

Directed Evolution: Novel and Improved Enzymes

Fei Wen*, *Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois*

Michael McLachlan*, *Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois*

Huimin Zhao, *Departments of Chemical and Biomolecular Engineering and Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois*

doi: 10.1002/9780470048672.wecb125

By mimicking Darwinian evolution in the test tube, directed evolution has become a powerful tool for engineering novel enzymes for basic and applied biology research and medicine. Unlike structure-based rational design, directed evolution is capable of altering single or multiple functional properties such as activity, specificity, selectivity, stability, and solubility of naturally occurring enzymes in the absence of detailed knowledge of enzyme structure, function, or mechanism. More recently, directed evolution has also been used to engineer metabolic pathways, viruses, and whole microorganisms, and to address fundamental problems in biology. The success of directed evolution has been largely fueled by the development of numerous molecular biology techniques that enable the creation of genetic diversity through random mutagenesis or homologous or nonhomologous recombination in the target genes and the development of powerful high throughput screening or selection methods as well as by novel applications. This review will highlight the key developments in directed evolution and focus on the design and engineering of novel enzymes through directed evolution and their implications in chemical biology.

Enzymes are truly remarkable catalysts that are essential to every biological process. They can catalyze a broad range of chemical transformations with exquisite selectivity (stereo-, regio-, and chemo-) and specificity. In addition, most enzymes are very efficient and operate at mild conditions. It is, therefore, not surprising that enzymes have been increasingly used as biological catalysts or therapeutic agents in various industries, including the chemical, pharmaceutical, agricultural, and food industries. However, the number and diversity of enzyme-based applications are still modest compared with the total number of enzymes identified so far (~5000 enzymes) (1). One main reason for this functional gap is that naturally occurring enzymes are the products of Darwinian evolution and are not designed for optimal industrial applications. To address this limitation, several enzyme engineering approaches have been developed

in the past few decades, among which directed evolution stands out as a particularly attractive approach. This entry discusses the brief history of directed evolution, the main methods of directed evolution, and their applications in engineering enzymes for basic and applied biology research. For more in-depth information on directed evolution, interested readers are referred to the Further Reading list.

A Primer for Directed Evolution

Before the advent of recombinant DNA technology in the 1970s, the ability to engineer novel enzymes was limited to chemical modification methods in which specific residues in an enzyme

Advanced Article

Article Contents

- A Primer for Directed Evolution
- Methods for Directed Evolution
- Applications of Directed Evolution
- Conclusions and Future Prospects

*These two authors contributed equally.

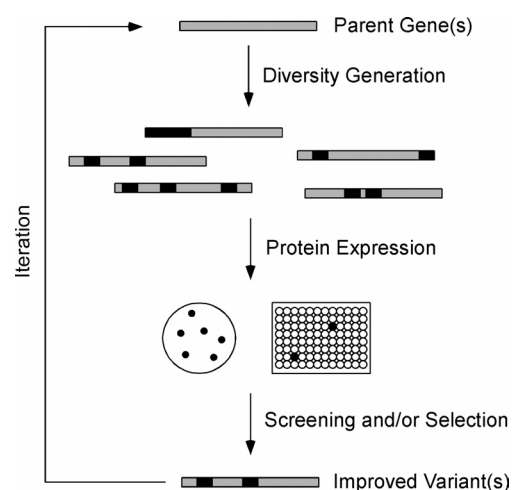


Figure 1 General scheme of directed evolution.

are modified by chemical agents. With the development of recombinant DNA technology, site-directed mutagenesis, and polymerase chain reaction (PCR) technology coupled with advances in X-ray crystallography, structure-based rational design became a dominant approach for engineering novel enzymes in the 1980s (2). Although rational design has achieved some notable successes, the requirement for extensive structural and mechanistic information on a target enzyme beset this method. Despite decades of research in protein science, it is still very difficult to identify the molecular determinants for the desired enzyme feature(s) even when the structure of the target enzyme is available, let alone the vast number of enzymes without crystal structures.

Directed evolution bypasses the bottleneck of rational design and mimics natural evolution in a test tube to evolve proteins without knowledge of their structures. What fundamentally differentiates directed evolution from natural evolution is its power to significantly accelerate the process of evolution. As shown in **Fig. 1**, directed evolution uses various methods to generate a collection of random protein variants, called a library, at the DNA level. Followed by screening/selection of the library, protein variants with improvement in desired phenotypes are obtained. Usually, the occurrence of these functionally improved protein variants is a rare event; thus, this two-step procedure has to be iterated several rounds until the goal is achieved or no further improvement is possible.

One of the very first directed evolution experiments can be traced back to as early as 1967 (3), but directed evolution did not become an established field until the mid-1990s. Advances in molecular biology have promoted rapid development of a wide variety of methods aimed at generating genetic diversities and at searching the molecular reservoir in a high throughput manner. In the past few years, directed evolution has been used to successfully engineer many enzymes for commercial and industrial applications (4), and the targets for directed enzyme evolution have been focused on activity, stability, specificity, and selectivity. It should be noted that the field of directed evolution is not limited only to enzyme engineering, but it can be applied to any single protein. In addition, more research

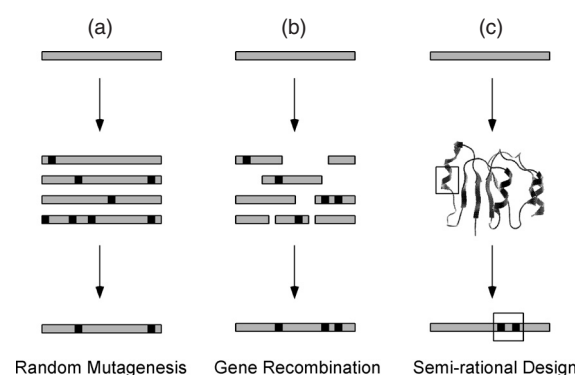


Figure 2 Comparison of (a) random mutagenesis, (b) gene recombination, and (c) semirational design.

has recently started addressing more complex systems, such as pathways (metabolic engineering), viruses, and even genomes.

Methods for Directed Evolution

A successful directed evolution experiment involves two key components: creating genetic diversity and developing a high throughput screening or selection method. In the past decade, many experimental methods and protocols for library construction and screening/selection have been developed. For more information on this topic, interested readers are referred to the two books edited by Arnold and Georgiou in the Further Reading list.

Library creation

Numerous molecular biology methods have been developed to introduce genetic diversity into the target gene, all of which can be grouped into three categories: methods of random mutagenesis, methods of gene recombination, and methods of semirational design. As shown in **Fig. 2**, random mutagenesis starts from a single parent gene and randomly introduces point mutations or insertions/deletions into the progeny genes. In comparison, gene recombination usually starts from a pool of mutants from a single gene or a pool of closely related or even non-related parental genes of different origin and creates blockwise exchange of sequence information among the parental genes. Finally, semirational design combines rational design and directed evolution by focusing mutagenesis on a few selected important residues or regions in a target gene.

Random mutagenesis

As a result of its simplicity and efficiency, error-prone polymerase chain reaction (EP-PCR) is the most widely used random mutagenesis method. It is essentially a variation of the standard PCR with slightly modified reaction conditions (5). There are many different protocols to implement EP-PCR, and the most popular one includes the following adjustments to normal PCR conditions: 1) use of nonproofreading DNA polymerases, such as *Taq* DNA polymerase; 2) use of low or unbalanced amount of dNTPs; 3) use of high concentration of Mg^{2+} (up to 10 mM);

and 4) incorporation of Mn^{2+} . The fourth modification has made EP-PCR more popular, because the error rate can be controlled precisely by the Mn^{2+} concentration (6). In general, 1–2 amino acid substitutions are introduced during each round of EP-PCR, which requires approximately 1–5 base mutations per kilobase of DNA. Higher mutagenic rates are not normally used because they often damage enzyme function and lead to an increased tendency to negate positive mutations. In addition, higher mutagenic rates result in a larger library size, which in turn requires an often unattainable robust screening/selection method to identify positive variants. On the other hand, a higher mutation rate increases the frequency of multiple mutations with synergistic effects, resulting in an overall enrichment of unique protein variants, and up to 30 mutations per gene have been reported (7).

The great success of EP-PCR in engineering all aspects of enzyme properties has established this method as a cornerstone in directed evolution. It should be noted, however, that this technique is not truly random and suffers a number of limitations. In addition to the intrinsic bias of DNA polymerases (transitions are favored over transversions), EP-PCR can only access 5–6 amino acid substitutions on average at each residue because of the degeneracy of genetic codons and the low probability of two mutations occurring right next to each other. Another limitation of EP-PCR is associated with the low mutation rates normally used, such that the progeny protein variants have similar phenotype to the parent. Thus, novel functions are difficult to evolve using this method alone even after several rounds of iteration. To search the sequence space more extensively, EP-PCR is used in combination with other DNA diversity generation methods, such as gene recombination.

Gene recombination

Gene recombination can be implemented both *in vivo* and *in vitro*. However, the latter is used much more often because of its simplicity, higher recombination efficiency, and flexibility. Therefore, only *in vitro* methods will be discussed here. Note that all the available *in vitro* gene recombination methods fall into two main categories: homology-dependent and homology-independent.

Homology-dependent gene recombination

As nature has found homologous recombination a useful evolving tool, biologists have also recognized its power of achieving “long jump” in adaptive molecular evolution (8). And the advances in molecular biology made it possible to mimic this process *in vitro*. The first and most frequently used gene recombination method, DNA shuffling, also known as “sexual PCR”, was developed by Stemmer in 1994 (9). As shown in **Fig. 3**, the target gene is digested by DNaseI into random fragments, of which 100–300 bp fragments are purified and reassembled in a self-priming (no primers are added) PCR reaction according to their sequence homology. Recombination occurs when a fragment derived from one sequence anneals to a fragment derived from another sequence. This method was later adapted to recombine a family of naturally occurring homologous proteins from diverse species under modified conditions, which is

called “family shuffling” (10). It was demonstrated that family shuffling significantly accelerated the rate of improvement of enzyme functions in comparison with EP-PCR and DNA shuffling.

As with every method, both DNA shuffling and family shuffling have their own limitations. First of all, both methods require relatively high homology, typically more than 70–75%, between the parental genes, because libraries created from more divergent sequences have a strong tendency to reassemble into parental genes. Various homology-independent methods have been developed to address this issue and will be discussed in the next session. Second, crossovers during template switching are favored in regions of high sequence identity, restricting the sequence space that can be explored. Third, fragments generated by DNaseI are not truly random, thus the diversity of the shuffled library is further decreased. Finally, there are also some nontechnical problems, such as limited access to natural sequence diversity and patent issues.

To address some of these limitations, a group of homologous gene recombination methods that do not involve DNA fragmentation but require addition of primers were developed, and staggered extension process (StEP) (11) was the first among them (**Fig. 3**). This method is essentially a modified PCR that uses very short extension time so that the elongation of short DNA fragments is staggered. During the subsequent rounds of DNA amplification, the fragments are repeatedly separated from the parental strand and prime a different one, resulting in multiple crossovers. StEP has several advantages over the original DNA shuffling method: 1) It needs only a small quantity of parental genes; 2) no digestion or DNA purification is needed, thus it is easy to be carried out; and 3) it avoids the DNaseI-induced bias. However, it should be noted that the StEP PCR conditions need to be optimized before a good library can be obtained, which might take a considerable amount of time.

Nonhomologous gene recombination

Incremental Truncation for the Creation of Hybrid enzymes (ITCHY) was the first developed homology-independent recombination method (12). Incremental truncation of two parental genes from both ends by exonuclease III under nonideal conditions generates a collection of all possible truncated fragments, which are subsequently blunt polished and ligated to give genes of various lengths. There are several limitations of ITCHY. First, the key to creating a successful ITCHY library is the tight control of the exonuclease digestion conditions, and aliquots of digestion mixture have to be taken at various time points to quench the reaction. Therefore, it can be very time-consuming and labor-intensive. To address this issue, the same group developed a modified method, called THIO-ITCHY (13). The incorporation of α -phosphothioate nucleotide analogs at low frequency in genes inhibits exonuclease III activity, thus avoiding the requirement for frequent removal of digestion samples. The second limitation of ITCHY is that because it is a single crossover process, the diversity of the created library is rather limited. Another method, named SCRATCHY (14), was developed by the same group to achieve multiple crossovers by shuffling two ITCHY libraries, thus increasing the diversity of the library. Third, the ITCHY library of hybrids is not full-length

and thus the two truncated genes are not necessarily fused at sites where the gene sequences align (15). It was shown previously that although insertions or deletions at the fusion portion of two parental genes might not necessarily have a deleterious effect on the enzyme function, the predominance of crossovers at positions of precise alignment in the selected active hybrids (12) indicates the importance of the alignment. This problem led to the birth of another method, sequence homology-independent protein recombination (SHIPREC) (15). In this method (Fig. 3), two parental genes are fused by a linker containing multiple restriction sites. After digestion by DNaseI at both ends of the fusion gene, full-length genes are selected, circularized, and digested by restriction enzyme in the linker region to give linear chimerical genes. The selection of a full-length gene helps maintain the sequence alignment of two genes and gives a larger fraction of functional hybrids. Finally, not only ITCHY, but all the methodologies discussed above, have one common limitation: only two parental genes can be recombined. Therefore, a few other multiple-parental homology-independent recombination methods have been developed, such as exon shuffling (16) and nonhomologous random recombination (NRR) (17).

Semirational design

Although rational design enables efficient targeting at critical protein sites, this approach is often hindered by limited availability of crystal structures and poor understanding of the structure–function relationship. To circumvent the limitations of rational design, directed evolution found its position as the “blind watchmaker.” However, as it is a “blind” searching process, the diversity pool must be as extensive as possible, which leads to the bottleneck of directed evolution: library screening. Therefore, any means to decrease the library redundancy would be beneficial. More importantly, when the engineering goal is to dramatically alter an enzyme function, it usually requires multiple close mutations in the active site, which are difficult to access by full-length gene random mutagenesis and require an even larger library to be screened. Therefore, to allow a more focused and more useful sequence space to be explored, the most logical way would be to combine the best features of the

two extreme methodologies. This process gave birth to the third library creation method, called semirational design.

The most popular semirational design strategy is targeted saturation mutagenesis. Functionally important residues are identified by analysis of protein crystal structures and mutated individually (18) or in combination (19) into the other 19 natural amino acids using degenerate primers (NNN or NNS, N = A/T/G/C, S = G/C). It should be noted that protein crystal structures are no longer the only source for identification of functionally important residues. When no protein structure information is available, key residues can be identified by EP-PCR, bioinformatics, or homology modeling. Another expanding area is *in silico* directed evolution, the ability of which to rationalize a huge protein database and to guide engineering experiments holds the possibility to create novel enzymes beyond the natural realm. Various algorithms have been developed recently to optimize library creation conditions, library design, and library prescreening. Interested readers are referred to a more comprehensive review (20) on computational protein design methods.

Library screening/selection

Advances in molecular biology have made it possible to generate protein variants at the DNA level, and a library size of greater than a billion members can be achieved. The real challenge lies in the ability to find the needle with desired properties in the haystack; therefore, a sensitive and high throughput assay is highly desirable for directed evolution. For each directed evolution experiment, the analysis method must be prudently chosen or developed, because of the first principle of directed evolution “you get what you select (screen) for.” There are two main categories of library analysis methods: screening and selection. Screening involves examining every mutant individually for the desired property, whereas selection is a method whereby only proteins with the desired property are carried through. Although various technologies have been developed in each category, a common principle underlying these assays exists: tagging the DNA (genotype) and the protein it encodes

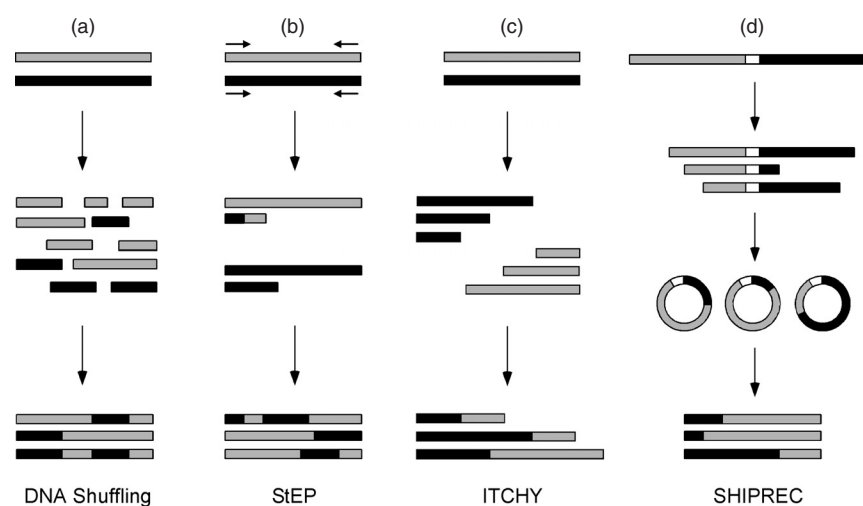


Figure 3 Comparison of various gene recombination methods, including (a) DNA shuffling, (b) StEP, (c) ITCHY, and (d) SHIPREC.

(phenotype) followed by screening/selection (phenotype analysis) that is compatible with the tagging. Physical linkage and spatial compartmentalization are two ways of tagging.

Screening technologies

As screening requires individual analysis of each protein variant, its throughput is relatively low and it can only be used to screen small libraries (up to a size of $\sim 10^4$). The 96-well plate is the most widely used screening format due to its versatility, although higher spatial density formats can be used, such as 384- or 1536-well microtiter plates or even protein microarrays. In a microtiter plate assay, not only are the protein and its encoding DNA compartmentalized in one well, but also the whole reaction; therefore, it is most suitable for enzyme activity assays. In addition, the enzyme is analyzed in the same way as in traditional biochemical assays: Each protein sample in the form of cell cultures, crude lysates, or purified proteins is transferred into one well and then examined, thus the reaction conditions can be controlled to mimic the final practical conditions as closely as possible. With the aid of an automatic colony picker and liquid handler, the assays can be easily adapted into this high throughput format and automated. Although the microtiter plates only provide compartmentalization for the DNA–protein pair, methods are needed to analyze the proteins. Currently, colorimetric or fluorometric assays are the most popular and convenient screening methods, whereby the positive variants can be easily identified by visual check or by measuring UV-Vis absorbance or fluorescence using a plate reader. However, they are not available for all enzymes. Other generic screening tools, such as HPLC, capillary electrophoresis, and thermistor arrays, have also been applied to engineering of enzymes (21).

To address the low throughput limitation associated with most screening methods, various fluorescence-activated cell sorting (FACS) based screening methods have been developed. Unlike the above mentioned screening methods, FACS can analyze and sort up to 100,000 cells per second in a quantitative manner (22). The first application of FACS to directed enzyme evolution was demonstrated by Georgiou and his coworkers in 2000 (23). By coupling with bacterial surface display (see the Surface Display subsection to follow), FACS was successfully used to engineer a protease variant with improved catalytic activity. A fluorescence resonance energy transfer (FRET) substrate was designed to assay the protease activity in which a fluorescent dye is quenched by its FRET quenching partner via a target scissile bond recognized by the protease. Enzymatic cleavage of the scissile bond results in the release of the FRET quenching partner while the fluorescent dye is retained on the cell surface, allowing isolation of active clones by FACS. Remarkably, this method achieved 5000-fold enrichment of active clones in a single FACS round.

Selection technologies

Compared with screening technologies, library selection applies certain selection pressure/criteria to the mutant library so that only positive variants are carried to the next round while unwanted variants are discarded. Therefore, a much larger library of enzyme variants (more than 10^{11}) can be assessed. However, the selection methods are normally developed for a specific

system or for analyzing a particular enzyme property. Many properties, such as enzyme activity at extreme temperatures or pH, or organic solvents, are not directly amenable to selection. As a result, screening is usually more applicable than selection. Based on the DNA–protein pair tagging method, selection methods can be divided into two categories: surface display and compartmentalization.

Surface display

Display technologies, employing nucleic acids, phage, yeast, or bacteria, were initially developed for binding assays and have made great success in engineering high affinity receptors, such as antibodies and T-cell receptors (24). Several inherent features of the display technologies made them suitable for directing enzyme evolution. First of all, display of proteins on the surface establishes a physical linkage between DNA and protein. Second, the proteins on the surface are accessible to external molecules, such as substrates or other target molecules. Finally, the DNA is restricted inside the phage particle or microbial cells, enabling easy tracking of the genotype. With the need for enzyme engineering growing, researchers recognized the potential of display technologies and progressively adapted them for enzyme engineering.

Phage display is the most commonly used technique for *in vitro* selection. Filamentous bacteriophages (e.g., M13) are used for protein display for their ability to infect host cells without killing them (25). In a practical enzyme phage display experiment, a phagemid DNA library is constructed first *in vitro* and then transformed into competent bacterial cells. The DNA that encodes the enzyme of interest is fused to one of the coat protein genes (pVIII for high copy display, pIII for low copy display), thus the enzyme is expressed as a fusion to the phage coat protein. During the phage assembly process, the target DNA is encapsulated inside the nascent phage particle as a part of its genome while its encoding enzyme is displayed on its surface; as a result, a physical linkage is established between the phenotype and genotype through the phage particle. Phage particles are then harvested as a batch and selected for those displayed enzymes with improved/novel functions. Phage display selection is naturally based on binding. More specifically, a phage library is selected by passing it through an affinity matrix whereby binding phages are captured while nonbinding phages are washed away. Therefore, to adapt the phage display technique to engineer enzyme properties such as activity, selectivity, and stability, the key is to couple enzyme properties to the capture or release of the phage from the affinity matrix, for example, by codisplaying the enzyme and substrate on the same phage particle. As shown in **Fig. 4**, upon catalysis, the product is displayed on the surface and recognized by the solid support. In contrast, phages displaying inactive enzymes cannot bind the affinity matrix and are washed away. Phage display has been successfully used to engineer enzymes with improved activity, altered substrate specificity, improved stability, and even novel function (25–28). However, it is almost impossible to develop a generic phage display system for all applications and phages lack posttranslation modification mechanisms that might be critical for functional expression of some enzymes.

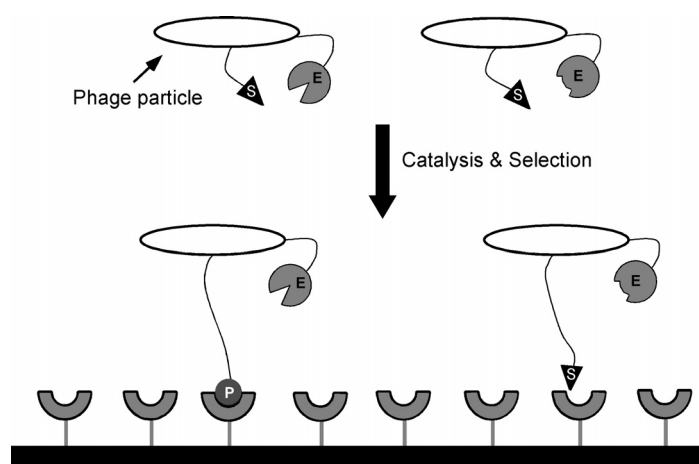


Figure 4 Schematic representation of a phage display-based selection method for directed enzyme evolution E: enzyme; S: substrate; P: product.

Compartmentalization

Just as in microtiter plate format based screening, compartmentalization is also used in selection methods; each DNA–protein pair is spatially isolated in an individual compartment, which is either a cell (*in vivo* selection) or a manmade compartment (*in vitro* selection) instead of individual wells.

Whenever accessible, *in vivo* selection is very powerful and can assess large numbers of mutants. The ultimate *in vivo* selection method would be, under a given selection pressure, only mutants harboring proteins with improvements could grow into colonies or show a significant phenotypic difference. Although it is a very powerful technique, the utility of *in vivo* selection is very limited, because most enzymes are of little direct biological relevance. Another reason is that the sophisticated genetic regulation networks of the host microorganism have evolved to encounter rapid changes in the environment, and thus the applied selection pressure may result in mutations out of the target genes. *In vitro* selection overcomes some of the limitations of *in vivo* selection. *In vitro* compartmentalization (IVC) (29) links the genotype and phenotype by colocalizing single genes together with necessary transcription and translation biochemical components in the aqueous compartments of a water-in-oil emulsion droplet. In most compartments, there is either no gene or only one gene that is later transcribed and translated *in vitro* within the same compartment. The enzymatic reaction is later carried out in the same droplet. To a certain extent, IVC is similar to microtiter plates but on a much smaller size scale with volumes close to those of bacteria (29). As the gene is transcribed and translated *in vitro*, general cloning is avoided and the library size is no longer limited by transformation efficiency. However, it seems that IVC can only be used to select enzymes that directly or indirectly act on DNA. For analyzing other enzyme properties, the droplets still need to be screened one by one, as in the case of 96-well plate screening. However, by combining with other technologies, such as FACS (30) or microbeads (31), IVC still holds promise for future enzyme engineering.

Applications of Directed Evolution

Directed evolution has been successfully used to alter existing enzyme properties and even to create novel enzyme functions. In addition to creating enzymes for specific industrial applications, directed evolution has also been increasingly used to address fundamental questions in biology, such as the evolutionary mechanisms of novel protein functions, protein structure–function relationship, and protein folding mechanisms.

Improving enzyme properties by directed evolution

Directed evolution has enjoyed great success in improving existing enzyme characteristics. In the following sections, only a few selected examples will be highlighted. Alterations have been made for almost all aspects of enzyme properties, such as substrate specificity, product specificity, selectivity, activity, stability, or folding/solubility. Such alterations are required for enzymes to become practically useful biocatalysts or therapeutics.

Substrate specificity

Although the analogy of lock and key is sometimes used to describe the relationship between an enzyme and its substrate, in reality, the level of specificity varies. A particular enzyme may perform similar reactions on a range of related substrates or it may show tremendous specificity to one molecule. This aspect of enzymes can be exploited to develop variants with altered substrate recognition. In some cases, it may be beneficial to expand the range of substrates acted on. For example, polychlorinated biphenyls (PCBs) are a class of organic compounds whose use is decreasing due to concerns over their long-term environmental persistence and health effects. Certain bacteria can degrade some of these compounds by oxygenation reactions. Shuffling of two biphenyl dioxygenases from different bacteria resulted in higher activity, and activity on novel substrates such as toluene (32). In another example, EP-PCR was used to convert *E. coli* aspartate aminotransferase into a valine

aminotransferase (33). A mutant enzyme with 17 amino acid substitutions was created that shows a 2.1×10^6 -fold increase in the catalytic efficiency for a non-native substrate, valine. Structural analysis of the mutant enzyme by protein crystallography indicated a remodeled active site and altered subunit interface caused by the accumulative effects of mutations. Most surprisingly, only one of the mutations directly contacts the substrate, which underscores our limited understanding of enzyme substrate specificity. These mutations would be difficult, if not impossible, to be identified and introduced to the mutant enzyme by a rational design approach.

Product specificity

In addition to altering an enzyme's substrate, the product of an enzymatic reaction can be modified by using directed evolution. One example of product specificity engineering that has received attention is that of carotenoid pathway enzymes. Farnesylgeranyl diphosphate synthase catalyzes the condensation of isopentenyl diphosphate into a C25 isoprenoid molecule. The chain length specificity of this enzyme was changed to produce C20 geranylgeranyl diphosphate (34). The conversion of this product into either lycopene or neurosporene by phytoene desaturase was investigated and shown to be amenable to almost a complete reversal of product specificity (35). Another example is directed evolution of γ -humulene synthase that acts on farnesyl diphosphate to produce over 50 sesquiterpenes via different cyclization reactions (36). Residues within the active site influence the reaction and were investigated by using saturation mutagenesis. Based on a model incorporating effects from individual sites, variants with multiple mutations were generated that showed increased specificity for particular products.

Selectivity

Chiral molecules have important roles in the pharmaceutical and chemical industries. Enzymes have the capability to be exquisitely enantioselective, and applications of directed evolution in this area have recently been reviewed (37). Pioneering work was carried out on a lipase from *Pseudomonas aeruginosa*, by EP-PCR and saturation mutagenesis. Using a model reaction, the hydrolysis of 2-methyldecanoic acid *p*-nitrophenyl ester, the enantioselectivity was increased from $E = 1.1$ to $E = 25.8$ (38). Carbohydrates are a large class of chiral molecules with essential roles in biology, and they can serve as useful precursors in chemical synthesis of complex organic molecules. Directed evolution has been used to alter the preferred stereoproduct of the condensation of dihydroxyacetone phosphate and glyceraldehyde 3-phosphate (39). Depending on the enzyme, these substrates can yield D-fructose-1, 6-bisphosphate or D-tagatose-1, 6-bisphosphate, which differ in the C4 stereochemistry. DNA shuffling of tagatose-1, 6-bisphosphate aldolase shifted the preference from >99:1 in favor of tagatose-1, 6-bisphosphate to 4:1 in favor of fructose-1, 6-bisphosphate, due to mutation of four residues within the substrate binding pocket.

Activity

Enzymes show a wide variety of reaction rates, which can be expressed in terms of either their turnover number or catalytic

efficiency. For practical purposes, a high reaction rate is desirable, and it can be achieved by increasing the k_{cat} or decreasing the K_m . A high throughput screening system was used with family shuffling of the thymidine kinase gene from herpes simplex virus I and II to increase the specificity of AZT phosphorylation (40). The authors used a robot to pick around 10,000 clones at each of four rounds of family shuffling, and they measured colony growth on different levels of AZT. Variants were found that conferred sensitivity to *E. coli* when exposed to 32-fold less AZT compared with HSV I thymidine kinase. These variants contained multiple crossovers and mutations affecting the binding site. Another high throughput screening system, *in vitro* compartmentalization, was used with site-directed saturation mutagenesis to screen libraries of phosphotriesterase for increased activity (31). Despite this enzyme already being very active, the k_{cat} was increased from 2280 s^{-1} to $144,300 \text{ s}^{-1}$ (63-fold). The k_{cat}/K_m was only increased slightly due to an increase in K_m , but at $1.76 \times 10^8 \text{ M}^{-1}\text{s}^{-1}$, it is approaching the diffusion-limited rate of catalysis.

Stability

A common aim of directed evolution is to increase the stability of an enzyme to conditions of practical use that may be very different from those the enzyme naturally functions in. Factors such as heat, altered pH, and the presence of oxidants or organic solvents can lead to denaturation or loss of enzyme function. Many researchers have successfully increased the stability of an enzyme to thermal denaturation (41, 42). Work with *p*-nitrobenzyl esterase increased the melting temperature 14°C after six rounds of EP-PCR and recombination without forfeiting enzyme activity (41). As another example, phosphite dehydrogenase catalyzes the formation of phosphate from phosphite, by reducing NAD^+ to NADH. However, the usefulness of this enzyme as a means of regenerating NADH cofactors for industry was impeded by the low stability of the wild-type enzyme isolated from *Pseudomonas stutzeri*. Four rounds of EP-PCR were used to identify 12 mutations that increased the half-life of the enzyme at 45°C by 7000-fold (42). Notably, family shuffling of 26 subtilisin genes produced variants with improved activity to either heat, pH 10, pH 5.5, or the presence of 35% dimethylformamide (43). Certain clones also showed better performance under combinations of these conditions.

Folding/solubility

Low solubility or improper folding may sometimes hamper the use of enzymes, particularly when expressed in a non-native host. A method of expressing proteins with a C-terminal GFP fusion to use fluorescence as a measure of the amount of correctly folded protein has been introduced (44). DNA shuffling produced variants of ferritin that showed increased solubility, even when they were recloned without the GFP fusion. This assay has been used to produce proteins for X-ray crystallography structure determination (45). The protein nucleoside diphosphate kinase from *Pyrobaculum aerophilum* is insoluble when expressed in *E. coli*, but after DNA shuffling, a functional variant with six mutations was found to have 90% solubility, which enabled its crystallization, and its structure was determined.

Creating enzymes with novel functions by directed evolution

One of the aspirations of directed evolution is to create new function in enzymes, which may be to carry out a reaction that has not been found in nature or may involve adding new control modalities. The challenge that the field is embracing is a significant one. Nature is conservative when it comes to the generation of new enzymes, as it retools existing structures for new functions rather than inventing a new scaffold for each different reaction. The (β/α)-8 barrel scaffold, for example, is the most common protein structure found in enzymes, and different enzymes with this scaffold carry out many different types of reactions. Protein engineers can take hope and inspiration from this in their attempts to create novel function in enzymes.

Novel substrate specificity

As already discussed, enzyme activity can be broadened to include substrates not previously acted on. But directed evolution can also yield enzymes with substrate recognition different from the starting point (46). DNA shuffling of two highly homologous triazine hydrolases produced variants that acted on triazines that neither parent had activity toward, which showed that examining small differences in sequence space can reveal new activities.

Altering substrate specificity sometimes follows a process of relaxing, followed by tightening. Collins et al. (47) used a clever dual selection strategy to alter the response of the LuxR transcription factor to different acyl-homoserine lactones. This type of research can produce modifiers of transcription with fine control by a desired ligand chosen so as to not interfere with other biological pathways. The response of LuxR was initially broadened from 3-oxo-hexanoyl-homoserine lactone (3OC6HSL) to accept a variety of straight-chain acyl-homoserine lactones. Negative selection was performed against response to 3OC6HSL, resulting in a variant that responded to straight-chain acyl-HSLs but not the original activator.

Novel activity

Novel functions can be incorporated within existing protein scaffolds that naturally have no activity for the desired reaction. Working within the $\alpha\beta/\alpha\beta$ -metallohydrolase enzyme scaffold, the activity of β -lactamase has been successfully introduced into glyoxalase II by a combination of rational design and directed evolution (48), which involved deletion of the original glyoxalase II substrate-binding domain, followed by the introduction of loops designed by examining metallo β -lactamases, EP-PCR, and DNA shuffling. The resulting enzyme had activity as a β -lactamase, albeit at much lower efficiency than seen for the native enzyme. New activity can also be incorporated into a noncatalytic protein scaffold, as demonstrated by the creation of triose phosphate isomerase activity within ribose-binding protein by computational design and EP-PCR (49). Current applications in this area rely on semirational design, with directed evolution typically used to increase the initial activity produced.

New ways of controlling enzyme function can also be introduced. Natural enzymes often exhibit some form of posttranslational regulation that affects their activity, which could take the

form of interaction with a small molecule to enhance or inhibit activity in a particular environment. The maltose-binding protein can function as a switch when inserted into a gene such as β -lactamase (50), which enables a level of control over the desired reaction, based on the presence or absence of a molecule such as maltose.

Understanding natural enzyme evolution

The power of directed evolution to create and analyze tens of millions of protein variants not only enables one to engineer enzymes with desired properties for practical applications and to study the structure and function of proteins, but it also provides researchers the means to understand natural evolutionary processes. Rather than being restricted to the snapshot of sequence space found in extant genes, researchers can conduct evolutionary experiments on catalytic mechanisms or protein structure to better understand how current genes arose. As little as one mutation has been shown to confer on an enzyme the ability to carry out a new reaction. Single mutations were discovered that allowed two members of the muconate lactonizing enzyme subgroup of the enolase superfamily to catalyze an additional reaction, that of the enzyme *o*-succinylbenzoate synthase (51). Other work has revealed the ease through which a promiscuous enzyme function can be improved by orders of magnitude, with comparatively little effect on the enzyme's main function (52). Proposed pathways of protein fold evolution have been examined for the DNA methyltransferase superfamily, showing that the circularly permuted variants seen in nature can be generated in the laboratory via intermediates that retain function (53).

Conclusions and Future Prospects

Directed evolution has been demonstrated to be very useful in modifying enzymes for practical applications, producing better stability, higher activity, and altered substrate specificity or product formation. Its influence will only increase, by producing enzymes for use as research tools in biology or therapeutics in medicine, and as a means of improving chemical syntheses or industrial processes. The future is likely to see an increased pairing of rational design and directed evolution, as researchers generate more protein structures and improve their ability to identify optimal target areas of proteins for randomization. Ambitious applications are also likely to continue, leading to new ways of controlling enzyme activity and examples of dramatic reconfiguration of the starting enzyme's function. It is still a long way off until researchers have the ability to design from first principles an enzyme for any given task, and, as such, directed evolution will continue to be an incredibly useful tool for many years to come.

References

1. Wandrey C, Liese A, Kihumbu D. Industrial biocatalysis: past, present, and future. *Org. Process Res. Dev.* 2000;4:286–290.
2. Brannigan JA, Wilkinson AJ. Protein engineering 20 years on. *Nat. Rev. Mol. Cell. Bio.* 2002;3:964–970.

3. Mills DR, Peterson RL, Spiegelman S. An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc. Natl. Acad. Sci. U.S.A.* 1967;58:217–224.
4. Johannes TW, Zhao H. Directed evolution of enzymes and biosynthetic pathways. *Curr. Opin. Microbiol.* 2006;9:261–267.
5. Leung D, Chen E, Goeddel D. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* 1989;1:11–15.
6. Zhao H, Moore JC, Volkov AA, Arnold FH. Methods for optimizing industrial enzymes by directed evolution. In: *Manual of Industrial Microbiology and Biotechnology*. 2nd edition. Davies JE, ed. 1999. ASM Press, Washington, D.C.
7. Drummond DA, Iverson BL, Georgiou G, Arnold FH. Why high-error-rate random mutagenesis libraries are enriched in functional and improved proteins. *J. Mol. Biol.* 2005;350:806–816.
8. Kauffman S, Levin S. Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.* 1987;128:11–45.
9. Stemmer WP. Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* 1994;370:389–391.
10. Cramer A, Raillard SA, Bermudez E, Stemmer WP. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* 1998;391:288–291.
11. Zhao H, Giver L, Shao Z, Affholter JA, Arnold FH. Molecular evolution by staggered extension process (StEP) *in vitro* recombination. *Nat. Biotechnol.* 1998;16:258–261.
12. Ostermeier M, Shim JH, Benkovic SJ. A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat. Biotechnol.* 1999;17:1205–1209.
13. Lutz S, Ostermeier M, Benkovic SJ. Rapid generation of incremental truncation libraries for protein engineering using alpha-phosphothioate nucleotides. *Nucleic Acids Res.* 2001;29:e16.
14. Lutz S, Ostermeier M, Moore GL, Maranas CD, Benkovic SJ. Creating multiple-crossover DNA libraries independent of sequence identity. *Proc. Natl. Acad. Sci. U.S.A.* 2001;98:11248–11253.
15. Sieber V, Martinez CA, Arnold FH. Libraries of hybrid proteins from distantly related sequences. *Nat. Biotechnol.* 2001;19:456–460.
16. Kolkman JA, Stemmer WP. Directed evolution of proteins by exon shuffling. *Nat. Biotechnol.* 2001;19:423–428.
17. Bittker JA, Le BV, Liu DR. Nucleic acid evolution and minimization by nonhomologous random recombination. *Nat. Biotechnol.* 2002;20:1024–1029.
18. Chockalingam K, Chen Z, Katzenellenbogen JA, Zhao H. Directed evolution of specific receptor-ligand pairs for use in the creation of gene switches. *Proc. Natl. Acad. Sci. U.S.A.* 2005;102:5691–5696.
19. Reetz MT, Bocla M, Carballeira JD, Zha D, Vogel A. Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test. *Angew. Chem. Int. Ed. Engl.* 2005;44:4192–4196.
20. Moore GL, Maranas CD. Computational challenges in combinatorial library design for protein engineering. *AICHE J.* 2004;50:262–272.
21. Wahler D, Reymond JL. High-throughput screening for biocatalysts. *Curr. Opin. Biotechnol.* 2001;12:535–544.
22. Shapiro H. *Practical Flow Cytometry*. 4th edition. 2003. Wiley & Sons, New York.
23. Olsen MJ, Stephens D, Griffiths D, Daugherty P, Georgiou G, Iverson BL. Function-based isolation of novel enzymes from a large library. *Nat. Biotechnol.* 2000;18:1071–1074.
24. Li M. Applications of display technology in protein analysis. *Nat. Biotechnol.* 2000;18:1251–1256.
25. Paschke M. Phage display systems and their applications. *Appl. Microbiol. Biotechnol.* 2006;70:2–11.
26. Jestin JL, Kristensen P, Winter G. A method for the selection of catalytic activity using phage display and proximity coupling. *Angew. Chem. Int. Ed. Engl.* 1999;38:1124–1127.
27. Pedersen H, Holder S, Sutherlin DP, Schwitter U, King DS, Schultz PG. A method for directed evolution and functional cloning of enzymes. *Proc. Natl. Acad. Sci. U.S.A.* 1998;95:10523–10528.
28. Demartis S, Huber A, Viti F, Lozzi L, Giovannoni L, Neri P, Winter G, Neri D. A strategy for the isolation of catalytic activities from repertoires of enzymes displayed on phage. *J. Mol. Biol.* 1999;286:617–633.
29. Tawfik DS, Griffiths AD. Man-made cell-like compartments for molecular evolution. *Nat. Biotechnol.* 1998;16:652–656.
30. Mastrobattista E, Taly V, Chanudet E, Treacy P, Kelly BT, Griffiths AD. High-throughput screening of enzyme libraries: *in vitro* evolution of a beta-galactosidase by fluorescence-activated sorting of double emulsions. *Chem. Biol.* 2005;12:1291–1300.
31. Griffiths AD, Tawfik DS. Directed evolution of an extremely fast phosphotriesterase by *in vitro* compartmentalization. *EMBO J.* 2003;22:24–35.
32. Kumamaru T, Suenaga H, Mitsuoka M, Watanabe T, Furukawa K. Enhanced degradation of polychlorinated biphenyls by directed evolution of biphenyl dioxygenase. *Nat. Biotechnol.* 1998;16:663–666.
33. Oue S, Okamoto A, Yano T, Kagamiyama H. Redesigning the substrate specificity of an enzyme by cumulative effects of the mutations of non-active site residues. *J. Biol. Chem.* 1999;274:2344–2349.
34. Lee PC, Mijts BN, Petri R, Watts KT, Schmidt-Dannert C. Alteration of product specificity of *Aeropyrum pernix* farnesylgeranyl diphosphate synthase (Fgs) by directed evolution. *Protein Eng. Des. Sel.* 2004;17:771–777.
35. Wang CW, Liao JC. Alteration of product specificity of *Rhodobacter sphaeroides* phytoene desaturase by directed evolution. *J. Biol. Chem.* 2001;276:41161–41164.
36. Yoshikuni Y, Ferrin TE, Keasling JD. Designed divergent evolution of enzyme function. *Nature* 2006;440:1078–1082.
37. Jaeger KE, Eggert T. Enantioselective biocatalysis optimized by directed evolution. *Curr. Opin. Biotechnol.* 2004;15:305–313.
38. Liebeton K, Zonta A, Schimossek K, Nardini M, Lang D, Dijkstra BW, Reetz MT, Jaeger KE. Directed evolution of an enantioselective lipase. *Chem. Biol.* 2000;7:709–718.
39. Williams GJ, Domann S, Nelson A, Berry A. Modifying the stereochemistry of an enzyme-catalyzed reaction by directed evolution. *Proc. Natl. Acad. Sci. U.S.A.* 2003;100:3143–3148.
40. Christians FC, Scapozza L, Cramer A, Folkers G, Stemmer WP. Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling. *Nat. Biotechnol.* 1999;17:259–264.
41. Giver L, Gershenson A, Freskgard PO, Arnold FH. Directed evolution of a thermostable esterase. *Proc. Natl. Acad. Sci. U.S.A.* 1998;95:12809–12813.
42. Johannes TW, Woodyer RD, Zhao H. Directed evolution of a thermostable phosphite dehydrogenase for NAD(P)H regeneration. *Appl. Environ. Microbiol.* 2005;71:5728–5734.
43. Ness JE, Welch M, Giver L, Bueno M, Cherry JR, Borchert TV, Stemmer WP, Minshull J. DNA shuffling of subgenomic sequences of subtilisin. *Nat. Biotechnol.* 1999;17:893–896.
44. Waldo GS, Standish BM, Berendzen J, Terwilliger TC. Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.* 1999;17:691–695.
45. Pedelacq JD, Piltch E, Liang EC, Berendzen J, Kim CY, Rho BS, Park MS, Terwilliger TC, Waldo GS. Engineering soluble proteins for structural genomics. *Nat. Biotechnol.* 2002;20:927–932.

46. Raillard S, Krebber A, Chen Y, Ness JE, Bermudez E, Trinidad R, Fullem R, Davis C, Welch M, Seffernick J, Wackett LP, Stemmer WP, Minshull J. Novel enzyme activities and functional plasticity revealed by recombining highly homologous enzymes. *Chem. Biol.* 2001;8:891–898.
47. Collins CH, Leadbetter JR, Arnold FH. Dual selection enhances the signaling specificity of a variant of the quorum-sensing transcriptional activator LuxR. *Nat. Biotechnol.* 2006;24:708–712.
48. Park HS, Nam SH, Lee JK, Yoon CN, Mannervik B, Benkovic SJ, Kim HS. Design and evolution of new catalytic activity with an existing protein scaffold. *Science* 2006;311:535–538.
49. Dwyer MA, Looger LL, Hellinga HW. Computational design of a biologically active enzyme. *Science* 2004;304:1967–1971.
50. Guntas G, Mansell TJ, Kim JR, Ostermeier M. Directed evolution of protein switches and their application to the creation of ligand-binding proteins. *Proc. Natl. Acad. Sci. U.S.A.* 2005;102:11224–11229.
51. Schmidt DM, Mundorff EC, Dojka M, Bermudez E, Ness JE, Govindarajan S, Babbitt PC, Minshull J, Gerlt JA. Evolutionary potential of (beta/alpha)₈-barrels: functional promiscuity produced by single substitutions in the enolase superfamily. *Biochemistry* 2003;42:8387–8393.
52. Aharoni A, Gaidukov L, Khersonsky O, McQ Gould S, Roodveldt C, Tawfik DS. The 'evolvability' of promiscuous protein functions. *Nat. Genet.* 2005;37:73–76.
53. Peisajovich SG, Rockah L, Tawfik DS. Evolution of new protein topologies through multistep gene rearrangements. *Nat. Genet.* 2006;38:168–174.

Further Reading

- Arnold FH, Georgiou G. Directed Evolution Library Construction: Methods and Protocols. 2003. Humana Press, Totowa, NJ.
- Arnold FH, Georgiou G. Directed Enzyme Evolution: Screening and Selection Methods. 2003. Humana Press, Totowa, NJ.
- Arnold FH. Design by directed evolution. *Acct. Chem. Res.* 1998;31:125–131.
- Bloom JD, Meyer MM, Meinhold P, Otey CR, MacMillan D, Arnold FH. Evolving strategies for enzyme engineering. *Curr. Opin. Struc. Biol.* 2005;15:447–452.
- Schmidt-Dannert C. Directed evolution of single proteins, metabolic pathways, and viruses. *Biochemistry* 2001;40:13125–13136.
- Rubin-Pitel SB, Zhao H. Recent advances in biocatalysis by directed enzyme evolution. *Comb. Chem. High Throughput Screen* 2006;9:247–257.
- Valetti F, Gilardi G. Directed evolution of enzymes for product chemistry. *Nat. Prod. Rep.* 2004;21:490–511.

See Also

Proteins: Computational Analysis