Appl Biochem Biotechnol (2007) 143:212–223
DOI 10.1007/s12010-007-8003-4

Developments in Directed Evolution for Improving Enzyme Functions

S. Sen · V. Venkata Dasu · B. Mandal

Published online: 18 August 2007

Abstract The engineering of enzymes with altered activity, specificity, and stability, using directed evolution techniques that mimic evolution on a laboratory timescale, is now well established. In vitro recombination techniques such as DNA shuffling, staggered extension process (StEP), random chimeragenesis on transient templates (RACHITT), iterative truncation for the creation of hybrid enzymes (ITCHY), recombined extension on truncated templates (RETT), and so on have been developed to mimic and accelerate nature's recombination strategy. This review discusses gradual advances in the techniques and strategies used for the directed evolution of biocatalytic enzymes aimed at improving the quality and potential of enzyme libraries, their advantages, and disadvantages.

Keywords Directed evolution · DNA shuffling · In vitro recombination · Mutagenesis · Random priming

Introduction

The field of enzyme biotechnology is undergoing rapid change and diversification [1]. Commercial enzyme products were originally developed as simple fermentation broths of naturally occurring organisms. Often, the resulting products were complex mixtures of secreted enzymes produced at relatively low yields of less than 10 g/l [2]. Today, over 90% of industrial enzymes are produced recombinantly to maximize product purity and

Submitted to Applied Biochemistry and Biotechnology

S. Sen · V. Venkata Dasu (✉) · B. Mandal
Centre for the Environment, Indian Institute of Technology (IIT)–Guwahati,
Guwahati, 781039 Assam, India
e-mail: veeranki@iitg.ernet.in

S. Sen
e-mail: shampa@iitg.ernet.in

B. Mandal
e-mail: bpmandal@iitg.ernet.in

productivity. Expression is carried out in fungal or bacterial hosts that have been modified to remove unwanted side activities and to maximize expression of heterologous genes, at times to levels significantly above 40 g/l [2]. The efficient application of biocatalysts requires the availability of suitable enzymes with high activity and stability under process conditions, desired substrate selectivity, and high enantioselectivity [3]. However, wild-type enzymes often need to be optimized to fulfill these requirements. Two rather contradictory tools can be used on a molecular level to create tailor-made biocatalysts: directed evolution and rational protein design [4]. Rational design usually requires both the availability of the structure of the enzyme and knowledge about the relationships between sequence, structure, and mechanism/function, and is, therefore, very information-intensive [5]. The concept of directed evolution was introduced as early as 1967 [6]. In the past several years, directed evolution has emerged as an alternative approach to rational design, enabling the improvement of structural and functional properties, such as stability and performance under different conditions (e.g., at extreme temperatures, pH, and in the presence of organic co-solvents), or changes in their reaction and substrate specificity [7]. Directed evolution is particularly well suited for 'tuning' enzyme function, that is, improving an activity that already exists at some albeit and low level. It was also shown to be useful for combining properties not necessarily found together in any naturally occurring enzyme [8]. Directed evolution implements an iterative Darwinian optimization process, whereby, the fittest variants are selected from an ensemble of random mutations [9]. Directed evolution differs from natural evolution in two key aspects, (1) natural evolution occurs under multiple and variable selection pressures, whereas directed evolution is accomplished under controlled selection pressure for predetermined functions. (2) in directed evolution, 'non-natural' functions of practical use can be obtained through the design of appropriate selection schemes, whereas natural evolution favors functions advantageous to the survival of the organism [10, 11]. An analysis of articles in the biomedical literature published between 1995 and 2004 using the National Institutes of Health Medline bibliographic database showed that the publication rate of 'directed evolution' articles has been increased steadily during the past decade from 4 per year in 1995 to about 60 per year in 2004 [1].

Although the production of an extensive and unbiased library is very important, the success of a directed evolution experiment highly depends on the method that is used to find the best mutant enzyme [12]. That is why, despite the significant advances to date on many industrially relevant enzymes, there still remains a need to improve directed evolution strategies and develop generic screening or selection tools that make the process of identifying novel enzyme activities more efficient, and also to access much greater changes to enzyme function [13]. This review discusses recent advances in the techniques and strategies used for the directed evolution of enzymes aimed at improving the quality and potential of enzyme libraries, as well as new screening tools that broaden the range of targets for directed evolution.

Why Directed Evolution

Methods for the creation of protein-encoding DNA libraries may be broadly divided into three categories. The first two categories encompass techniques that directly generate sequence diversity in the form of point mutations, insertions, or deletions. These can be divided in turn into methods where changes are made at random along a whole gene and methods that involve randomization at specific positions within a gene sequence. The first category, randomly targeted methods, encompasses most techniques in which the copying

of a DNA sequence is deliberately disturbed. These methods, which include the use of physical and chemical mutagens, mutator strains, and some forms of insertion and deletion mutagenesis, as well as the various forms of error-prone polymerase chain reaction (epPCR), generate diversity at random positions within the DNA being copied. A statistical analysis showed that existing gene mutagenesis methods are limited and highly biased viz., an average amino acid substitution per residue of only 3.15–7.4 can be achieved with current random mutagenesis methods [14].

The second category of methods targets a controlled level of randomization to specific positions within the DNA sequence. Most work using rational design focuses on mutations close to the active site, whereas directed evolution experiments often find mutations far from the active site. The enzymes in which the mutations caused the largest increase in enantioselectivity are a phosphotriesterase, horseradish peroxidase (HRP) and a lipase [15–17]. In phosphotriesterase, the most successful single mutation was Gly60Ala (amino acid residue at position 60 substituted from glycine to alanine), which increased the enantioselectivity from 21 to 11,000 (520-fold) for hydrolysis of *p*-nitrophenyl ethyl phenyl phosphate [15]. This mutation was chosen rationally from the crystal structure of phosphotriesterase with a bound substrate analog to decrease the size of the small binding pocket of the active site. For HRP, a Phe41Leu mutation increased enantioselectivity up to 35-fold for sulfoxidation of thioethers [16]. This mutation was also chosen rationally based on structures of related peroxidases to increase the accessibility of the substrate to the active site. The crystal structure of HRP has been solved and showed that Phe41 is 7.7 Å from a bound inhibitor. For the *Pseudomonas aeruginosa* lipase, the best mutation was Leu162Gly, identified through directed evolution. This single mutation increased the enantioselectivity from 1.1 to 34 towards *p*-nitrophenyl 2-methyldecanoate [17]. This amino acid residue was only 6.4 Å from the stereocenter of the substrate. Rational design is clearly information-intensive, and a significant drawback of this approach is that the structural and mechanistic data required are available for only a small number of enzymes. In addition, many of the early attempts to tailor enzymes were unsuccessful due to the fact that our knowledge of the structural basis of enzymatic catalysis is still incomplete. The power of rational redesign has been demonstrated by the generation of a faster superoxide dismutase, already one of the fastest known enzymes in nature and complete inversion of coenzyme specificities for both isocitrate and isopropylmalate dehydrogenases [18].

The original method available for altering protein properties, covalent chemical modification, has now reemerged as a powerful complementary approach to site-directed mutagenesis and directed evolution for tailoring proteins and enzymes [19]. All the methods for the redesign of enzyme structure and function are typically limited to the 20 most abundant proteinogenic amino acids. The use of chemical modification overcomes this limitation to allow virtually unlimited alteration of amino acid side chain structures [20]. Chemical modification of amino acid side chains allows a greater, almost unlimited variety of groups to be introduced, but the reactions used for their introduction are typically non-specific in nature. Thus, despite the many potential advantages, many classical methods used for protein modification create mixtures of proteins as a result of poorly discriminating or insufficiently efficient chemistry [20]. This is also exacerbated by the limited variety and multiple copy numbers of chemical functional groups in proteins. Thus, there is strong requirement for novel strategies that exploit selective and efficient protein chemistry.

While directed evolution is a powerful method to overcome some of the limitations of biocatalysts, it can take considerable time to implement [21]. On the other hand, improvements can be made through process techniques more rapidly, although they may be rather more modest. Much of the appeal of directed evolution of proteins lies in the fact

that the coding information is held in a molecular medium which is straightforward to amplify, read and manipulate, whereas the functional molecule, the protein, has a rich chemistry that provides a wide range of possible activities. The most popular targets for directed enzyme evolution to date have been activity, substrate specificity, thermal and oxidative stability, enantio-selectivity or enantio-specificity, pH range, and tolerance to solvent [21].

Directed Evolution Strategies

Error-Prone PCR (epPCR)

The most commonly used random mutagenesis method is error-prone PCR, which introduces random mutations during PCR by reducing the fidelity of DNA polymerase [22]. The fidelity of DNA polymerase can be reduced by adding manganese ions or by biasing the dNTP concentration. Use of the compromised DNA polymerase causes misincorporation of incorrect nucleotides during the PCR reaction, yielding randomly mutated products. To convert the product to a suitable form for transformation of a host strain, at least three steps are required: digestion of the product with restriction enzymes, separation of the fragments by agarose gel electrophoresis and ligation into a vector. Error-prone rolling circle amplification (RCA) method consists of only one RCA step followed by direct transformation of the host strain, and yields mutants with an adequate mutation frequency for in vitro evolution experiments (3–4 mutations per kilo base) [23]. This method has been used to increase the ceftazidime resistance of TEM-1 β-lactamase. The plasmid pUC19, which has TEM-1 β-lactamase gene, was mutated by errorprone RCA, and *Escherichia coli* DH5α transformed with the RCA product was cultured on a ceftazidime plate. Of the seven mutant pUC19 plasmids with improved ceftazidime resistance, all had mutations at R164 (to H, G or C) or D179 (to G). Both of these amino acids are located at the root of the Ω loop of the TEM-1 β-lactamase structure, which forms part of the substrate-binding domain, and mutations in these residues are known to improve ceftazidime resistance [23].

Error-prone PCR was used to expand the functionality of lipase from *Rhizopus arrhizus* (RAL). The optimum temperature of the mutant lipase was higher by 10°C than that of the wild-type RAL (WT-RAL). In addition, the thermostability characteristic of the mutant was also improved as the result of directed evolution. The half-life ($T_{1/2}$) at 50°C of the mutant exceeded those of WT-RAL by 12-fold [24]. A glutaryl acylase was engineered toward an adipyl acylase by randomising the gene using epPCR followed by a selection on a mimic substrate to obtain a single mutant having a 15-times-improved catalytic efficiency, resulting from an improved K_m and a k_{cat} similar to the k_{cat} of wild-type enzyme and was almost good enough to be used in an industrial process to produce 7-aminodesacetoxy-cephalosporanic acid (7-ADCA) from adipyl-7-ADCA [25]. The enantioselectivity of an epoxide hydrolase from *Aspergillus niger* in the hydrolytic kinetic resolution of glycidyl phenyl ether was doubled from 4.6 to 10.8 by just one round of epPCR [26].

DNA Shuffling

The process of engineering molecules with desired properties is the keystone of molecular evolution. It is achieved through a repeating process of changes in the DNA, followed by screening or selection of the optimum sequence. The changes in the DNA occur in

evolution as a result of recombination and mutagenesis. DNA shuffling describes a combination of in vitro recombination of related sequences and a low rate of random point mutagenesis in one experiment. The method originally involved digesting a large gene with DNase I to a pool of random DNA fragments [27]. These fragments can be reassembled into a full-length gene by repeated cycles of annealing in the presence of DNA polymerase. The fragments prime each other based on homology, and recombination occurs when fragments from one copy of a gene prime on another copy, causing a template switch. Wild-type subtilisin E and its thermostable mutant 1E2A genes were randomly recombined by DNA shuffling [28]. Miyazaki used the DNA fragments generated by endonuclease V digestion instead of DNase to obtain more reproducible DNA shuffling [29]. Synthetic green fluorescent protein (GFPuv; GenBank accession no. U62636 with T357C, T811A, and C812G base substitutions) was assembled from synthetic oligonucleotides by consensus shuffling protocols to reduce errors in the resultant DNA populations [30].

Staggered Extension Process

The process consists of polymerase-catalyzed priming of the DNA templates, followed by repeated cycles of denaturation and short annealing-extension cycles. During each cycle, the fragments increase in size and anneal to different templates based on sequence complementarity. The extension is interrupted by heat denaturation and then resumes during the next annealing-extension step. At this step the partially extended primers can anneal to different templates. This switching of templates generates a library of chimeric sequences. In an effort to reduce mutational bias, two different low-fidelity DNA polymerases were combined, Taq and Mutazyme, which have opposite mutational spectra [31]. As a first step, random mutants of the *Bacillus thuringiensis* cry9Ca1 gene were generated by separate error-prone polymerase chain reactions (PCRs) with each of the two polymerases. Subsequent shuffling by staggered extension process (StEP) of the PCR products resulted in intermediate numbers of AT and GC substitutions, compared to the Taq or Mutazyme error-prone PCR libraries.

Random-Priming Recombination

In this process random sequence primers are used to generate a large number of short DNA fragments complementary to different sections of the template sequence(s). Due to base misincorporation and mispriming, these short DNA fragments also contain a low level of point mutations. The short DNA fragments can prime one another based on homology, and be recombined and reassembled into full-length genes by repeated thermocycling in the presence of thermostable DNA polymerase. These sequences can be further amplified by conventional PCR and cloned into a vector for expression, followed by screening or selection. Random-priming recombination (RPR) and screening or selection can be repeated over multiple cycles in order to evolve the desired properties. Thermal stability was enhanced by the RPR method by the mutagenesis and recombination of genes RC1 and RC2 encoding thermostable *Bacillus subtilis* subtilisin E variants [32].

Heteroduplex Recombination

In this process the emphasis is on recombination occurring between non-homologous regions rather than the conventional method, in which recombination occurs between regions of homology. Whole plasmids or target fragments are used to create plasmid

heteroduplexes. The plasmids are linearized with specific restriction enzymes whose recognition sites are located outside the target region. This allows the plasmids to be differentiated from one another. The heteroduplex form (circular) is repaired by the mismatch-repair system and is transformed much more efficiently than the linear homoduplex form. For short target sequences (where the restriction sites cannot be used to identify the desired sequence), asymmetric PCR sequences are generated from one of the two strands. These single stranded sequences are used for-annealing, increasing the efficiency of recombination. The genes encoding truncated green fluorescent protein (GFP) were recombined with this method to reconstruct the full-length gene that restored the characteristic fluorescence [33].

Random Chimeragenesis on Transient Templates

RACHITT begins with DNase I fragmentation and size fractionation of single-stranded DNA, and hybridization in the absence of polymerase to a complementary single-stranded scaffold [34]. Any overlapping fragments leave single-stranded overhangs that are trimmed down. The gaps between fragments are filled in; the fragments are then ligated, yielding a pool of full-length, diversified single strands hybridized to the scaffold. This scaffold, which is synthesized so as to include uracil (U), can be efficiently fragmented so as to preclude its amplification, it is replaced with PCR by a new strand that is complementary to the diversified strand, and the whole is amplified [35]. RACHITT was used to evolve the gene *dszC*, encoding dibenzothiophene mono-oxygenase (DBT-MO), a key enzyme of the diesel biodesulfurization pathway used in refinery-level biodesulfurization of fossil fuels. Genes, *dszC*, from two different bacteria, one with a higher substrate affinity and/or substrate range, and the other with a higher specific reaction rate for a particular substrate type were successfully recombined. The two genes were 89.9% identical and encoding proteins with 38 amino acid differences. After screening relatively few clones (175 of the 35,000 clone library) without selection, chimeras with 60–320% improvement in flux through the enzyme pathway were isolated [36].

Recombined Extension on Truncated Templates

Recombined extension on truncated template (RETT) generates random recombinant gene library by template switching of unidirectionally growing polynucleotides from primers in the presence of unidirectional single-stranded DNA (ssDNA) fragments pool used as templates. This method does not use DNA endonucleases for generation of shuffling blocks. Instead, it makes unidirectional ssDNA fragments by either DNA polymerase in the presence of random primers or serial deletion with exonuclease. These unidirectional ssDNA fragments only act as templates in PCR, not as primers. Chimeric chitinases that were more thermostable than the parental enzyme were obtained by RETT-based recombination [37].

The Incremental Truncation for the Creation of Hybrid Enzymes

Incremental truncation is a method for creating a combinatorial library containing one base pair deletions of a gene or gene fragment of interest. In this protocol, truncations are introduced in opposite directions on fragments from two different genes in two separate reactions [38]. The sets of truncated DNA molecules from each digestion are ligated to each other with DNA ligase. The resulting “fusions” are cloned as chimeric molecules. The

library of cloned fusions is transformed into bacteria and used for further experiments (e.g., phage display, enzymatic activity assay, etc.). An improvement over incremental truncation for the creation of hybrid enzyme (ITCHY) is SCRATCHY (ITCHY combined with DNA shuffling) [39]. The approach combines two methods for recombining genes: ITCHY and DNA shuffling. First, ITCHY is used to create a comprehensive set of fusions between fragments of genes in a DNA homology-independent fashion. This artificial family is then subjected to a DNA-shuffling step to augment the number of crossovers. SCRATCHY libraries were created from the glycinamide–ribonucleotide formyltransferase (GART) genes from *E. coli* (purN) and human (hGART) [38].

Degenerate Oligonucleotide Gene Shuffling

The degenerate oligonucleotide gene shuffling (DOGS) procedure requires the design of perfectly complementary pairs of primers [40]. Each primer has a nondegenerate core flanked by both 5′ and 3′ degenerate ends, referred to as complementary degenerate-end primers (CDE primers). The 3′ degenerate core gives each CDE primer their template-binding specificity, whereas the nondegenerate region acts as a stabilizing clamp in subsequent rounds of the PCR. The 5′ degenerate end is not required to contribute to the binding efficiency of the CDE primer during PCR. However, it plays a pivotal role in allowing efficient binding and subsequent overlap extension of separate PCR products (gene segments) generated using, respectively, the forward or the reverse CDE primers. The nondegenerate core of individual CDE primers is generally based upon the corresponding coding sequence of one gene designated the parental gene for shuffling. This results in the formation of chimeric fragments that retain parental sequence at the points of segment overlap. This technique was illustrated with a diverse family of beta-xylanase genes that possess widely different G and C contents [41].

Random Drift Mutagenesis

The RNDM has been developed to determine whether a phenotype that is derived from the interaction of multiple amino acids might require the accumulation and interaction of neutral mutations (neutral in isolation) and adaptive mutations. The procedure uses iterative misincorporation mutagenesis but no screening for adaptive mutations occurs. Instead, screening is only done for retained ability (whether unchanged, improved, or reduced) to catalyze the hydrolysis of a substrate. This procedure is intended to provide high-speed screening of mutants for retained activity without tedious assay procedures and allows a comprehensive examination of sequence space for superior mutants. It is a platform technology applicable to any protein for which there is a colorimetric or fluorescent assay. For example, the precipitation of indigo as the result of hydrolysis of an indoxyl group as the sorting signal combined with flow cytometric analysis [41]. Bergquist et al. [41] explained the utility of RNDM for directed evolution of β-glucosidase of *Caldicellu losiruptor saccharolyticus* using the substrate Imagene Green.

Sequence Saturation Mutagenesis

The SeSaM is a four-step method that can saturate every single nucleotide position of the target sequence with all four standard nucleotides. In the first step, a pool of DNA fragments with a random length distribution is generated. In the second step, DNA fragments are ‘tailed’ at 3′-termini with universal bases using terminal transferase. In the

third step, the elongated DNA fragments are extended in PCR to the full-length genes by using a single-stranded template and a reverse primer. In the fourth step, a concluding PCR is used to replace the universal bases with standard nucleotides. Using enhanced green fluorescence protein as the model system and deoxyinosine as the universal base, the concept of the SeSaM method was proved by sequencing 100 genes and achieved a random distribution of mutations with the mutational bias expected for deoxyinosine [42].

Nucleotide Exchange and Excision Technology

A gene library is amplified by PCR using deoxy uridine triphosphate (dUTP) as a fragmentation defining exchange nucleotide with thymidine, together with the three other nucleotides. The incorporated uracil bases are excised using uracil–DNA–glycosylase and the DNA backbone subsequently cleaved with piperidine. These end-point reactions require no adjustments. Polyacrylamide urea gels are used to demonstrate adjustable fragmentation size over a wide range. The oligonucleotide pool is reassembled by internal primer extension to full length with a proofreading polymerase to improve yield over Taq. The technique was demonstrated by shuffling chloramphenicol acetyltransferase gene libraries. A 33% dUTP PCR resulted in shuffled clones with an average parental fragment size of 86 bases even without employment of a fragment size separation and revealed a low mutation rate (0.1%) [43].

Pros and Cons of Different Techniques of Directed Evolution

Error-prone PCR methods remain one of the most popular approaches for generating libraries for directed evolution experiments. The ease with which mutations can be generated by modifying PCR conditions (addition of Mn^{2+}, biasing of dNTP concentrations or addition of dNTP analogues) makes these methods appealing for any laboratory that is approaching directed evolution as a means to an end. However, most of these methods produce libraries with a bias in the type of nucleotide mutations (error bias) [44], a bias in the types of amino acid changes seen in the protein (codon bias) [45], and a bias in the distribution of specific sequences in the library (amplification bias) [43].

DNA shuffling, a method of in vitro recombination was developed as a technique to generate mutant genes that would encode proteins with improved or unique functionality. DNA shuffling can be applied to sequences >1 kb, has a mutagenesis rate similar to error-prone PCR, and also works with pools of unknown sequence. Repeated cycles of any mutagenesis strategy lead to an accumulation of neutral mutations, which, for example, may make a protein immunogenic. Only with DNA shuffling it is possible to remove such neutral mutations by backcrossing with excess parental or wild-type DNA [27]. However, when this method is applied to regions of low sequence homology, recombination is relatively inefficient and only a small number of variants result. Although DNA shuffling has traditionally been used to create diversity through the combinatorial shuffling of mutations in a population, DNA shuffling also creates a subpopulation of sequences with a reduction in diversity, as correct fragments can recombine to produce error-free sequences [30].

In StEP, fragments are added in steps to the end of a growing strand. The growing strand is prevented from reaching its full length by keeping the extension time very short. This results in only partial elongation of a strand in any one-extension step. StEP can be harder for an inexperienced user to set up than DNA shuffling, as full-length templates are included in the StEP reaction [45]. Once optimized for a specific thermal cycler, primers,

and template, StEP can be easier to perform than DNA shuffling as fewer steps are involved. The major difficulty with both DNA shuffling and StEP is that they rely on the annealing of a growing DNA strand to a template. Sequences can, therefore, only be recombined when they are similar enough to allow annealing, and crossovers will occur preferentially where the template sequences are most similar.

Compared to DNA shuffling, the RPR technique has several advantages. (a) RPR can use single-stranded polynucleotide templates without an intermediate step of synthesizing the whole second strand. (b) Gene reassembly is generally easier with the RPR technique, which employs random priming synthesis to obtain the short DNA fragments. Furthermore, because DNase I hydrolyzes double-stranded DNA preferentially at sites adjacent to pyrimidine nucleotides, its use in template digestion may introduce a sequence bias into the recombination. (c) The synthetic random primers are uniform in their length and lack sequence bias. The sequence heterogeneity allows them to form hybrids with the template DNA strands at many positions so that, at least in principle, every nucleotide of the template should be copied or mutated at a similar frequency during extension. (d) The random-priming DNA synthesis is independent of the length of the DNA template. DNA fragments as small as 200 bases can be primed equally well, as large DNA molecules such as linearized plasmids or λ DNA. This offers unique potential for engineering small peptides. (e) Because the parent polynucleotide serves solely as the template for the synthesis of nascent, single-stranded DNA, 10–20 times less parent DNA is needed as compared to DNA shuffling [43].

Heteroduplex recombination provides a convenient addition to existing DNA recombination methods (DNA shuffling) and should be particularly useful for recombining large genes or entire operons [33]. This method can be used to create libraries of chimeric polynucleotides and proteins for directed evolution to improve their properties or to study structure–function relationships. This method neither suffers the limitations of PCR-based approaches nor requires transformation with multiple gene fragments.

RACHITT is a technique that is conceptually similar to StEP and DNA shuffling but is designed to produce chimera with a much larger number of crossovers. RACHITT exploits a bottom-strand template from only one parent and only top-strand fragments of other parents. This prevents parental fragments from reannealing to their own complementary strands. In sexual PCR, priming among fragments of the same gene is favored and is known to decrease the number of crossovers and increase the proportion of nonchimeric parents regenerated in recombinant libraries. RACHITT helps to obtain undetectable levels of siblings, relatively few inactive proteins, and no unshuffled parental clones. It also yields an unprecedented predominance of more highly mosaic clones and the ability to effect frequent recombination between close and even adjacent alleles [34].

RETT has several advantages over other in vitro recombination techniques [37]. Compared with DNA shuffling and RACHITT, which use enzymatic cleavage (most commonly DNase I digestion) to generate shuffling blocks, RETT may increase randomness in recombination process because random-priming synthesis and exonuclease III digestion do not cause sequence bias in generating unidirectional ssDNA fragments. In RETT process, the recombinational synthesis and template-switching of growing primers are carried out under normal PCR conditions of annealing and DNA elongation unlike StEP, which uses extremely abbreviated annealing and DNA elongation conditions. RETT may also generate more random library than StEP because it is not influenced by sequence-specific pause sites met by DNA polymerase like StEP process.

In contrast to DNA shuffling and related methods, ITCHY does not rely on the parental genes containing regions of DNA sequence homology to create crossovers. Instead,

incremental truncation libraries are generated by digestion of the parental genes with exonuclease III under controlled conditions. ITCHY allows one to create comprehensive fusion libraries between fragments of genes without any sequence dependency. However, the main drawback of the method, as well as similar techniques is that members of these libraries contain only one crossover per gene. As suggested, the DNA shuffling of ITCHY libraries could potentially introduce multiple crossovers between the genes of interest by preserving ITCHY crossovers (prepositioned crossovers) in the starting material and by recombining regions of homology between genes [38]. This combination of ITCHY and DNA shuffling has been named SCRATCHY.

The DOGS procedure demonstrates that it is possible to shuffle members of a gene family that are not particularly closely related and still generate chimeric molecules at a high-enough frequency that comprehensive and time consuming screens are not necessary [41]. This procedure has the advantage of avoiding the use of endonucleases for gene fragmentation before shuffling and allows the use of random mutagenesis of selected segments of the gene as part of the procedure. This procedure can be combined with RNDM for wider exploration of the sequence space of shuffled genes [41].

Compared with currently used epPCR methods, SeSaM has the advantage of being completely independent of the mutational bias of DNA polymerases [42]. Moreover, SeSaM can target a nucleotide species of the selected sequence, and each nucleotide species can be exchanged in a controlled manner, as SeSaM regulates the mutational spectra through a universal base. In addition, the fragment distribution of a DNA library is controllable.

The key advantages of NExT method are (1) calculable experimental setup aided by a computer program, (2) reproducible end-point reactions without adjustments, (3) no gel purification required, (4) efficient reassembly with a proofreading polymerase, (5) gene recombination including very short fragments of only a few bases, (6) low error rate, and (7) practically no contamination with unshuffled clones [43].

Future Aspects

Till date, whatever methods have been used for directed evolution, none of them are without disadvantage or difficulty. In the well-established protocol of Stemmer, DNase is used to fragment DNA requiring careful optimization of the digest conditions, e.g., time, temperature, amount of nuclease, and DNA [27]. Other methods such as the staggered extension process [45] and random priming [32] are limited by the DNA composition, and matters are complicated further by the lack of controllability of the range of fragment sizes generated. Methods such as RACHITT [35] also require DNase digests and are even more labor intensive. The race for the best method is still on. Simple comparisons can be helpful but need to be taken with caution, as the gene length, the homology of the shuffled gene libraries and the intended crossover rate would have to be taken into account. Besides the homology-dependent methods, which are related to the presented data, homology-independent methods have also been developed based on DNA fragment fusion (e.g., thio-ITCHY) [38]. Even with new methods for the creation of diversity in variant libraries and the firmly established methods of saturation mutagenesis and DNA shuffling, the screening or selection of those variants with the new desired or improved activity remains the most critical step in directed evolution experiments. New screens are required to enable the identification of improved enzymes from larger libraries and also to obtain the desired properties with generic methods that measure it directly. The use of phage- and ribosome-

display methods also has the potential to search much larger variant libraries. While, mechanisms for the affinity-based capture of active enzyme variants are continually being developed and improved, there is still some way to go in widening the general applicability of these methods to more useful enzyme activities. Computational approaches are also improving rapidly and will become very useful in either creating novel enzymes as starting points for directed evolution, or for defining smarter libraries that contain fewer redundant enzyme variants. Whereas for simple reactions the capability of computational methods is approaching that of genetic techniques, there is still considerable effort required to extend its use toward obtaining novel enzymes with more complex catalytic mechanisms.

Till date, whatever work has been done in this regard is very discrete. One strategy is developed and used to improve one characteristic of an enzyme, whereas some other technique is developed and used for some other property improvement for some other enzyme. In 2003, Cherry and Fidantsef [2] reviewed critically directed evolution applied to different industrially important enzymes. Further studies are to be done to establish the most efficient strategy for enzyme modification. There may not be any particular strategy or technique that will be commonly suitable for all purpose. For example, one technique may be suitable to enhance stability for one enzyme, whereas some other may be proven to be more promising for enhancing a particular enzymatic activity. The decision can be made only after applying the speediest and cheapest techniques to improve the industrially important enzymes and comparing the results critically.

References

1. Nikolaos, E. L. (2005). *Biomolecular Engineering, 22*, vii–ix.
2. Cherry, J. R., & Fidantsef, A. L. (2003). *Current Opinion in Biotechnology, 14*, 438–443.
3. Anwar, A., & Saleemuddin M. (1998). *Bioresource Technology, 64*, 175–183.
4. Bylina, E. J., Coleman, W. J., Grek, C. L., Yang, M. M., & Youvan, D. C. (2000). *Biotechnology Et Alia, 7*, 1–6.
5. Bornscheuer, U. T., & Pohl, M. (2001). *Current Opinion in Chemical Biology, 5*, 137–143.
6. Mills, D. R., Peterson, R. L., & Spiegelman, S. (1967). *Proceedings of the National Academy of Sciences of the United States of America, 58*, 217–224.
7. Tao, H., & Cornish, V. W. (2002). *Current Opinion in Chemical Biology, 6*, 858–864.
8. Schmidt-Dannert, C., & Arnold, F. H. (1999). *Trends Biotechnology, 17*, 135–136.
9. Roodveldt, C., Aharoni, A., & Tawfik, D. S. (2005). *Current Opinion in Structural Biology, 15*, 50–56.
10. Schmidt-Dannert, C. (2001). *Biochemistry, 40*, 13125–13136.
11. Williams, G. J., Nelson, A. S., & Berry (2004). *Cellular and Molecular Life Sciences, 61*, 3034–3046.
12. You, L., & Arnold, F. H. (1996). *Protein Engineering, 9*, 77–83.
13. Hibbert, E. G., & Dalby, P. A. (2005). *Microbial Cell Factories, 4*, 29.
14. Wong, T. S., Tee, K. L., Hauer, B., & Schwaneberg, U. (2004). *Nucleic Acids Research, 32*, 3e26.
15. Chen-Goodspeed, M., Sogorb, M. A., Wu, F., & Raushel, F. M. (2001). *Biochemistry, 40*, 1332–1339.
16. Ozaki, S. I., & Ortiz de, M. P. R. (1994). *Journal of the American Chemical Society, 116*, 4487–4488.
17. Reetz, M. T., Wilensek, S., Zha, D., & Jaeger, K. E. (2001). *Angewandte Chemie. International Edition in English, 40*, 3589–3591.
18. Chen, R. (2001). *Trends Biotechnology, 19*, 13–14.
19. DeSantis, G., & Jones, J. B. (1999). *Current Opinion in Biotechnology, 10*, 324–330.
20. Davis, B. G. (2003). *Current Opinion in Biotechnology, 14*, 379–386.
21. Hibbert, E. G., Baganz, F., Hailes, H. C., Ward, J. M., Lye, G. J., Woodley, J. M., et al. (2005). *Biomolecular Engineering, 22*, 11–19.
22. Pritcharda, L., Corneb, D., Kella, D., Rowland, J., & Winson, M. (2005). *Journal of Theoretical Biology, 234*, 497–509.
23. Fujii, R., Kitaoka, M., & Hayashi, K. (2004). *Nucleic Acids Research, 32*, 19 e145.
24. Niu, W., Li, Z., Zhang, D., Yu, M., & Tan, T. (2006). *Journal of Molecular Catalysis, B: Enzymatic, 43*, 33–39.

25. Otten, L. G., Sio, C. F., Vrielink, J., Cool, R. H., & Quax, W. J. (2002). *Journal of Biological Chemistry, 277*, 42121–42127.
26. Reetz, M. T., Torre, C., Eipper, A., Lohmer, R., Hermes, M., Brunner, B., et al. (2004). *Organic Letters, 6*, 177–180.
27. Stemmer, W. P. C. (1994). *Proceedings of the National Academy of Sciences of the United States of America, 91*, 10747–10751.
28. Zhao, H., & Arnold, F. H. (1997). *Nucleic Acids Research, 25*, 1307–1308.
29. Miyazaki, K. (2002). *Nucleic Acids Research, 30*, 24 e139.
30. Binkowski, B. F., Richmond, K. E., Kaysen, J., Sussman, M. R., & Belshaw, P. J. (2005). *Nucleic Acids Research, 33*, 6 e55.
31. Vanhercke, T., Ampe, C., Tirry, L., & Denolf, P. (2005). *Analytical Biochemistry, 339*, 9–14.
32. Shao, Z., Zhao, H., Giver, L., & Arnold, F. H. (1998). *Nucleic Acids Research, 26*, 681–683.
33. Volkov, A. A., Shao, Z., & Arnold, F. H. (1999). *Nucleic Acids Research, 27*, 18 e18.
34. Pelletier, J. N. (2001). *Nature Biotechnology, 19*, 314–315.
35. Lee, S. H., Ryu, E. J., Kang, M. J., Wang, E., Piao, Z., Choi, Y. J., et al. (2003). *Journal of Molecular Catalysis, B: Enzymatic, 26*, 119–129.
36. Coco, W. M., Levinson, W. E., Michael, J. C., Harm, J. H., Aldis, D., Pienkos, P. T., et al. (2001). *Nature Biotechnology, 19*, 354–359.
37. Lutz, S., Ostermeier, M., & Benkovic, S. J. (2001). *Nucleic Acids Research, 29*, 4 e16.
38. Lutz, S., Ostermeier, M., Moore, G. L., Maranas, C. D., & Benkovic, S. J. (2001). *PNAS, 98*, 11248–11253.
39. Zhao, H., Chockalingam, K., & Chen, Z. (2002). *Current Opinion in Biotechnology, 13*, 104–110.
40. Gibbs, M. D., Nevalainen, K. M. H., & Bergquist, P. L. (2001). *Gene, 271*, 13–20.
41. Bergquist, P. L., Reeves, R. A., & Gibbs, M. D. (2005). *Biomolecular Engineering, 22*, 63–72.
42. Wong, T. S., Roccatano D., Zacharias, M., & Schwaneberg, U. (2006). *Journal of Molecular Biology, 355*, 858–871.
43. Muller, K. M., Stebel, S. C., Knall, S., Zipf, G., Bernauer, H. S., & Arndt, K. M. (2005). *Nucleic Acids Research, 33*(13), e117.
44. Miyazaki, K., & Arnold, F. H. (1999). *Journal of Molecular Evolution, 49*, 716–720.
45. Neylon, C. (2004). *Nucleic Acids Research, 32*(4), 1448–1459.