Review

Journal of PeptideScience

Received: 15 December 2009 Revised: 24 February 2010 Accepted: 2 March 2010 Published online in Wiley Online Library: 28 June 2010

(wileyonlinelibrary.com) DOI 10.1002/psc.1233

Evolutionary mechanism as a template for protein engineering‡

Simone Eisenbeis and Birte Höcker*

The goal of a protein engineer is to adjust a protein to a specified new function. This is exactly what natural evolution has achieved many times. By studying evolutionary mechanisms, we can learn about ways to use the adaptability of proteins and to build new proteins. In fact, many techniques used in engineering are successfully mimicking evolutionary processes. We introduce the fundamental evolutionary mechanisms, take a closer look at duplication and fusion, recombination, and circular permutation and discuss their influence on protein engineering. Some important techniques are presented and illustrated with examples. Copyright © 2010 European Peptide Society and John Wiley & Sons, Ltd.

Keywords: recombination; duplication; fusion; circular permutation; diversification

Introduction

The objectives of a protein engineer are very similar to what nature has achieved over millions of years of evolution. In both cases, the ability of a protein to adapt to a specific new task is targeted. In a natural environment, for example, the cellular enzyme machinery has to adjust to changing conditions. Similar mechanisms can be used to engineer an enzyme for scientific or industrial purposes.

The protein universe is tremendously diverse. However, when taking a closer look it becomes apparent that proteins also display considerable similarities. This is due to the fact that they originate from the same basic set of autonomously folding units, the protein domains. These structural forms are classified into folds based on the spatial arrangement of their major secondary structural elements and their topological connections [1]. Protein folds are still complex and it has been assumed that they evolved by recombination from smaller but intrinsically stable peptide fragments [2]. Thus, the origin of the protein network that we observe today and how it developed have been the subject of manifold discussions.

A very efficient way to develop proteins with new specificities or catalytic mechanisms is the conversion of an already existing protein scaffold. In fact, it is well known that a lot of enzymes are already promiscuous for other activities [3] and thus only need minor changes to be optimized for these other substrates, reaction mechanisms or reaction conditions. A major factor in evolution is horizontal gene transfer, which allows the recruitment of already existing proteins from other organisms [4,5]. Gene duplication and diversification also play a crucial role. It is estimated that approximately 50% of all microorganism genes originated from gene duplication events [6,7].

Proteins then change through random drift and natural selection, thereby using a broad repertoire of mechanisms to produce selectable variants. Some of the mechanisms are more common than others and therefore have a greater contribution to the evolution of protein diversity. Frequently observed changes are point mutations, which only rarely result in a significantly different protein structure. More dramatic changes are caused by rearrangements of the protein sequence with the most prominent processes being duplication and fusion, recombination, and circular permutation (CP; Figure 1). We therefore want to have a closer look at these three mechanisms:

Duplication and fusion are fundamental processes in molecular evolution that comprise the single or multiple duplication of protein-coding DNA regions that through fusion events can lead to hybrid genes.

Recombination is a natural process of breaking and rejoining DNA strands to produce new combinations of genes and generate genetic variation.

Circular permutation (CP) is a rearrangement of the amino acid sequence, so that the *N*- and *C*-terminal ends are linked and new *N*- and *C*-termini are created.

All these mechanisms have been applied in various ways in the past. Nowadays, they are often combined with new methodologies that arise in the field of bioinformatics. Rapidly growing databases for sequences and protein structures provide plenty of information for predictive computer algorithms that can be used in protein design. For example, sequence alignments are used in consensus design approaches to improve protein stability, while rules about side-chain packing, for instance, are deduced from structural data that in the form of rotamer libraries and force fields can be used for structure prediction. In this article, however, we will focus on the evolutionary mechanism that more directly served as a template for protein engineering.

* *Correspondence to: Birte Höcker, Max Planck Institute for Developmental Biology, Spemannstr. 35, 72076 Tübingen, Germany. E-mail: birte.hoecker@tuebingen.mpg.de*

Max Planck Institute for Developmental Biology, Spemannstr. 35, 72076 Tübingen, Germany

‡ *Special issue devoted to contributions presented at the Symposium 'Probing Protein Function through Chemistry', –September 20–23, 2009, Ringberg Castle, Germany.*

Journal of PeptideScience

Biography

Simone Eisenbeis studied biology at Eberhard-Karls-Universität in Tübingen with a focus on microbiology. In 2006 she joined the group of Birte Höcker at the Max Planck Institute for Developmental Biology to work on her PhD and study the evolutionary relationship of protein folds.

Birte Höcker studied biology at the University of Göttingen and at Carleton University in Ottawa. In 2003 she completed her PhD in biochemistry at Köln university with Prof. Reinhard Sterner. From 2003 to 2005 she worked as a postdoctoral fellow on computational protein design at Duke University Medical Center, USA. Since 2006 she is a group leader at the MPI for Developmental Biology in Tübingen. Her main research interests are the evolution and design of protein folds and functions.

The Role of Duplication and Fusion in Natural Evolution

It is undisputed that gene duplication is a key force in molecular evolution. Genome duplication provides genetic material on which mutation, drift, and selection can act upon [8]. For example, a study of the evolutionary relationship of *Escherichia coli* and *Haemophilus influenzae* revealed that nearly all of the duplicated genes found in both bacteria today were already present as duplicates in their last common ancestor [9]. Duplication also had a strong influence on the evolution of metabolic pathways as has been investigated in detail for the histidine biosynthesis pathway [10]. On the single gene level, duplication removes the immediate selective pressure thereby allowing accumulation of mutations that could lead to the establishment of a new, related, or more complex function.

An equally important process for evolution is gene fusion. Covalent linkage of protein domains leads to a decrease of translational and rotational entropy of the unfolded states and therefore to a mutual stabilization [11–13]. A similar effect of entropy reduction in the unfolded state of monomeric proteins can be observed after the introduction of cross-links [14]. Gene fusion is thought to have an important impact on the evolution of protein interfaces. After the fusion of two genes coding for noninteracting proteins, the domains of the resulting fusion-protein might interact only unspecifically. The interactions at this interface can then gradually become optimized by successive mutations [15]. This model is supported by the observation that contact areas within proteins (intramolecular contacts) resemble contact areas between proteins (intermolecular contacts). The physical association of catalytic and regulatory structures also offers the advantage of promoting the channeling of unstable or low concentrated reactants. Thus, it is not surprising that the fusion of protein-coding DNA sequences has often been observed in metabolic pathway evolution.

Duplication and fusion events in combination are considered to also have played a crucial role in the early evolution of protein folds. Multiple duplication and fusion events can give rise to greater structural variability, especially when the repetition unit encompasses a region of defined secondary structure [2]. For instance, such events have resulted in the formation of repeat proteins, e.g. leucine-rich, tetratricopeptide, and ankyrin repeats. That these processes are still ongoing can be appreciated through evolutionary studies on the β-propeller fold, which consists of repeated four-stranded β-meanders [16].

Duplication and Fusion in Engineering

A protein family that over the years has attracted many evolutionary discussions are the $(\beta\alpha)_8$- or TIM (triose phosphate isomerase)-barrel proteins. In this fold, the repeating unit becomes apparent

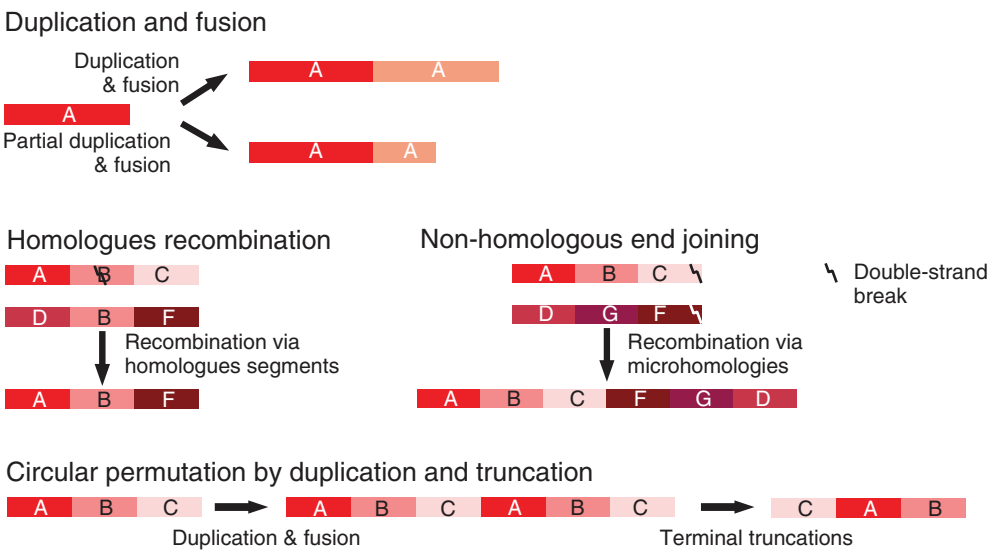

Figure 1. Schematic representation of duplication and fusion, homologous and nonhomologous recombination, and CP.

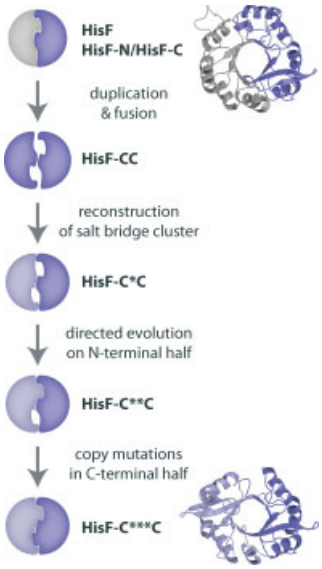

Figure 2. Scheme depicting the building process of a stable $(\beta\alpha)_8$-barrel from the *C*-terminal half of HisF (in blue). Duplication and fusion of the half-barrel is followed by successive optimization of the interface using rational as well as directed evolution approaches. The structures of HisF (pdb-id 1thf) and the symmetric HisF-C***C (pdb-id 2 w6r) are shown as cartoons colored according to the scheme on the left.

only on the structural level. However, an internal sequence symmetry was detected in two closely related $(\beta\alpha)_8$-barrel proteins of the histidine biosynthesis pathway named HisA and HisF. Their crystal structures revealed an astonishing similarity of their halves, which supports the hypothesis that the two structures evolved from an ancestral half-barrel via a twofold duplication event [17]. Further evidence for this hypothesis provided the observation that separately produced halves of HisF behave as independent folding units that form a functional heterodimer [18]. These studies prompted more experiments that tried to reconstruct such an evolutionary scenario. Thus, a stable and symmetric $(\beta\alpha)_8$-barrel was reconstructed from the *C*-terminal half of HisF (HisF-C) through a combination of rational design and directed evolution (Figure 2). First HisF-C was duplicated and fused yielding the protein HisF-CC. The interactions at the new interface were optimized by the introduction of two residues that reconstruct a salt bridge cluster in the inner barrel as it is observed in wild-type HisF [19]. To evolve this optimized variant (HisF-C*C) further, it was subjected to directed evolution in which it was selected for improved solubility upon expression in *E. coli*. For technical reasons only the first half of HisF-C*C was randomized. The selection identified two mutations close to the interface that improved solubility as well as multiple mutations in the loop connecting the two halves [20]. Shortening of the loop as well as the introduction of the two mutations into the second half improved solubility as well as stability. This new variant called HisF-C***C could be crystallized. Its solution structure provides insights into how the rational and randomly introduced changes influence the packing of this highly symmetric structure [21].

Another fold that has been engineered following evolutionary principles is the already mentioned repeat proteins. They are typically nonglobular proteins that are involved in numerous biological processes, especially mediating specific target interactions comparable to immunoglobulins. They are comprised of successive homologous structural units of 20–40 amino acids that stack to form elongated structures. Their stability increases with the number of repeating units as has been found in a series of consensus design experiments. In this method, stabilizing point mutations are introduced based on a multiple sequence alignment of homologous repeats. This approach has been applied successfully to ankyrin [22,23], tetratricopeptide [24], leucine-rich [25] and armadillo repeats [26]. Some of these idealized repeat proteins have served as scaffolds for directed evolution experiments that very successfully implemented new interaction surfaces for specific protein–protein interactions [27,28].

As shown, duplication and fusion events are important basic steps used in protein engineering and are an integral part of diversification mechanisms such as recombination and CP as is mentioned below.

Natural Recombination and its Role in the Evolution of Proteins

The purpose of homologous recombination in nature is to assure the integrity of genomic DNA after the occurrence of double-stranded breaks. Hence, natural recombination plays an important role in the DNA repair machinery. A second mechanism to fix breaks is nonhomologous end joining (NHEJ). The use of these two mechanisms differs between species [29]. Eukaryotic cells possess all features to apply both pathways: recombination between similar (homologous) and dissimilar DNA molecules (NHEJ). For NHEJ, short homologous single-stranded overhangs at the end of a double-stranded break are sufficient to guide repair, while in homologous recombination the sequences have to share a certain degree of homology. Originally, it was believed that prokaryotes only exhibit double-stranded break repair via homologous recombination. However, it was discovered that at least some bacteria are using NHEJ as described for *Bacillus subtilis* [30].

Homologous recombination is an extremely important evolutionary force to generate diversity. This was, for example, observed through band pattern and recombination detection methods within the *Vibrio* species [31]. Consequently, natural homologous recombination was applied to produce a combinatorial chicken antibody library using the recE recombination system of *E. coli* to generate diversity of two homologous genes [32].

Likewise, combinatorial assembly and/or exchange of smaller gene segments are believed to have played a crucial role in the evolution of protein domains [33]. This theory is supported by the work of Riechmann and Winter [34], who created stable chimeric protein domains by a combination of polypeptide fragment shuffling with a phage-based selection system. The *N*-terminal half of the five-stranded β-barrel domain of the cold shock protein CspA from *E. coli* was fused to random fragments of uniform length from genomic *E. coli* DNA. Stable proteins were selected via phage-display: the chimeric polypeptides were displayed on the surface of a filamentous bacteriophage and folded polypeptides selected by the ability to survive proteolysis. The structural part used from CspA was chosen based on its inability to fold independently; it corresponds to the 36 *N*-terminal residues and the first three β-strands of the domain. The size of the random fragments was limited to around 120 bp or 40 amino acids, which is also too small to form an independently folded domain. Several rounds of proteolytic digestion were performed to select for folded peptides. In the end, 11 different fragments were received, and 7 of these had the same open-reading frame as the *E. coli* gene where they originated from. Only one of the fused fragments showed detectable homology with the donor structure. Overall,

Journal of PeptideScience

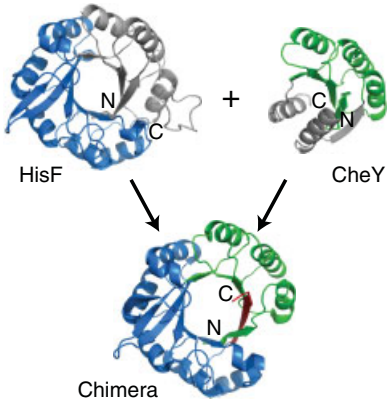

Figure 3. Generation of the chimera CheYHisF (pdb-id 3cwo) by fusion of fragments from HisF (in blue, pdb-id 1thf) and CheY (in green, pdb-id 1tmy). The newly formed ninth strand is shown in red.

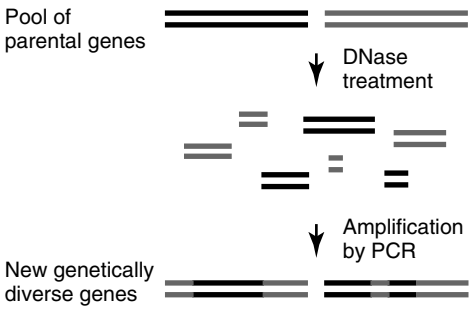

Figure 4. Schematic representation of DNA shuffling.

the experiment shows that domain diversity can be generated by recombination of random fragments, which is further supported by the ability of the CspA fragment to work as a folding promoter for other completely unstructured peptide segments.

As the main selective pressure in nature is functionality, the authors used the same technique in a second experiment in which they additionally screened for binding affinity. The folding template in this test was a subdomain fragment from hen egg lysozyme, which is part of the binding site in the folded protein. The second selection step that was based on the interaction with anti-lysozyme antibodies led to a chimera with improved affinity [35]. As the shuffled segments show no significant sequence homology to the applied fragments, this method demonstrates how the mechanism of nonhomologous recombination broadens the spectrum of reachable diversity considerably.

How natural recombination could have contributed to the diversification of protein folds can also be appreciated from rational recombination experiments of fold fragments. Proteins of the frequently observed $(\beta\alpha)_8$-barrel fold, which is believed to have evolved by duplication and fusion from a precursor half its size (see above), were build by a combination of fragments. For example, the combination of halves of the related HisA and HisF enzymes from histidine biosynthesis led to a very stable well-folded protein [19,36]. Subsequent directed evolution of the chimeric protein toward the TrpF activity, which is similar to the HisA activity, produced an extremely efficient enzymatic catalyst through the introduction of only few mutations [36]. This recombination experiment suggested an additional dimension for the diversification of $(\beta\alpha)_8$-barrel enzymes, namely their evolution by exchange of half-barrel domains with distinct functional properties. In addition, such proteins were built from fragments of two different folds. Starting from the question whether there are any proteins other than $(\beta\alpha)_8$-barrels that contain half-barrel-like structures, striking structural similarities between halves of the $(\beta\alpha)_8$-barrel HisF and proteins with the flavodoxin-like fold were observed [37]. Based on this observation, a chimera was constructed by replacing the *N*-terminal part of the $(\beta\alpha)_8$-barrel HisF with the structurally corresponding part of the flavodoxin-like protein CheY. The fragments $\beta1$ and $\alpha2-\beta5$ of CheY were fused to $\alpha4-\alpha8$ of HisF resulting in a very stable chimera. The crystal structure of the chimera at 3.1 Å revealed that the overall structure of the fragments is extremely similar to their structure in the parent proteins (Figure 3) [38]. Because the interfaces are not optimal, the *C*-terminus of the protein inserted into the β-barrel forming an additional ninth strand. This illustrates how random fragments can adapt to a scaffold and lead to variations of existing folds.

Recombination Techniques used in Engineering

The development of DNA shuffling [39], which is an efficient way to create *in vitro* recombination events, led to a multitude of applications. The method enables to combine, eliminate and redistribute sequences from one individual gene with other members of a gene pool. One starts with a target gene or a pool of parental genes that are randomly digested through DNase treatment. The resulting fragments are then reassembled in a subsequent PCR reaction (Figure 4). One limitation of this method, however, is the amount of diversity that can be created. As it is based on homologous recombination, shuffling of genes with less than 70% sequence identity is very challenging.

To solve this problem, more specified shuffling methods have been invented such as family shuffling [40] and gene sequence optimization [41]. Both lower the required sequence identity to around 50%. Family shuffling uses bridging intermediate gene sequences, whereas gene sequence optimization uses the computational tools and the degenerative genetic code to redesign an optimal sequence identity.

To build recombinant proteins in which the location of the crossover region is independent of sequence homology, a method called chimeragenesis was developed [42]. Here, chimeras are constructed based on two parental gene sequences that are truncated on either the one or the other end and are subsequently fused. The possibility to combine fragments without significant sequence identity opens up new applications that also allow far-reaching changes in protein structure as well as in function.

One such specialized method is Incremental Truncation for Creation of HYbrid enzymes (ITCHY) [43]. It uses two parental genes whose sequences are truncated from the end corresponding to either the *N*- or the *C*-terminal side. Randomly picked *N*- or *C*-terminal fragments from either parental protein are fused. The resulting library comprises fusion constructs of different sizes due to partial duplication and deletion with some members having the same length as the parent sequences. The method SCRATCHY combines ITCHY with DNA shuffling, which enables the introduction of multiple fusion points also in regions with lower sequence identity thereby raising the number of homology-independent crossovers.

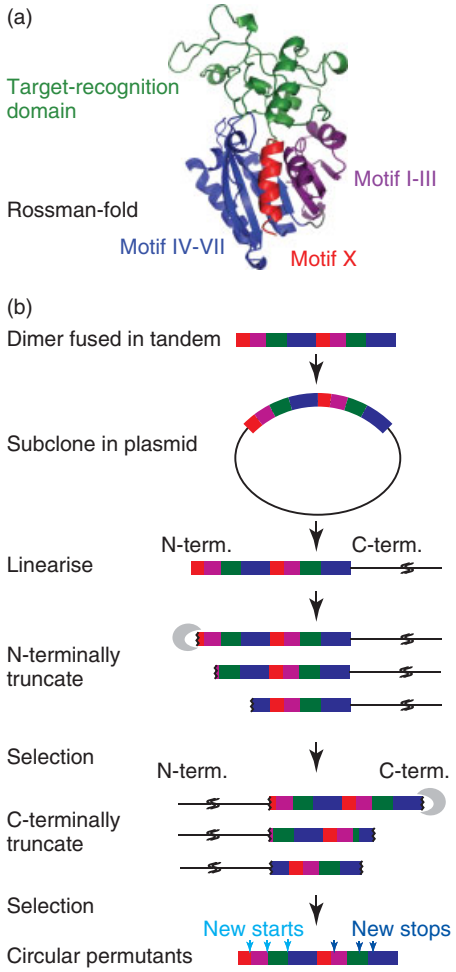

Figure 5. Mimicking CPs of DNA-methyltransferases. (a) The structure of M.HaeIII (pdb-id 1dct) is shown as a cartoon with domains and subdomains color-coded. (b) Schematic representation of the strategy used to create CPs of M.HaeIII. The colors of the blocks that display the protein-coding DNA correspond to the colors in (a).

CPs Observed in Nature

Once considered to be rare, today, more and more cases of CPs have been observed. With the increasing availability of data combined with improvements in sequence alignment and structure comparison tools, the numbers have enlarged further. In a systematic database search by Jung and Lee [44], 412 of 3035 (14%) protein domains were found to have arisen via CP. Known CPs are compiled and classified in the recently established database *Carcinogenic Potency Database*, which also provides a number of CP site prediction tools [45].

One reason why CPs are not frequently found and also not in all protein families is that very distant protein termini are not easily linked. On the other hand, CPs can occur with very little effect on the overall structure, stability and function, especially if the *N*- and *C*-termini of a protein are close enough to enable joining. It has even been proposed that CPs can be advantageous in cases where the original termini do not allow the insertion of domains into a multidomain context. This has been observed with PDZ domains [46].

There are several mechanisms by which CPs arise in nature. The first one identified was CP via posttranslational modification: the original *N*- and *C*-termini are connected via a peptide bond while the chain is cleaved at another position [47]. In this case, there is no change in the actual gene sequence. However, much more often CPs arise due to rearrangements on the genetic level through duplication and partial deletion, probably a multistep arrangement. On the contrary, CPs are usually engineered by ligation of the 5′ and 3′ ends of a gene and opening it at another site resulting in a protein with new termini.

One well-characterized example is the bacterial SAM (S-adenosylmethionine)-dependent DNA-methyltransferase family. The proteins in this family contain five permutation units, a SAM-binding domain, a catalytic domain, two halves of the target recognition domain and a *C*-terminal helix, which are found in alternating linear orders. Peisajovich *et al*. [48] simulated the postulated permutation by duplication using *HaeIII* methyltransferase (M.HaeIII) as the starting point (Figure 5a). Their aim was to observe all intermediates needed during the multistep rearrangement, because in principle all intermediates should retain some basic activity that enables the organism to survive. In a first step, a gene duplication and a fusion event were mimicked by cloning two copies of the M.HaeIII gene in tandem connected via a five amino acid linker (Figure 5b). This construct was fully active and protected against endonuclease digestion *in vivo*, which is the selection assay used. Then a library of truncated variants was generated by random introduction of stop or start codons using ITCHY (see recombination). From a potential repertoire of approximately 320 truncated intermediates three *N*- and eight *C*-terminal variants were selected. The locations of the new *N*- and *C*-termini map onto regions in the structure that link subdomains or domains. The new variants were then randomly truncated from the opposite end, which led to two different classes of M.HaeIII CPs. In a second approach, CPs were directly selected without an intermediate step. Because the resulting variants are similar, the authors conclude that the same structural constraints account for both intermediate and end product. Based on the location of the new termini, four modules were identified that in part behave as independent folding units. This indicates that the achievable topologies are restricted to intrinsic protein modularity.

Another example for naturally occurring CPs has been observed in the cradle-loop metafold. Alva *et al*. [49] introduced the metafold concept to classify protein folds based on their topology and natural descent. They illustrated the concept starting from the double-psi β-barrel that consists of a duplicated $\beta\beta\alpha\beta$ unit and extending it to the metafold of the cradle-loop barrels. An intensive bioinformatics and structural study allowed the identification of multiple homologs that are related by duplication, fusion, and CP events even though the proteins are not classified together. The metafold definition therefore provides a frame for a classification that reflects the relations of folds in natural evolution.

CP used in Engineering

As mentioned earlier, the classical goal in protein engineering is to improve predefined properties of the target protein. But as important as identifying improved variants, it is equally valuable to understand the reasons behind the changes. The first studies using artificially generated CPs were performed to identify recombinatory units and to learn about protein folding and stability. Only later it was realized that CP of an enzyme could also lead to improved activity.

One of these early protein folding studies was performed on the $(\beta\alpha)_8$-barrel protein phosphoribosyl anthranilate isomerase

Journal of PeptideScience

(PRAI) from *Saccharomyces cerevisiae* [50]. New *N*- and *C*-termini were introduced to test the influence of the order of the secondary structural elements on stability and folding. Two circular permutants were created: cPRAI-1 starts in the loop connecting helix 6 and strand 7, whereas cPRAI-2 starts in a loop close to the active site between strand 6 and helix 6. The wild-type termini in both versions have been linked by a new oligopeptide loop. Biophysical characterization indicated that the CP variants were structurally very similar to the wild type. Both were also catalytically active, although only cPRAI-1 was in a comparable range to the wild type. Thus, it was concluded that there are multiple folding pathways and that the change in termini location did not hinder the protein to reach a stable fold.

An approach to generate a library of CPs in the laboratory is to circularize the DNA sequence of the target protein via a linker omitting the start and stop codons. Then, the low levels of nonspecific endonuclease are used to randomly digest the cyclized DNA, ideally introducing one break per double-stranded DNA. In the last step, after the generation of blunt ends, these variants are cloned into an expression vector with already existing start and stop codons in all three possible reading frames [51].

Libraries of CPs can be used to identify sites in a given protein that are suitable for the insertion of other protein sequences. When generating a library for the target protein and selecting for stable CPs, the locations of the new termini are indicating possible regions to insert or fuse other target sequences. This approach has been used, for example, for the engineering of protein switches. In general, switches consist of a regulatory site that recognizes a signal, and a spatial distinct active site whose activity is modulated via the input signal of the recognition site. Ostermeier *et al*. generated such switches by combining *E. coli* maltose-binding protein (MBP) with the TEM1 β-lactamase (BLA) enzyme [52,53]. They created a family of enzymatic switches that confer resistance to β-lactam antibiotics through modulation of the maltodextrin concentration in the media.

Protein engineering via CP has been frequently used in folding studies; a very extensive one was performed on dihydrofolate reductase [54]. Additionally, it has also been exploited for many applications including the improvement of catalytic activity [55], protein stability [56], crystallizability [57] and oligomeric state modification [58].

Concluding Remarks

The diversity of the protein universe we observe today has developed through a multitude of small changes on existing schemes. Gene duplication has provided the material for mutation, drift and selection, whereas recombination and CP have played important roles in creating diversity. These mechanisms are the basis for many successful engineering approaches. The study of protein evolution therefore appears to be a good template for the design of protein engineering experiments.

References

1 Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 1995; **247**: 536–540.
2 Söding J, Lupas AN. More than the sum of their parts: on the evolution of proteins from peptides. *Bioessays* 2003; **25**: 837–846.
3 Hult K, Berglund P. Enzyme promiscuity: mechanism and applications. *Trends Biotechnol.* 2007; **25**: 231238.
4 Jain R, Rivera MC, Moore JE, Lake JA. Horizontal gene transfer in microbial genome evolution. *Theor. Popul. Biol.* 2002; **61**: 489–495.
5 Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. *Nature* 2000; **405**: 299–304.
6 Fani R, Mori E, Tamburini E, Lazcano A. Evolution of the structure and chromosomal distribution of histidine biosynthetic genes. *Orig Life Evol Biosph* 1998; **28**: 555–570.
7 Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. *Science* 2000; **290**: 1151–1155.
8 Crow KD, Wagner GP. Proceedings of the SMBE Tri-National Young Investigators' Workshop 2005. What is the role of genome duplication in the evolution of complexity and diversity? *Mol. Biol. Evol.* 2006; **23**: 887–892.
9 de Rosa R, Labedan B. The evolutionary relationships between the two bacteria *Escherichia coli* and *Haemophilus influenzae* and their putative last common ancestor. *Mol. Biol. Evol.* 1998; **15**: 17–27.
10 Fondi M, Emiliani G, Fani R. Origin and evolution of operons and metabolic pathways. *Res. Microbiol.* 2009; **160**: 502–512.
11 Terwilliger TC. Engineering the stability and function of gene V protein. *Adv. Protein Chem.* 1995; **46**: 177–215.
12 Nagi AD, Regan L. An inverse correlation between loop length and stability in a four-helix-bundle protein. *Fold Des.* 1997; **2**: 67–75.
13 Erickson HP. Co-operativity in protein–protein association. The structure and stability of the actin filament. *J. Mol. Biol.* 1989; **206**: 465–474.
14 Matsumura M, Becktel WJ, Levitt M, Matthews BW. Stabilization of phage T4 lysozyme by engineered disulfide bonds. *Proc. Natl. Acad. Sci. U. S. A.* 1989; **86**: 6562–6566.
15 Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, Eisenberg D. Detecting protein function and protein–protein interactions from genome sequences. *Science* 1999; **285**: 751–753.
16 Chaudhuri I, Soding J, Lupas AN. Evolution of the beta-propeller fold. *Proteins* 2008; **71**: 795–803.
17 Lang D, Thoma R, Henn-Sax M, Sterner R, Wilmanns M. Structural evidence for evolution of the beta/alpha barrel scaffold by gene duplication and fusion. *Science* 2000; **289**: 1546–1550.
18 Höcker B, Beismann-Driemeyer S, Hettwer S, Lustig A, Sterner R. Dissection of a (betaalpha)8-barrel enzyme into two folded halves. *Nat. Struct. Biol.* 2001; **8**: 32–36.
19 Höcker B, Claren J, Sterner R. Mimicking enzyme evolution by generating new (betaalpha)8-barrels from (betaalpha)4-half-barrels. *Proc. Natl. Acad. Sci. U. S. A.* 2004; **101**: 16448–16453.
20 Seitz T, Bocola M, Claren J, Sterner R. Stabilisation of a (betaalpha)8-barrel protein designed from identical half barrels. *J. Mol. Biol.* 2007; **372**: 114–129.
21 Höcker B, Lochner A, Seitz T, Claren J, Sterner R. High-resolution crystal structure of an artificial (betaalpha)(8)-barrel protein designed from identical half-barrels. *Biochemistry* 2009; **48**: 1145–1147.
22 Kohl A, Binz HK, Forrer P, Stumpp MT, Plückthun A, Grutter MG. Designed to be stable: crystal structure of a consensus ankyrin repeat protein. *Proc. Natl. Acad. Sci. U. S. A.* 2003; **100**: 1700–1705.
23 Mosavi LK, Minor DL, Jr, Peng ZY. Consensus-derived structural determinants of the ankyrin repeat motif. *Proc. Natl. Acad. Sci. U. S. A.* 2002; **99**: 16029–16034.
24 Main ER, Xiong Y, Cocco MJ, D'Andrea L, Regan L. Design of stable alpha-helical arrays from an idealized TPR motif. *Structure* 2003; **11**: 497–508.
25 Stumpp MT, Forrer P, Binz HK, Plückthun A. Designing repeat proteins: modular leucine-rich repeat protein libraries based on the mammalian ribonuclease inhibitor family. *J. Mol. Biol.* 2003; **332**: 471–487.
26 Parmeggiani F, Pellarin R, Larsen AP, Varadamsetty G, Stumpp MT, Zerbe O, Caflisch A, Plückthun A. Designed armadillo repeat proteins as general peptide-binding scaffolds: consensus design and computational optimization of the hydrophobic core. *J. Mol. Biol.* 2008; **376**: 1282–1304.
27 Binz HK, Amstutz P, Kohl A, Stumpp MT, Briand C, Forrer P, Grütter MG, Plückthun A. High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotechnol.* 2004; **22**: 575–582.
28 Binz HK, Amstutz P, Plückthun A. Engineering novel binding proteins from nonimmunoglobulin domains. *Nat. Biotechnol.* 2005; **23**: 1257–1268.
29 Barzel A, Kupiec M. Finding a match: how do homologous sequences get together for recombination? *Nat. Rev.* 2008; **9**: 27–37.

30 Weller GR, Kysela B, Roy R, Tonkin LM, Scalan E, Della M, Devine SK, Day JP, Wilkinson A, d'Adda di Fagagna F, Devine KM, Bowater RP, Jeggo PA, Jackson SP, Doherty AJ. Identification of a DNA nonhomologous end-joining complex in bacteria. *Science* 2002; **297**: 1686–1689.
31 Thompson CC, Thompson FL, Vicente AC, Swings J. Phylogenetic analysis of vibrios and related species by means of atpA gene sequences. *Int. J. Syst. Evol. Microbiol.* 2007; **57**: 2480–2484.
32 Wang PL, Lo BK, Winter G. Generating molecular diversity by homologous recombination in *Escherichia coli*. *Protein Eng. Des. Sel.* 2005; **18**: 397–404.
33 Gilbert W, de Souza SJ, Long M. Origin of genes. *Proc. Natl. Acad. Sci. U. S. A.* 1997; **94**: 7698–7703.
34 Riechmann L, Winter G. Novel folded protein domains generated by combinatorial shuffling of polypeptide segments. *Proc. Natl. Acad. Sci. U. S. A.* 2000; **97**: 10068–10073.
35 Riechmann L, Winter G. Early protein evolution: building domains from ligand-binding polypeptide segments. *J. Mol. Biol.* 2006; **363**: 460–468.
36 Claren J, Malisi C, Höcker B, Sterner R. Establishing wild-type levels of catalytic activity on natural and artificial (beta alpha)8-barrel protein scaffolds. *Proc. Natl. Acad. Sci. U. S. A.* 2009; **106**: 3704–3709.
37 Höcker B, Schmidt S, Sterner R. A common evolutionary origin of two elementary enzyme folds. *FEBS Lett.* 2002; **510**: 133–135.
38 Bharat TA, Eisenbeis S, Zeth K, Höcker B. A beta alpha-barrel built by the combination of fragments from different folds. *Proc. Natl. Acad. Sci. U. S. A.* 2008; **105**: 9942–9947.
39 Crameri A, Raillard SA, Bermudez E, Stemmer WP. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* 1998; **391**: 288–291.
40 Chang CC, Chen TT, Cox BW, Dawes GN, Stemmer WP, Punnonen J, Patten PA. Evolution of a cytokine using DNA family shuffling. *Nat. Biotechnol.* 1999; **17**: 793–797.
41 Moore GL, Maranas CD. eCodonOpt: a systematic computational framework for optimizing codon usage in directed evolution experiments. *Nucleic Acids Res.* 2002; **30**: 2407–2416.
42 Ostermeier M, Shim JH, Benkovic SJ. A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat. Biotechnol.* 1999; **17**: 1205–1209.
43 Ostermeier M, Lutz S. The creation of ITCHY hybrid protein libraries. *Methods Mol. Biol.* 2003; **231**: 129–141.
44 Jung J, Lee B. Circularly permuted proteins in the protein structure database. *Protein Sci.* 2001; **10**: 1881–1886.
45 Lo WC, Lee CC, Lee CY, Lyu PC. CPDB: a database of circular permutation in proteins. *Nucleic Acids Res.* 2009; **37**: D328–D332.
46 Vogel C, Morea V. Duplication, divergence and formation of novel protein topologies. *Bioessays* 2006; **28**: 973–978.
47 Carrington DM, Auffret A, Hanke DE. Polypeptide ligation occurs during post-translational modification of concanavalin A. *Nature* 1985; **313**: 64–67.
48 Peisajovich SG, Rockah L, Tawfik DS. Evolution of new protein topologies through multistep gene rearrangements. *Nat. Genet.* 2006; **38**: 168–174.
49 Alva V, Koretke KK, Coles M, Lupas AN. Cradle-loop barrels and the concept of metafolds in protein classification by natural descent. *Curr. Opin. Struct. Biol.* 2008; **18**: 358–365.
50 Luger K, Hommel U, Herold M, Hofsteenge J, Kirschner K. Correct folding of circularly permuted variants of a beta alpha barrel enzyme in vivo. *Science* 1989; **243**: 206–210.
51 Graf R, Schachman HK. Random circular permutation of genes and expressed polypeptide chains: application of the method to the catalytic chains of aspartate transcarbamoylase. *Proc. Natl. Acad. Sci. U. S. A.* 1996; **93**: 11591–11596.
52 Guntas G, Mansell TJ, Kim JR, Ostermeier M. Directed evolution of protein switches and their application to the creation of ligand-binding proteins. *Proc. Natl. Acad. Sci. U. S. A.* 2005; **102**: 11224–11229.
53 Guntas G, Ostermeier M. Creation of an allosteric enzyme by domain insertion. *J. Mol. Biol.* 2004; **336**: 263–273.
54 Iwakura M, Nakamura T, Yamane C, Maki K. Systematic circular permutation of an entire protein reveals essential folding elements. *Nat. Struct. Biol.* 2000; **7**: 580–585.
55 Qian Z, Horton JR, Cheng X, Lutz S. Structural redesign of lipase B from *Candida antarctica* by circular permutation and incremental truncation. *J. Mol. Biol.* 2009; **393**: 191–201.
56 Whitehead TA, Bergeron LM, Clark DS. Tying up the loose ends: circular permutation decreases the proteolytic susceptibility of recombinant proteins. *Protein Eng. Des. Sel.* 2009; **22**: 607–613.
57 Schwartz TU, Walczak R, Blobel G. Circular permutation as a tool to reduce surface entropy triggers crystallization of the signal recognition particle receptor beta subunit. *Protein Sci.* 2004; **13**: 2814–2818.
58 Chalton DA, Musson JA, Flick-Smith H, Walker N, McGregor A, Lamb HK, Williamson ED, Miller J, Robinson JH, Lakey JH. Immunogenicity of a Yersinia pestis vaccine antigen monomerized by circular permutation. *Infect. Immun.* 2006; **74**: 6624–6631.

