Perspective

Biotechnology and Bioengineering

Directed Evolution of Novel Protein Functions

Huimin Zhao[1,2]

[1]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801; telephone: 217-333-2631; fax: 217-333-5052; e-mail: zhao5@uiuc.edu

[2]Departments of Chemistry and Bioengineering, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801

Received 9 March 2007; accepted 29 March 2007

Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bit.21455

ABSTRACT: Directed evolution has been successfully used to engineer proteins for basic and applied biological research. However, engineering of novel protein functions by directed evolution remains an overwhelming challenge. This challenge may come from the fact that multiple simultaneously or synergistic mutations are required for the creation of a novel protein function. Here we review the key developments in engineering of novel protein functions by using either directed evolution or a combined directed evolution and rational or computational design approach. Specific attention will be paid to a molecular evolution model for generation of novel proteins. The engineered novel proteins should not only broaden the range of applications of proteins but also provide new insights into protein structure–function relationship and protein evolution.

Biotechnol. Bioeng. 2007;98: 313–317.

© 2007 Wiley Periodicals, Inc.

KEYWORDS: proteins; directed evolution; rational design; promiscuous functions

Introduction

By harnessing the power of natural evolution in the test tube, directed evolution has emerged as a powerful tool for protein engineering, and fundamental studies of protein structure–function relationship (Arnold, 2001; Brannigan and Wilkinson, 2002). Proteins in nature have evolved, through selective pressure, to specific biological tasks. Thus, they are often not optimized for specific, non-natural applications, and require further tailoring. In the early 1980s, a structure-based rational design approach that involves rational alterations of selected residues in a protein via site-directed mutagenesis was developed to engineer proteins (Brannigan and Wilkinson, 2002). However, due to our limited understanding of protein folding, structure, function, and dynamics, this approach met with limited success. An alternative approach, directed evolution, was developed to overcome the limitations of rational design in the early 1990s (Brannigan and Wilkinson, 2002). Unlike rational design, directed evolution does not require any mechanistic, and structural information, and relies on high throughput screening or selection of a pool of protein variants generated by random mutagenesis and/or gene recombination. This approach has been successfully used to engineer a wide variety of protein functions such as activity, stability, selectivity, specificity, and affinity (Johannes and Zhao, 2006). It should be noted that directed evolution is not limited to protein engineering, and can also be used to engineer operons, pathways, viruses, and whole organisms (Schmidt-Dannert, 2001).

However, most of the directed evolution studies, or more broadly speaking, protein engineering studies, are concerned with improvement of existing protein functions, whereas creation of novel protein functions remains an overwhelming challenge (Brannigan and Wilkinson, 2002). This challenge may come from the fact that the engineering of a completely new function into an existing protein scaffold may require multiple coupled or synergistic mutations whereas individual mutations may have no effect (Bolon et al., 2002). Identifying such coupled mutations is difficult for both rational design and directed evolution. In this perspective, we highlight the key achievements in using directed evolution or a combined directed evolution

Correspondence to: H. Zhao

Contract grant sponsor: National Science Foundation
Contract grant number: BES-0348107
Contract grant sponsor: Biotechnology Research and Development Consortium (BRDC)
Contract grant number: Project 2-4-121
Contract grant sponsor: Office of Naval Research
Contract grant number: N00014-02-1-0725
Contract grant sponsor: National Institutes of Health, Department of Energy, and the DuPont Company

© 2007 Wiley Periodicals, Inc.

and rational design approach to create novel protein functions and discuss the implications of a molecular evolution model for the engineering of novel protein functions.

A Directed Evolution Approach

Due to its stepwise and incremental nature, directed evolution is well suited to improving an existing protein function. However, with the development of in vitro gene recombination methods that can recombine multiple homologous or non-homologous genes and ultra-high throughput screening or selection methods such as mRNA display, a much larger sequence space becomes accessible via a directed evolution approach, which may lead to the discovery of proteins with novel functions. As an early successful example, 50 rounds of DNA shuffling coupled with selection for ampicillin resistance were used to create a protein with novel ampicillin-resistant activity in *E. coli* from a functionally unrelated DNA fragment isolated from a hyperthermophilic archaeon *Pyrococcus furiosus* (Yano and Kagamiyama, 2001). The mechanism of the ampicillin resistance of this protein remains a mystery. DNA shuffling was also used to recombine two highly homologous triazine hydrolases. Screening a library of 1,600 variants against a synthetic library of 15 triazines yielded a set of variants that hydrolyzed five of the eight substrates that were not hydrolyzed by either of the parental enzymes (Raillard et al., 2001). The surprisingly diverse patterns of reactions catalyzed by these chimeric enzymes suggested the high evolvability of these two enzymes, which also seems to be the case for a DNA polymerase (Sauter and Marx, 2006). In the latter example, error-prone PCR was used to evolve reverse transcriptase activity in an N-terminal, shortened form of DNA polymerase from *Thermus aquaticus*. Positive mutants were identified after screening only approximately 2,000 variants (Sauter and Marx, 2006). In another example, a non-homologous gene recombination method was used to recombine the genes encoding for *E. coli* maltose binding protein (MBP) and TEM1 β-lactamase to create β-lactamase enzymes whose catalytic activity was regulated by maltose (Guntas et al., 2005).

As a demonstration of the true power of directed evolution in creating novel protein functions, mRNA display was used to create a library of 6×10^{12} proteins each containing 80 continuous random amino acids, followed by in vitro selection for ATP-binding (Keefe and Szostak, 2001). Repeated rounds of selection and amplification coupled with error-prone PCR yielded four new ATP-binding proteins that appear to be unrelated to each other or to anything found in the current protein databases (Keefe and Szostak, 2001). Further structure determination of one of these four proteins by protein crystallography indeed indicated a novel ADP- and zinc-binding fold (Lo Surdo et al., 2004). In a related study, by using the same mRNA display method, a zinc-finger domain of a human steroid receptor was converted to an ATP-binding protein with moderate affinity but high specificity (Cho and Szostak, 2006). Notably, the ability to isolate new functional proteins from a relatively small sampling of the entire sequence space shows that functional proteins are sufficiently common in protein sequence space (roughly 1 in 10^{11}) that they may be discovered by entirely stochastic methods. However, this frequency is still too low for protein engineers who attempt to design new protein functions. To increase this frequency, focused libraries of de novo proteins rather than libraries of completely randomly generated de novo proteins may be constructed by using a simple ‘‘binary code’’ which specifies only whether a given position is hydrophobic or hydrophilic (Hecht et al., 2004).

As an alternative strategy to accessing a larger sequence space and addressing the library size limitations in most directed evolution experiments, we recently developed an in vitro coevolution method that allows creation of novel protein functions (Chen and Zhao, 2005a,b). This process mimics the process of natural coevolution in the test tube, and involves the design of a hypothetical evolutionary pathway for the target function followed by stepwise directed evolution of the corresponding protein along the pathway (Fig. 1). Since the target function cannot be screened experimentally for the wild-type protein, the hypothetic pathway is designed such that each intermediate function can be screened experimentally. With the progression of directed evolution, these evolved intermediate functions will eventually bridge the functional gap between the wild-type protein function and the novel protein function. As proof of concept, this approach was successfully used to create a novel corticosterone activity in the human estrogen receptor (Chen and Zhao, 2005b). We are extending this approach to other proteins such as homing endonucleases with a goal of creating novel functions (Chen and Zhao, 2005a).

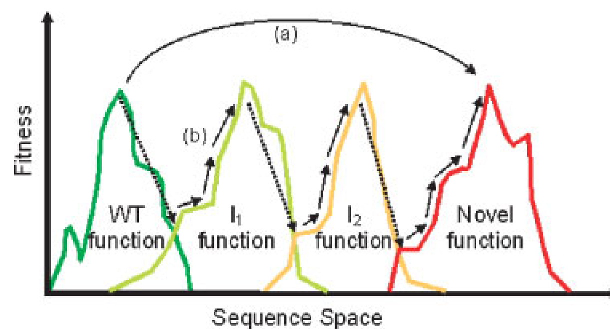

Figure 1. Scheme of the in vitro coevolution approach for creating novel protein functions. The engineering of a novel protein function represents a large jump in the sequence space (**a**) which may requires multiple coupled mutations. However, the addition of intermediate functions (I_1 and I_2) that are amenable to directed evolution creates a molecular bridge between the wild-type function and the novel function. **b**: The arrows illustrate a potential evolutionary path leading to the novel protein function.

DOI 10.1002/bit

A Combined Directed Evolution and Rational or Computational Design Approach

One of the main limitations of directed evolution is the inability to exhaustively search the vast sequence space of a protein. For a protein of typical size (300 amino acids), the number of variants containing three simultaneous mutations exceeds 10^{10}, which is often too large to be screened experimentally. Conversely, it is also an overwhelming challenge for rational or computational design to identify multiple mutations needed for creation of a novel protein function. Thus, a combined directed evolution and rational or computational design approach has been increasingly used to create novel protein functions.

As an impressive demonstration of this combined approach, the activity of β-lactamase was successfully introduced into the αβ/βα metallohydrolase scaffold of glyoxalase II (Park et al., 2006). The engineering of this novel activity into an existing scaffold involved deletion of the original glyosalase II substrate binding domain, followed by the introduction of loops designed by examining the structures of metallo β-lactamases, error-prone PCR, and DNA shuffling. The resulting enzyme had activity as a β-lactamase, albeit with much lower efficiency than the native enzyme. New activity can also be introduced into a non-catalytic scaffold, as demonstrated in the creation of triose phosphate isomerase activity within a ribose binding protein by using computational design and error-prone PCR (Dwyer et al., 2004).

In addition to introducing new enzymatic activity into an existing scaffold, new ways of controlling enzyme functions can also be introduced either by using a directed evolution approach as demonstrated in the above-mentioned maltose-regulated β-lactamase enzymes (Guntas et al., 2005) or by using a combined directed evolution and rational design approach. Using the latter strategy, Liu and coworkers (Buskirk et al., 2004) first inserted the ligand binding domain of human estrogen receptor into a minimal intein, which as expected, destroyed the splicing activity of the intein. Several rounds of directed evolution were then used to restore the splicing activity, resulting in a novel ligand-regulated intein. When this evolved intein was inserted into four unrelated proteins in living cells, the functions of those target protein also became dependent on the presence of the small molecule. Thus, the combined rational design and directed evolution approach created a general tool for controlling the function of arbitrary proteins in living cells using a small molecule.

In another impressive demonstration of this combined approach, Schultz and coworkers engineered an orthogonal transfer RNA/aminoacyl-tRNA synthetase pair that expands the number of genetically encoded amino acids in *E. coli* (Wang et al., 2001). Based on a structural model of the target synthetase, a library of synthetase variants was generated by randomizing five residues in the substrate binding pocket (Fig. 2a). To identify synthetase variants that specifically recognize the target unnatural amino acid (*O*-methyl-L-tyrosine) and no endogenous host amino acid, the synthetase library was subjected to alternating rounds of positive, and negative selections. When introduced into *E. coli*, the evolved pair led to the in vivo incorporation of *O*-methyl-L-tyrosine into protein in response to an amber nonsense codon. The same protein engineering approach was used to create multiple orthogonal transfer RNA/aminoacyl-tRNA synthetase pairs that enable the in vivo incorporation of more than 30 unnatural amino acids in *E. coli*, yeast, or mammalian cells (Wang et al., 2006). Most of these non-natural amino acids have useful functional groups such as spectroscopic probes, metal chelators, photoaffinity labels, and posttranslational modifications (Fig. 2b). Though yet to be demonstrated, introduction of such unnatural amino acids into the active site of an enzyme may result in a variant with new catalytic activity and chemistry.

A Model for Evolution of Novel Protein Functions

As demonstrated in the above-highlighted examples as well as many others (O'Brien and Herschlag, 1999), the number of

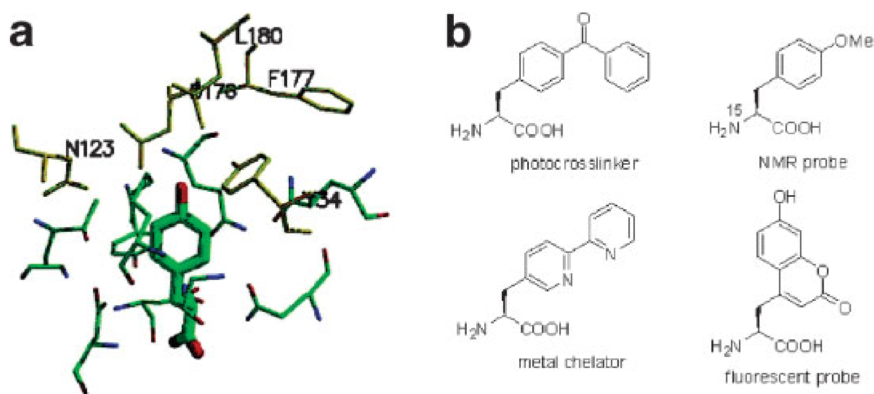

Figure 2. Engineering of orthogonal transfer RNA/aminoacyl-tRNA synthetase pairs to expand the genetic code. **a**: A library of aminoacyl-tRNA synthetase variants was generated by randomizing five residues in the substrate binding pocket followed by alternating rounds of positive and negative selections. **b**: A few representative unnatural amino acids that have been added to the genetic codes of *E. coli*, yeast, or mammalian cells.

mutations that endow a new function to an existing protein seems to be rather small. However, the mechanism of evolving novel protein functions both in the laboratory and in nature is still unclear. A generally accepted model for molecular evolution suggests that an evolving protein can initially acquire increased fitness for a new function without losing its original function (a promiscuous function), which is then followed by gene duplication and the divergence of a completely new function (Aharoni et al., 2005). A corollary of this model is that protein evolution involves the transition from a specialized protein into a general intermediate and ultimately a new re-specialized protein. Such model may have important implications for directed evolution of novel protein functions.

First, since the number of mutations required to create a new protein function is relatively small, the task of creating novel proteins, though still challenging, may not be as difficult as previously thought. It was believed that although distant mutations play an important role in protein functions as demonstrated in numerous directed evolution experiments, the active site residues seem to be more important determinants for altering of a protein function in an existing scaffold than remote residues (Morley and Kazlauskas, 2005). Several recent studies further confirmed this hypothesis by demonstrating the effectiveness of manipulating active site residues in creating new protein functions (Norrgard et al., 2006; Yoshikuni et al., 2006). In addition, thanks to recent advances in structural genomics and bioinformatics, identifying the active site residues for a protein of interest has become relatively straightforward and simple. Thus, a combined directed evolution and rational design approach that first explores locations with higher probability of success (i.e., the active site residues) seems to be the best strategy for creating novel protein functions. We recently used such approach to create an orthogonal receptor–ligand pair for use in gene switches and synthetic biology (Chockalingam et al., 2005).

Second, due to the dense collection of catalytic groups in an active site, a large number of proteins seem to have promiscuous activities (Aharoni et al., 2005; Bornscheuer and Kazlauskas, 2004). Such proteins should be excellent targets for creating novel protein functions. For example, the proteins with the $(\beta/\alpha)_8$-barrel, or TIM barrel, fold, which represent one of the largest superfamilies, exhibit a wide array of enzymatic functions. These proteins may be readily re-engineered to create new protein functions (Schmidt et al., 2003). In addition, it should be noted that new functions can sometimes also be discovered in well-characterized proteins. One impressive example is the recent discovery that *E. coli* alkaline phosphatase is also a phosphite-dependent hydrogenase (Yang and Metcalf, 2004).

Conclusion and Future Prospects

Creating proteins with novel functions has long been a goal of protein engineering and protein science, and advances have been made in recently years with methods using directed evolution or a combined directed evolution, and rational design approach. It has become increasingly accepted that many proteins show promiscuous functions and also only a limited number of amino acid substitutions are needed for the introduction of a new function in an existing protein scaffold. Such awareness should encourage more ambitious attempts and also stimulate the development of new approaches to engineering novel protein functions. Further developments in functional genomics, computational biology, and miniaturization, and automation technologies will shorten the time needed to create novel proteins for commercial applications and fundamental studies. And the combined use of directed evolution with rational or computational design will continue to be an incredibly useful tool for protein engineering.

We gratefully acknowledge financial support from National Science Foundation (BES-0348107), National Institutes of Health, Biotechnology Research and Development Consortium (BRDC) (Project 2-4-121), Office of Naval Research (N00014-02-1-0725), Department of Energy, and the DuPont Company.

References

Aharoni A, Gaidukov L, Khersonsky O, Gould SM, Roodveldt C, Tawfik DS. 2005. The 'evolvability' of promiscuous protein functions. Nat Genet 37:73–76.

Arnold FH. 2001. Combinatorial and computational challenges for biocatalyst design. Nature 409:253–257.

Bolon DN, Voigt CA, Mayo SL. 2002. De novo design of biocatalysts. Curr Opin Chem Biol 6:125–129.

Bornscheuer UT, Kazlauskas RJ. 2004. Catalytic promiscuity in biocatalysis: Using old enzymes to form new bonds and follow new pathways. Angew Chem Intl Edit 43:6032–6040.

Brannigan JA, Wilkinson AJ. 2002. Protein engineering 20 years on. Nat Rev Mol Cell Biol 3:964–970.

Buskirk AR, Ong YC, Gartner ZJ, Liu DR. 2004. Directed evolution of ligand dependence: Small-molecule-activated protein splicing. Proc Natl Acad Sci USA 101:10505–10510.

Chen Z, Zhao HM. 2005. A highly sensitive selection method for directed evolution of homing endonucleases. Nucleic Acids Res 33:e154.

Chen Z, Zhao HM. 2005. Rapid creation of a novel protein function by in vitro coevolution. J Mol Biol 348:1273–1282.

Cho GS, Szostak JW. 2006. Directed evolution of ATP binding proteins from a zinc finger domain by using mRNA display. Chem Biol 13:139–147.

Chockalingam K, Chen ZL, Katzenellenbogen JA, Zhao HM. 2005. Directed evolution of specific receptor–ligand pairs for use in the creation of gene switches. Proc Natl Acad Sci USA 102:5691–5696.

Dwyer MA, Looger LL, Hellinga HW. 2004. Computational design of a biologically active enzyme. Science 304:1967–1971.

Guntas G, Mansell TJ, Kim JR, Ostermeier M. 2005. Directed evolution of protein switches and their application to the creation of ligand-binding proteins. Proc Natl Acad Sci USA 102:11224–11229.

Hecht MH, Das A, Go A, Bradley LH, Wei YN. 2004. De novo proteins from designed combinatorial libraries. Protein Sci 13:1711–1723.

Johannes TW, Zhao HM. 2006. Directed evolution of enzymes and biosynthetic pathways. Curr Opin Microbiol 9:261–267.

Keefe AD, Szostak JW. 2001. Functional proteins from a random-sequence library. Nature 410:715–718.

Lo Surdo P, Walsh MA, Sollazzo M. 2004. A novel ADP- and zinc-binding fold from function-directed in vitro evolution. Nat Struct Mol Biol 11:382–383.

Morley KL, Kazlauskas RJ. 2005. Improving enzyme properties: when are closer mutations better? Trends Biotechnol 23:231–237.

DOI 10.1002/bit

Norrgard MA, Ivarsson Y, Tars K, Mannervik B. 2006. Alternative mutations of a positively selected residue elicit gain or loss of functionalities in enzyme evolution. Proc Natl Acad Sci USA 103:4876–4881.

O'Brien PJ, Herschlag D. 1999. Catalytic promiscuity and the evolution of new enzymatic activities. Chem Biol 6:R91–R105.

Park HS, Nam SH, Lee JK, Yoon CN, Mannervik B, Benkovic SJ, Kim HS. 2006. Design and evolution of new catalytic activity with an existing protein scaffold. Science 311:535–538.

Raillard S, Krebber A, Chen Y, Ness JE, Bermudez E, Trinidad R, Fullem R, Davis C, Welch M, Seffernick J. et al. 2001. Novel enzyme activities and functional plasticity revealed by recombining highly homologous enzymes. Chem Biol 8:891–898.

Sauter KBM, Marx A. 2006. Evolving thermostable reverse transcriptase activity in a DNA polymerase scaffold. Angew Chem Intl Edit 45:7633–7635.

Schmidt-Dannert C. 2001. Directed evolution of single proteins, metabolic pathways, and viruses. Biochemistry 40:13125–13136.

Schmidt DMZ, Mundorff EC, Dojka M, Bermudez E, Ness JE, Govindarajan S, Babbitt PC, Minshull J, Gerlt JA. 2003. Evolutionary potential of (beta/alpha)(8)-barrels: Functional promiscuity produced by single substitutions in the enolase superfamily. Biochemistry 42:8387–8393.

Wang L, Brock A, Herberich B, Schultz PG. 2001. Expanding the genetic code of *Escherichia coli*. Science 292:498–500.

Wang L, Xie J, Schultz PG. 2006. Expanding the genetic code. Annu Rev Biophys Biomed 35:225–249.

Yang KC, Metcalf WW. 2004. A new activity for an old enzyme: *Escherichia coli* bacterial alkaline phosphatase is a phosphite-dependent hydrogenase. Proc Natl Acad Sci USA 101:7919–7924.

Yano T, Kagamiyama H. 2001. Directed evolution of ampicillin-resistant activity from a functionally unrelated DNA fragment: A laboratory model of molecular evolution. Proc Natl Acad Sci USA 98:903–907.

Yoshikuni Y, Ferrin TE, Keasling JD. 2006. Designed divergent evolution of enzyme function. Nature 440:1078–1082.