CHAPTER #

Current Approaches for Engineering Proteins with Diverse Biological Properties

Isaac T Li, Elizabeth Pham and Kevin Truong*

Abstract

In the past two decades, protein engineering has advanced significantly with the emergence of new chemical and genetic approaches. Modification and recombination of existing proteins not only produced novel enzymes used commercially and in research laboratories, but furthermore, they revealed the mechanisms of protein function. In this chapter, we will describe the applications and significance of current protein engineering approaches.

Introduction

The ultimate goal of protein engineering is to create proteins that are perfectly suited to particular biological applications. Proteins, polymers consisting of 20 distinct amino acids, play a pivotal role in many biological activities such as catalysis, signal transduction, structural reinforcement and cell motility, to name a few. The diversity in the choice of amino acids and total length of proteins (ranging from tens to thousands of amino acids) allows for a tremendous combinatorial sequence space. Through evolutionary pressure over time, organisms have found specific sequences of amino acids that fold into unique molecular structures that determine their biochemical functions. The molecular structures of over thirty thousand proteins have been solved and this number continues to grow steadily alongside new initiatives in structural proteomics. Nevertheless, the total number of distinct proteins found in nature represents only a small fraction of the possible polypeptides that could be permutated with the 20 amino acids. By exploring the combinatorial sequence space around these natural proteins, many current approaches in protein engineering can create novel proteins with desirable properties. In this chapter, we will discuss these approaches and their diverse biological applications.

Random Mutagenesis

Random mutagenesis is the process of introducing random point mutations to generate a diverse gene library for screening desired protein properties. It has been successfully used to identify key residues in proteins as well as to improve or alter protein activities.[1-12] Random mutagenesis samples the local sequence space of the starting protein (Fig. 1A). As a result, the properties of the mutants will not diverge far from their parents. If the mutation rate is too low, the diversity in the resulting library may be insufficient, as quite often, biological advantages are conferred from multiple cooccurring mutations. On the other hand, if the mutation rate is too high, potentially advantageous mutations may be silenced by mutations that destroy the protein function altogether. Thus, the rate of mutation is a very important parameter to study

*Corresponding Author: Kevin Truong—Institute of Biomaterials and Biomedical Engineering, University of Toronto, 164 College Street, Toronto, Ontario, M5S 3G9, Canada; Email: kevin.truong@utoronto.ca

Bio-Applications of Nanoparticles, edited by Warren C.W. Chan. ©2007 Landes Bioscience.

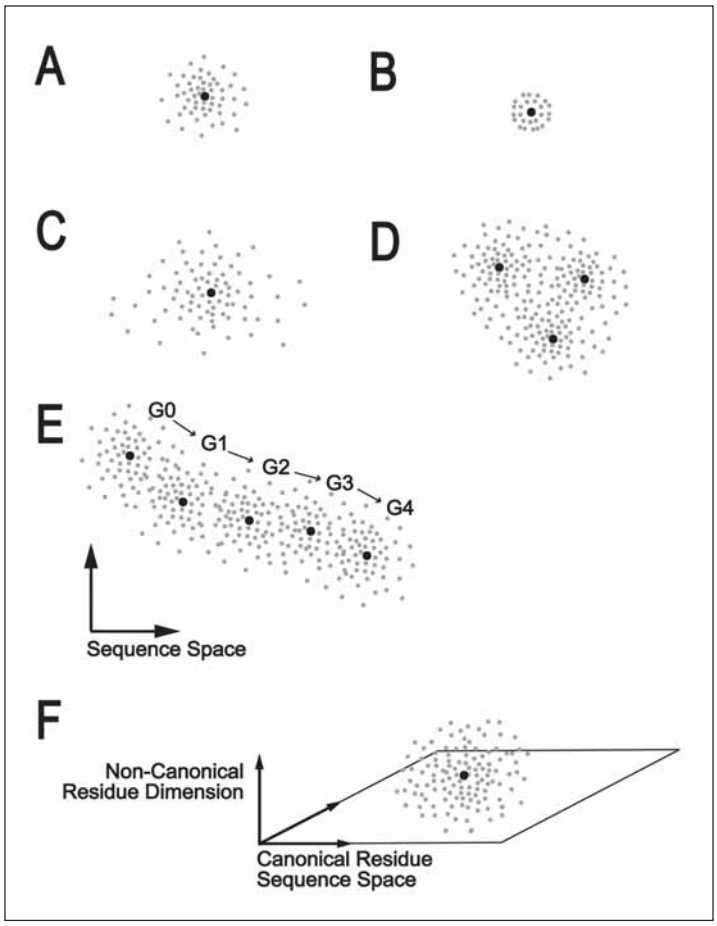

Figure 1. Sequence space representations of different protein engineering approaches. The solid black dots represent the locations of parent genes; the grey dots, mutants. Distance between the dots represents extent of divergence between their corresponding sequences. (A) Random mutagenesis samples the sequence space isotropically to generate a library of mutants with a Gaussian distribution centred at the parent gene. (B) Site-directed mutagenesis samples small and predefined regions close to the parent gene. (C) DNA recombination on mutagenesis libraries results in a larger divergence of mutants around the parent gene than compared to random mutagenesis alone. (D) DNA recombination of three distinct genes results in even greater sequence space coverage by including the sequence space between regions. (E) Directed evolution can reach far from parent gene by following the path of generations of improved sequences. Starting from the parent gene at G0 (Generation 0), mutagenesis and recombination is used to generate a library of mutants. The mutant with the most improved properties is selected as the parent gene for the next generation - G1. This process is repeated for subsequent generations until the desired improvement of enzymatic properties is achieved. (F) Introduction of noncanonical amino acids adds new dimensions to the existing sequence space. The horizontal plane shows the canonical sequence space, where the vertical axis illustrates the new dimension created by noncanonical amino acids.

and fine tune in mutagenesis experiments. In this section, we summarize three commonly used techniques in random mutagenesis: error-prone polymerase chain reaction (PCR), bacteria mutator strains and chemical mutagens.

Error-prone PCR is an in vitro technique to increase the DNA copy error rate by altering the PCR conditions. For example, high Mg^{2+} concentration stabilizes base-pair mismatch, the addition of Mn^{2+} reduces base-pair specificity, imbalanced nucleotides (dNTP) enhance the mismatching probability during PCR and a higher concentration of *Taq* DNA polymerase encourages elongation of mistakenly primed termini. Statistics show that under normal conditions, PCR reactions using *Taq* DNA polymerase has an error rate of ~0.01%, while under the error-prone PCR conditions, the error rate increases to ~1%. Of the total mutations, 90% are nucleotide substitutions and of these mutations, ~30% are silent and ~70% result in amino acid substitutions. The remaining 10% of mutations are insertions or deletions, which usually result in frame shifts and subsequently, incorrect gene expressions. Multiple cycles of error-prone PCR can further control the amount of mutations introduced to the template. This technique has recently helped identify key residues in the M3 muscarinic acetylcholine receptor, epidermal growth factor receptor (EGFR) and *Saccharomyces cerevisiae* Hal3 salt tolerance regulator.[1,2,4] From the engineering/design aspect, PCR random mutagenesis has been used to improve enzymatic properties of subtilisins E, *Thermus aquaticus* amylomaltase, cyclodextrin glucanotransferase and *Rhizopus niveus* lipase.[5-8]

Random mutations can also be introduced in vivo by mutator bacteria strains such as mutD5-FIT and XL1-Red (Stratagene).[13,14] These *Escherichia coli* strains have mutations in genes (such as the *mutD*, *mutS*, *mutL*, *mutH* and *mutT*) found in the DNA repair pathway which increase the rate of mutation to over a thousand fold higher than the wild type strain.[15,16] It is estimated that the error rate after 30 cell generations is ~0.05%. The advantage of this technique is in its simplicity - transform and harvest. Although the mutation rate is not as high as error-prone PCR, it is proven sufficient to generate diversity in many studies.[3,9,17-19] Random mutagenesis with mutator strains was used to identify the essential residues responsible for the function of Erm(B) rRNA methyltransferase.[3] A 5-fold improvement of enzymatic activity using this technique was demonstrated for polyhydroxyalkanoate (PHA) synthase.[9] In another example, mutator strains also helped alter protein properties as seen in the conversion of oxidosqualene-cycloartenol synthase specificity to that of oxidosqualene-lanosterol cyclase.[12]

A third way to introduce random mutations in DNA is by chemical mutagens. One class of chemical mutagens are DNA base analogs such as 5-bromo-deoxyuridine (5BU), which exist in two tautomeric forms, mimicking cytosine (C) and thymine (T). Therefore, when mutagens in this class are introduced to the cells during DNA replication, the mutagens are likely to be incorporated into the DNA, causing mutations to occur. Another class of mutagens are alkylators such as ethyl methane sulfonate (EMS), which reacts with certain bases on the DNA. Other mutagens such as sodium bisulphite converts cytosine to uracil, achieving substitution directly on the target DNA.[20]

Site-Directed Mutagenesis

Site-directed mutagenesis allows the introduction of a mutation to a specific location on a gene. It is commonly used for studying the effects of individual point mutations and designing novel proteins with key mutations (Fig. 1B). There are in vivo and in vitro site-directed mutagenesis techniques and among them, two common techniques will be summarized here: oligonucleotide mismatch mutagenesis and PCR based mutagenesis.[21-23] The in vivo oligonucleotide mismatch mutagenesis was originally proposed by Michael Smith in 1979, who later won the Nobel Prize in 1993. This technique incorporates a point mutation to an oligonucleotide whose sequence is otherwise complementary to the region of mutation on the gene of interest. This oligonucleotide is then annealed to the gene of interest and the full-length plasmid DNA is synthesized with mismatching double strands at the point of mutation. Then, cells are transformed with this plasmid DNA, which replicate the mismatching gene into two

species of proteins - one containing the mutation and the other retaining the natural form (Fig. 2). Site-directed mutagenesis can also be performed in vitro by PCR. Primers with desired point mutations first amplify regions of the parental DNA on both sides of the mutation site, which creates two portions of the parental DNA with overlapping regions (Fig. 3). Then a PCR is performed using the mixture of the two overlapping portions to create the full length DNA, which now contains the point mutation.

Site-directed mutagenesis studies provide valuable information on the structure-function relationship of a protein as shown in numerous recent examples.[24-29] In addition to learning about proteins, site-directed mutagenesis is applied to the rational design of proteins based on the knowledge of structure-function relationships. For example, improvements on catalytic activities were shown in the *Aspergillus niger* NRRL 3135 phytase and *Bacillus* protease by targeted mutations of residues in the catalytic sites revealed by their molecular structures.[30,31] Alteration of enzymatic properties has been demonstrated in the 150-fold change of relative substrate specificity of the oncoprotein v-Fps by mutating the recognition element at the arginine residue at position 1130 (Arg1130).[32] Similarly, the excitation and emission spectra of green fluorescent protein (GFP) was red-shifted by replacing the Thr203 with Tyr or His to promote aromatic interactions with the fluorophore.[33]

Non-Canonical Amino Acid Substitution

The introduction of noncanonical amino acids provides a new perspective to engineering proteins by expanding the combinatorial sequence space (Fig. 1F). Most natural organisms can only synthesize and incorporate the 20 canonical amino acids as the building blocks for their proteins. By substituting noncanonical amino acids for natural residues, we overcome this limitation. In general, there are two types of noncanonical amino acid substitutions: residue-specific and site-specific. In residue-specific substitution, a specific type of amino acid is replaced by a noncanonical one. For instance, the replacement of tryptophan in the barstar

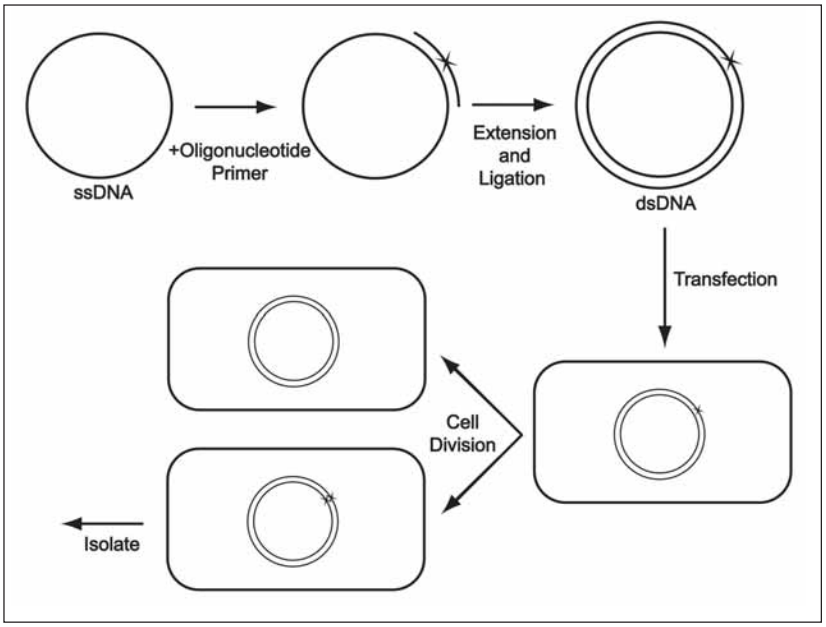

Figure 2. Oligonucleotide mismatch mutagenesis. The cross "×" donates a point mutation on a primer or strand of DNA.

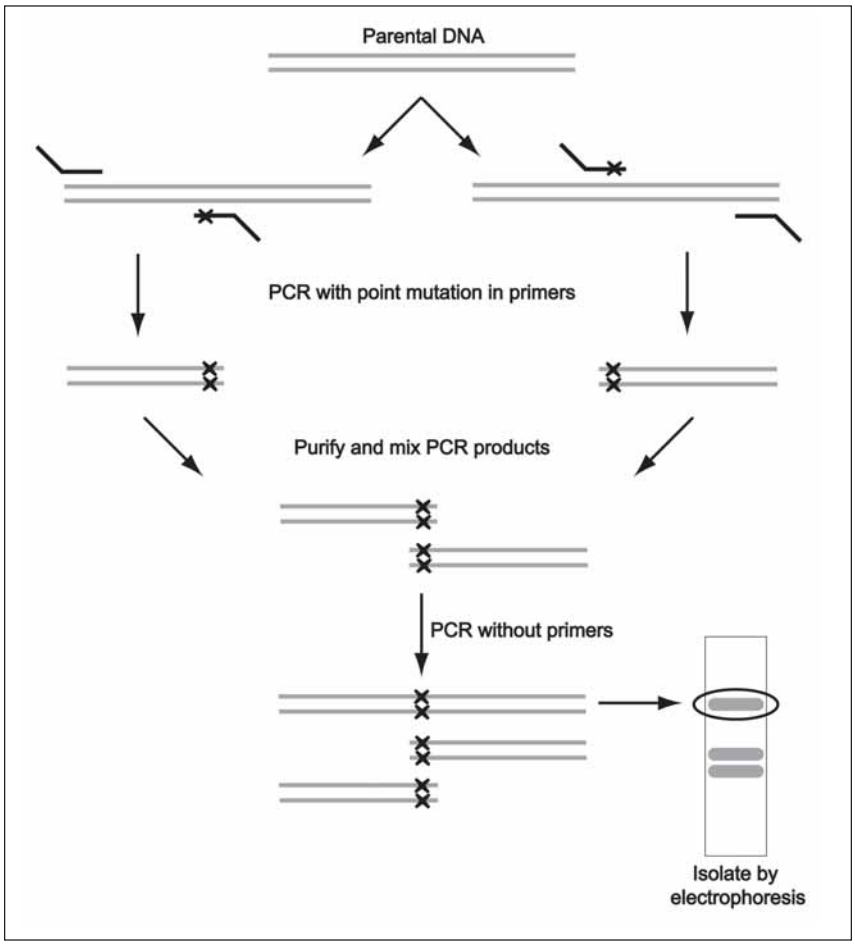

Figure 3. PCR-based site-directed mutagenesis. The cross "×" donates a point mutation a primer or strand of DNA.

protein by a noncanonical pH sensitive aminotryptophan created a pH sensor.[34] In site-specific substitution, the amino acid at a particular location is replaced by a noncanonical amino acid. Similar to site directed mutagenesis, this facilitates protein engineering based on structure-function relationships. The challenge of noncanonical amino acid substitution lies in the synthesis of the modified proteins. Measures must be taken when synthesizing proteins with noncanonical residues in vivo to ensure the proteins of the expression host do not acquire the noncanonical amino acids, which could be fatal to the host cell. Furthermore, in site-specific substitution, one must also ensure that only specific amino acids on the protein are targeted.

Common approaches in residue-specific substitution rely on gene expression in an auxotroph grown in a media containing the noncanonical amino acid.[35] Auxotrophic bacteria lack the ability to synthesize a certain amino acid, hence relying on the supply of that amino acid in the growth media. To maximize expression of target protein with noncanonical amino acids, auxotroph bacteria are first grown in media with the natural amino acids. During this stage, the expression of the target protein is strongly repressed to minimize its expression with natural amino acids. After the cell density threshold is reached, the expression of the target protein is

induced in a new growth media containing the noncanonical amino acids. Using this technique, noncanonical amino acids were incorporated into enhanced cyan fluorescent protein (ECFP) to produce "gold" fluorescent protein (GdFP) that had an emission spectrum red-shifted by 69 nm.[36] The noncanonical protein expression level can be further improved by over-expressing the corresponding aminoacyl-tRNA synthase (aaRS) to the replaced natural amino acid. The 20 aaRS, one for each type of amino acid, create tRNA amino acid complexes between particular amino acids and tRNAs according to the genetic code. Since the rate at which aaRS incorporates noncanonical amino acids is over a thousand-fold less than that for natural amino acids, over-expression of the particular aaRS would produce more tRNA amino acid complexes and hence allow the cell to incorporate noncanonical amino acids in target proteins more efficiently.

There are other techniques for incorporating noncanonical amino acids to target proteins including modifications to aaRS that favor recognition of noncanonical amino acids and inaccurate hydrolytic editing.[37] Alternatively, heterologous aaRS and tRNA from other organisms that have different codon usage can be introduced into the host organism.[38] Hence, the codon degeneracy is broken as long as the noncanonical amino acids and tRNA do not interfere with the tRNA synthesis pathway of the host. This technique allows for a codon-specific substitution. Global noncanonical amino acid substitution can also be achieved by in vitro post-translational chemical modifications.[39]

To achieve site-specific substitution of a noncanonical amino acid, one approach is by breaking the codon degeneracy using 4-base or 5-base codons to code for the noncanonical amino acid through in vitro protein synthesis.[40-44] The specificity and efficiency is increased by using these 4- and 5-base codon systems. One example is the incorporation of a fluorescent noncanonical amino acid to a specific site of streptavidin by 4-base codons.[40,41] The fluorescence of the incorporated amino acid is highly sensitive to the binding of biotin, which can be used for low concentration detection. In another example, the Tyr66 of GFP was mutated to 18 distinct noncanonical aromatic amino acids by the 4-base CCCG codon tRNA, resulting in two blue-shifted mutants.[43]

DNA Recombination

Novel protein characteristics can be created by swapping the regions of one gene with another in a process called DNA recombination. This process samples a larger sequence space compared to mutagenesis alone (Fig. 1D). In many cases, mutagenesis and recombination are used together to create gene libraries with even greater diversity. It can enhance or alter protein activities by preserving advantageous mutations while removing disadvantageous ones.[45,46] Here, we discuss two DNA recombination techniques: DNA shuffling and the staggered extension process.

In 1994, William Stemmer introduced the DNA shuffling technique to create a chimera of β-lactamase with a 32,000-fold increase in tolerance to its inhibitor.[47,48] In DNA shuffling, a library of homologous genes is first randomly digested by DNase I (Fig. 4). Then, DNA fragments with sizes from 10 to 50 base pairs are isolated. PCR amplification of the gene library with these isolated fragments as primers will produce many PCR products from different homologous templates. For this step to be effective there must be significant homology (> 70%) in the gene library. Next, an additional PCR step is performed to extract the desired genes using primers that define the start and end of a full-length gene. Finally, these PCR products are inserted into vectors, expressed and screened for desired properties. Using 3 cycles of DNA shuffling on a GFP mutant library, a combination of three mutations were discovered that improved folding properties and increased fluorescence intensity by 45-fold.[49] A family of serum paraoxonases were recombined and screened to create a variant with 40-fold higher enzymatic activity and over 2000-fold specificity.[50] Applying DNA shuffling to eight cefE homologous genes in combination with directed evolution created chimeras with 118-fold increased activity.[51] In another example, Lipase B protein homologs from three organisms were

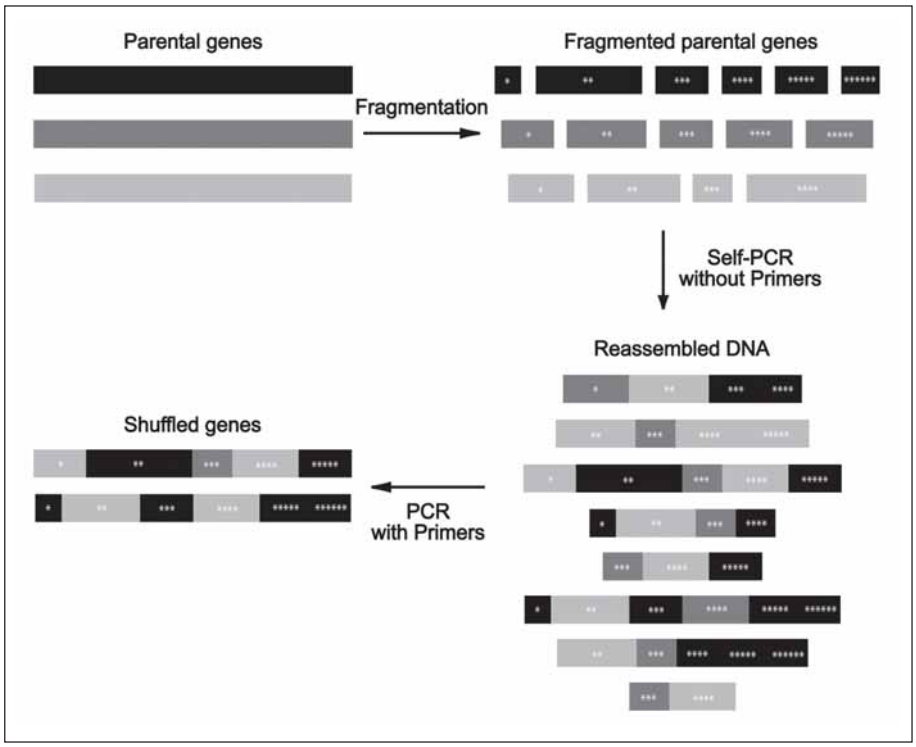

Figure 4. DNA shuffling. There are three homologous parental genes identified by their distinct shades. The white dots labelled on each gene fragments represent their relative order to the parental gene, which is preserved after reassembly from the self-PCR process. Only the full length reassembled genes can be amplified using primers that specify the two ends of the gene.

subjected to DNA shuffling, resulting in a chimera with improved stability and a 20-fold higher hydrolysis activity towards diethyl 3-(3',4'-dichlorophenyl)glutarate.[52]

In 1998, Zhao and colleagues proposed another in vitro recombination technique called the staggered extension process (StEP).[53] The recombination was achieved in one PCR reaction with optimized conditions. Similar to DNA shuffling, a library of homologous genes is used as the template. The annealing and extension steps in each PCR cycle are replaced by an extremely short annealing step (~1s) (Fig. 5). This short annealing step amplifies only a fraction of the template. Since a mixture of templates is used, the partly extended primer will randomly anneal to a different template in the next thermal cycle. Consequently, the primers are extended to full length by fragments from each of the homologous genes. Next, these fragments are inserted into the expression vector and screened. StEP was used to recombine the aroA gene from *Salmonella typhimurium* and *Escherichia coli*, creating 4 chimeras with a 2 to 10 fold increase in activity.[54] In another study, by combining random mutagenesis and StEP, the 1,6 regioselectivity of AgaB gene from *Bacillus stearothermophilus* was removed and 1,3 regioselectivity was retained.[55]

Directed Evolution

DNA mutagenesis and recombination alone are often insufficient to yield desired enzymatic properties because the sequence space is sampled around the parent genes. However, the coding sequence of the desired enzymes may be significantly divergent from the parent.

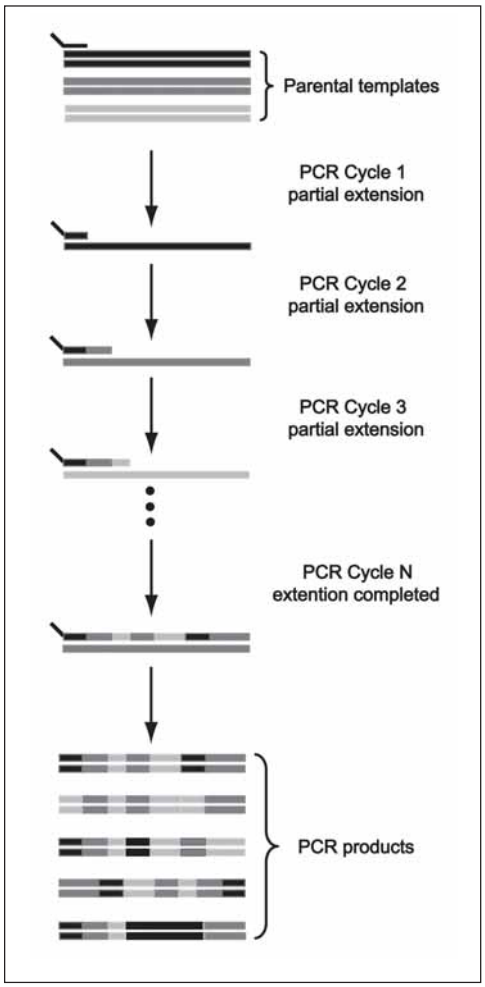

Figure 5. Staggered Extension Process (StEP). There are three double-stranded homologous parental genes that are distinguished by their shade and used as PCR templates. In the first PCR cycle, the primer anneals to the single-stranded parental template after melting. Following annealing, the extension time of the PCR reaction is short enough to allow only partial extension of the primer on the parental gene. The same process is repeated in subsequent cycles until the extended primer reaches full length as the parental genes.

Directed evolution creates a path in the sequence space to guide mutagenesis and recombination (Fig. 1E). There are four essential steps involved in directed evolution (Fig. 6). First, a large and diverse gene library is generated by mutagenesis or recombination from parent genes. Second, the library of genes is expressed by cells or phages. Third, cells exhibiting more desired properties are selected through either a manual process of choosing colonies from agar plates or an automated process such as flow cytometry. Note here that it is important to devise an evolutionary pressure in the selection process that is strong enough to reduce the selection cycles, yet sensitive enough to discriminate between proteins with small differences in properties. Lastly, the genes from the selected cells are amplified and extracted before they are used as parents in

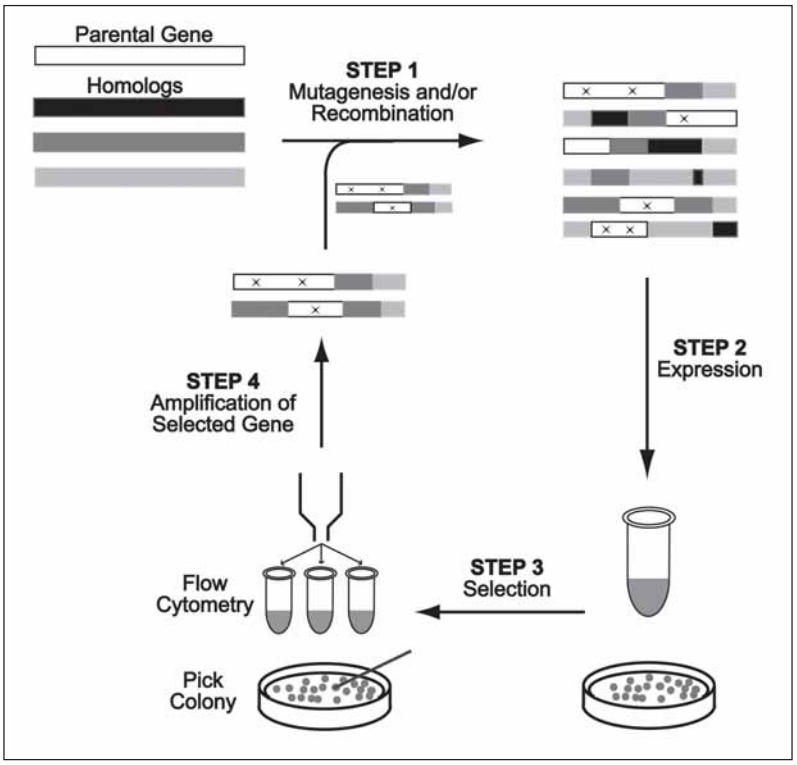

Figure 6. Directed evolution. The cross "×" donates a point mutation; a grey dot, colony on a culture plate. The first generation of genes contains the parental gene and/or its homologs as identified by distinct shades. Then, mutants are created, expressed, selected and amplified for subsequent generations.

the next cycle. Several cycles of this process may be necessary to generate proteins with significantly enhanced properties.

Directed evolution has been successfully applied to enhance or create enzymatic activity, cellular pathways, molecular switches and fluorescence properties. Recently, directed evolution has been used to enhance enzymatic activity and thermal stability.[56-63] In other examples, directed evolution was applied to alter properties of proteins involved in cellular pathways. For instance, modified proteins were introduced into cells to create cellular pathways that form disulfide bonds; in another case, proteins in the carotenoid biosynthetic pathways were altered to create new carotenoid products.[64,65] Molecular switches were also created by evolving the hybrid of maltose binding protein and β-lactamase.[66-68] After applying mutagenesis and 3 generations of directed evolution to the *EcoRV* restriction endonuclease, a variant was created with altered DNA substrate specificity.[11] Recently, a superfolder GFP was created with a 50-fold brighter fluorescence and a higher tolerance to circular permutation and chemical denaturants.[45] This superfolder GFP was created by screening for brighter GFP mutants that were fused with a poorly folded peptide, which effectively targeted the folding properties of GFP.

Fusion Proteins

Fusion proteins are created by joining two or more proteins at their N- and C- termini (Fig. 7A), thereby forming a single protein with the combined functions of the components. By

exploiting the intramolecular interactions among components, fusion proteins can create biosensors and molecular switches. As genetic engineering of fusion proteins can be tedious, a fluorescent cassette technique was introduced to make fusion proteins more efficiently.[69]

Fluorescent protein biosensors represent a major class of fusion protein applications. For example, the N- or C-terminal fusion of a fluorescent protein (FP) to a target protein (Fig. 7B) allows tracking of the target protein's expression, localization and movement inside cells. By fusing a sensor protein to the middle of a FP gene, researchers have created split-FP biosensors whose fluorescence is modulated by stimulus-induced conformational change of the sensor protein (Fig. 7C). For example, when maltose is bound, the adjacent N- and C-terminal of maltose binding protein (MBP) allows two halves of FP to fold together, hence fluorescence increases; when maltose is released, the large conformational change of MBP separates the two halves of FP and decreases fluorescence.[70] Another class of biosensors is based on the principle of fluorescence resonance energy transfer (FRET) where the distance and orientation between a donor and an acceptor FP governs their fluorescence emission spectral properties. Typically, the donor and acceptor FP are fused to the N- and C-termini of a sensor protein that responds to a biological event with a large protein conformational change. This conformational change alters the relative position between the donor and acceptor FP and hence a change in FRET efficiency is correlated to the biological event (Figs. 7D, E). FRET biosensors based on the intrinsic conformational change of the sensor proteins have been demonstrated using IP_3 receptors, cGMP-dependent protein kinase I, MBP and bacterial periplasmic binding proteins (Fig.7D).[71-74] Alternatively, conformational change caused by the conditional binding of a protein to its peptidal substrate can also induce FRET efficiency changes (Fig. 7E). A classic example is the Ca^{2+} biosensor. Calmodulin and its Ca^{2+}-dependent binding peptide M13 are fused and sandwiched between cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP). When Ca^{2+} increases, the binding of calmodulin to M13 shortens the distance between CFP and YFP and hence FRET efficiency increases.[75-77] Using similar concepts, FRET biosensors have also been created to detect protein kinase activities.[78-81]

Protein fusion was also used to create molecular switches. Since the function of a protein is closely related to its conformation, the simplest approach to engineer a protein switch is by an allosteric effect (Fig. 7F). This can be achieved by fusing an effector protein with a sensor protein. For example, circularly permutated β-lactamase (BLA) was inserted into specific locations in MBP such that the enzymatic site of BLA is distorted by the conformational change of MBP upon the release of maltose, resulting in the loss of BLA activity.[66-68]

One drawback of fusion proteins is that the functions of proteins are not always preserved. Fusing proteins together may cause one or more of the fusion partners to fold improperly. It has been experimentally observed that well folded proteins make the fusion protein fold better, whereas poorly folded proteins could destabilize the entire fusion. Steric restrictions between fusion partners may also cause the loss of function. In addition, fusion proteins are more prone to aggregate due to their large size and flexibility, destroying their activities as well. Besides the limited guidelines provided by semi-rationally studying the protein structure of each fusion partner, fusion protein engineering is still a trial-and-error process.

Circular Permutation

Circular permutation of a protein creates new N- and C- termini at nonnatural locations on the protein by fusing together the natural N- and C- termini (Fig. 8A). This can be achieved easily by fusing two copies of the target gene in tandem and then amplifying at specific circular permutation locations by PCR (Fig. 8B).[82] Circular permutation has been used to study the biochemical properties of proteins such as folding and functional elements. In one example, dihydrofolate reductase was circularly permutated systematically to break its backbone at each amino acid and the resulting variants were tested for functionalities. The correlation between the circular permutation location and the activity of the variant helped to determine the regions crucial for folding and function.[83,84] In addition to identifying critical regions of a

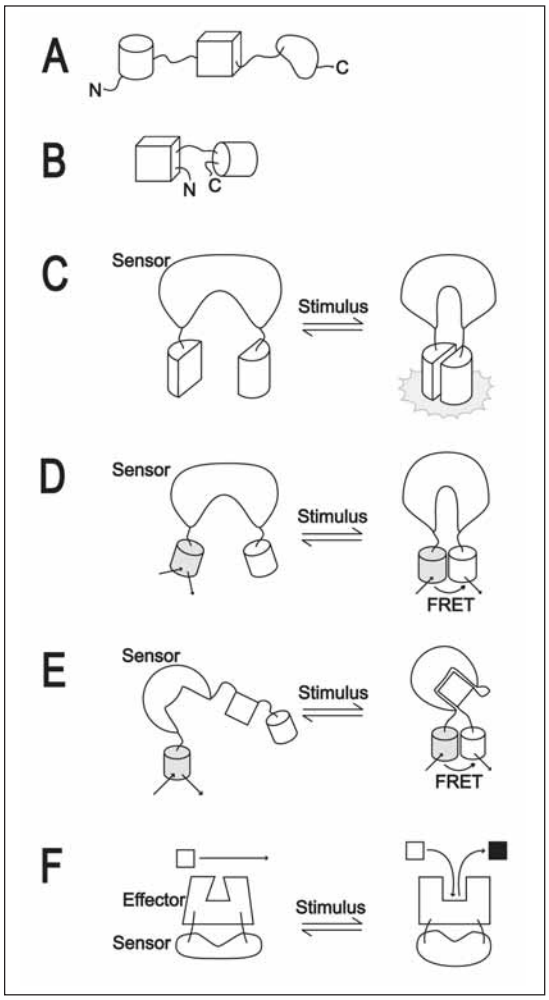

Figure 7. Fusion protein and application mechanisms. The grey cylinder represents CFP, while the white, YFP. Arrows entering the cylinder represent excitation light, while arrows leaving the cylinder represent emission light. (A) The concept of a fusion protein. (B) FP tagged fusion protein. (C) Split-FP biosensor modulates fluorescence by the conformational change of the sensor protein. (D) FRET of the biosensor is increased by the conformational change of the sensor protein that brings the FRET pairs together. (E) FRET of the biosensor is created when the sensor binds to the substrate and vice versa. (F) An example of the protein switching mechanism. The white and black boxes are the substrate and converted substrate of the effector enzyme. Stimulation changes the conformation of the sensor protein, which alters the enzymatic site of the effector enzyme that allows enzymatic activities to be carried out and hence turns it "on".

protein, circular permutation was used to engineer proteins with enhanced properties. The catalytic activity of *Candida antarctica* Lipase B was improved in its circularly permutated variants due to better active site accessibility and backbone flexibilities.[85] In another example, circular permutation of yellow fluorescent protein improved the dynamic range of FRET Ca^{2+} biosensor because it changes the orientation factor between the donor and acceptor fluorescent proteins.[86]

Conclusions and Perspectives

Nature tinkers with new protein designs by random processes that alter genomic DNA such as mutations, insertions, deletions, inversions, gene fusions and translocations. Then through natural selection arising from environmental conditions, the best protein designs are discovered. These natural evolutionary mechanisms have inspired the techniques discussed in this chapter. Random mutagenesis and DNA recombination can create random gene libraries that can then be screened through successive generations by directed evolutionary techniques for incrementally better designs. Protein fusion, another natural evolutionary mechanism, can combine functions from two different proteins. Furthermore, by hijacking the natural processes for protein translation, we can expand protein design by introducing noncanonical amino acids. Finally, with a strong understanding of protein structure-function relationships, site directed mutagenesis provides the most direct route to protein design.

These techniques, however, are restricted by the starting templates - the natural proteins. This limits the functions of engineered proteins to those similar to natural proteins. In order to engineer proteins with truly novel structures and customized functions, we need to be able to predict the structure and function of a protein from its primary structure. This requires overcoming two major hurdles in understanding: first, how proteins fold into unique three-dimensional structures and second, how structure determines biochemical function. As a deeper insight is gained into sequence-structure-function relationships, rational protein design will become more feasible. Looking ahead, protein engineering will play increasingly important roles in making new molecular tools, in designing better industrial enzymes, in producing biomaterials with novel properties and in rewiring cellular pathways related to human diseases.

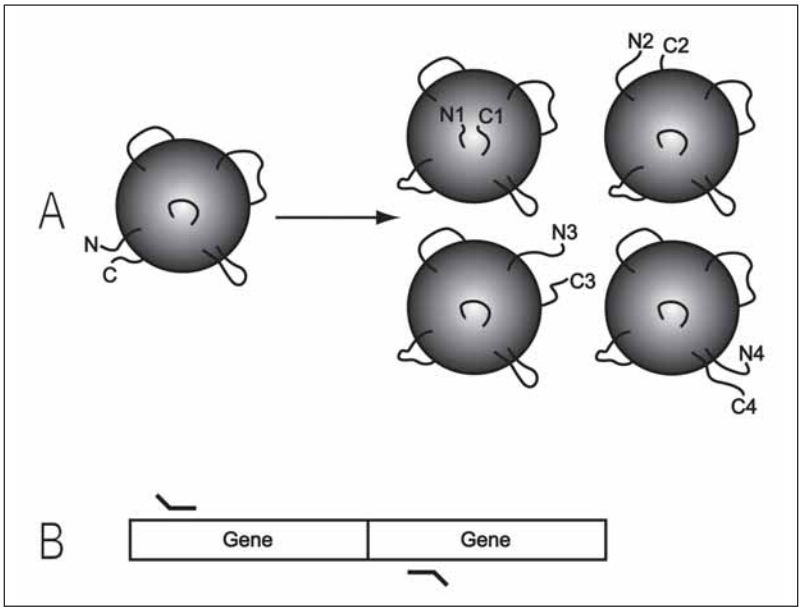

Figure 8. Circular permutation. A) Four possible circular permutations from the original protein. N1-4 and C1-4 are the newly created termini. B) A technique to generate circular permutations of a gene. Each pair of primers with the same shade will be used in a PCR reaction to create circular permutation at that location.

Acknowledgements

This work was supported by grants from the Canadian Foundation of Innovation (CFI) and the National Science and Engineering Research Council (NSERC).

References

1. Li B, Nowak NM, Kim SK et al. Random mutagenesis of the M3 muscarinic acetylcholine receptor expressed in yeast: Identification of second-site mutations that restore function to a coupling-deficient mutant M3 receptor. J Biol Chem 2005; 280(7):5664-5675.
2. Chao G, Cochran JR, Wittrup KD. Fine epitope mapping of anti-epidermal growth factor receptor antibodies through random mutagenesis and yeast surface display. J Mol Biol 2004; 342(2):539-550.
3. Farrow KA, Lyras D, Polekhina G et al. Identification of essential residues in the Erm(B) rRNA methyltransferase of Clostridium perfringens. Antimicrob Agents Chemother 2002; 46(5):1253-1261.
4. Munoz I, Ruiz A, Marquina M et al. Functional characterization of the yeast Ppz1 phosphatase inhibitory subunit Hal3: A mutagenesis study. J Biol Chem 2004; 279(41):42619-42627.
5. Takahashi M, Hasuura Y, Nakamori S et al. Improved autoprocessing efficiency of mutant subtilisins E with altered specificity by engineering of the pro-region. J Biochem (Tokyo) 2001; 130(1):99-106.
6. Fujii K, Minagawa H, Terada Y et al. Use of random and saturation mutageneses to improve the properties of Thermus aquaticus amylomaltase for efficient production of cycloamyloses. Appl Environ Microbiol 2005; 71(10):5823-5827.
7. Shim JH, Kim YW, Kim TJ et al. Improvement of cyclodextrin glucanotransferase as an antistaling enzyme by error-prone PCR. Protein Eng Des Sel 2004; 17(3):205-211.
8. Kohno M, Enatsu M, Funatsu J et al. Improvement of the optimum temperature of lipase activity for Rhizopus niveus by random mutagenesis and its structural interpretation. J Biotechnol 2001; 87(3):203-210.
9. Amara AA, Steinbuchel A, Rehm BH. In vivo evolution of the Aeromonas punctata polyhydroxyalkanoate (PHA) synthase: Isolation and characterization of modified PHA synthases with enhanced activity. Appl Microbiol Biotechnol 2002; 59(4-5):477-482.
10. Delagrave S, Hawtin RE, Silva CM et al. Red-shifted excitation mutants of the green fluorescent protein. Biotechnology (NY) 1995; 13(2):151-154.
11. Lanio T, Jeltsch A, Pingoud A. Towards the design of rare cutting restriction endonucleases: Using directed evolution to generate variants of EcoRV differing in their substrate specificity by two orders of magnitude. J Mol Biol 1998; 283(1):59-69.
12. Wu TK, Griffin JH. Conversion of a plant oxidosqualene-cycloartenol synthase to an oxidosqualene-lanosterol cyclase by random mutagenesis. Biochemistry 2002; 41(26):8238-8244.
13. Greener A, Callahan M, Jerpseth B. An efficient random mutagenesis technique using an E. coli mutator strain. Mol Biotechnol 1997; 7(2):189-195.
14. Greener A, Callahan M, Jerpseth B. An efficient random mutagenesis technique using an E. coli mutator strain. Methods Mol Biol 1996; 57:375-385.
15. Kunkel TA, Erie DA. DNA mismatch repair. Annu Rev Biochem 2005; 74:681-710.
16. Brakmann S, Lindemann BF. Generation of mutant libraries using random mutagenesis. In: Brakmann S, Schwienhorst A, eds. Evolutionary methods in biotechnology: Clever tricks for directed evolution. Weinheim: Wiley-VCH, 2004:5-12.
17. Bornscheuer UT, Altenbuchner J, Meyer HH. Directed evolution of an esterase: Screening of enzyme libraries based on pH-indicators and a growth assay. Bioorg Med Chem 1999; 7(10):2169-2173.
18. Bornscheuer UT, Altenbuchner J, Meyer HH. Directed evolution of an esterase for the stereoselective resolution of a key intermediate in the synthesis of epothilones. Biotechnol Bioeng 1998; 58(5):554-559.
19. Henke E, Bornscheuer UT. Directed evolution of an esterase from Pseudomonas fluorescens. Random mutagenesis by error-prone PCR or a mutator strain and identification of mutants showing enhanced enantioselectivity by a resorufin-based fluorescence assay. Biol Chem 1999; 380(7-8):1029-1033.
20. Shortle D, Nathans D. Local mutagenesis: A method for generating viral mutants with base substitutions in preselected regions of the viral genome. Proc Natl Acad Sci USA 1978; 75(5):2170-2174.
21. Shimada A. PCR-based site-directed mutagenesis. Methods Mol Biol 1996; 57:157-165.
22. Storici F, Lewis LK, Resnick MA. In vivo site-directed mutagenesis using oligonucleotides. Nat Biotechnol 2001; 19(8):773-776.
23. Hutchison IIIrd CA, Phillips S, Edgell MH et al. Mutagenesis at a specific position in a DNA sequence. J Biol Chem 1978; 253(18):6551-6560.

24. Saint-Joanis B, Souchon H, Wilming M et al. Use of site-directed mutagenesis to probe the structure, function and isoniazid activation of the catalase/peroxidase, KatG, from Mycobacterium tuberculosis. Biochem J 1999; 338(Pt 3):753-760.
25. Ohtsubo K, Imajo S, Ishiguro M et al. Studies on the structure-function relationship of the HNK-1 associated glucuronyltransferase, GlcAT-P, by computer modeling and site-directed mutagenesis. J Biochem (Tokyo) 2000; 128(2):283-291.
26. Farh L, Hwang SY, Steinrauf L et al. Structure-function studies of Escherichia coli biotin synthase via a chemical modification and site-directed mutagenesis approach. J Biochem (Tokyo) 2001; 130(5):627-635.
27. Alam M, Vance DE, Lehner R. Structure-function analysis of human triacylglycerol hydrolase by site-directed mutagenesis: Identification of the catalytic triad and a glycosylation site. Biochemistry 2002; 41(21):6679-6687.
28. Sviridov D, Hoang A, Huang W et al. Structure-function studies of apoA-I variants: Site-directed mutagenesis and natural mutations. J Lipid Res 2002; 43(8):1283-1292.
29. Gupta RP, He YA, Patrick KS et al. CYP3A4 is a vitamin D-24- and 25-hydroxylase: Analysis of structure function by site-directed mutagenesis. J Clin Endocrinol Metab 2005; 90(2):1210-1219.
30. Mullaney EJ, Daly CB, Kim T et al. Site-directed mutagenesis of Aspergillus niger NRRL 3135 phytase at residue 300 to enhance catalysis at pH 4.0. Biochem Biophys Res Commun 2002; 297(4):1016-1020.
31. Tobe S, Shimogaki H, Ohdera M et al. Expression of Bacillus protease (Protease BYA) from Bacillus sp. Y in Bacillus subtilis and enhancement of its specific activity by site-directed mutagenesis-improvement in productivity of detergent enzyme. Biol Pharm Bull 2006; 29(1):26-33.
32. Konkol L, Hirai TJ, Adams JA. Substrate specificity of the oncoprotein v-Fps: Site-specific mutagenesis of the putative P+1 pocket. Biochemistry 2000; 39(1):255-262.
33. Ormo M, Cubitt AB, Kallio K et al. Crystal structure of the Aequorea victoria green fluorescent protein. Science 1996; 273(5280):1392-1395.
34. Budisa N, Rubini M, Bae JH et al. Global replacement of tryptophan with aminotryptophans generates noninvasive protein-based optical pH sensors. Angew Chem Int Ed Engl 2002; 41(21):4066-4069.
35. Cohen GN, Cowie DB. Total replacement of methionine by selenomethionine in the proteins of *Escherichia coli*. C R Hebd Seances Acad Sci 1957; 244(5):680-683.
36. Hyun Bae J, Rubini M, Jung G et al. Expansion of the genetic code enables design of a novel "gold" class of green fluorescent proteins. J Mol Biol 2003; 328(5):1071-1081.
37. Kobayashi T, Sakamoto K, Takimura T et al. Structural basis of nonnatural amino acid recognition by an engineered aminoacyl-tRNA synthetase for genetic code expansion. Proc Natl Acad Sci USA 2005; 102(5):1366-1371.
38. Kwon I, Kirshenbaum K, Tirrell DA. Breaking the degeneracy of the genetic code. J Am Chem Soc 2003; 125(25):7512-7513.
39. Plettner E, Khumtaveeporn K, Shang X et al. A combinatorial approach to chemical modification of subtilisin Bacillus lentus. Bioorg Med Chem Lett 1998; 8(17):2291-2296.
40. Taki M, Hohsaka T, Murakami H et al. A novel fluorescent nonnatural amino acid that can be incorporated into a specific position of streptavidin. Nucleic Acids Res Suppl 2002; (2):203-204.
41. Murakami H, Hohsaka T, Ashizuka Y et al. Site-directed incorporation of fluorescent nonnatural amino acids into streptavidin for highly sensitive detection of biotin. Biomacromolecules. Spring 2000; 1(1):118-125.
42. Sisido M, Hohsaka T. Introduction of specialty functions by the position-specific incorporation of nonnatural amino acids into proteins through four-base codon/anticodon pairs. Appl Microbiol Biotechnol 2001; 57(3):274-281.
43. Kajihara D, Hohsaka T, Sisido M. Synthesis and sequence optimization of GFP mutants containing aromatic nonnatural amino acids at the Tyr66 position. Protein Eng Des Sel 2005; 18(6):273-278.
44. Hohsaka T, Sisido M. Incorporation of nonnatural amino acids into proteins by using five-base codon-anticodon pairs. Nucleic Acids Symp Ser 2000; (44):99-100.
45. Pedelacq JD, Cabantous S, Tran T et al. Engineering and characterization of a superfolder green fluorescent protein. Nat Biotechnol 2006; 24(1):79-88.
46. Castle LA, Siehl DL, Gorton R et al. Discovery and directed evolution of a glyphosate tolerance gene. Science 2004; 304(5674):1151-1154.
47. Stemmer WP. Rapid evolution of a protein in vitro by DNA shuffling. Nature 1994; 370(6488):389-391.
48. Stemmer WP. DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution. Proc Natl Acad Sci USA 1994; 91(22):10747-10751.

49. Crameri A, Whitehorn EA, Tate E et al. Improved green fluorescent protein by molecular evolution using DNA shuffling. Nat Biotechnol 1996; 14(3):315-319.
50. Aharoni A, Gaidukov L, Yagur S et al. Directed evolution of mammalian paraoxonases PON1 and PON3 for bacterial expression and catalytic specialization. Proc Natl Acad Sci USA 2004; 101(2):482-487.
51. Hsu JS, Yang YB, Deng CH et al. Family shuffling of expandase genes to enhance substrate specificity for penicillin G. Appl Environ Microbiol 2004; 70(10):6257-6263.
52. Suen WC, Zhang N, Xiao L et al. Improved activity and thermostability of Candida antarctica lipase B by DNA family shuffling. Protein Eng Des Sel 2004; 17(2):133-140.
53. Zhao H, Giver L, Shao Z et al. Molecular evolution by staggered extension process (StEP) in vitro recombination. Nat Biotechnol 1998; 16(3):258-261.
54. He M, Yang ZY, Nie YF et al. A new type of class I bacterial 5-enopyruvylshikimate-3-phosphate synthase mutants with enhanced tolerance to glyphosate. Biochim Biophys Acta 2001; 1568(1):1-6.
55. Dion M, Nisole A, Spangenberg P et al. Modulation of the regioselectivity of a Bacillus alpha-galactosidase by directed evolution. Glycoconj J 2001; 18(3):215-223.
56. Vamvaca K, Butz M, Walter KU et al. Simultaneous optimization of enzyme activity and quaternary structure by directed evolution. Protein Sci 2005; 14(8):2103-2114.
57. Gould SM, Tawfik DS. Directed evolution of the promiscuous esterase activity of carbonic anhydrase II. Biochemistry 2005; 44(14):5444-5452.
58. Wong DW, Batt SB, Lee CC et al. High-activity barley alpha-amylase by directed evolution. Protein J 2004; 23(7):453-460.
59. Kim YW, Lee SS, Warren RA et al. Directed evolution of a glycosynthase from Agrobacterium sp. increases its catalytic activity dramatically and expands its substrate repertoire. J Biol Chem 2004; 279(41):42787-42793.
60. Johannes TW, Woodyer RD, Zhao H. Directed evolution of a thermostable phosphite dehydrogenase for NAD(P)H regeneration. Appl Environ Microbiol 2005; 71(10):5728-5734.
61. Hoseki J, Yano T, Koyama Y et al. Directed evolution of thermostable kanamycin-resistance gene: A convenient selection marker for Thermus thermophilus. J Biochem (Tokyo) 1999; 126(5):951-956.
62. Hao J, Berry A. A thermostable variant of fructose bisphosphate aldolase constructed by directed evolution also shows increased stability in organic solvents. Protein Eng Des Sel 2004; 17(9):689-697.
63. Zhang N, Suen WC, Windsor W et al. Improving tolerance of Candida antarctica lipase B towards irreversible thermal inactivation through directed evolution. Protein Eng 2003; 16(8):599-605.
64. Masip L, Pan JL, Haldar S et al. An engineered pathway for the formation of protein disulfide bonds. Science 2004; 303(5661):1185-1189.
65. Umeno D, Tobias AV, Arnold FH. Diversifying carotenoid biosynthetic pathways by directed evolution. Microbiol Mol Biol Rev 2005; 69(1):51-78.
66. Guntas G, Mitchell SF, Ostermeier M. A molecular switch created by in vitro recombination of nonhomologous genes. Chem Biol 2004; 11(11):1483-1487.
67. Guntas G, Ostermeier M. Creation of an allosteric enzyme by domain insertion. J Mol Biol 2004; 336(1):263-273.
68. Guntas G, Mansell TJ, Kim JR et al. Directed evolution of protein switches and their application to the creation of ligand-binding proteins. Proc Natl Acad Sci USA 2005; 102(32):11224-11229.
69. Truong K, Khorchid A, Ikura M. A fluorescent cassette-based strategy for engineering multiple domain fusion proteins. BMC Biotechnol 2003; 3:8.
70. Jeong J, Kim SK, Ahn J et al. Monitoring of conformational change in maltose binding protein using split green fluorescent protein. Biochem Biophys Res Commun 2006; 339(2):647-651.
71. Fehr M, Frommer WB, Lalonde S. Visualization of maltose uptake in living yeast cells by fluorescent nanosensors. Proc Natl Acad Sci USA 2002; 99(15):9846-9851.
72. Nikolaev VO, Gambaryan S, Lohse MJ. Fluorescent sensors for rapid monitoring of intracellular cGMP. Nat Methods 2006; 3(1):23-25.
73. Remus TP, Zima AV, Bossuyt J et al. Biosensors to measure inositol 1,4,5-trisphosphate concentration in living cells with spatiotemporal resolution. J Biol Chem 2006; 281(1):608-616.
74. Deuschle K, Okumoto S, Fehr M et al. Construction and optimization of a family of genetically encoded metabolite sensors by semirational protein engineering. Protein Sci 2005; 14(9):2304-2314.
75. Miyawaki A, Llopis J, Heim R et al. Fluorescent indicators for Ca^{2+} based on green fluorescent proteins and calmodulin. Nature 1997; 388(6645):882-887.
76. Truong K, Sawano A, Mizuno H et al. FRET-based in vivo Ca^{2+} imaging by a new calmodulin-GFP fusion molecule. Nat Struct Biol 2001; 8(12):1069-1073.
77. Miyawaki A, Griesbeck O, Heim R et al. Dynamic and quantitative Ca^{2+} measurements using improved cameleons. Proc Natl Acad Sci USA 1999; 96(5):2135-2140.

78. Braun DC, Garfield SH, Blumberg PM. Analysis by fluorescence resonance energy transfer of the interaction between ligands and protein kinase C delta in the intact cell. J Biol Chem 2005; 280(9):8164-8171.
79. Schleifenbaum A, Stier G, Gasch A et al. Genetically encoded FRET probe for PKC activity based on pleckstrin. J Am Chem Soc 2004; 126(38):11786-11787.
80. Sato M, Umezawa Y. Imaging protein phosphorylation by fluorescence in single living cells. Methods 2004; 32(4):451-455.
81. Wang Y, Botvinick EL, Zhao Y et al. Visualizing the mechanical activation of Src. Nature 2005; 434(7036):1040-1045.
82. Chiang JJ, Li I, Truong K. Creation of circularly permutated yellow fluorescent proteins using fluorescence screening and a tandem fusion template. Biotechnol Lett 2006; 28(7):471-475.
83. Nakamura T, Iwakura M. Circular permutation analysis as a method for distinction of functional elements in the M20 loop of Escherichia coli dihydrofolate reductase. J Biol Chem 1999; 274(27):19041-19047.
84. Iwakura M, Nakamura T, Yamane C et al. Systematic circular permutation of an entire protein reveals essential folding elements. Nat Struct Biol 2000; 7(7):580-585.
85. Qian Z, Lutz S. Improving the catalytic activity of Candida antarctica lipase B by circular permutation. J Am Chem Soc 2005; 127(39):13466-13467.
86. Nagai T, Yamada S, Tominaga T et al. Expanded dynamic range of fluorescent indicators for Ca^{2+} by circularly permuted yellow fluorescent proteins. Proc Natl Acad Sci USA 2004; 101(29):10554-10559.