# Chapter 3. Directed Evolution Tools in Bioproduct and Bioprocess Development

**Sheryl B. Rubin-Pitel[a], Catherine M-H. Cho[b], Wilfred Chen[b]*, and Huimin Zhao[a,c]***

[a]Department of Chemical and Biomolecular Engineering, University of Illinois, Urbana, IL 61801

[c]Department of Chemistry, University of Illinois, Urbana, IL 61801

[c]Center for Biophysics and Computational Biology, University of Illinois, Urbana, IL 61801

[b]Department of Chemical and Environmental Engineering, University of California, Riverside, CA 92521

## 1. INTRODUCTION

Darwinian evolution is often regarded as a negative process, with natural selection merely purging the less competitive among an uneven population. George Gaylord Simpson, in his 1944 treatise *Tempo and Mode in Evolution*, countered that selection does not "simply kill off or permit to live fixed types of organisms delivered to it…Selection also determines which among the millions of possible types of organisms will actually arise, and it is therefore a truly creative factor in evolution" [1, 2]. Man has harnessed the creative power of selection for thousands of years through the process of classical breeding, thereby molding a plethora of livestock, crops, and companion animals to fulfill collective needs or desires. Only in the past decades have researchers exploited the positive nature of selection at the scale of biological macromolecules or single cells rather than an entire organism. Evolutionary methods have been applied to achieve improved or novel characteristics in nucleic acids, proteins, viruses, and bacterial strains. The general strategy of mimicking natural evolution in the laboratory is termed "directed (molecular) evolution" or "*in vitro* evolution" [3]. Since it was first described in the 1970s, directed evolution has grown in popularity and found a wide range of applications across industry, academia, and medicine.

One of the earliest examples of "directed evolution" was *in vitro* evolution of nucleic acids carried out by Mills *et al* [4]. However, it was not until several decades later that the concept of directed evolution was applied for the *in vitro* engineering of proteins on the molecular level [5–7]. More recently, directed evolution techniques have been applied to the

engineering of more complex subjects such as metabolic pathways, viruses, and bacterial genomes [8–14].

The method of directed evolution involves an iterative strategy. The procedure begins by determining a target biomolecule, metabolic pathway, or organism, and a desired phenotypic goal. A diverse library of mutants is generated *in vivo* or *in vitro* through methods that mirror the strategies of traditional evolution: introduction of random mutations in the genetic material and/or "sexual" gene recombination. A high-throughput screening or selection method is used to identify improved progeny among the library, which are subsequently used as parents in a second round of the cycle. The process is repeated until the phenotypic goal is achieved, or when no further improvement of the phenotype is observed despite repeated iterations.

Microorganisms and the enzymes they hold have been exploited by man for thousands of years, for example, in the production of food products through fermentation. Recent decades have seen an expanding role for enzymes and microbes in the development of bioproducts and bioprocesses in industry, organic synthesis, and medical therapies. While existing enzymes may hold great potential, their use is often hindered by the low stability, lack of specificity, and low catalytic efficiency encountered when exporting these highly evolved biological entities into non-natural environments and applications [3]. Directed evolution provides the means to enhance the performance of enzymes under requisite process conditions and customize the reactions they catalyze. Directed evolution tools have been used to improve synthesis yields of desired products, limit or expand substrate specificity, alter cofactor specificity, and improve stability over a wider range of temperature and pH.

The methods, applications, and achievements of directed evolution have been described in many recent review articles and books [3, 15–18]. This review will focus only on the strategies for diversity generation that are applicable to the development of bioproducts and bioprocesses via directed evolution. The application of directed evolution to functional nucleic acids is of limited relevance compared to the engineering of protein catalysts and improved strains, and so will not be addressed herein; interested readers are referred to several recent review articles [19–21]. Additionally, high-throughput screening and selection methods for sorting through diverse mutant libraries will not be discussed in this chapter.

## 2. DIRECTED EVOLUTION TOOLS FOR DIVERSITY GENERATION

By natural evolution, the Earth began with an ancient unicellular ancestor and filled its skies, land, and oceans with a vast array of organisms. Damage to genetic material by irradiation or oxidation, failures of DNA replication, recombination, or repair, and invasion by parasitic DNA elements led to substitutions, deletions, insertions, duplications, inversions, and translocation of DNA segments from one chromosome to another [22]. These events—predominantly accidents or mistakes—led fortuitously to the existence of human life and the amazing diversity we experience. It must be noted, however, that evolution is a creative but sluggish process. The *in vivo* mechanisms of evolution mentioned above are highly inefficient, producing negligible changes in gene structure or function after thousands or even millions of years. For organisms possessing more advanced DNA replication and repair

machinery, it has been suggested that a typical protein (of 400 amino acids) would suffer a random amino acid change in the germline approximately once every 200,000 years [22]. Thus, while nature has created a bountiful variety of life, it should not be surprising what can be accomplished when one has four billion years to tinker.

To recreate evolution in the laboratory, the mechanisms of natural evolution must be accelerated such that meaningful diversity can be created and selected in a much shorter timeframe, mere days to weeks being favored. This defines the two-fold strategy of directed evolution: rapid generation of a functionally diverse collection of mutants, and rapid identification of the best performers among them [3]. The two natural evolutionary processes which have been adapted for *in vitro* evolution are gene recombination and random mutagenesis. Gene recombination refers to the exchange of blocks of genetic material among two or more DNA strands, and is often considered the "sexual" component of evolution. Recombination can be divided into four main types: (i) homologous recombination, where recombination occurs between two genes with high sequence identity, (ii) non-homologous recombination, where recombination occurs between two DNA sequences with little or no sequence identity, (iii) reciprocal recombination, in which a symmetrical exchange of genetic material occurs between two DNA strands, and finally (iv) site-specific recombination, in which specialized nucleotide sequences exhibiting some degree of target site specificity are moved between nonhomologous sites within a genome [22, 23]. Stemmer introduced DNA shuffling [5, 6], the first *in vitro* homologous recombination method, in 1994. Since that time, numerous other homologous recombination methods have been developed, as well as methods for recombination of genes lacking sequence identity.

Random mutagenesis refers to changes in the genome resulting from improper DNA replication or inadequate repair of DNA damage following events such as irradiation, exposure to oxidative or alkylating agents, and natural deamination of cytosine. Random mutation can be divided into five categories: (i) transitions, which involve substitution of a purine nucleotide by another purine, or a pyrimidine by a second pyrimidine, (ii) transversions, which involve substitution of a purine nucleotide by a pyrimidine, or vice-versa, (iii) deletions, in which one or more nucleotides are eliminated from a gene, (iv) insertions, in which one or more extra nucleotides are incorporated into a gene, and (v) inversions, which involve the 180º rotation of a double-stranded DNA segment of two base pairs or longer [3, 24]. *In vitro* random mutagenesis methods have been developed to generate substitutions, deletions, and insertions. One of the simplest and most popular directed evolution tools, error-prone polymerase chain reaction (PCR) takes advantage of the fallibility of DNA polymerase to generate random base pair substitutions. Similarly, mutator strains of *E. coli* exploit defective DNA repair machinery and also create random point mutations.

Random mutagenesis and gene recombination methods are compared in Fig. 1. Random mutagenesis methods use a single gene as a starting point, and introduce mutations along the entire gene or in predefined sites or regions. Nucleotides may be substituted randomly, generating point mutations, inserted into the sequence, or deleted. As many point mutations will be deleterious, a low mutation rate is necessary to preserve protein function. In contrast, gene recombination typically begins with a collection of parent molecules and exploits the

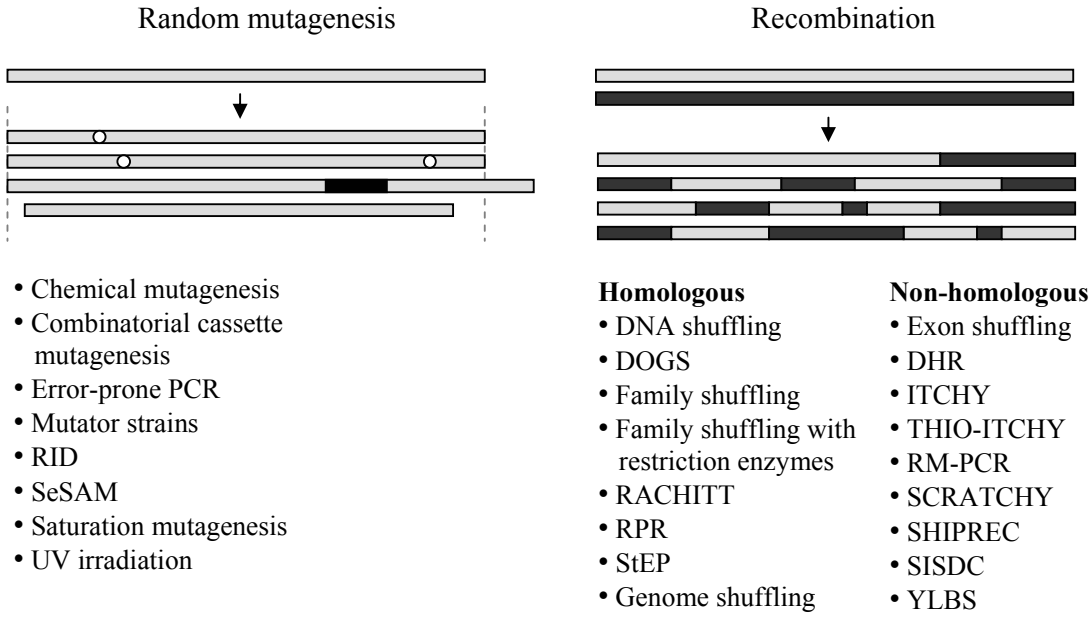


Fig. 1. Comparison of (a) random mutagenesis and (b) recombination strategies.

existing variation among them to create novel sequences. The pool of parent genes could be an assortment of mutant progeny resulting from random mutagenesis of a single parent (DNA shuffling), or a set of closely related genes from different strains or species (family shuffling). Typically the parent sequences are fragmented and the resulting short strands are pieced together into complete genes. The chimeric progeny are created with contributions from at least two parents. Unlike random mutagenesis, in which mutation events are restricted, maximal recombining of the genes, or crossover, may be desired. Some recently developed gene shuffling strategies merge gene recombination and random mutagenesis by using PCR to generate full-length progeny sequences from the gene fragments and also to amplify them [3]. In this way, misincorporations by DNA polymerase can provide additional diversity in the recombination library. As shown in Fig. 1, numerous experimental protocols have been formulated for each diversification strategy. These protocols will be described in further detail below.

### 2.1. *In vitro* mutagenesis methods

Random mutagenesis strategies are relatively simple and popular methods for generating molecular diversity. Early mutagenesis protocols involved creation of point mutations in a parent gene by damaging the DNA strand, for example by treatment with chemical mutagens including hydroxylamine [25], nitrous acid [25], methoxylamine [26], and sodium bisulfite [27], or by ultraviolet irradiation [25]. These methods tend to be inefficient, because they are typically discontinuous and can cause substantial cell damage if performed *in vivo* [28]. Point

mutations can also be induced by error-prone PCR [29–32] or mutator strains of *E. coli* [28, 33–36]. The aforementioned mutagenesis methods will generate point mutations across the entire length of the parent gene. Other schemes have been developed that allow mutations to be focused in specific sites or regions of the parent DNA sequence. Some of the most common random mutagenesis methods are listed in Table 1.

Table 1
Random mutagenesis methods

| Method | Advantages | Disadvantages | References |
|---|---|---|---|
| Chemical mutagenesis | Simplicity | Accumulates deleterious mutations<br>Low mutation level<br>Low efficiency<br>Limited amino acid substitutions<br>Cannot control mutation rate | [26, 27] |
| Mutator strains | Simplicity | Low mutation level<br>Accumulates deleterious mutations<br>Progeny must be transferred to DNA repair-competent strain for screening<br>Limited amino acid substitutions<br>Cannot control mutation rate | [28, 33–36] |
| Error-prone PCR | Simplicity | Accumulates deleterious mutations<br>Limited amino acid substitutions<br>Polymerase bias | [29, 30] |
| Saturation mutagenesis | Simplicity<br>Mutate specific site(s) in a gene<br>Access all 20 amino acids | Limited diversity generation<br>Gene sequence required | [44] |
| Sequence saturation mutagenesis (SeSAM) | Overcomes polymerase bias<br>Target a specific nucleotide in a sequence | Small fragments not mutagenized<br>Four PCR reactions needed to remove bias<br>Limited amino acid substitutions | [41] |
| Random insertion / deletion (RID) | Flexible<br>Insert or remove an amino acid randomly<br>Access all 20 amino acids | Point mutations may occur<br>Time-consuming and technically challenging | [45] |

*2.1.1. Mutagenic strains*

Propagating a gene of interest in a mutational strain represents the simplest method of random mutagenesis. Mutator strains of *E. coli* are deficient in one or more DNA repair genes, leading to single base substitutions at a rate of approximately 1 mutation per 1000 base pairs and mutation cycle [36]. This mutation rate is fairly low, and mutations may occur outside of the gene of interest, across the plasmid vector and bacterial genome. To generate a mutant library, the gene of interest is cloned into a plasmid or phagemid and propagated in mutator *E. coli* cells through a limited number of replications [33, 34]. The plasmid or phagemid library is then rescued from the mutator strains and stably expressed in a DNA-repair competent strain for amplification and selection of the mutant progeny; if necessary, the procedure of mutation, amplification, and selection is repeated until the desired phenotype is achieved [34]. The process is relatively easy, and commercial mutator strains such as XL1-Red (Stratagene, La Jolla, California) are available. Mutator *E. coli* strains find only modest use today, despite comparable methods being more time-consuming, difficult to implement, and expensive [33]. Rather, error-prone PCR is by far the most popular random mutagenesis method.

*2.1.2. Error-prone polymerase chain reaction*

Error-prone PCR relies on the misincorporation of nucleotides by DNA polymerase to generate point mutations in a gene sequence. The accuracy of DNA polymerase can be adjusted *in vitro* by addition of manganese ion into the PCR reaction mixture [37]. Additionally, PCR mutagenesis protocols have been designed which incorporate nucleotide analogs or use "mutagenic polymerases" [38–40]. Any one of these strategies, or a combination, can be incorporated into a PCR reaction to achieve a specific mutation rate. The relative simplicity and versatility of error-prone PCR have propelled it to become the most widely used mutagenesis strategy, but it suffers from several drawbacks. First, due to the redundancy of the genetic code, error-prone PCR methods are limited in their ability to create diversity at the protein level. From a single amino acid, an average of less than six other amino acids can be obtained, rather than all 19 possible substitutions [41]. Second, DNA polymerases used in PCR reactions have mutational biases that limit diversity. *Taq* polymerase and Mutazyme (Stratagene, La Jolla, California) will preferentially induce mutations at AT base pairs over GC base pairs [41]. Further, the majority of mutations are transitions, and amino acid substitutions, when present, tend to preserve the characteristics of the original residue [3, 41]. Third, in order to maintain adequate numbers of functionally active progeny, the mutation rate is kept low, generally only 1-3 mutations per 1000 base pairs [38]; these few mutations are unlikely to occur next to each other [41]. Finally, nucleotide analogs are not incorporated by DNA polymerases efficiently, and their incorporation tends to occur at certain favored sites [41]. Thus, nucleotide analog methods may result in low mutation frequencies, limited diversity, and low product yield [41–43].

*2.1.3. Saturation mutagenesis*

The limitations of error-prone PCR mutagenesis may be overcome by site-directed mutagenesis and saturation mutagenesis methods. Site-directed mutagenesis uses an

oligonucleotide primer to introduce a single-base pair substitution at a specified position in a gene [46]. Saturation mutagenesis involves the substitution of all possible amino acids randomly at a predetermined residue or continuous series of residues in the protein of interest [3]. Several strategies of saturation mutagenesis have been developed, including combinatorial cassette mutagenesis [47, 48], recursive ensemble mutagenesis [49, 50], scanning saturation mutagenesis [51–53], and codon cassette mutagenesis [54, 55]. More recently, Wong *et al.* [41] described the method of sequence saturation mutagenesis (SeSaM), which is able to randomize a DNA sequence at every nucleotide position through use of a universal base.

*2.1.4. Mutagenesis by random insertion or deletion*

Random mutagenesis can also be accomplished by insertion or deletion of nucleotides from a target gene sequence. Random insertion or deletion leads to a net change in the length of the gene of interest, opening a new realm of diversity that cannot be reached by point mutation alone. In the past random insertion has been accomplished by exploiting naturally occurring transposable elements or by random elongation mutagenesis, in which peptide "tails" are fused to a gene [56–58]. Transposable elements have several advantages for random mutagenesis: transposons can be designed to carry selectable markers such as antibiotic resistance or phage immunity; the occurrence of transposon insertion can be controlled; mutagenesis is highly efficient; and the occurrence of secondary mutations is low [25]. However, transposons cannot be used to create random deletions. Random elongation mutagenesis can also create a functionally diverse library of mutants, but is limited to fusing additional peptides to the C-terminus of a protein, and also cannot facilitate random deletions. A more recent method developed by Murakami *et al.* [45] can introduce both insertions and deletions at any position in a gene sequence. Random insertion/deletion (RID) mutagenesis allows the deletion of up to 16 bases from random sites on the target gene and subsequent insertion of a random or predetermined sequence of any number of bases at the same position [45]. This method can be used to replace three randomly selected base pairs by a specific codon, a mixture of codons, a restriction site, or by four-base codons for non-natural amino acids [45]. Though a more versatile method, RID mutagenesis is also technically challenging, time consuming, requires a large amount of template DNA, and is difficult to iterate [3].

Because most mutations will be neutral or deleterious, a low mutation rate is maintained in random mutagenesis methods. As a result random mutagenesis uncovers diversity in a very small region of sequence space, and is unlikely to foster detection of synergistic effects of multiple beneficial mutations in a single gene [3]. Furthermore, the small evolutionary steps taken by random mutagenesis may not be sufficient to allow the wholesale changes required, for example, to evolve a novel activity in a target gene. Neutral or deleterious point mutations may also accumulate in a library of progeny. Such nonessential mutations may make the resulting protein immunogenic [6]. Finally, random mutagenesis methods are restricted by the use of a single parent as a starting point. Although it can be clearly defined as to which of a collection of existing enzymes has the most favorable characteristics, it is impossible to predict which enzyme has the greatest potential for improvement through directed evolution. Use of only a single parent represents a fundamental flaw of random mutagenesis methods,

and limits the evolutionary potential of progeny [59]. This shortcoming of random mutagenesis is overcome by recombination methods.

### 2.2. *In vitro* homologous recombination methods

Homologous recombination methods mimic the "sexual" recombination of genetic material that rearranges maternal and paternal chromosomes in germ cell DNA. Such recombination increases the genetic variation among a population and is vital to the continued evolution of organisms in response to an ever-changing environment [22]. Unlike mutagenesis methods, which create novel diversity at the molecular level, recombination methods simply rearrange existing gene sequences to exploit the diversity that naturally exists among a population. While the results of random point mutations are unpredictable and often deleterious, recombination provides the advantage that all diversity present in a mutant sequence was drawn from folded and fully functional proteins. Recombination also makes it possible to remove neutral or deleterious mutations, which accumulate during random mutagenesis, by backcrossing progeny with excess parental or wild-type DNA [5]. Table 2 compares the advantages and disadvantages of various homologous recombination methods.

#### *2.2.1. DNA shuffling and family shuffling*

Stemmer introduced DNA shuffling, the first homologous recombination method, in 1994 [5, 6]. DNA shuffling involves the digestion of a gene by DNaseI into random fragments, and the reassembly of those fragments into a full-length gene by primerless PCR: the fragments prime on each other based on sequence homology, and recombination occurs when fragments from one copy of a gene anneal to fragments from another copy, causing a template switch, or crossover event. This method was used to fragment and recreate a single gene, to recombine a group of point mutants, and to recombine several related genes. The reassembly process introduces point mutations at a rate similar to error-prone PCR, due to misincorporations by the DNA polymerase. These mutations add to the diversity of the mutant library, and any unnecessary mutations can later be eliminated by backcrossing to parent or wild-type sequences. If necessary, use of a high fidelity DNA polymerase allows the rate of random point mutations to be reduced drastically [60]. Several years after the introduction of DNA shuffling, the method was applied to the recombination of a family of related genes from various species. This new method, termed family shuffling, applied DNA shuffling to a group of naturally occurring homologous genes rather than laboratory-created mutants. Crameri *et al* found that family shuffling significantly accelerated the rate of functional enzyme improvement in a single recombination-selection cycle [61]. Although they are powerful methods, DNA shuffling and family shuffling are not without limitations. Shuffling methods require the presence of zones of relatively high sequence homology surrounding regions of diversity [6]. Additionally, significant biases are found in where crossover events occur and in which parents are involved: crossover tends to occur in regions of higher homology, and among parents which share greater sequence identity [62]. Bias is also introduced by nonrandom gene fragmentation by the DNaseI enzyme [63]. All of these factors limit the diversity created in a shuffled library. In extreme cases, lack of homology

among parents can lead to the majority of reconstructed "shuffled" sequences entirely representing a single parent [64].

Table 2
Homologous recombination methods

| Method | Advantages | Disadvantages | References |
|---|---|---|---|
| DNA shuffling | Robust, flexible<br>Back-crossing to parent removes non-essential mutations | Biased to crossovers in high homology regions<br>Low crossover rate<br>High percentage of parent | [5, 6] |
| Family shuffling | Exploit natural diversity<br>Accelerates functional enzyme improvement | Biased to crossover in high homology regions<br>Need high sequence homology in the gene family<br>High percentage of parent | [61] |
| Family shuffling using restriction endonucleases | Lower representation of parent in a library | Point mutations<br>Low crossover rate | [65] |
| DOGS | Reduced parental genes in a shuffled library<br>Lower homology required<br>Can bias representation of parent in library | Point mutations<br>Frameshifts may occur<br>Relatively low crossover rate | [64] |
| RACHITT | No parent genes in a shuffled library<br>Higher rate of recombination<br>Recombine genes of low sequence homology | Complex<br>Requires synthesis and fragmentation of single-stranded complement DNA | [66] |
| RPR | Compatible with ssDNA<br>DNase I-independent<br>Removes sequence bias<br>Independent of template length<br>Less parent DNA needed | Need gene sequence<br>Biased point mutations also occur | [67] |
| StEP | Simplicity | Need high homology<br>Low crossover rate<br>Need tight control of PCR | [68, 69] |
| Synthetic shuffling | Greater flexibility<br>Increased diversity | Chemical synthesis of many degenerate oligonucleotides | [70] |
| Genome shuffling | Improve complex, poorly understood phenotypes<br>Adapt to multiple phenotypic goals<br>New strains not GMOs | Possibility of novel antibiotic resistance or pathogenicity<br>Genome flexibility restricted by metabolic network rigidity | [10, 11, 13] |

Numerous homologous gene recombination methods have been designed to address the limitations of family shuffling. Kikuchi *et al* described a method for gene shuffling using endonuclease digestion at restriction sites, rather than DNaseI digestion; however, sequence homology surrounding the digested restriction sites is still required for overlap extension to occur [64, 65]. Degenerate oligonucleotide gene shuffling (DOGS) utilizes a PCR reaction with degenerate-end, complementary primer pairs to shuffle genes with limited sequence similarity and G + C content [64]. Additionally, by modifying primer extension conditions the progeny can be biased towards one or more of the parent genes [64].

*2.2.2. Oligonucleotide- and oligonucleotide primer-based methods*

Several other alternatives to DNA shuffling have been established, including random-priming *in vitro* recombination (RPR) [67], the staggered extension process (StEP) [69], and synthetic shuffling [70] . Recombination by RPR utilizes elongation from random sequence primers to generate a collection of small DNA fragments complementary to different areas of the template sequence(s) [67]. The method of RPR is shown in Fig. 2. Similar to DNA shuffling, fragments prime each other based on sequence homology and are reconstructed into a full length sequence by DNA polymerase-catalyzed elongation [67]. StEP also utilizes primer elongation to generate small DNA fragments for recombination. In StEP recombination, flanking primers are annealed to a denatured template and allowed to extend for a very short time period; cycles of denaturation and short annealing/elongation are repeated [68, 69]. Crossover occurs when partially extended primers anneal randomly to different templates based on homology, and extend further [68, 69]. The cycle of denaturation/annealing/elongation is continued until full-length genes are created, and if necessary, a traditional PCR amplification can be used to increase the yield of chimeric progeny [68, 69]. In synthetic shuffling, the fragments to be shuffled are degenerate oligonucleotides that are chemically synthesized and encode all the variations in a family of homologous genes [70]. Compared to fragmentation-based DNA shuffling formats, synthetic shuffling is more flexible in the construction of permutated protein libraries and also introduces more diversity into these libraries. For example, this method does not require physical starting genes and can incorporate optimal codon usage or known beneficial mutations.

*2.2.3. Random Chimeragenesis on Transient Templates (RACHITT)*

In contrast to the above methods, RACHITT does not utilize thermocycling, strand switching, or staggered extension of primers [66]. Instead, a uracil-containing parent gene is made single-stranded to serve as a scaffold for the ordering of top-strand fragments of additional, homologous parent gene(s), and recombination occurs when fragments from different parent genes hybridize to the scaffold. *Pfu* DNA polymerase 3'-5' exonuclease activity removes the unhybridized 5' or 3' overhanging "flaps" created by fragment annealing, and also fills gaps between the annealed fragments using the transient scaffold as a template. The template strand is then eliminated by treatment with uracil-DNA-glycosylase before applying the template-chimera hybrid to PCR, resulting in amplification of double stranded, homoduplex chimerical gene sequences. The process of RACHITT recombination

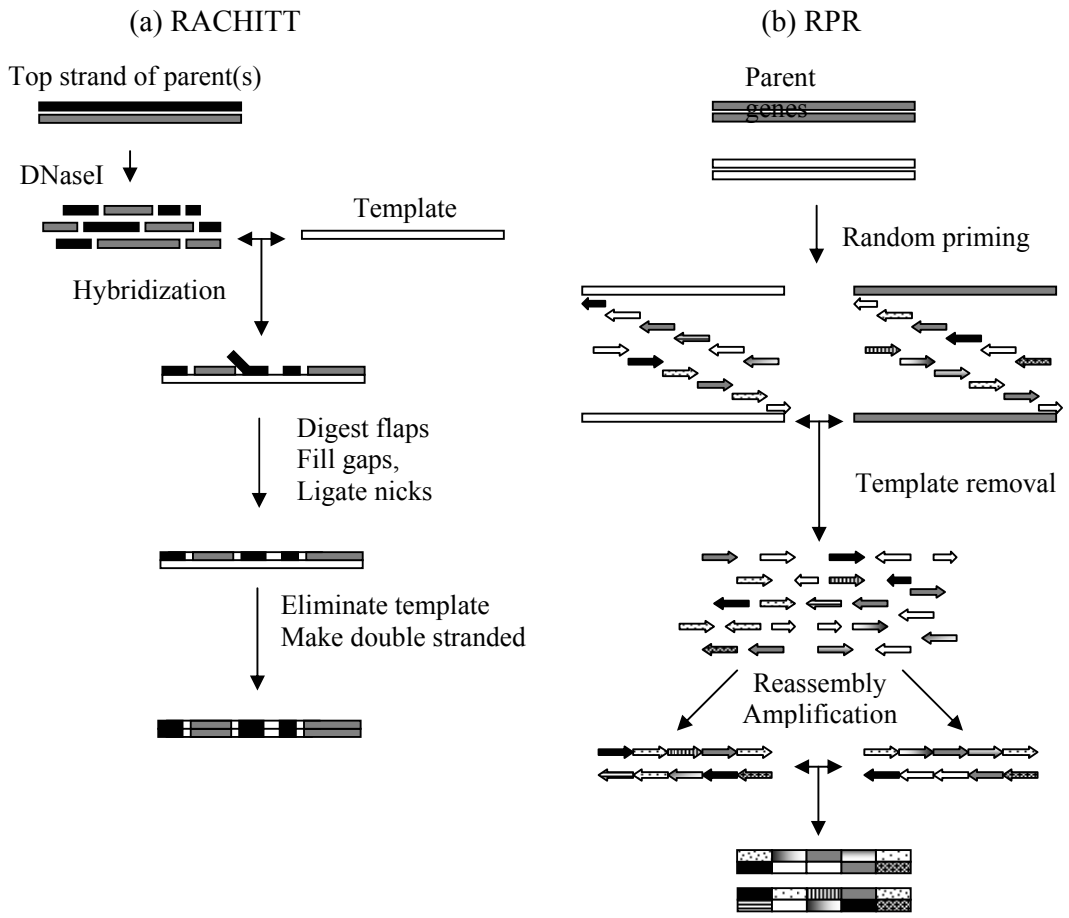


Fig. 2. Random homologous DNA recombination by (a) RACHITT and (b) RPR.

is illustrated in Fig. 2. RACHITT provides a significantly higher rate of crossover compared to other family shuffling methods, with an average of 14 crossovers per gene versus one to four crossovers for most other methods. RACHITT also generates 100% chimerical progeny with no duplications of recombination pattern in chimerical genes. Although the benefits of this method are obvious, its use may be limited by its complexity and the requirement to create single stranded gene fragments as well as single stranded, uracil-DNA template.

*2.2.4. Genome shuffling*

The technique of genome shuffling emerged recently as an alternative method for the optimization of industrial production strains [10–13]. Strain optimization is typically achieved by classical strain improvement techniques, which involve rounds of recombination and/or mutagenesis followed by screening for a desired phenotype, selective breeding, and rational schemes of metabolic engineering. Like other recombination methods, genome shuffling exploits the diversity that already exists among a population of organisms and allows back-crossing of progeny to parents to eliminate non-essential or deleterious gene changes that may accumulate during rounds of random mutagenesis. In genome shuffling,

homologous recombination of genomes is achieved by protoplast fusion. The process of protoplast fusion in bacteria was reviewed by Gokhale *et al.* [71]. Protoplast fusion first involves the isolation of protoplasts from cells by digestion of the cell wall in the presence of osmotic stabilizers. Isolation of protoplasts from gram negative organisms is generally more difficult than gram positive due to their complex cell wall. Fusion is achieved by mixing of the parental protoplasts and addition of a fusogen, such as polyethylene glycol (PEG). PEG stimulates aggregation of protoplasts, and fusion events occur after the PEG is diluted or washed away. The PEG-treated protoplasts are subsequently plated onto appropriate media and the fused protoplasts are identified by selection. Protoplast fusion has also been described for the production of improved yeast strains [72].

The technique of genome shuffling by protoplast fusion offers several advantages. Protoplast fusion is a well-established technique that is applicable to an array of organisms including bacteria and both lower and higher eukaryotes. Protoplast fusion also provides simultaneous changes at different positions throughout the entire genome, without the requirement of genome sequence data [11]. This technique is therefore particularly applicable to the engineering of complex or poorly understood phenotypes, engineering of multiple phenotypic goals simultaneously, and engineering of organisms with limited availability of molecular biological tools and sequence information. Additionally, strains engineered by protoplast fusion, a form of natural homologous recombination, are not considered to be "genetically modified" [13], and therefore avoid the additional regulations and public distaste reserved for genetically modified organisms (GMOs). Genome shuffling by protoplast fusion has already shown promise in the improvement of industrial production strains. Zhang *et al.* showed the utility of genome shuffling to *Streptomyces* species [13], which are commonly employed in the commercial production of antibiotics. Genome shuffling of existing *Streptomyces fradiae* industrial strains was used to create a new strain with higher production of the polyketide antibiotic tylosin. By only two rounds of genome shuffling, strain improvement was equivalent to the results achieved after 20 rounds of classical strain improvement (CSI; sequential random mutagenesis and screening). Patnaik *et al* applied genome shuffling to the improvement of acid tolerance of *Lactobacillus* species, which are exploited in the commercial production of lactic acid [10]. The improved strain produced by genome shuffling showed faster growth and higher lactic acid production at a lower pH value, with tolerance to acidic pH approximately 5-fold higher than the wild type.

### 2.3. *In vitro* non-homologous recombination methods

The requirement of high sequence identity among parent genes limits the application of homologous recombination methods. In many situations it may be desirable to shuffle genes with low or even no evident sequence identity. The increasing availability of protein structures has also indicated that many enzymes with little or no sequence homology can have high protein structural homology; it may also be useful to shuffle such proteins, but would be inefficient with homologous recombination methods [73]. The intron-exon organization of eukaryotic genomes also facilitates non-homologous gene recombination [74]. A single exon or a collection of exons often encodes a distinct protein domain, and it is advantageous to swap domains and create combinatorial libraries of proteins. By recombining genes within

non-homologous introns, exchange of protein domains is permitted while still ensuring the integrity of the coding DNA sequence, the exons. Such "exon shuffling" reflects a mechanism of natural evolution which swapped exons among unrelated genes, and led to existent proteins from distant families sharing conserved functional domains.

Table 3
Non-homologous recombination methods

| Method | Advantages | Disadvantages | References |
|---|---|---|---|
| Exon shuffling | Preserves exon function | Requires known intron-exon organization of target gene<br>Limited diversity | [73] |
| ITCHY | Eliminate recombination bias<br>Structural knowledge not needed | Limited to two parents<br>Significant fraction of progeny out-of-frame<br>Complex, labor-intensive | [74] |
| THIO-ITCHY | Same advantages as ITCHY<br>Combines recombination and random mutagenesis<br>Simplified ITCHY method | Same disadvantages as ITCHY<br>Incorporated dNTP analogs may complicate further experimentation | [75] |
| SCRATCHY | Eliminate recombination bias<br>Structural knowledge not needed | Limited to two parents<br>Significant fraction of progeny out-of-frame<br>Complex, labor-intensive | [74] |
| DHR | High recombination rate<br>Eliminate recombination bias | Synthesize numerous complementary oligonucleotides<br>Gene sequence needed | [76] |
| RM-PCR | Unbiased incorporation of variable size DNA fragments | Frame shifts may occur<br>Mutants may be longer or shorter than expected | [77] |
| SHIPREC | Crossovers occur at structurally related sites | Limited to two parents<br>Single crossover per gene | [78] |
| SISDC | Recombines fragments without bias<br>Ligates fragments in a desired order | Gene sequence needed<br>Must engineer endonuclease sites into parent genes<br>Must synthesize numerous oligonucleotide primers | [79] |
| YLBS | Recombines variable size DNA fragments<br>Shuffles large fragments such as exons or domains | Non-stoichiometric incorporation of DNA fragments<br>Frame shifts may occur<br>Low product recovery | [80] |

*2.3.1. Exon shuffling*

The method of *in vitro* exon shuffling has been described by Kolkman and Stemmer [73]. The general scheme of this recombination method is shown in Figure 3. Exon shuffling requires the creation of DNA fragments containing exons or combinations of exons that encode a protein domain. The exon fragments are amplified with a mixture of synthetic chimeric oligonucleotides, causing the fragments to be spliced together randomly. These spliced fragments are then assembled by primerless PCR, where individual fragments prime against each other to recreate a full-length gene. Recombination occurs when a chimeric oligonucleotide connects an exon from one parent gene to a second exon from a different parent gene. The diversity in an exon shuffling library is controlled by the number of modules which are recombined, and the number of homologs that are included for each module; in some cases, the availability of homologous domains may limit the creation of a shuffled library. The diversity of an exon shuffling library can also be controlled experimentally through the design of the chimeric oligonucleotides, facilitating certain connections between domains but not others, or by modifying the molar ratio of domain-encoding fragments to control the stoichiometry of the individual domains in the progeny. As with other recombination methods, additional diversity can be created in the library by introducing random point mutations, insertions, or deletions. Rearranging the order of domain-encoding exons also creates novel diversity.

*2.3.2. Incremental truncation methods*

Several non-homologous recombination methods have been designed to facilitate the shuffling of genes with insufficient sequence identity for homologous shuffling techniques. Ostermeier *et al* introduced the technique of incremental truncation for the creation of hybrid enzymes (ITCHY) [74], in which random fusion of domains from two parent enzymes is used to generate novel chimeras. Because it is difficult to predict at what locations two protein domains should be fused for maximal performance or novel activity, ITCHY libraries contain every combination. This is achieved through controlled digestion of DNA by exonuclease III to generate a collection of all possible truncated fragments of the parent genes, followed by blunt-end ligation of the fragments to form hybrid proteins. Tight control of exonuclease activity is required in addition to frequent removal of digested fragments and quenching of the reaction, in order to collect a variety of fragment lengths. Thus, ITCHY becomes a time-consuming and laborious method. ITCHY is also limited by other factors, including that only two parents can be used, gene length is not conserved by random fusion of fragments, recombination predominantly occurs at sites which are not structurally related, and only a fraction of crossover events connect fragments from two parent genes at sites where the sequences align [78]. A modified incremental truncation method, termed THIO-ITCHY, introduces a simpler procedure for creating fragment libraries from the parent genes [75]. THIO-ITCHY entails the random, low-frequency incorporation (spiking) of α-phosphothioate nucleotide analogs into the parent genes. The α- phosphothioate nucleotides protect the DNA from exonuclease activity, and so ensure the desired variation in truncation length without timed removal and quenching of digestion aliquots. If a DNA polymerase is used to incorporate nucleotide analogs, then random mutagenesis can also be integrated into

the library via error-prone PCR conditions. Additional diversity can also be created by shuffling of two ITCHY libraries. This method, termed SCRATCHY, was described by Ostermeier *et al* [74].

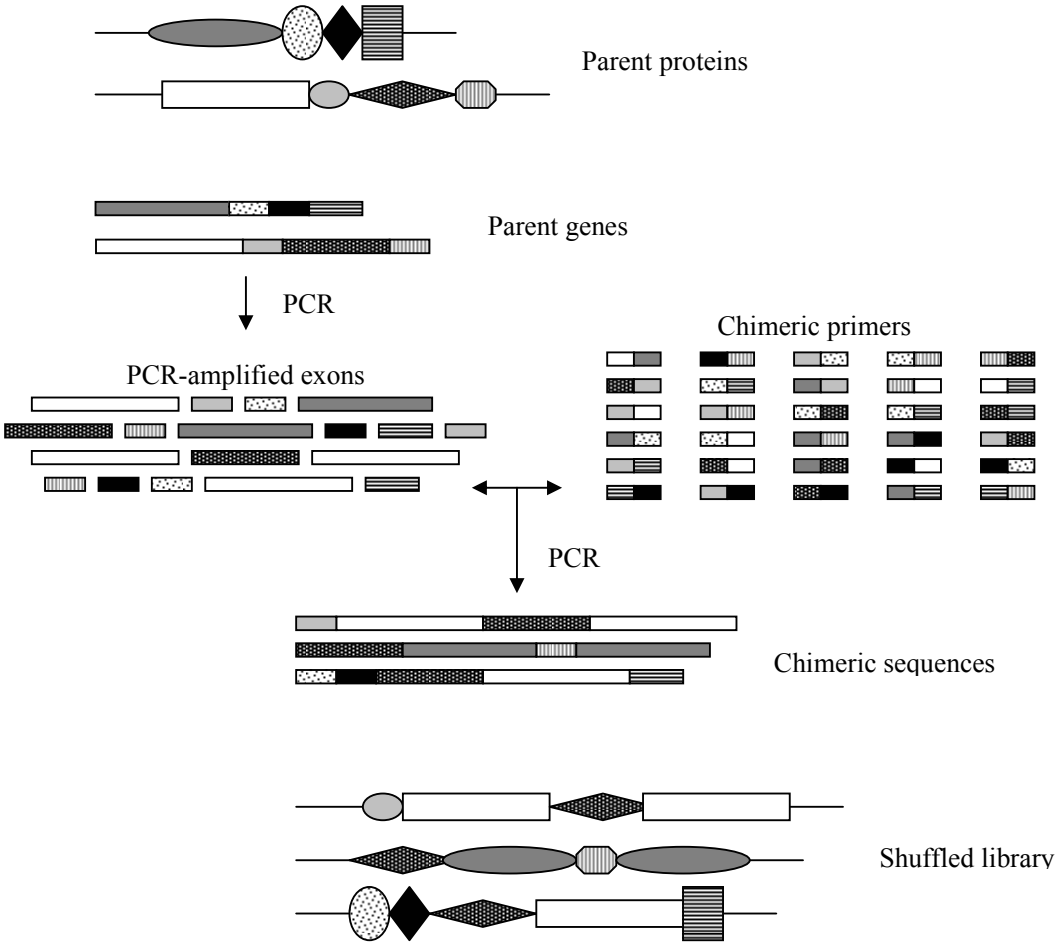


Fig. 3. Method of non-homologous recombination by exon shuffling.

*2.3.3. Sequence Homology-Independent Protein Recombination (SHIPREC)*

Another method conceptually similar to ITCHY is sequence homology-independent protein recombination (SHIPREC), which was used by Sieber *et al.* to create a library of interspecies hybrids from a membrane-bound human cytochrome P450 and a soluble bacterial P450 from *Bacillus megaterium* [78]. SHIPREC also involves the fusion of two parent genes and creation of a library of random length fragments. Two parent genes are joined in the first step, with a linker between them containing a unique restriction site. The fusion product is then digested with DNase I to form a library of random fragments, and fragments of length corresponding to the size of either parent gene are isolated and treated with S1 nuclease to

produce blunt ends. The fragments are then circularized by blunt-end ligation and relinearized by digestion at the restriction site within the linker sequence; by this method, the gene at the 5' end of the dimer will now be at the 3' end and provide the C-terminus of the hybrid protein. SHIPREC is superior in its ability to create fusion hybrids where sequence alignment is maintained, but is limited to only one crossover event and also permits only two parent genes. Other methods for recombination of genes with limited sequence identity include degenerate homoduplex recombination (DHR) [76], random multirecombinant PCR (RM-PCR) [77], sequence independent site-directed chimeragenesis (SISDC) [79], and Y-ligation based shuffling (YLBS) [80]. A comparison of the advantages and disadvantages of these methods is provided in Table 3.

## 3. APPLICATIONS OF DIRECTED EVOLUTION TOOLS

### 3.1. Applications in enzyme engineering

Enzyme biocatalysis is increasingly viewed as a competitive and cost-effective alternative for the manufacturing of fine chemicals, pharmaceuticals, and agrochemical intermediates. Enzymes have major appeal for catalysis because of their high turnover number and refined level of selectivity, particularly in the synthesis of single-enantiomer compounds. Until recently, most of the successful industrial applications of enzymes have been limited to hydrolytic enzymes such as lipases, esterases, acylases, and hydantoinases. This situation is changing with the emergence of enzymes that perform a wide range of transformations, including asymmetric reduction, oxidation, and carbon–carbon bond formation [81–84].

Historically, microbial culture has been the most important route for enzyme discovery, even though only a small fraction of all microbes can be sampled by this method [85]. This classical strategy has rapidly been replaced by high-throughput methods based on genomic sequence discovery [86]. However, even these strategies are limited by the natural ability of enzymes to perform only a well-defined set of transformations. Directed evolution has been used with great success in recent years for the diversification of gene sequences and optimization of enzyme phenotypes [15, 87]. By surveying the available gene sequence space, specific traits are created through screening of libraries consisting of $10^4$–$10^{10}$ individuals. In all cases, optimal assay development is critical to the success in optimizing the fitness landscape of these enzymes.

#### *3.1.1. Improving catalytic activity/stability*

One of the most popular applications of directed evolution is to improve enzyme activity or stability under well-defined process conditions. By screening for initial activity and residual activity at an elevated temperature, both the thermostability and activity of mesophilic subtilisin E [88] and *p*-nitrobenzyl esterase [89] were significantly increased. Similarly, a directed evolution approach was successfully used to enhance the specific activity of a thermophilic 3-isopropylmalate dehydrogenase at lower temperatures [90], demonstrating the flexibility of this method in tailoring desirable enzymatic traits. In addition to thermal properties, enzymes with enhanced activity have also been created. In one example, directed evolution was used to improve the hydrolysis rate of organophosphorus hydrolase for several

poorly degraded pesticides (25 to 700 fold) [91, 92], suggesting that this approach may be useful in generating other variants that could rapidly decontaminate structurally similar chemical warfare agents. Directed evolution approaches have also been used to enhance catalytic activities in non-natural environments such as organic solvents, for organic-phase syntheses. Moore and Arnold [93] created several *p*-nitrobenzyl esterase variants that were up to 60-fold more active in 30% dimethylformamide. Another recent work using error-prone PCR was described to achieve a five-fold improvement in the amylase activity at pH 10, an alkaline pH required for the paper industry and as a detergent additive [94].

*3.1.2. Expanding specificity*

Another application of directed evolution is to fine-tune the specificity of enzymes. Many successful examples have been demonstrated that are useful for the production of important industrial products. The *E. coli* D-2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase, which catalyzes the highly specific reversible aldol reaction on D-configurated KDPG substrates, was subjected to DNA shuffling and screening, and one variant was isolated capable of accepting both D- and L-glyceraldehyde as substrates in a non-phosphorylated form [95]. More recently, the P450 BM-3 monoxygenase, normally specific for medium-chain fatty acids, has been evolved to accept small hydrocarbon substrates and convert them at very high rates [96].

Perhaps the most dramatic success in this area is the use of directed evolution to create novel specificity and activity. Sun *et al.* [97] used combinatorial mutagenesis to change the substrate specificity of galactose oxidase to use glucose as a substrate. One variant (with only three point mutations) exhibited activity against D-glucose and oxidized other primary and secondary alcohols. Family shuffling of two homologous biphenyl dioxygenases created several variants with enhanced substrate specificity towards *ortho*-substituted polychlorinated biphenyls [98] and other aromatic compounds such as benzene [99], suggesting the feasibility to expand the biodegradability of other highly recalcitrant pollutants.

In addition to substrate specificity, product specificity can also be altered by directed evolution. Wild-type toluene 4-monooxygenase (T4MO) of *Pseudomonas stutzeri* OX1 oxidizes toluene to *p*-cresol (96%) and oxidizes benzene sequentially to phenol, catechol, and 1,2,3-trihydroxybenzene. To synthesize novel dihydroxy and trihydroxy derivatives of benzene and toluene, DNA shuffling of the alpha-hydroxylase fragment of T4MO (TouA) and saturation mutagenesis of the TouA active site residues were used to generate random mutants [100]. Several variants were isolated to form 4-methylresorcinol, 3-methylcatechol, and methylhydroquinone from *o*-cresol, whereas wild-type T4MO formed only 3-methylcatechol. These variants also formed catechol, resorcinol, and hydroquinone from phenol, whereas wild-type T4MO formed only catechol. These reactions show the potential synthesis of important intermediates for pharmaceuticals.

*3.1.3. Changing stereo- and enantio-selectivity*

Often the production of enantiomerically pure compounds is of extreme importance, particularly in the pharmaceutical industry. In this respect, directed evolution has been useful in creating enzymes with desirable enantioselectivity. May *et al*. were the first to demonstrate

the feasibility to invert the enantioselectivity of D-hydantoinase to generate an enzyme that has enhanced selectivity towards L-5-(2-methylthioethyl)hydantoin [101]. Similarly, inversion of enantioselectivity of a lipase was achieved towards (R)-selectivity with E =30 (comparing to E = 1.1 for the wild type enzyme) [100]. Perhaps the best industrial success was demonstrated with the synthesis of *cis*-(1S, 2R)-indandial, a key precursor of an inhibitor of HIV protease, by toluene dioxygenase [102]. In three rounds of screening, several variants with up to three-fold decrease in production of the undesirable 1-indenol (only 20% from 60%) were obtained. In addition to enantioselectivity, the steroselectivity can be easily altered by directed evolution. Williams *et al*. [103] demonstrated that stereospecificity of tagatose-1,6-bisphosphate aldolase can be altered by 100-fold via three rounds of DNA shuffling and screening. The resulting mutant catalyzes the formation of carbon-carbon bonds with unnatural diastereoselectivity, where the >99:<1 preference for the formation of tagatose 1,6-bisphosphate was switched to a 4:1 preference for the diastereoisomer, fructose 1,6-bisphosphate.

### 3.2. Applications in pathway engineering

Metabolic pathway engineering is a rapidly growing area with great potential to impact industrial biocatalysis [104]. As enzymes are the central components in metabolic pathways, the strategy for the generation of sequence diversity and the screening/selection methods can be readily applied for pathway engineering. Directed evolution can be used to optimize an existing pathway, but the ability of this evolutionary approach to create new pathways that are capable of synthesizing novel compounds may be the most promising aspect for the future.

Carotenoids are important antioxidants and food additives that have been attracting commercial attention in recent years. Unfortunately, the synthesis of useful quantities from conventional chemical routes or from natural microorganisms is often costly and limited. The colorful nature of carotenoids makes them easy to detect via high-throughput screening. As a result, gene clusters for carotenoid synthesis have been introduced into *E. coli* and by performing directed evolution on two phytoene desaturases and two lycopene cyclases, several novel carotenoids were produced [105]. More recently, the C30 carotene synthase CrtM from *Staphylococcus aureus* was subjected to one round of mutagenesis and screening, and variants capable of synthesizing C-40 carotenoids were identified [106]. This plasticity of CrtM with respect to its substrate and product range highlights the potential in creating further new carotenoid backbone structures. As a result, previously unknown C-45 and C-50 carotenoid backbones were obtained from the appropriate isoprenyldiphosphate precursors [107]. Similar strategies have been applied successfully to evolve pathways for porphyrin synthesis [108].

Polyketides belong to a second class of important bioactive compounds and efforts have been directed towards the generation of novel structures for uses as antibiotics or anti-cancer agents. The modular nature of the polyketide synthases (PKS) renders polyketide synthesis inherently amenable to directed evolution strategy, particularly in the engineering of novel polyketide structures. Typically a given PKS can generate only one product. However, Shen *et al*. [109] reported that a minimal PKS from *Streptomyces coelicolor* is capable of generating more than 30 different structures, suggesting the flexibility in engineering a large

number of useful structures by a single PKS. By systematically deleting domains of the erythromycin PKS or exchanging domains with other PKS modules, several variants were obtained that are capable of generating more than 50 different polyketides [110]. These examples imply the feasibility of creating entirely novel products via directed evolution of metabolic pathways.

Directed evolution can also be used as a powerful tool in optimizing an entire metabolic pathway. Functional evolution of an arsenic resistance operon has been accomplished by three rounds of shuffling and selection, resulting in cells that grew in 0.5 M arsenate, a 40-fold increase in resistance [111]. Ten mutations were located in *arsB*, encoding the arsenite membrane pump, resulting in a 4-fold to 6-fold increase in arsenite resistance. While *arsC*, the arsenate reductase gene, contained no mutations, its expression level was increased, and the rate of arsenate reduction was increased 12-fold.

Directed evolution has also been shown to enable the construction of artificial networks of transcriptional control elements in living cells [112]. By applying directed evolution to genes comprising a simple genetic circuit, a nonfunctional circuit containing improperly matched components can evolve rapidly into a functional one. Such an approach is likely to result in a library of genetic devices with a range of behaviors that can be used to construct more complex genetic circuits.

## 4. ALTERNATIVES TO DIRECTED EVOLUTION

### 4.1. Rational approaches to enzyme evolution

In addition to combinatorial approaches to enzyme evolution, many different methods for rational protein design have been devised. The strengths of directed evolution and rational design are highly complementary and may be combined to provide significant advantages over the use of a single approach.

#### *4.1.1. Rational computational design*

Computational methods [113] represent a widely used approach for rational protein design. These methods rely on the use of a force field to identify amino acid sequences that are optimal for stabilizing a protein backbone. The major drawback of these methods is that the number of possible sequences will often exceed what could be exhaustively searched by existing computing power. However, recent developments in powerful search algorithms have generated new excitement in this area [114]. Specifically, experimental data are incorporated to iteratively improve the empirical force field calculations [115]. In one example, a novel active site for activated ester hydrolysis was computationally designed into the scaffold of thioredoxin [116]. Ranking of different active site designs on the basis of substrate binding resulted in an enzyme with the ability to catalyze the predicted reaction. Even though the activity is quite modest, this example demonstrates the utility of the computational approach in designing proteins with the desired catalytic functions.

Computational methods have also been used to guide experimental design for directed evolution. There have been several studies to optimize the mutation or recombination rate with respect to the number of mutants that can be screened [117]. Other methods focused on

algorithms that target the diversity of regions that will preserve structures during the evolutionary process. Inverse folding algorithms were used to predict the protein sequences that are amenable to mutagenesis without perturbing the overall protein structure [118].

*4.1.2. Site-directed mutagenesis*

Rational design by site-directed mutagenesis has enjoyed some success in the past, mostly in term of engineering enzyme specificity. Our ability to redesign enzyme mechanisms or completely new reactions, however, remains a difficult task. With the ever increasing knowledge of protein structure and function, site-directed mutagenesis could become a powerful complementary approach to directed evolution.

Active site substitution by site-directed mutagenesis based on structural information has been the conventional approach in protein engineering. This strategy has been used successfully in reshaping substrate or co-factor specificity and reactivity [119]. Another powerful tool in enzyme design is based on the use of structural homology to graft the desired properties from one enzyme into another via site-directed mutagenesis. Very often, new catalytic residues are introduced to alter enzyme mechanism and function. The introduction of a Ser-His-Asp triad into a peptidyl-prolyl isomerase resulted in a remarkably efficient proline-specific endopeptidase [120]. In addition, homology-driven design provides information for more accurate and detailed physical models for future rational enzyme design.

The ability to carry out novel, unique chemistries has been achieved by incorporating new catalytic groups using related proteins with similar folding structures. For example, four substitutions were enough to confer an oleate-hydroxylase activity on an oleate-desaturase [121].

**4.2. Semi-rational approaches to enzyme evolution**

*4.2.1. SISDC: Sequence-independent site directed chimeragenesis*

Conventional methods for gene shuffling are useful only if the parental genes share high levels of sequence identity (usually 70%). This means that including a relatively diverse pool of parental sequences, potentially allowing the exploration of more vast sequence space, is often difficult to accomplish. As mentioned above, several methods have been reported for creating chimeric protein libraries independent of homology, such as ITCHY, SCRATCHY, and SHIPREC. These methods, however, generate large numbers of non-functional sequences, and therefore diverse libraries of functional proteins have not been demonstrated convincingly. Recently, the Arnold group presented a simple and general method called sequence-independent site-directed chimeragenesis (SISDC) that allows for recombination of distantly related proteins at multiple discrete sites with little sequence bias and in which all targeted fragments were recombined in the desired order [79]. If desired, various modifications, such as insertion, deletion, and rearrangement, can be incorporated easily. A complementary computational algorithm called SCHEMA was also developed to estimate the disruption by the inheritance of amino acid from different parents upon recombination [122].

*4.2.2. GSSM: Gene site saturation mutagenesis*

Gene site saturation mutagenesis (GSSM) technology is a unique method for rapid laboratory evolution of proteins whereby each amino acid of a protein is replaced with each of the other 19 naturally occurring amino acids [123]. This is accomplished at the genetic level through the use of degenerate primer sets, comprising either 32 or 64 codon variants, for each amino acid residue. Subsequent use of standard methods for DNA replication generates a library of genes possessing all codon variations required for saturation mutagenesis of the original gene. A unique application of this method was demonstrated to evolve a nitrilase as a process-scale enantioselective biocatalyst [124]. Comprehensive mutagenesis and screening using LC-MS resulted in a nitrilase variant with high enantiomeric excess (*ee*) at high substrate concentrations. The essential mutation required two base changes in a single codon, which is difficult to achieve through other random mutagenesis methods.

## 5. CONCLUSION

Directed evolution tools have been increasingly used to engineer new or improved enzymes, metabolic pathways, and whole genomes for various bioprocessing applications. In the past decade, numerous molecular biology techniques have been developed to create genetic diversity through random mutagenesis and/or homologous or non-homologous recombination in the target genes, pathways and genomes. Coupled with the development of powerful high-throughput screening or selection methods, these evolutionary techniques have been successfully used to solve challenging problems in protein engineering and metabolic engineering. For the foreseeable future, directed evolution will not only remain a powerful tool for bioproduct and bioprocess development, but also a powerful research tool for solving fundamental biological problems such as the protein structure-function relationship and protein folding. In addition, directed evolution is highly complementary to rational design which capability is rapidly growing due to recent advances of structural genomics and computational biology. It seems that the combination of directed evolution and rational design represents the most powerful tool for protein engineering and metabolic engineering, and will likely become a fertile ground for innovations in the coming years.

## ACKNOWLEDGEMENTS

We thank the Office of Naval Research (N000140210725 to H.Z.), National Science Foundation (BES-0348107 to H.Z. and Graduate Research Fellowship to S.R.-P.), and National Institute of Health (CMBTG Fellowship to S. R.-P.) for supporting our work on development and applications of new directed evolution tools for protein engineering and metabolic engineering.

## REFERENCES

[1] W.M. Fitch and F.J. Ayala, Proc Natl Acad Sci U S A, 91(1994) 6717.
[2] G.G. Simpson, Tempo and Mode in Evolution, Columbia University Press, New York, N.Y., 1944.

[3] H. Zhao and W. Zha, (A. Svendsen), Evolutionary Methods for Protein Engineering, Marcel Dekker, Inc., New York, 2004, 353.
[4] D.R. Mills, R.L. Peterson, and S. Spiegelman, Proc Natl Acad Sci U S A, 58(1967) 217.
[5] W.P. Stemmer, Nature, 370(1994) 389.
[6] W.P. Stemmer, Proc Natl Acad Sci U S A, 91(1994) 10747.
[7] K. Chen and F.H. Arnold, Proc Natl Acad Sci U S A, 90(1993) 5618.
[8] N.W. Soong, et al., Nat Genet, 25(2000) 436.
[9] M. Kramer, et al., Metab Eng, 5(2003) 277.
[10] R. Patnaik, et al., Nat Biotechnol, 20(2002) 707.
[11] R. Petri and C. Schmidt-Dannert, Curr Opin Biotechnol, 15(2004) 298.
[12] G. Stephanopoulos, Nat Biotechnol, 20(2002) 666.
[13] Y.X. Zhang, et al., Nature, 415(2002) 644.
[14] M. Kleerebezem, et al., J Biotechnol, 98(2002) 199.
[15] F. Valetti and G. Gilardi, Nat Prod Rep, 21(2004) 490.
[16] H. Tao and V.W. Cornish, Curr Opin Chem Biol, 6(2002) 858.
[17] D.E. Robertson and B.A. Steer, Curr Opin Chem Biol, 8(2004) 141.
[18] K.A. Powell, et al., Angew Chem Int Ed, 40(2001) 3948.
[19] D.S. Wilson and J.W. Szostak, Annu Rev Biochem, 68(1999) 611.
[20] A. Jaschke and B. Seelig, Curr Opin Chem Biol, 4(2000) 257.
[21] G.F. Joyce, Annu Rev Biochem, 73(2004) 791.
[22] B. Alberts, et al., Molecular Biology of the Cell, Garland Science, New York, 2002.
[23] F.E. Wurgler, Mutat Res, 284(1992) 3.
[24] W.-H. Li and D. Graur, Fundamentals of Molecular Evolution, Sinauer Associates, Inc, Sunderland, MA, 1991.
[25] D. Botstein and D. Shortle, Science, 229(1985) 1193.
[26] J.T. Kadonaga and J.R. Knowles, Nucleic Acids Res, 13(1985) 1733.
[27] D. Shortle and D. Botstein, Methods Enzymol, 100(1983) 457.
[28] O. Selifonova, F. Valle, and V. Schellenberger, Appl Environ Microb, 67(2001) 3645.
[29] D.W. Leung, E. Chen, and D.V. Goeddel, Technique, 1(1989) 11.
[30] R.C. Cadwell and G.F. Joyce, PCR Methods Appl, 3(1994) S136.
[31] M. Fromant, S. Blanquet, and P. Plateau, Anal Biochem, 224(1995) 347.
[32] R.A. Beckman, A.S. Mildvan, and L.A. Loeb, Biochemistry, 24(1985) 5810.
[33] X. Lu, et al., Journal of Virological Methods, 94(2001) 37.
[34] G. Coia, P.J. Hudson, and R.A. Irving, J Immunol Methods, 251(2001) 187.
[35] E. Henke and U.T. Bornscheuer, Biol Chem, 380(1999) 1029.
[36] U.T. Bornscheuer, J. Altenbuchner, and H.H. Meyer, Biotechnol Bioeng, 58(1998) 554.
[37] H. Zhao, et al., (J.E. Davies), Methods for optimizing industrial enzymes by directed evolution, ASM Press, Washington DC, 1999, 597.
[38] J. Cline and H.H. Hogrefe, Strategies, 13(2000) 157.
[39] O.P. Kuipers, Methods Mol Biol, 57(1996) 351.
[40] M. Zaccolo, et al., J Mol Biol, 255(1996) 589.
[41] T.S. Wong, et al., Nucleic Acids Res, 32(2004) e26.
[42] J.H. Spee, W.M. de Vos, and O.P. Kuipers, Nucleic Acids Res, 21(1993) 777.
[43] M.A. Innis, et al., Proc Natl Acad Sci U S A, 85(1988) 9436.
[44] R.M. Myers, L.S. Lerman, and T. Maniatis, Science, 229(1985) 242.
[45] H. Murakami, T. Hohsaka, and M. Sisido, Nat Biotechnol, 20(2002) 76.
[46] R.B. Wallace, et al., Nucleic Acids Res, 9(1981) 3647.
[47] J.A. Wells, M. Vasser, and D.B. Powers, Gene, 34(1985) 315.
[48] J.F. Reidhaar-Olson and R.T. Sauer, Science, 241(1988) 53.
[49] S. Delagrave, E.R. Goldman, and D.C. Youvan, Protein Eng, 6(1993) 327.
[50] A.P. Arkin and D.C. Youvan, Proc Natl Acad Sci U S A, 89(1992) 7811.
[51] J.A. Maynard, et al., Methods Mol Biol, 182(2002) 149.
[52] G. Chen, et al., Protein Eng, 12(1999) 349.
[53] E.A. Burks, et al., Proc Natl Acad Sci U S A, 94(1997) 412.
[54] D.M. Kegler-Ebo, G.W. Polack, and D. DiMaio, Methods Mol Biol, 57(1996) 297.

[55] D.M. Kegler-Ebo, C.M. Docktor, and D. DiMaio, Nucleic Acids Res, 22(1994) 1593.
[56] B. Hallet, D.J. Sherratt, and F. Hayes, Nucleic Acids Res, 25(1997) 1866.
[57] Y. Kang, et al., J Bacteriol, 186(2004) 4921.
[58] T. Matsuura, et al., Nat Biotechnol, 17(1999) 58.
[59] J.E. Ness, et al., Nat Biotechnol, 17(1999) 893.
[60] H. Zhao and F.H. Arnold, Nucleic Acids Res, 25(1997) 1307.
[61] A. Crameri, et al., Nature, 391(1998) 288.
[62] J.M. Joern, P. Meinhold, and F.H. Arnold, J Mol Biol, 316(2002) 643.
[63] J. Sambrook, E.F. Fritsch, and T. Maniatis, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
[64] M.D. Gibbs, K.M. Nevalainen, and P.L. Bergquist, Gene, 271(2001) 13.
[65] M. Kikuchi, K. Ohnishi, and S. Harayama, Gene, 236(1999) 159.
[66] W.M. Coco, et al., Nat Biotechnol, 19(2001) 354.
[67] Z. Shao, et al., Nucleic Acids Res, 26(1998) 681.
[68] H. Zhao, Methods Enzymol, 388(2004) 42.
[69] H. Zhao, et al., Nat Biotechnol, 16(1998) 258.
[70] J.E. Ness, et al., Nat Biotechnol, 20(2002) 1251.
[71] D.V. Gokhale, U.S. Puntambekar, and D.N. Deobagkar, Biotech Advs, 11(1993) 199.
[72] S. Scheinbach, Biotech Advs, 1(1983) 289.
[73] J.A. Kolkman and W.P. Stemmer, Nat Biotechnol, 19(2001) 423.
[74] M. Ostermeier, A.E. Nixon, and S.J. Benkovic, Bioorg Med Chem, 7(1999) 2139.
[75] S. Lutz, M. Ostermeier, and S.J. Benkovic, Nucleic Acids Res, 29(2001) E16.
[76] W.M. Coco, et al., Nat Biotechnol, 20(2002) 1246.
[77] T. Tsuji, M. Onimaru, and H. Yanagawa, Nucleic Acids Res, 29(2001) E97.
[78] V. Sieber, C.A. Martinez, and F.H. Arnold, Nat Biotechnol, 19(2001) 456.
[79] K. Hiraga and F.H. Arnold, J Mol Biol, 330(2003) 287.
[80] K. Kitamura, et al., Protein Eng, 15(2002) 843.
[81] H.E. Schoemaker, D. Mink, and M.G. Wubbolts, Science, 299(2003) 1694.
[82] A. Schmid, et al., Nature, 409(2001) 258.
[83] O. Kirk, T.V. Borchert, and C.C. Fuglsang, Curr Opin Biotechnol, 13(2002) 345.
[84] S. Panke and M.G. Wubbolts, Curr Opin Biotechnol, 13(2002) 111.
[85] N.J. Turner, Trends Biotechnol, 21(2003) 474.
[86] M. Lafferty and M.J. Dycaico, Methods Enzymol, 388(2004) 119.
[87] H. Zhao, K. Chockalingam, and Z. Chen, Curr Opin Biotechnol, 13(2002) 104.
[88] H. Zhao and F.H. Arnold, Protein Eng, 12(1999) 47.
[89] A. Gershenson, et al., Biochemistry, 39(2000) 4658.
[90] T. Suzuki, et al., Protein Eng, 14(2001) 85.
[91] C.M.H. Cho, A. Mulchandani, and W. Chen, Appl Environ Microb, 68(2002) 2026.
[92] C.M.H. Cho, A. Mulchandani, and W. Chen, Appl Environ Microb, 70(2004) 4681.
[93] J.C. Moore and F.H. Arnold, Nat Biotechnol, 14(1996) 458.
[94] C. Bessler, et al., Protein Sci, 12(2003) 2141.
[95] S. Fong, et al., Chem Biol, 7(2000) 873.
[96] A. Glieder, E.T. Farinas, and F.H. Arnold, Nat Biotechnol, 20(2002) 1135.
[97] L. Sun, et al., Chembiochem, 3(2002) 781.
[98] F. Bruhlmann and W. Chen, Biotechnol Bioeng, 63(1999) 544.
[99] H. Suenaga, et al., J Bacteriol, 183(2001) 5441.
[100] G. Vardar and T.K. Wood, Appl Environ Microbiol, 70(2004) 3253.
[101] O. May, P.T. Nguyen, and F.H. Arnold, Nat Biotechnol, 18(2000) 317.
[102] N. Zhang, et al., Metab Eng, 2(2000) 339.
[103] G.J. Williams, et al., Proc Natl Acad Sci U S A, 100(2003) 3143.
[104] L. Rohlin, M.K. Oh, and J.C. Liao, Curr Opin Microbiol, 4(2001) 330.
[105] C. Schmidt-Dannert, D. Umeno, and F.H. Arnold, Nat Biotechnol, 18(2000) 750.
[106] D. Umeno, A.V. Tobias, and F.H. Arnold, J Bacteriol, 184(2002) 6690.
[107] D. Umeno and F.H. Arnold, J Bacteriol, 186(2004) 1531.
[108] S.J. Kwon, et al., Chembiochem, 5(2004) 1069.

[109] Y. Shen, et al., Proc Natl Acad Sci U S A, 96(1999) 3622.
[110] R. McDaniel, et al., Proc Natl Acad Sci U S A, 96(1999) 1846.
[111] A. Crameri, et al., Nat Biotechnol, 15(1997) 436.
[112] Y. Yokobayashi, R. Weiss, and F.H. Arnold, Proc Natl Acad Sci U S A, 99(2002) 16587.
[113] A.G. Street and S.L. Mayo, Structure Fold Des, 7(1999) R105.
[114] J.R. Desjarlais and N.D. Clarke, Curr Opin Struct Biol, 8(1998) 471.
[115] D.B. Gordon, S.A. Marshall, and S.L. Mayo, Curr Opin Struct Biol, 9(1999) 509.
[116] D.N. Bolon and S.L. Mayo, Proc Natl Acad Sci U S A, 98(2001) 14274.
[117] C.A. Voigt, S. Kauffman, and Z.G. Wang, Adv Protein Chem, 55(2000) 79.
[118] C.A. Voigt, et al., Proc Natl Acad Sci U S A, 98(2001) 3778.
[119] F. Cedrone, A. Menez, and E. Quemeneur, Curr Opin Struct Biol, 10(2000) 405.
[120] E. Quemeneur, et al., Nature, 391(1998) 301.
[121] P. Broun, et al., Science, 282(1998) 1315.
[122] M.M. Meyer, et al., Protein Sci, 12(2003) 1686.
[123] J.M. Short, Patent US6171820 (2001).
[124] G. DeSantis, et al., J AM Chem Soc, 125(2003) 11476.