fragile folded state, the additional requirements of having to fold in the first place, and the need of maintaining or even reengineering a catalytic site that is affected at some level by virtually any modification yields a design problem of such complexity that any rational design effort will require enormous inputs of structural, mechanistic, and dynamic information. Information that is available for but a tiny fraction of interesting catalysts. Even if one trait is successfully designed (e.g., enhanced stability), it is virtually impossible to predict the cost to another (e.g., catalytic activity or expression level). The relatively few examples where rational design has yielded *useful* enzymes do not negate the view that rational enzyme design is often a fruitless exercise.

Extending Natural Diversity by Laboratory Evolution

All these hurdles to the rational design of enzymes are bypassed by evolution. The power of evolution as an algorithm for molecular design is perhaps best appreciated by studying its products. By constructing the evolutionary histories of today's proteins we have learned that they are highly adaptable molecules, at least on evolutionary time scales. Well illustrated by the panoply of α/β-barrel enzymes (1), enzymes catalyzing very different reactions have evolved divergently from a common ancestral protein of the same general structure, acquiring diverse capabilities by processes of random mutation, recombination, and natural selection. We also know that enzymes sharing a common function (for example, all catalyzing a particular

step in a metabolic pathway) in addition to three-dimensional structure can exhibit widely different properties (stability, solubility, tolerance to pH, etc.), depending on where they are found.

Enzymes evolve, and adapt, at the molecular level. The structures of protein modules are conserved (although the modules themselves are often shuffled to create new, multifunctional proteins). Function, however, can vary. Specific features such as substrate specificity or thermostability vary significantly. Amino acid sequences can vary to such an extent that evolutionary relationships may no longer be apparent from sequences alone (1).

Evolution is a powerful algorithm with proven ability to alter enzyme function and especially to "tune" enzyme properties. It is also an algorithm that can be implemented in the laboratory for redesign. The challenge is to collapse the time scale to months, or even weeks.

CHOOSING AN EVOLUTIONARY STRATEGY

Evolutionary mechanisms at work in nature assure adaptability to ever-changing environments. Evolution does not work toward any particular direction, nor is there a goal; the underlying processes occur spontaneously during reproduction and survival of the whole organism. In contrast, a directed molecular evolution experiment has a defined goal, and the key processes (mutation, recombination, and screening or selection) are controlled by the experimenter. Although there may be multiple ways to reach a defined goal (i.e., a desired enzyme function), the approach that minimizes the effort is preferred.

The major steps in a typical directed enzyme evolution experiment are outlined in Figure 1. The genetic diversity for evolution is created by mutagenesis and/or recombination of one or more parent sequences. These altered genes are inserted into a plasmid for expression in a suitable host organism (bacteria or yeast). Clones expressing improved enzymes are identified in a high-throughput screen or by selection, and the gene(s) encoding those improved enzymes are isolated and recycled to the next round

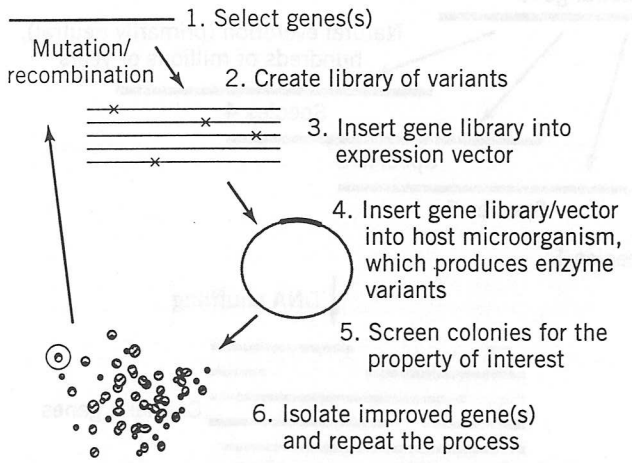

Figure 1. Key steps of a typical directed enzyme evolution experiment.

of directed evolution. Approaches for carrying out these key steps are discussed in some detail in the "Methods of Directed Evolution" section. Here we focus on more fundamental considerations that help to define workable strategies for directed evolution.

To appreciate the challenge of designing and carrying out a successful directed protein evolution experiment, it is important to underscore the powerful combinatorial features of this system. A typical enzyme is a linear chain of N amino acids (N is usually several hundred), and there are 20 possible amino acids at each position in the chain. Thus the "sequence space" of possible proteins is huge beyond the imagination (20^N). Even in 4 billion years, nature has had a chance to explore but a tiny fraction of these possibilities. A laboratory exploration of this vast space of sequences and their corresponding functions must obviously be severely limited and carefully guided (2). Because much of sequence space will be devoid of the desired function, and probably even of folded proteins, it is best to direct the evolution of an existing enzyme rather than look for function in random peptide libraries.

Evolution is often referred to as a hill-climbing exercise in the fitness landscape of sequence space (3,4). The fitnesses (performance, for laboratory evolution) of the proteins in sequence space make up this landscape, whose most basic features are still quite unknown. The landscape for laboratory evolution will be different for each property or collection of properties undergoing evolution. An uphill climb in a protein landscape is more likely to be successful if it can take place in small steps (one or two amino acid substitutions). The high dimensionality of the surface (there are $19N$ one-mutant neighbors of any given sequence) offers many opportunities to find improved mutants. Although we may never reach the "global optimum," the improvements achieved by taking even a simple random up-hill walk via single amino acid mutations often yield useful results. A widely effective evolutionary strategy, illustrated in Figure 2, is one in which the steps are small (preferably one or two amino acid substitutions in each generation), and multiple such mutations are accumulated either sequentially (5) or by recombination (6,7) to acquire the desired function. Such an approach is compatible with a low level of random mutagenesis over the entire gene. An alternative approach is to direct a much higher level of random mutation to a relatively small region of the gene (8). Both approaches have their advantages. Mutagenesis over the entire gene allows discovery of unanticipated solutions (a common experience). More intense, directed mutation, however, may yield novel combinations of amino acid substitutions, combinations that would be inaccessible by single-step walks because the intermediates are unfavorable. Details of the evolutionary exploration are ultimately dictated by a combination of *(1)* the power of the search tool, *(2)* the frequency of beneficial mutations (usually small!), and *(3)* the choice of starting point(s).

A different approach to creating diversity for directed evolution is the in vitro shuffling of homologous genes, or "family shuffling" (9), illustrated in Figure 3. Here, recombination of two or more parent genes yields a chimeric gene library for evolution of the desired features. Because the recombined sequences are related through divergent evo-

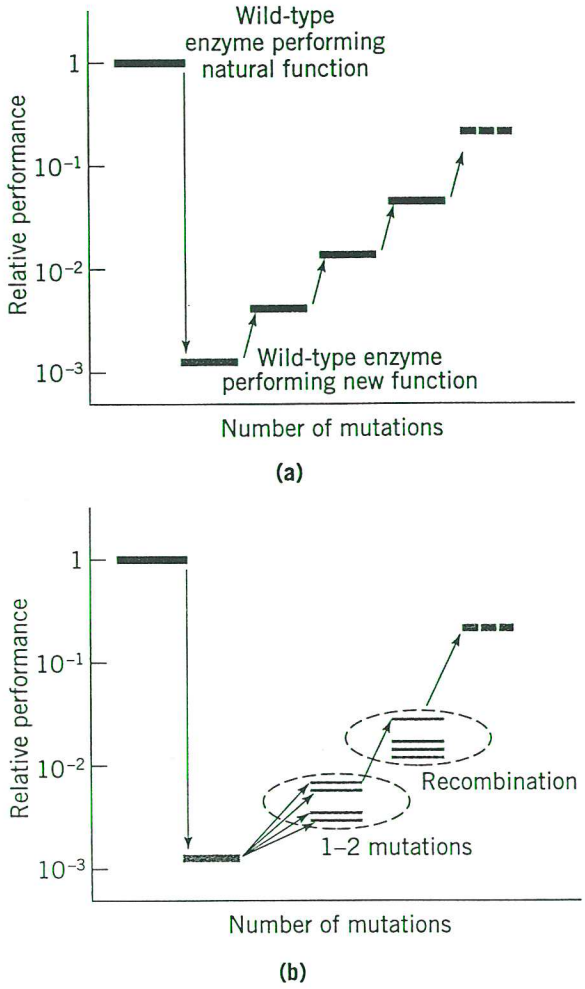

Figure 2. An effective strategy for directed enzyme evolution is one in which small changes associated with one or two amino acid substitutions are accumulated sequentially (**a**) or by recombination of improved genes (**b**).

lution from a common ancestor of similar structure and function (and therefore the sequence differences are to some extent neutral with respect to structure and function), it appears that very large jumps in sequence space can yield functional proteins (9,10). In vivo recombination can also yield interesting new chimeric enzymes (11–14).

Good Problems for Directed Enzyme Evolution

There are four requirements for successful directed evolution. *(1)* The desired function must be physically possible. *(2)* The function must also be biologically, or evolutionarily, feasible. In practice, this means that there exists a mutational pathway to get from here to there through ever-improving variants (see earlier discussion). Although we cannot know a priori that the path exists, a good experiment will maximize the likelihood. *(3)* One must be able to make libraries of mutants complex enough to contain rare beneficial mutations. This usually means functional expression in a suitable microorganism such as *Escherichia coli* or *Saccharomyces cerevisiae*. *(4)* One must have a rapid screen or selection that reflects the desired function. Just how rapid the screen must be depends on how rare mutations leading to the desired property are and how many must be accumulated to achieve the desired result.

Whether directed evolution will solve a particular problem depends to some extent on how hard natural evolution has already worked at it. If a particular trait is already under selective pressure (e.g., catalytic activity), it is unlikely that further improvements can be obtained in the laboratory by small mutational steps. However, if biological function has imposed additional constraints, for example the trait is coupled to another trait that is also under selective pressure (e.g., high thermostability), then this balance can be altered during laboratory evolution (15). While selected traits are often difficult to improve, they should be relatively easy to remove (e.g., product inhibition). Many traits are not under selective pressure; they may be changing as a result of random genetic drift,

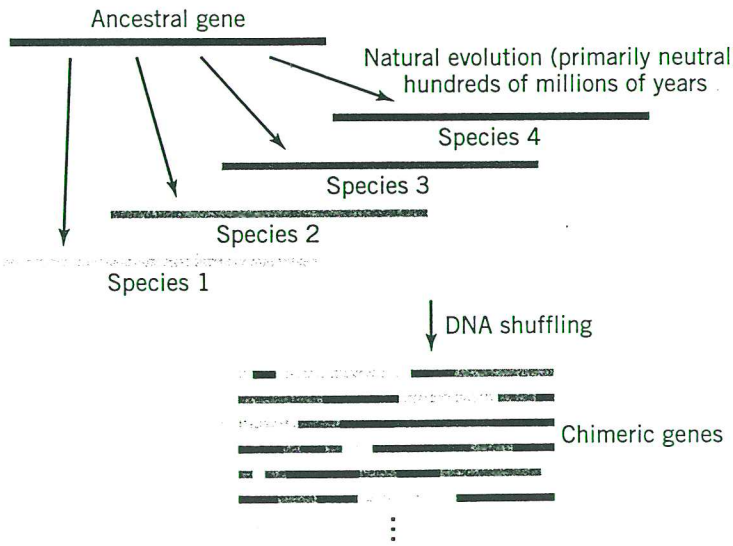

Figure 3. Homologous proteins are descended by divergent evolution from a common ancestral protein and share its overall three-dimensional structure. Recombining homologous genes creates chimeras, some fraction of which should also fold into that structure. Such family-shuffled libraries could be rich in novel function.

they may be vestigial—reflecting the enzyme's history—or they may be coupled to selected traits (16). In general, nonselected traits are easier to improve, but may be more difficult to remove. It is especially easy to improve traits never required for biological function, such as stability or activity in a nonnatural environment or activity toward a new substrate; there is often much room for improvement, and small changes in sequence and function can be accumulated. As expected for any hill-climbing exercise, the number of pathways leading uphill diminishes as a peak is approached. Thus the ease with which improved mutants are identified (the frequency of improved clones) should eventually decrease as the sequences move closer to an optimum.

The preceding discussion focuses only on altering existing enzyme traits. Evolving a completely new function is a risky venture because we rarely know how far in sequence space we have to go in order to create the new function, or how frequent the solutions will be. If there is good reason to believe that a new function, for example, activity toward a substrate not accepted at all by the wild-type enzyme, can be obtained (at a level measurable during a high-throughput screen) by making one or two amino acid substitutions, then evolution makes sense. However, if the new function requires the simultaneous placement of multiple new amino acid residues, it is unlikely to appear in a random library of mutants. Such a problem is probably a better candidate for a combination of rational design and combinatorial tuning (17).

Table 1 summarizes the key features of selected directed enzyme evolution experiments. In nearly all cases, the desired trait(s) was at a measurable, albeit low, level in the starting enzyme(s). The problems can be roughly divided into a few major categories: improving function in nonnatural or extreme environments (where activities or stabilities are low), improving activity toward a new substrate, tuning specificity (enantioselectivity), and increasing functional expression in a heterologous host.

METHODS OF DIRECTED EVOLUTION

Methods for Generating Genetic Diversity

The first step in a directed evolution experiment is the creation of a mutant library containing an appropriate degree of molecular diversity. For this, the mutation rate must be tuned to the power of the sorting method. Useful, reasonably sized libraries can be created by introducing multiple mutations in a particular region or across a limited number of positions (e.g., by combinatorial mutagenesis using oligonucleotide cassettes [42]). However, such an approach excludes the many useful solutions found in unexpected places. We therefore usually try to evolve the entire gene, rather than target particular positions. Whole-gene evolution requires a correspondingly lower mutation rate (see later).

Diversity is created by two operations: point mutagenesis and recombination (Fig. 4). Most random point mutagenesis methods create mutations in single bases. Due to the degeneracy of the genetic code, this provides access to only about 6 amino acid substitutions instead of 19, significantly reducing the potential diversity. In addition, many mutagenesis methods are not really random, further limiting the number of amino acid substitutions actually accessible in a given experiment. For example, a commonly used method, error-prone PCR, shows a strong bias for transitions over transversions (43).

Recombination methods combine sequences from multiple parent genes. In vitro recombination methods include DNA shuffling (6,7,44), random-priming recombination (45), and the staggered extension process (StEP) (46). In yeast, in vivo recombination is particularly simple (11,14). All these methods have a (controllable) level of associated point mutagenesis. Directed evolution can begin from multiple, closely related starting points rather than a single sequence. Recombination of existing functional sequences (i.e., homologous enzymes) will create another level of diversity that point mutagenesis cannot generate (9).

Point Mutagenesis. The most important factors in random mutagenesis are the mutation frequency and mutation bias. Mutation frequency is the average number of mutations per gene and is usually reported as a percentage. The optimal target mutation frequency can be calculated from the length of the DNA coding sequence and an estimate of the desired number of mutations per sequence. In deciding on the optimal number of mutations, one must balance the cost of combinatorially searching large libraries against the likelihood of finding mutants with improved properties. The number of possible variants increases drastically with the number of mutations introduced. The number of possible variants that can be produced by simultaneously substituting M amino acids in a protein of length N is $19^M[N!/(N - M)!M!]$ (47). For example there are only 3,800 possible single mutants, but 7,183,900 double mutants, and 9,008,610,600 triple mutants of a protein of only 200 amino acids. The single-mutant library of an enzyme is generally within our ability to screen exhaustively. Double-mutant libraries are already on the outer limit of current screening technology, and triple-mutant libraries are well beyond our capability for the foreseeable future. To these numerical considerations we add the biochemical fact that beneficial mutations are rare, and combinations of beneficial mutations are extremely rare. In light of these considerations it appears that, at present, single-mutant libraries usually represent the best compromise between search effort and diversity: searching single-mutant libraries maximizes the chance of finding beneficial mutations in a reasonable amount of time. If high-throughput screening technology is available, or if significant fractions of possible mutations are neutral, then double-mutant libraries may be the optimal choice. An average of one amino acid substitution per sequence corresponds to approximately three base substitutions per gene. Thus for a 1-kb sequence, the mutation frequency should be ~0.3–0.5%.

Although numerous methods for making random DNA mutations exist, error-prone PCR is often preferable because the procedure is simple, rapid, robust, and most importantly, the mutation frequency can be precisely controlled. Error-prone PCR does have some intrinsic bias as to the location of mutations (mutations at AT base pairs

Table 1. Selected Examples of Directed Enzyme Evolution

Target enzyme	Target function	Change effected	Approach	Organism	Reference
Kanamycin nucleotidyltransferase	Thermostability	>200-fold increase in half-life at 60–65 °C	Mutator strain + selection	*B. stearothermophilus*	18
Subtilisin E	Activity in organic solvents	~170-fold increase in −60% dimethylformamide	Error-prone PCR + screening	*B. subtilis*	5,19
β-Lactamase	Activity toward new substrate	32,000-fold greater resistance to cefotaxime	DNA shuffling + selection	*E. coli*	6
Subtilisin BPN′	Stability in the absence of Ca^{2+}	1,000-fold increase in half-life	Loop removed +casette mutagenesis + screening	*B. subtilis*	20
para-Nitrobenzyl esterase	Activity toward pNB esters; activity in organic solvent	60 to 150-fold increase	Error-prone PCR and DNA shuffling + screening	*E. coli*	21,22,23
Thymidine kinase	Substrate specificity (gene therapy)	43-fold increase in sensitivity to gancyclovir in hamster cells	Cassette mutagenesis + selection and screening	*E. coli*	8
β-Galactosidase	Activity toward new substrate; substrate specificity	66-fold increased activity; 1,000-fold increase in substrate specificity	DNA shuffling + screening	*E. coli*	24
Subtilisin E	Expression level; activity in organic solvents	500-fold increase in total activity	Error-prone PCR + screening	*B. subtilis*	25
*O*6-Alkylguanine-DNA alkyltransferase	Protection against alkylating agents (gene therapy)	10-fold increased protection against toxic methylating agent	Cassette mutagenesis + selection	*E. coli*	26
Arsenate detoxification pathway	Arsenic resistance	12-fold increased rate of arsenate reduction	DNA shuffling + screening	*E. coli*	27
Aminoacyl-tRNA synthetase	Aminoacylation of a modified tRNA	55-fold increase in activity	DNA shuffling + selection	*E. coli*	28
Aspartate aminotransferase	Activity toward β-branched amino and 2-oxo-acids	10^5 increase	DNA shuffling + selection	*E. coli*	29
Lipase	Wash performance	Improved performance in one-cycle wash	Mutagenesis and in vivo recombination + screening	*S. cerevisiae*	30
Lipase	Enantioselectivity in hydrolysis of *p*-nitrophenyl 2-methyldecanoate	Increase in enantiomeric excess from 2% to 81%	Error-prone PCR + screening	*Pseudomonas aeruginosa* PAOI	31
Lipases	Activity toward long chain *p*-nitrophenyl esters	3-fold increase	In vivo recombination of homologous genes + screening	*E. coli*	13

pNB esterase	Thermostability	14 °C increase in T_m + increased activity at all temperatures	Error-prone PCR, DNA shuffling + screening	*E. coli*	15
Esterase	Enantioselectivity of hydrolysis of a sterically hindered 3-hydroxy ester	Increase in enantiomeric excess from 0% to 25%	Mutator strain + selection	*E. coli*	32
Subtilisin E	Thermostability	17 °C increase in T_m + increased activity at all temperatures	Error-prone PCR, DNA shuffling + screening	*B. subtilis*	33
Subtilisin E	Thermostability	50-fold increase in half-life at 65 °C	DNA shuffling + screening	*B. subtilis*	34
B. lentus subtilisin	Expression level (total activity of secreted enzyme)	50% increase	Error-prone PCR + enrichment in hollow fibers	*B. subtilis*	35
Subtilisin BPN′	Activity at 10 °C	2-fold increase	Chemical mutagenesis + screening	*B. subtilis E. coli*	36
3-Isopropylmalate dehydrogenase	Thermostability	3.4-fold increase in activity at 70 °C	Spontaneous mutations + selection	*T. thermophilus*	37
Cephalosporinases	Activity toward moxalactam	270 to 540-fold increased resistance	DNA shuffling of homologous genes + selection	*E. coli*	9
Chorismate mutase	Conversion to monomeric enzyme (solubility)	Functional monomeric enzyme	Oligonucleotide-directed codon mutagenesis + selection	*E. coli*	17
Biphenyl dioxygenases	Degradation of polychlorinated biphenyls (PCBs)	Gained activity toward substrates, poorly degraded by native enzymes, improved activity toward various substrates	DNA shuffling of homologous genes + screening	*E. coli*	38
FLP recombinase	In vivo recombination efficiency at elevated temperatures in *E. coli* and mammalian cells; in vitro thermostability	Improved recombination efficiency in *E. coli* and mammalian cells	Error-prone PCR and DNA shuffling + screening	*E. coli*	39
Echinocandin B deacylase	Activity	3-fold increase	Error-prone PCR, random-priming recombination and screening	*Streptomyces lividans*	40
Horseradish peroxidase	Functional expression in *E. coli*	11-fold increase in expressed activity	Error-prone PCR and screening	*E. coli*	41

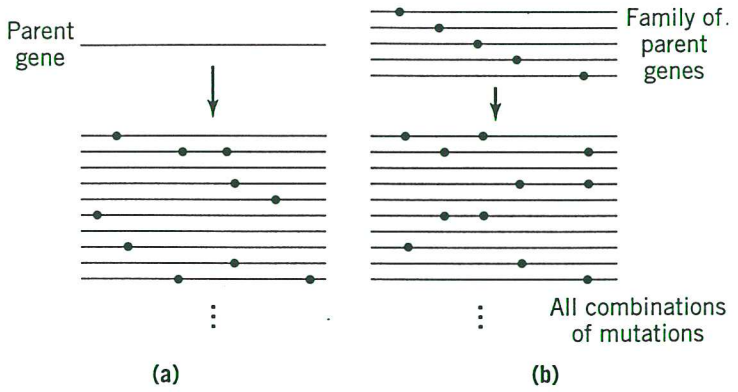

Figure 4. Two general approaches for generating molecular diversity: random point mutagenesis (**a**) and in vitro recombination (**b**).

occur much more frequently than mutations at GC base pairs) as well as their type (A is more frequently substituted with G). PCR modifications reduce bias, but do not eliminate it (48).

The error-prone PCR method routinely used was originally outlined by Leung et al. (49) and further examined by Cadwell and Joyce (43) and Shafikhani et al. (50). Mutation frequencies ranging from 0.11 to 2% have been obtained under different reaction conditions. In particular, the mutation frequency can be controlled simply by adjusting the concentration of manganese ions in the reaction mixture (48).

Because polymerase fidelity also depends on the nature of the target sequences, different mutation rates will be observed on different sequences, even when the exact same reaction conditions are used. A straightforward way to assess the overall mutation frequency and the nature of mutations is to sequence a few random clones from the amplified population. However, sequencing is time-consuming and expensive. A simple and efficient alternative is to estimate the mutation frequency from the fraction of active clones from the amplified population (51). Activity profiles of clones sampled from libraries of subtilisin E variants produced by error-prone PCR with different $MnCl_2$ concentrations are plotted in Figure 5. The clones are sorted and plotted in descending order to create local profiles, or landscapes, of the enzyme's fitness. The higher the Mn^{2+} concentration, the higher the mutation rate, which manifests itself in a higher proportion of inactivated clones. Even though different enzymes show different activity profiles for different mutation rates, the fraction of active clones is a convenient index of mutation frequency and can be used as a diagnostic check for the successful creation of the desired randomly mutated library.

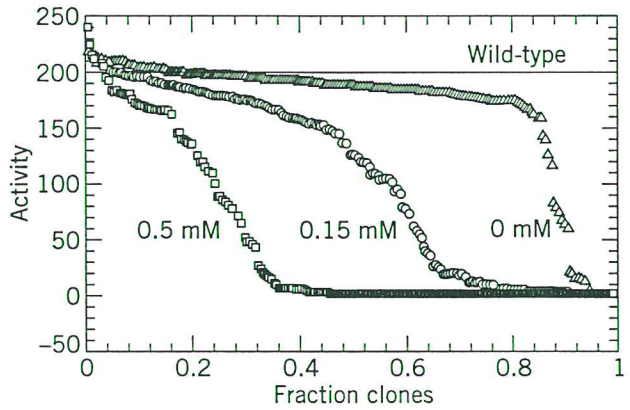

Figure 5. Activity profiles of subtilisin E mutant libraries generated by error-prone PCR with different concentrations of $MnCl_2$. Mutants are sorted according to activity in descending order. Horizontal line indicates the activity of the wild-type enzyme. *Source*: Ref. 51.

Recombination. When an enzyme is evolved by sequential generations of random mutagenesis and screening, only the best variant identified in each generation is used to parent the next generation. Other improved variants are set aside and must be rediscovered in subsequent generations in order to become incorporated. This is wasteful because screening is time-consuming. If the mutation rate is high, this approach can also accumulate deleterious mutations, possibly limiting the fitness that can be reached. "DNA shuffling" (6,7,44) and related in vitro recombination methods (45,46,48) can overcome these limitations of the sequential approach. Recombining parental genes to produce libraries of different mutation combinations, and screening for improved variants quickly accumulates the beneficial mutations, while removing deleterious ones.

The goal is to create gene libraries containing novel combinations of mutations present in the parent genes (Fig. 4). This library can be screened in order to find the combinations of mutations giving rise to the best enzymes. Screening eliminates unfavorable combinations of mutations. For example, if two point mutations that are beneficial by themselves become harmful when combined, the variant carrying this combination will be removed during screening. We now discuss some general considerations for recombination experiments before describing different recombination methods.

Statistics of Random Recombination. When recombining genes from multiple improved variants it is important to consider the recombined library size. The various in vitro recombination methods can recombine any number of parent genes. However, the resulting libraries may be so impossibly large that the probability of finding large improvements in function is effectively zero.

For the recombination of N sequences and M total mutations, the probability of generating progeny sequences containing μ mutations is equal to the number of ways a μ mutation sequence can be generated (C_μ^M) multiplied by the probability of generating any single μ-mutation sequence (22)

$$P_\mu = C_\mu^M \left(\frac{1}{N}\right)^\mu \left(\frac{N-1}{N}\right)^{M-\mu}$$
$$= \frac{M!}{(M-\mu)!\mu!} \left(\frac{1}{N}\right)^\mu \left(\frac{N-1}{N}\right)^{M-\mu}$$

For recombination of N single mutants the probability of generating single-mutant parents or wild-type grandparents is approximately 75%, and only 25% of the library consists of new sequences. The probability of generating individual sequences decreases precipitously with increasing numbers of parents. The least-frequent sequences (often the most desirable) are those containing the majority of mutations from the parent population. The rarest sequence will be the one containing all mutations ($\mu = M$). The probability P_M of generating this sequence is $1/N^m$.

Some degree of oversampling is required in practice to maximize the chance of discovering a given variant. The sampling S required in order to achieve a given level of confidence of having sampled the rarest variant in a library is given by (22)

$$(1 - P_M)^S < 1 - \text{(confidence limit)}$$

Generally, for 95% certainty that a specific clone has been sampled, the oversampling is between 2.6 and 3.0.

Recombination Methods. Stemmer described the first method for in vitro recombination of DNA sequences, called DNA shuffling (6,44). This method, illustrated in Figure 6, involves enzymatic digestion of the parental DNA into short fragments, followed by reassembly of the fragments into full-length genes. Because the fragments are free to associate with complementary fragments from other, similar genes, mutations from one parent can be combined with mutations from other parent(s) to generate novel combinations. The fidelity of this process is not perfect, and new point mutations not present in any of the parents will occur at a finite rate. The original method has been modified to simplify it and to yield better control over the associated mutagenic rate (48,52). The mutagenic frequency can be controlled over a wide range, from about 0.05 to 0.7%, by the inclusion of Mn^{2+} or Mg^{2+}, by the choice of DNA polymerase, and/or by using restriction enzyme digestion to prepare the starting DNA (48). The finite error frequency associated with DNA shuffling has been used by Stemmer and coworkers to supply point mutations for recombination and evolution. When starting with a single sequence, we prefer to use error-prone PCR under controlled conditions to generate libraries of variants. Improved sequences identified during screening can then be recombined.

An alternative method for recombination was developed by Shao et al. (45). In this method the template DNA sequences are primed with random-sequence primers, which are then extended by DNA polymerase to generate a pool of short DNA fragments. After removal of the template, these fragments can be reassembled, as with the Stemmer shuffling method, to give a library of full-length sequences containing information from multiple parents. Yet another method for in vitro recombination, the staggered extension process (StEP), does not require fragmentation and reassembly of the parent sequences (34). This method is illustrated in Figure 7 for recombination of two parent genes. In StEP recombination the template genes are primed and very briefly extended before denaturation and reannealing. The growing fragments can reanneal to different templates and therefore pick up sequence information from different parents as they grow into full-length sequences over hundreds of very brief cycles. StEP (in principle) involves only a single relatively short protocol as opposed to the multiple steps of fragmentation, isolation, and annealing required in DNA shuffling.

"Family Shuffling": Recombination of Homologous Genes. The approaches already described (point mutation and recombination of improved variants) allow a local exploration of sequence space. Only sequences that are quite close to the parent sequence are sampled. One would like to sample distant regions of sequence space to search for new functions, but this is not readily accomplished using these methods. The simultaneous introduction of more than a few random mutations into a wild-type sequence will almost always result in variants that are inactive or unfolded. The probability that an improved enzyme will be found is usually much smaller than the screening capacity. An alternative approach to making large jumps in sequence space is by "family shuffling" (9) (see Fig. 3). In this approach, homologous genes from different organisms are recombined to create a library of chimeric molecules. Family shuffling exploits the fact that the sequence differences between two homologous proteins are *not* random. Consider the two proteases subtilisin E and subtilisin thermitase. These two enzymes differ at 157 of their amino acid positions, and yet they fold to the same overall three-dimensional structure and catalyze the same proteolytic reaction. The amino acid substitutions in these and other naturally occurring subtilisins have been preselected to be largely neutral with respect to folding and gross function. Mutations resulting in a misfolded or inactive enzyme have been eliminated by natural selection. Accordingly, recombination of this pool of largely neutral mutations will be far more likely to result in folded, active proteins than recombination of purely random mutations. The use of family shuffling to access remote unexplored regions of sequence space may yield proteins with a variety of desirable, but not yet found, functions.

Searching for Improved Enzymes

In general, the most time-consuming and expensive part of directed enzyme evolution is setting up and implementing the search for improved variants. Developing an optimal search strategy is crucial for success. A fundamental rule of directed evolution, "you get what you screen for," specifies that the screen (or selection) should reflect the desired result as closely as possible.

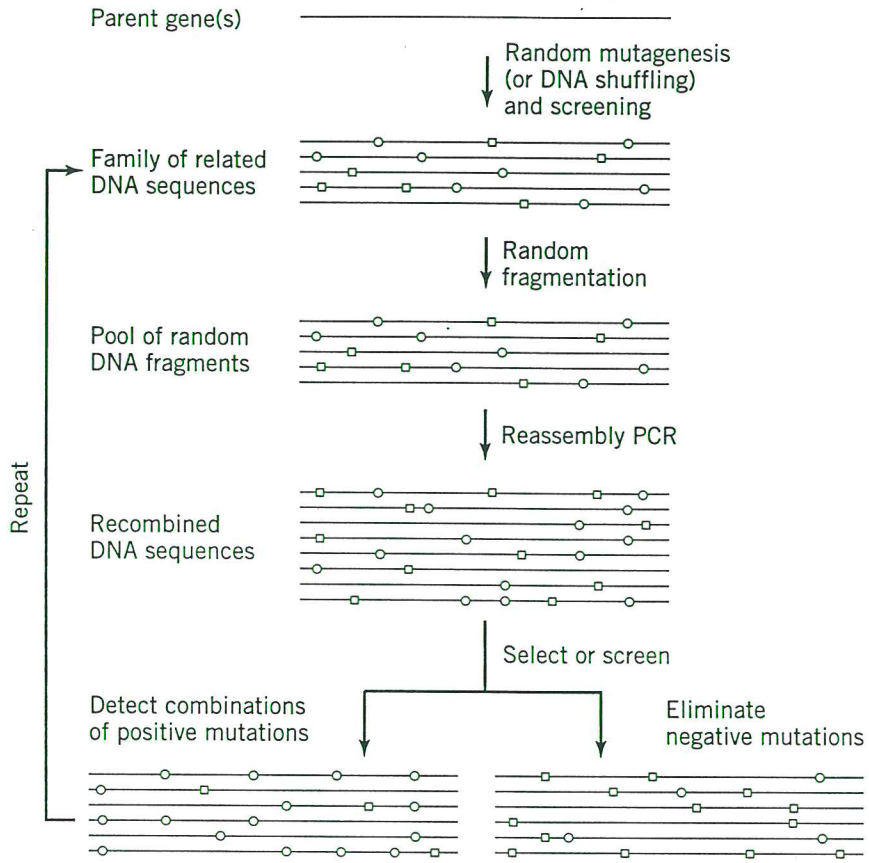

Figure 6. Recombination of parent genes by DNA shuffling. *Source*: Ref. 44.

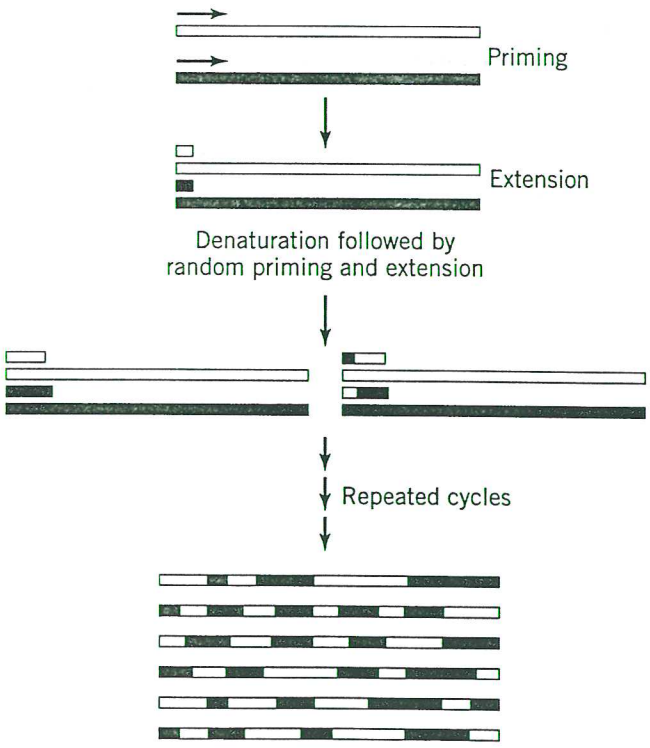

Figure 7. Recombination of two parent genes by StEP. *Source*: Ref. 34.

Mutant libraries are searched by screening or selection. In a genetic selection, the ability of an enzyme to perform the desired function is directly coupled to the survival and reproduction of the host organism. A straightforward example is selection for drug resistance: clones expressing an enzyme capable of degrading an antibiotic are selected for their ability to grow on antibiotic-containing plates (53,54). Another example is the selection of functional mutants of a human methyltransferase. Active mutants were selected based on their ability to protect alkyltransferase-deficient *E. coli* cells from a methylating agent (26). Such approaches are attractive in principle because they allow one to search larger libraries than does screening (10^6–10^7 for selection, as opposed to 10^4–10^5 for screening). Kast and Hilvert (55) recently reviewed genetic selections in directed evolution.

However, selections suffer serious drawbacks. The types of new properties that can be searched for are limited. For example, enzymes tolerate a number of environments that cannot sustain life (e.g., organic solvents). It becomes extremely difficult to tie the desired function to the survival or growth of an organism when the organism cannot survive. The main drawback, however, is the complexity of the organisms themselves and the surprising solutions that spontaneously arise during adaptation. Given our limited understanding of even simple, well-studied organisms such as *E. coli*, it is difficult to ensure that improving a particular enzyme is the *only* way for an organism to adapt to the selective pressure applied. One may go

through several rounds of selection only to find that the organism has found some alternate way to solve the problem.

Various in vitro selections have been proposed based on column binding or "panning" (56). In these methods, the proteins are displayed on the surface of a phage or a cell. Very large libraries ($>10^9$) can be searched based on the ability of the displayed protein to bind to a substrate immobilized on a column. The greatest limitation to this approach is that it is not generally applicable to screening for properties other than binding.

For many problems, and especially those of practical interest, libraries of variants must be screened rather than selected. A typical screening strategy involves the construction of an arrayed enzyme library and application of a rapid assay of the desired property. The screen can be more or less sensitive, depending on the willingness of the researcher to accept false positives (and to apply additional tests) (57).

Prerequisites for an effective screening strategy include the following:

1. The screen should be sufficiently rapid that large numbers of clones can be searched in a reasonable amount of time. It should be feasible to screen at least a few hundred, and preferably a few thousand, in a day. For obvious reasons, the screen should not require purification of the enzyme. A second-level assay can eliminate false positives.
2. Only rarely do very large improvements result from a single amino acid substitution. Directed evolution most often succeeds by accumulating a number of modest improvements over several generations. Therefore a screen must be accurate enough to allow modest improvements to be identified against the background variation.
3. The phenotype (the enzyme property that is being screened) must be physically coupled to the genotype (the DNA or RNA sequence coding for the enzyme). This is most often accomplished trivially by the fact that proteins are expressed in host organisms: the same cell expressing the enzyme contains the gene coding for its sequence. Coupling can also be accomplished by display on the surface of a virus containing the genetic information in the form of DNA or RNA (58). Recently developed in vitro translation systems such as ribosome display (59) or RNA–peptide fusions (60) link the enzyme directly to the genetic information in the form of RNA without using cells. Tawfik and Griffiths (61) have encapsulated the in vitro translation apparatus and genes within reverse micelles.

Simple visual screens are widely used when the function of interest can generate a visible signal. Various oxidases, for example, produce hydrogen peroxide as a by-product of their oxidation reaction. As illustrated in Figure 8, coupling this to a second enzyme that is easily assayed colorimetrically (horseradish peroxidase) provides a visual screen for the oxidase activity. Clones secreting active proteases produce a zone of clearing or "halo," the size of which is proportional to the hydrolytic activity when the organism is grown on agar plates containing casein or skim milk proteins. This simple method has been used extensively in screening protease libraries (25,62).

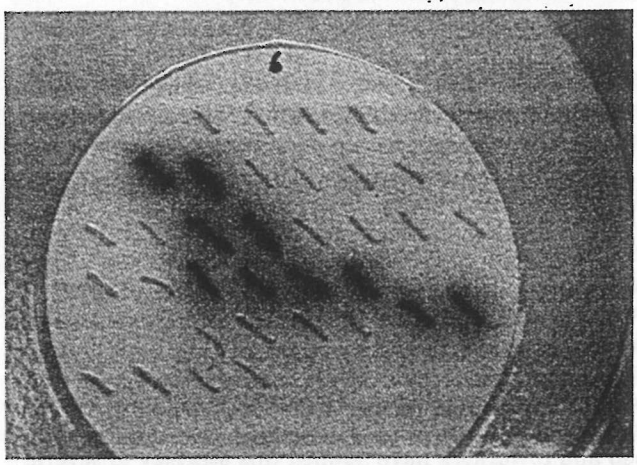

Figure 8. Example data from a visual screen for galactose oxidase activity toward D-galactose. Colonies are grown on agar plates and transferred to a membrane. Active colonies generate a purple color when incubated with 2,2′-azinobis(3-ethylbenzothiazline-6-sulfonic acid), horseradish peroxidase, and D-galactose. *Source*: M. Yagasaki, unpublished results, 1998.

Although visual screening based on color or halo formation is rapid and efficient, it is also nonquantitative and often relatively insensitive. Digital imaging spectroscopy has been developed to increase the sensitivity and throughput of filter- and agar-plate-based screens (63). However, the 96-well microtiter plate remains the standard format for automated, high-throughput screening. Screening automation and quantification of the results are highly desirable; 96- and 384-well plates appear to be the format most compatible with currently available robotic arms, liquid handling systems, and plate readers. Data collection and analysis are greatly facilitated by computerized data acquisition (computerized data are easily sorted into local enzyme fitness landscapes such as those shown in Fig. 5). Computerized data acquisition will certainly facilitate exploitation of the enormous amounts of information available in protein libraries. Many conventional enzyme assays can be readily converted into automated formats.

If activity against the desired substrate does not result in an easily detectable signal (such as a change in absorbance or fluorescence), it is sometimes possible to screen against a surrogate substrate that does. Naturally, this approach requires caution. It must be verified that activity against the replacement substrate reflects activity against the real substrate under the conditions of interest. A successful application of this approach is described in the example provided later of the evolution of a *p*-nitrobenzyl esterase (21).

Enzyme properties other than activity are often targeted for improvement. A convenient approach to screening for thermostable variants of irreversibly inactivated

enzymes is shown in Figure 9. Duplicate 96-well plates are made from a single master plate. One is incubated at high temperature for a fixed period. Clones in both are then assayed for activity, and the ratio of activity after heating to activity without heating provides a measure of enzyme stability at the temperature of incubation.

An attractive approach to screening large libraries is to couple functional complementation with screening. By requiring that at least some of the biological function of the protein is retained, functional complementation can greatly reduce the subsequent screening requirements. The initial selection asks an essentially binary question (is there any activity? yes or no), and only in the later screening stage are differences in the level of activity detected. Loeb and coworkers have applied this strategy to several enzymes of interest in gene therapy and other applications (8,26). However, the major limitation is finding or constructing an appropriate complementation system. Furthermore, retention of biological function may preclude acquisition of other functions.

Extracting Useful Information from the Results of Evolution

Two major types of information can be extracted from the results of directed evolution experiments. The first and most obvious is the information contained in the sequences of the evolved proteins. The second consists of the overall activity profiles, or local fitness landscapes, that can be generated by analysis of the entire screened library.

Clearly, we would like to identify the specific amino acid substitutions that are responsible for conferring new properties. In contrast to comparisons of evolutionarily related proteins found in nature, this identification can be accomplished relatively easily in directed evolution experiments. In nature, related enzymes that are separated by millions of years of evolution will generally have accumulated a large number of neutral mutations in addition to those that are responsible for changes in specific properties. In contrast, in laboratory evolution we strive to allow only functional mutations to survive and carry on to the next generation. A striking example of the difference between natural and laboratory evolution can be seen in a recent study in which subtilisin E was converted into a functional equivalent of the thermophilic enzyme thermitase (33). Subtilisin E differs from thermitase at 157 positions, but only eight substitutions were required to turn it into a thermitase-like enzyme, with an 18 °C increase in its temperature optimum and a 250-fold increase in its half life at 65 °C.

However, it is not always possible to identify functional mutations simply by examining the sequences of the evolved enzymes. In the experimental procedures for introducing random point mutations, only the *average* error rate can be controlled. Thus, some variants will accumulate multiple mutations in a single generation. If a given variant has multiple mutations, only one of which is functional, the remaining neutral mutations can "hitchhike" their way into subsequent generations and complicate analysis of the final evolved sequences. A method for rapidly identifying functional mutations by back-crossing with wild type has been described (64).

It is hoped that detailed structural analysis of evolved enzymes will provide insight into the underlying physical–chemical mechanisms responsible for protein adaptation for new functions or environments. However, the results of single point mutations can be very subtle, and the analysis is not straightforward even when the three-dimensional structure is known. For example, both subtilisin E and pNB esterase have been successfully evolved for activity in aqueous organic solvents, but functional mutations in

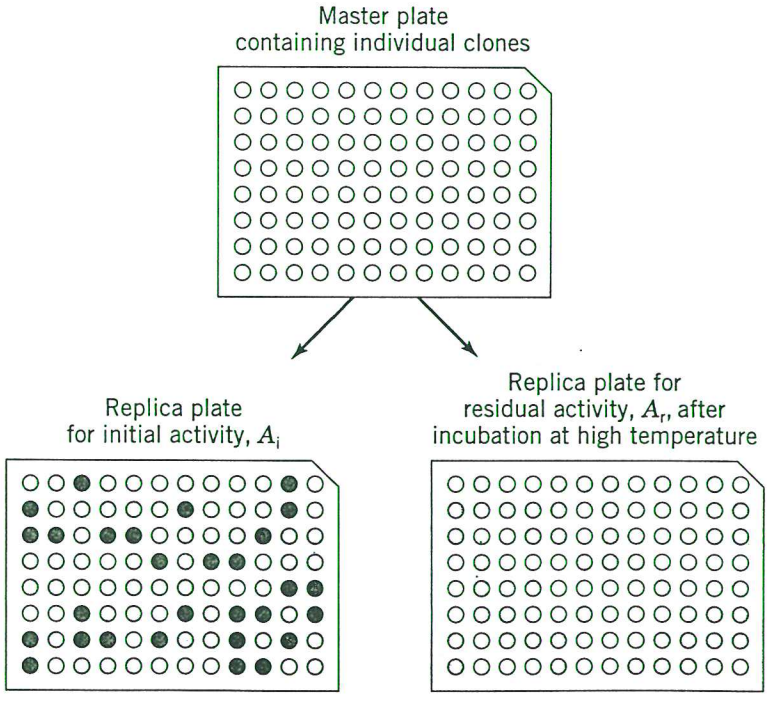

Figure 9. A rapid screen for enzyme thermostability based on catalytic activity before and after incubation at high temperature. *Source*: Ref. 34.

the evolved enzymes follow no discernible pattern: large residues are replaced by small, small by large, charged by uncharged, uncharged by charged (5,21,25). Further, mutations are found both close to and far from the enzyme active sites. Numerous studies now attest to the subtlety of the effects of amino acid substitutions on enzyme structure and properties (8,24,38,39).

The distribution of fitnesses (defined in terms of activity, stability, etc.) in a mutant library can provide useful information for designing evolutionary strategies. One way to access this information is by presenting the data in a local fitness landscape such as those shown in Figure 5. As mentioned previously, the mutation rate can be estimated from the fraction of inactive mutants. More generally, such a plot represents (at least for single- or double-mutant libraries) the distribution of fitness in the region of sequence space immediately surrounding the wild-type sequence. Most mutants will be equally or less fit than the wild type; a small number may be better. The fraction of improved mutants can be used to estimate the evolvability of the property of interest. When two or more properties are being evolved simultaneously, plots of one versus the other can be informative. Typically, single mutations that simultaneously improve two independent properties are extremely rare. However, individuals may be recombined to yield an enzyme with multiple improved properties (Fig. 10) (65).

EXAMPLES OF DIRECTED EVOLUTION APPLICATIONS

Evolution of Lipases with High Activity in Detergent Solutions

New lipase variants that give greatly improved residual lipid removal after one wash cycle have been developed by directed evolution (30). Lipases are used as detergent enzymes to remove lipid or fatty stains from clothes and other textiles. A drawback of the known detergent lipases is that they exert the best lipid-removing effect after more than one wash cycle, presumably because they are more active during a certain period of the drying process than during the wash itself. Therefore at least two wash cycles, separated by a sufficiently long drying period, are required to remove fatty stains.

An intense protein engineering effort involving both site-directed and random mutagenesis yielded a large number of variants of *Thermomyces lanuginosa* lipase that were strongly improved in multicycle wash performance, but not significantly improved in one-cycle wash tests. Random mutant libraries were screened using filter assays containing detergent and low calcium concentrations, to mimic wash conditions. Consecutive rounds of mutagenesis and screening with increasing amounts of detergent were performed.

New variants with a strong one-cycle wash effect were obtained by recombining 20 variants with the best performance in multicycle wash tests and screening. The recombination was performed using a simple in vivo method in *S. cerevisiae* in which PCR fragments from the 20 variants mixed with the opened vector are used to transform competent *S. cerevisiae* cells. Upon transformation, the vector and fragments recombine and shuffle the variants (14). Screening yielded seven variants with a strong one-cycle wash effect; the best of these is now commercially available.

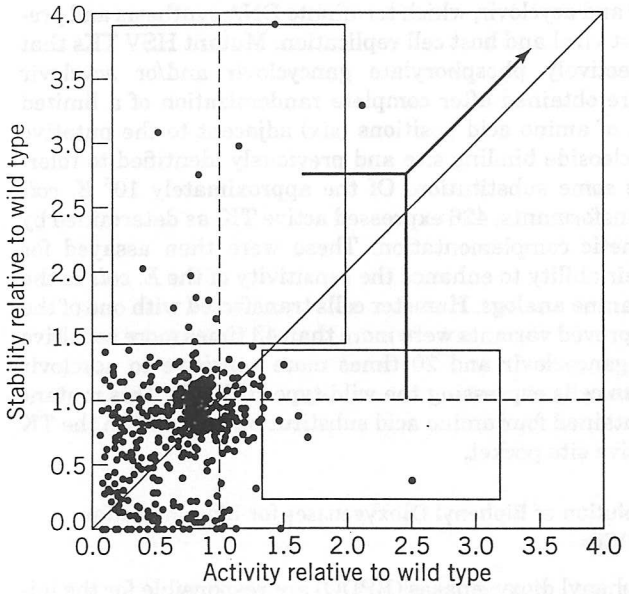

Figure 10. Activity and stability to hydrogen peroxide of a randomly mutagenized enzyme library. The individual populations shown in boxes can be evolved in parallel and recombined to obtain enzymes that are both more active and more stable. *Source*: Reproduced with permission from Ref. 65.

Evolution of an Esterase for Deprotection of *p*-Nitrobenzyl Esters and Enhanced Thermostability

By screening microbial cultures, scientists at Eli Lilly identified an enzyme in *Bacillus subtilis* that could remove the *p*-nitrobenzyl ester protecting group used during the large-scale synthesis of certain β-lactam antibiotics. However, the enzyme's relatively poor activity, particularly in the solvents required to solubilize the ester substrate, made it a poor competitor to the chemical catalyst for deprotection of a loracarbef synthetic intermediate. The natural function of *B. subtilis* pNB esterase is unknown.

Moore and Arnold were able to evolve highly active pNB esterases that also function well in mixed aqueous–organic solvents (21,22). A surrogate, chromogenic substrate (the antibiotic *p*-nitrophenyl ester) was used in a rapid screen to identify potential positives. Variants identified during the rapid screen could be verified during a second-level screen on the *p*-nitrobenzyl ester using HPLC. Four generations of PCR mutagenesis and two rounds of recombination by DNA shuffling yielded a clone with more than 100 times the total activity of wild type in 15–20% dimethylformamide (DMF) toward loracarbef-*p*-nitrobenzyl ester. The enzyme's catalytic efficiency increased more than 50-fold (Fig. 11). The total activity toward the screening substrate (loracarbef-*p*-nitrophenyl ester) increased more than 150-fold. Although the contributions of individual effective amino acid substitutions to enhanced activity were small (usually approximately two-fold increases in activity), the accumulation of multiple mutations over a num-

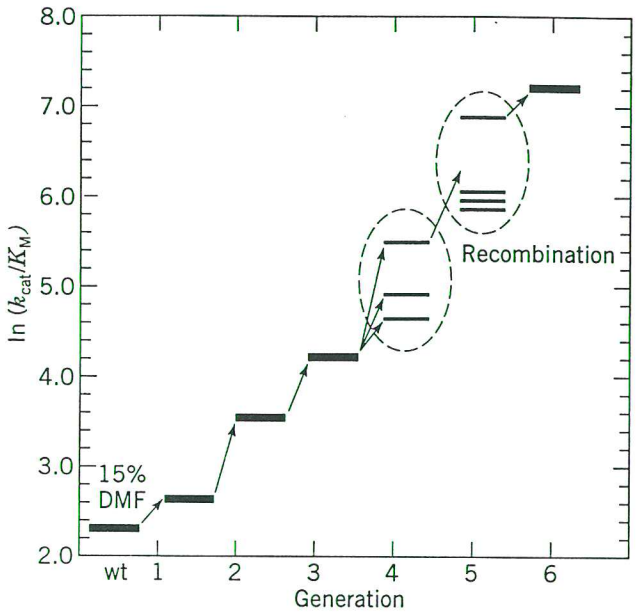

Figure 11. Evolutionary progression of catalytic efficiency of pNB esterase (toward LCN-pNB) in 15% DMF through four generations of random mutagenesis, followed by two rounds of recombination of circled populations. *Source*: Ref. 66.

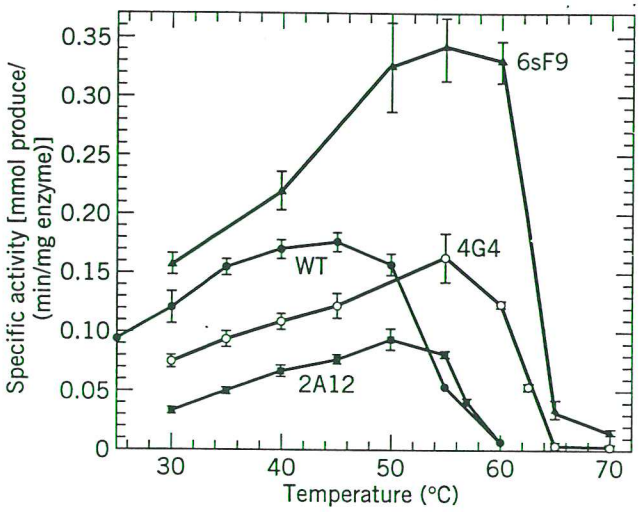

Figure 12. Activities of wild-type and evolved pNB esterases as a function of temperature. *Source*: Ref. 15 and A. Gershenson, unpublished results.

ber of generations significantly improved the biocatalyst for this nonnatural reaction. Moore and Arnold concluded that none of the mutations accumulated in the evolved enzyme are in direct contact with the substrate. Some are as far away as 20 Å. Limiting the search for possible solutions to residues that line the substrate binding site would have overlooked important beneficial mutations (and may not have succeeded).

Directed evolution has also significantly increased the thermostability of the *B. subtilis* pNB esterase (15). A simple screen was developed based on retention of catalytic activity after incubation at high temperature. Positives were then subjected to differential scanning calorimetry (DSC) to verify an increase in melting temperature (T_m) relative to the parent enzyme. Accumulating mutations over eight generations of random mutagenesis and DNA shuffling yielded an increase in T_m of 17 °C (15, A. Gershenson, unpublished results). This large increase in thermostability is equivalent to the difference between proteins from mesophiles and many thermophilic organisms. Furthermore, because only those variants that retained their catalytic activity as well were chosen, the increased thermostability was accompanied by a very significant increase in enzyme activity at elevated temperatures (Fig. 12). Reflecting the fact that the thermostability screen involved activity toward *p*-nitrophenyl acetate, the resulting enzymes are highly active toward this substrate and less active toward the antibiotic substrate and *p*-nitrobenzyl esters. The most thermostable pNB esterase variant has 13 amino acid substitutions (out of 490).

Evolution of Herpes Simplex Virus Thymidine Kinase for Gene Therapy

The Loeb group has pioneered the use of directed evolution for developing improved enzymes for cancer gene therapy. Here, greater activity toward nonnatural substrates is desired, with the goal of selectively protecting bone marrow cells (26) or sensitizing tumor cells (8) to toxic agents. This group consistently directs intense mutagenesis of a limited region of the gene and genetic selection to identify the sequences that are still biologically functional. Screening this much smaller subset of sequences identifies those with the desired properties (altered substrate specificity, ability to protect or sensitize cells).

In one study (8), directed evolution was used to create herpes simplex virus thymidine kinase (HSV TK) variants that enhance cell killing by the guanine analogs gancyclovir and acyclovir, which terminate DNA synthesis and prevent viral and host cell replication. Mutant HSV TKs that selectively phosphorylate gancyclovir and/or acyclovir were obtained after complete randomization of a limited set of amino acid positions (six) adjacent to the putative nucleoside binding site and previously identified to tolerate some substitution. Of the approximately 10^6 *E. coli* transformants, 426 expressed active TK, as determined by genetic complementation. These were then assayed for their ability to enhance the sensitivity of the *E. coli* to the guanine analogs. Hamster cells transfected with one of the improved variants were more than 43 times more sensitive to gancyclovir and 20 times more sensitive to acyclovir than cells expressing the wild-type HSV TK. This mutant contained four amino acid substitutions, all within the TK active site pocket.

Evolution of Biphenyl Dioxygenases for Bioremediation of PCBs

Biphenyl dioxygenases (BPDO) are responsible for the initial oxidation of polychlorinated biphenyls (PCBs) during their biodegradation by various organisms. In an effort to create BPDOs with enhanced ability to oxidize a wider range of PCB congeners, Furukawa and coworkers used DNA shuffling to create a chimeric library from two genes

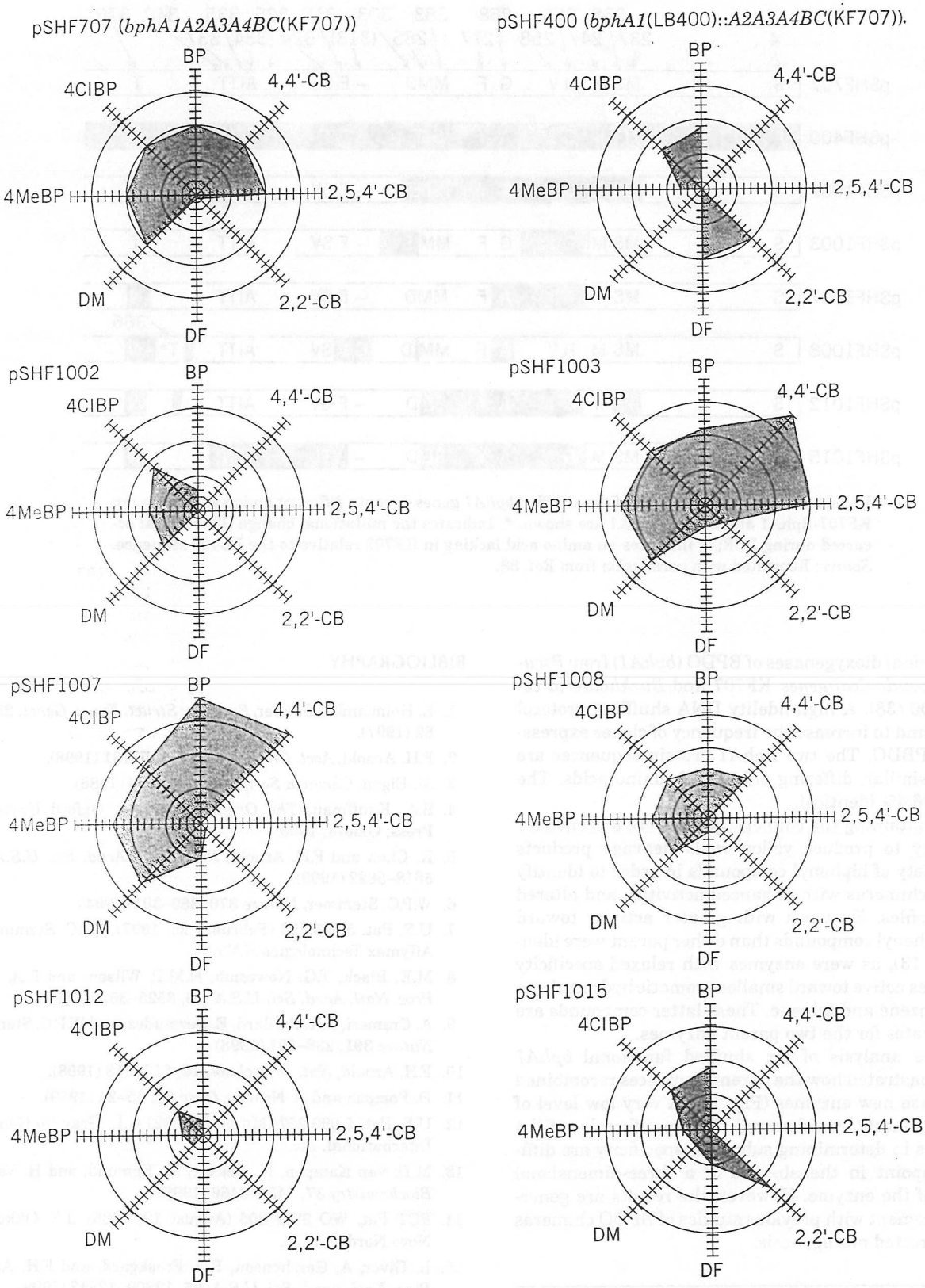

Figure 13. Formation of *meta*-cleavage yellow products from a variety of biphenyl compounds by *E. coli* expressing chimeric biphenyl dioxygenases. BP, biphenyl; 4CIBP, 4-chlorobiphenyl; 2.2′-CB, 2,2′-dichlorobiphenyl; 4,4′-CB, 4,4′-dichlorobiphenyl; 2,5,4′-CB, 2,5,4′-trichlorobiphenyl; MeBP, 4-methylbiphenyl; DM, diphenylmethane; DF, dibenzofuran. The relative values of yellow products formed from each compound are plotted for *E. coli* strains carrying the evolved enzymes. *Source*: Reprinted with permission from Ref. 38.

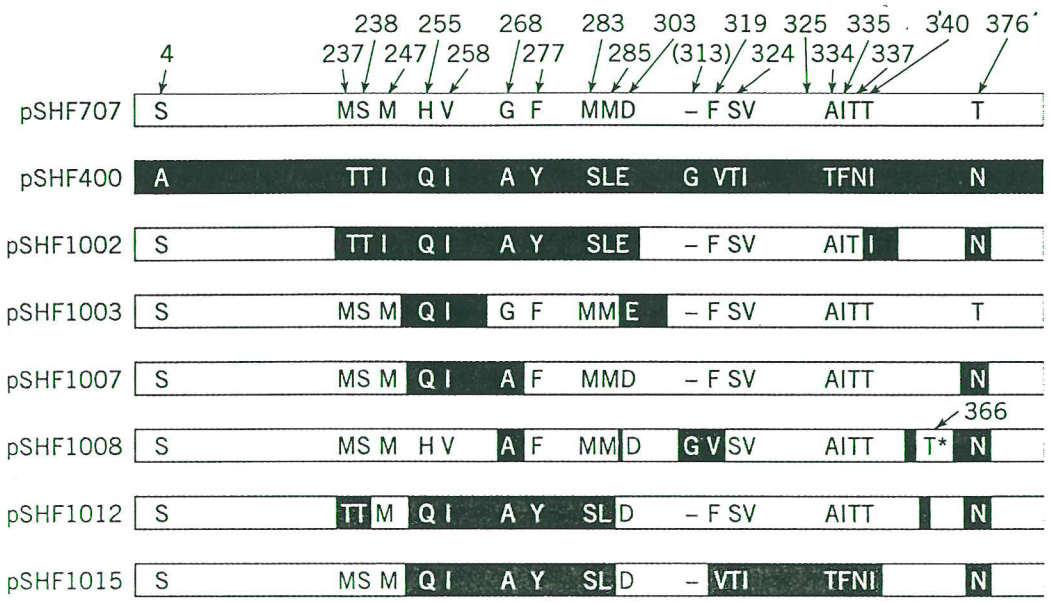

Figure 14. Sequence analysis of the shuffled *bphA1* genes. Twenty different amino acids between KF707-BphA1 and LB400-BphA1 are shown. *, Indicates the mutational change (I366T) that occurred during PCR; –, indicates an amino acid lacking in KF707 relative to the LB400 sequence. *Source*: Reprinted with permission from Ref. 38.

for the terminal dioxygenases of BPDO (*bphA1*) from *Pseudomonas pseudoalcaligenes* KF707 and *Burkholderia cepacia* LB400 (38). A high-fidelity DNA shuffling protocol (48) was found to increase the frequency of clones expressing active PBDO. The two BphA1 protein sequences are extremely similar, differing only at 20 amino acids. The genes are 96.4% identical.

E. coli containing the chimeric genes were screened for their ability to produce yellow *meta*-cleavage products from a variety of biphenyl compounds in order to identify functional chimeras with enhanced activities and altered activity profiles. Enzymes with greater activity toward various biphenyl compounds than either parent were identified (Fig. 13), as were enzymes with relaxed specificity and enzymes active toward smaller aromatic hydrocarbons such as benzene and toluene. These latter compounds are poor substrates for the two parent enzymes.

Sequence analysis of six shuffled functional *bphA1* genes demonstrated how the parent sequences recombined to yield these new enzymes (Fig. 14). A very low level of point mutagenesis was observed. The specific roles of the amino acids in determining substrate specificity are difficult to pinpoint in the absence of a three-dimensional structure of the enzyme. However, the results are generally in agreement with previous studies of BPDO chimeras and site-directed mutagenesis.

ACKNOWLEDGMENTS

F.H.A. wishes to thank the many talented undergraduate, graduate, and postdoctoral students who have contributed to directed evolution research at Caltech. The financial support of the Office of Naval Research, the Department of Energy, and the Army Research Office is gratefully acknowledged.

BIBLIOGRAPHY

1. L. Holm and C. Sander, *Proteins: Struct. Funct. Genet.* **28**, 72–82 (1997).
2. F.H. Arnold, *Acct. Chem. Res.* **31**, 125–131 (1998).
3. M. Eigen, *Chemica Scripta* **26B**, 13–26 (1986).
4. S.A. Kauffman, *The Origins of Order*, Oxford University Press, Oxford, 1993.
5. K. Chen and F.H. Arnold, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 5618–5622 (1993).
6. W.P.C. Stemmer, *Nature* **370**, 389–391 (1994).
7. U.S. Pat. 5,605,793 (February 25, 1997) W.P.C. Stemmer (to Affymax Technologies N.V.).
8. M.E. Black, T.G. Newcomb, H.M.P. Wilson, and L.A. Loeb, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 3525–3529 (1996).
9. A. Crameri, S.A. Raillard, E. Bermudez, and W.P.C. Stemmer, *Nature* **391**, 288–291 (1998).
10. F.H. Arnold, *Nat. Biotechnol.* **16**, 617–618 (1998).
11. D. Pompon and A. Nicolas, *Gene* **83**, 15–24 (1989).
12. U.S. Pat. 5,093,257 (March 3, 1992) G.L. Gray (to Genencor International, Inc.).
13. M.D. van Kampen, N. Dekker, M. Egmond, and H. Verheij, *Biochemistry* **37**, 3459–3466 (1998).
14. PCT Pat. WO 97/07205 (August 12, 1996), J.S. Okkels (to Novo Nordisk A/S).
15. L. Giver, A. Gershenson, P.O. Freskgard, and F.H. Arnold, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12809–12813 (1998).
16. S.A. Benner, *Chem Rev.* **89**, 789–806 (1989).
17. G. MacBeath, P. Kast, and D. Hilvert, *Science* **279**, 1958–1961 (1998).
18. H. Liao, T. McKenzie, and R. Hageman, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 576–580 (1986).

19. U.S. Pat. 5,316,935 (May 31, 1994), F.H. Arnold and K. Chen (to California Institute of Technology).
20. S.L. Strausberg, P.A. Alexander, D.T. Gallagher, G.L. Gilliland, B.L. Barnett, and P.N. Bryan, *Biotechnology* **13**, 669–673 (1995).
21. J.C. Moore and F.H. Arnold, *Nat. Biotechnol.* **14**, 458–467 (1996).
22. J.C. Moore, H.M. Jin, O. Kuchner, and F.H. Arnold, *J. Mol. Biol.* **272**, 336–347 (1997).
23. U.S. Pat. 5,741,691 (April 21, 1998), F.H. Arnold and J.C. Moore (to California Institute of Technology).
24. J-H. Zhang, G. Dawes, and W.P.C. Stemmer, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4504–4509 (1997).
25. L. You and F.H. Arnold, *Protein Eng.* **9**, 77–83 (1996).
26. F.C. Christians and L.A. Loeb, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6124–6128 (1996).
27. A. Crameri, G. Dawes, E. Rodrigues, Jr., S. Silver, and W.P.C. Stemmer, *Nat. Biotechnol.* **15**, 436–438 (1997).
28. D.R. Liu, T.J. Magliery, M. Pasternk, and P.G. Schultz, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10092–10097 (1997).
29. T. Yano, S. Oue, and H. Kagamiyama, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5511–5515 (1997).
30. PCT Pat. WO 97/07202 (February 27, 1997), J.S. Okkels, M. Thellersen, D.A. Petersen, A. Svendsen, S.A. Patkar, and K. Borch (to Novo Nordisk A/S).
31. M.T. Reetz, A. Zonta, K. Schimossek, K. Liebeton, and K.E. Jaeger, *Angew. Chem. Int. Ed. Engl.* **36**, 2830–2832 (1997).
32. U.T. Bornscheuer, J. Altenbuchner, H.H. Meyer, *Biotechnol. Bioeng.* **58**, 554–555 (1998).
33. H. Zhao and F.H. Arnold, *Protein Eng.* **12**, 47–53 (1999).
34. H. Zhao, L. Giver, Z. Shao, J.A. Affholter, and F.H. Arnold, *Nat. Biotechnol.* **16**, 258–235 (1998).
35. D. Naki, C. Paech, G. Ganshaw, and V. Schellenberger, *Appl. Microbiol. Biotechnol.* **49**, 290–294 (1998).
36. S. Taguchi, A. Ozaki, and H. Momose, *Appl. Environ. Microbiol.* **64**, 492–495 (1998).
37. S. Akanuma, A. Yamagushi, N. Tanaka, and T. Ooshima, *Protein Sci.* **7**, 698–705 (1998).
38. T. Kumamaru, H. Suenaga, M. Mitsuoka, T. Watanabe, and K. Furukawa, *Nat. Biotechnol.* **16**, 663–666 (1998).
39. F. Buchholz, P.O. Angrand, and A.F. Steward, *Nat. Biotechnol.* **16**, 657–662 (1998).
40. Z. Shao, M. Callahan, and F.H. Arnold, *Appl. Environ. Microbiol.* (in press).
41. Z. Lin, T. Thorsen, and F.H. Arnold, *Biotechnol. Prog.* (in press).
42. J. Reidhaar-Olson and R. Sauer, *Science* **241**, 53–57 (1988).
43. R. Cadwell, and G. Joyce, *PCR Methods Appl.* **3**, S136–S140 (1994).
44. W.P.C. Stemmer, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 10747–10751 (1994).
45. X. Shao, H. Zhao, L. Giver, and F.H. Arnold, *Nucleic Acids Res.* **26**, 681–683 (1998).
46. H. Zhao, L. Giver, X. Shao, J. Affholter, and F.H. Arnold, *Nat. Biotechnol.* **16**, 258–261 (1998).
47. F.H. Arnold, *Chem. Eng. Sci.* **51**, 5091–5102 (1996).
48. H. Zhao and F.H. Arnold, *Nucleic Acids Res.* **25**, 1307–1308 (1997).
49. D. Leung, E. Chen, and D. Goeddel, *Technique* **1**, 11–15 (1989).
50. S. Shafikhani, R. Siegel, E. Ferrari, and V. Schellenberger, *Biotechniques* **23**, 304–310 (1997).
51. H. Zhao, J. Moore, and F.H. Arnold, in *ASM Manual of Industrial Microbiology and Biotechnology*, ASM Press, Washington D.C., 1999, pp. 597–604.
52. I.A. Lorimer and I. Pastsn, *Nucleic Acids Res.* **23**, 3067–3068 (1995).
53. J. Petrosino and T. Palzkill, *J. Bacteriol.* **178**, 1821–1828 (1996).
54. W. Huang, J. Petrosino, M. Hirsch, P. Shenkin, and Y. Palzkill, *J. Mol. Biol.* **258**, 688–703 (1996).
55. P. Kast and D. Hilvert, *Curr. Opin. Struct. Biol.* **7**, 470–479 (1997).
56. C. Wang, Q. Yang, and C. Craik, *Methods Enzymol.* **267**, 52–68 (1996).
57. H. Zhao and F.H. Arnold, *Curr. Opin. Struct. Biol.* **7**, 480–485 (1996).
58. K. O'Neil and R. Hoess, *Curr. Opin. Struct. Biol.* **5**, 443–449 (1995).
59. J. Hanes and A. Pluckthun, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4937–4949 (1997).
60. R. Roberts and J. Szostak, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12297–12302 (1997).
61. D. Tawfik and A. Griffiths, *Nat. Biotechnol.* **16**, 652–656 (1998).
62. S. Sidhu and T. Borgford, *J. Mol. Biol.* **257**, 233–245 (1996).
63. D. Youvan, E. Goldman, S. Elgrave, and M. Yang, *Methods Enzymol.* **246**, 732–749 (1995).
64. H. Zhao, and F.H. Arnold, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 7997–8000 (1997).
65. O. Kuchner and F.H. Arnold, *TIBTECH* **15**, 523–530 (1997).
66. F.H. Arnold and J. Moore, *Adv. Biochem. Eng. Biotechnol.* **58**, 2–14 (1997).

See also ENZYMES, BAKING, BREAD MAKING; ENZYMES, DETERGENT; GOOD MANUFACTURING PRACTICE (GMP) AND GOOD INDUSTRIAL; PROTEIN EXPRESSION, SOLUBLE; PROTEIN SECRETION, *SACCHAROMYCES CEREVISIAE*.