Protein Structure Design and Engineering

Steven M Berry, *University of Minnesota Duluth, Duluth, Minnesota, USA*

Yi Lu, *University of Illinois at Urbana-Champaign, Illinois, USA*

Advanced article

Article Contents

Online posting date: 15th February 2011

Protein structure design and engineering is a research endeavour in which proteins with predicted structure and function are constructed in the laboratory through rational design, combinatorial selection or combination of both approaches. It is built upon our knowledge about the structure and function of proteins and can be accomplished either from scratch (*de novo* design) or based on native scaffolds (redesign). The area of protein design is an exciting and rapidly growing field, advancing from the design of simple protein structures, to those that are more complicated and recently to the designs of functional proteins. Design of artificial proteins containing unnatural amino acids, backbone linkages or cofactors have also been reported, making it possible to prepare proteins with structural and functional properties beyond those of native proteins. These advances bring us closer to realising the dream of tailor-made artificial enzymes with high catalytic efficiency and selectivity for biotechnological and pharmaceutical applications.

Introduction

Proteins play essential roles in biology. Among their most exciting roles is that as highly specific and efficient biocatalysts or enzymes. Scientists have devoted a considerable amount of time and effort to study these proteins using biochemical and biophysical techniques to understand how these proteins form specific structures and then how these structures confer functions to the protein. Armed with the knowledge from these studies, scientists are beginning to design and engineer artificial proteins that mimic native proteins structurally and functionally. Protein design is a research endeavour that is an ultimate test of our understanding; the more we know about protein structure and function, the more likely a design will succeed. An equally important outcome is that protein design and engineering may allow us to produce artificial proteins that are smaller, cheaper and more stable than the naturally occurring proteins for use in many biotechnological and pharmaceutical applications (Lu, 2006a). **See also**: Engineered Enzymes

The protein design process can be divided into two steps: the design of the overall scaffold and design of the active site (Lu *et al*., 2009b). Although the designs of the overall scaffold and that of the active sites are linked, it is helpful to distinguish between these two steps because our knowledge about each step and their interplay is still limited at this early stage of protein design. We begin this article by describing a major challenge in the protein design and engineering field, the protein folding problem. This problem has given rise to many technologies that are used to find small proteins with a desired stable structure (Koder and Dutton, 2006). We will distinguish between two different extreme approaches used in protein design, namely rational and combinatorial approaches, and give examples of where these approaches have been used individually as well as in combination, to design stable proteins. Next, we summarise recent progress in designing proteins with non-natural amino acids and cofactors. Finally, we discuss recent progress in the area of designing functional proteins.

ELS subject area: Structural Biology

How to cite:
Berry, Steven M; and Lu, Yi (February 2011) Protein Structure Design and Engineering. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester.
DOI: 10.1002/9780470015902.a0002983.pub2

Protein Folding Problem

Proteins, or polypeptides, are linear polymers of the 20 naturally occurring amino acids and can be found in many different combinations. The linear polypeptide chains

(called the primary structure) fold into recurring small structural motifs (called the secondary structure). Alpha (α) helices and beta (β) sheets are the two primary types of secondary structure and may form through interaction of amino acids that are near one another. These secondary structural units assemble to form the basic overall structure of the protein (called the tertiary structure) through the spatial arrangement and interactions of amino acids that are far apart in the linear sequence. Finally, association of two or more polypeptide chains, each with its own tertiary structure, can result in even more complex structures called quaternary structure.

The protein folding problem refers to the vast number of possible structures that a single polypeptide chain can adopt. For example, if we consider a relatively small protein of 100 amino acids and assume that each residue will sample only two conformations, the number of possible conformations for this protein would be 2^{100} or 10^{30}. Even though the protein can sample each conformation very rapidly, for example, in picoseconds ($1\,\text{ps} = 10^{-12}\,\text{s}$), it would still take approximately 10^{10} years to achieve all the possible conformations. It is well known, however, that a protein can fold into a single conformation in a much shorter time (minutes or seconds). The fact that many natural proteins fold reliably and quickly to their native state despite the vast number of possible configurations is known as Levinthal's paradox. To further complicate this paradox, a protein in a cell often folds sequentially as it is synthesised one residue at a time from the ribosome. Chaperones may also aid in the proper folding of certain proteins. Finally, cofactors (such as metal ions) and post-translational modifications (such as phosphorylation or glycosylation) may influence the folding of the final structure. These are challenges scientists face in the field of protein structure design and engineering. **See also**: Protein Denaturation and the Denatured State; Protein Structure Prediction; Protein Structure Prediction and Databases

Protein Design and Engineering: Rational Approach

A protein can be designed rationally, on the basis of known information about protein structure and folding. This can be accomplished by design from scratch (*de novo* design) or through redesign based on a native protein scaffold. *De novo* design, a bottom-up approach, entails designing an entirely new protein, one amino acid at a time. This approach represents protein design in its purest and most challenging form, owing to the protein folding problem discussed earlier. On the contrary, redesign bypasses the folding problem by using native scaffolds that are known to fold properly and focuses mainly on improving the proteins stability or on the design of the active site, which is the final goal of protein design. The latter approach of redesign is based on the observation that the tens of thousands of known protein structures can all be classified into a few hundred scaffold types with recurring tertiary structures. Thus, nature is good at using a limited number of scaffolds to design proteins with different active sites for various functions. Learning this 'trick', separating the scaffold design from the active site design, is an important tool in the protein design process.

De novo rational design

De novo rational design relies on the existing knowledge of structural features responsible for the formation of stable secondary and tertiary structures as well as of the active sites of proteins. Many computer programs sample protein structures by fixing the backbone and varying residue side-chains (Suarez and Jaramillo, 2009). But recent programs are venturing out and introducing backbone flexibility in their calculations (Ventura and Serrano, 2004). In addition, new computational algorithms and theories have been developed to compute stable structures taking into account not only amino acid and protein backbone interactions but also solvent and other electrostatic effects (Vizcarra and Mayo, 2005). For example, the recently developed Rosetta program has found use in a wide range of modelling applications (Das and Baker, 2008). Furthermore, research in designing functional proteins has been made with goals of creating new binding sites, fostering protein–protein interactions and designing new functions (Mandell and Kortemme, 2009). **See also**: Protein Design

Currently, small stable proteins can be generated by *de novo* methods (DeGrado *et al*., 1999; Kang and Saven, 2007). However, these techniques are still being perfected and they often give rise to a starting protein structure that requires further rational refinement or combinatorial variation to stabilise the fold. Multiple α helical proteins, consisting of two, three, and four helices, and β sheet structures have been generated (DeGrado *et al*., 1999). Other protein structures are often redesigned to confer stability or functionality (Jestin and Pecorari, 2007). A major accomplishment was made with the computational design of a 93-residue mixed α/β protein that has a structure which is unknown in nature (**Figure 1**) using the Rosetta program (Kuhlman *et al*., 2003). In addition, work has been performed on designing lipid membrane stable peptides (Slusky *et al*., 2009). Chimeras of synthetic α helical structures with naturally occurring α helical membrane scaffolds have been constructed (Ye *et al*., 2004). By building on these successes, *de novo* designed proteins of folds already known in nature, as well as those with unknown folds, are gradually becoming a reality.

Furthermore, progress has been made in designing active sites into these *de novo* designed scaffolds. For example, haem-, dizinc-, diiron-, mercury-, cadmium- and iron–sulfur cluster-binding sites have been successfully engineered into proteins of four helical bundles (Calhoun *et al*., 2005; DeGrado *et al*., 1999; Ghosh and Pecoraro, 2005). However, active sites are still more often designed into naturally occurring scaffolds that are more frequently

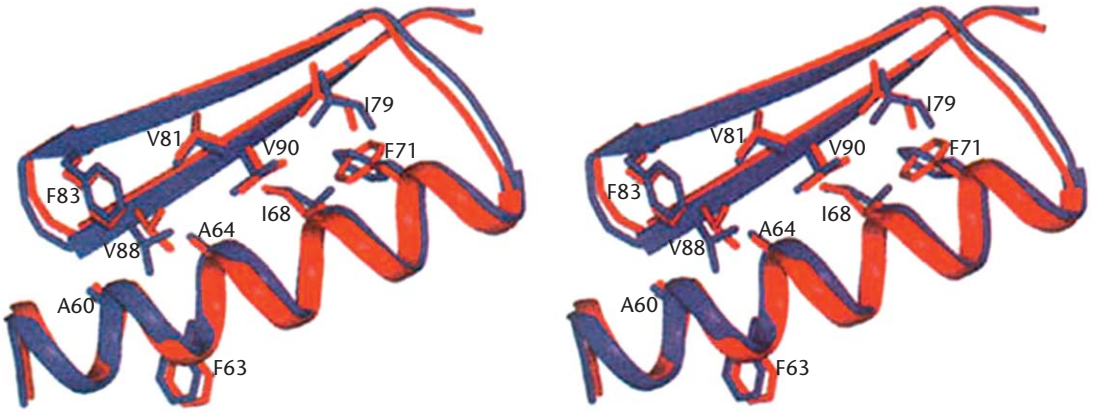

Figure 1 Stereoview of the X-ray crystal structure (red) of the unique *de novo* designed 93-residue α/β protein called Top7, which was determined to be very similar to the computationally designed model (blue) determined by the Rosetta program. Reproduced with permission from Kuhlman *et al.* (2003).

found to be stable to mutagenesis. This is changing, however, as more stable *de novo* designed scaffolds are made. Finally, successes in designing functional attributes into proteins are being explored and may result in the complete *de novo* design of proteins with tailor-made catalytic activities in the near future (Kaplan and DeGrado, 2004; Nanda and Koder, 2010).

Rational design based on native scaffolds

Rational design based on native scaffolds (redesign) usually entails a careful examination of known protein structures, followed by site-directed mutagenesis, large loop-directed mutagenesis or even domain swapping to introduce structural features to achieve the desired structure and function (Jestin and Pecorari, 2007; Lu *et al.*, 2009b; Nanda and Koder, 2010). Similar to *de novo* design, redesign requires a detailed knowledge of the protein active site, which can often be enhanced by computational algorithms to help identify specific structural elements required by the site.

Protein redesign has resulted in many proteins with improved or new functional activities (Jestin and Pecorari, 2007; Lu *et al.*, 2009b; Nanda and Koder, 2010). The rational redesign of existing cofactor-binding sites (including metal-binding sites, see **Figure 2**) to new cofactor sites with dramatically different structure and function has been achieved in many cases. Computational packages, such as Metal Search, Dezymer and Rosetta, can also be used to insert metal sites into protein scaffolds (Nanda and Koder, 2010; Wang *et al.*, 2010). Furthermore, the rational alteration of substrate specificity, reaction stereochemistry and solvent exposure can be achieved. Finally, studies have demonstrated that active site properties, such as redox potential, can be altered and tuned through secondary coordination sphere variations that leave the active site structure essentially unchanged as demonstrated by the successful tuning of a small copper protein across a 700-mV range (Marshall *et al.*, 2009). Many of these redesigned proteins have found practical applications in biochemistry and biotechnology.

Proteins Design and Engineering: Combinatorial Approach

Although tremendous progress has been made recently in rational design or redesign, one of the drawbacks is that our knowledge about protein structure and function is still limited. Furthermore, it is extremely difficult to predict long-range effects of residues far from the active site on the structure and function. For example, as aforementioned, most computer programs fix the protein backbone during structure optimisations. Only recently have programs begun to explore backbone movements (Suarez and Jaramillo, 2009) as these long-range effects are important in protein design.

In contrast to rational design, design through combinatorial and evolutionary methods requires little prior knowledge of the protein structure. Proteins with desired structure and function can be isolated from a large pool through various selection processes. The pools of proteins can be generated randomly or rationally (Wei and Hecht, 2004). Random pools are the most challenging to work with, due to the small percentage of 'quality' folds found within a pool and because only a limited amount of the vast sequence space can be searched. For example, a pool of typically $\sim 10^6$–10^{13} sequences can be generated, which is a minute fraction of the possible number of sequences, for example, found in a 100- ($20^{100} = 10^{130}$) or even 15-amino acid long ($20^{15} = 10^{19}$) peptide. Therefore, rational pools of random proteins are often more successful. For example, 'binary patterned' random peptides of alternating polar (P) and nonpolar (N) amino acids (Bradley *et al.*, 2007) have been used to generate proteins that are largely α helical

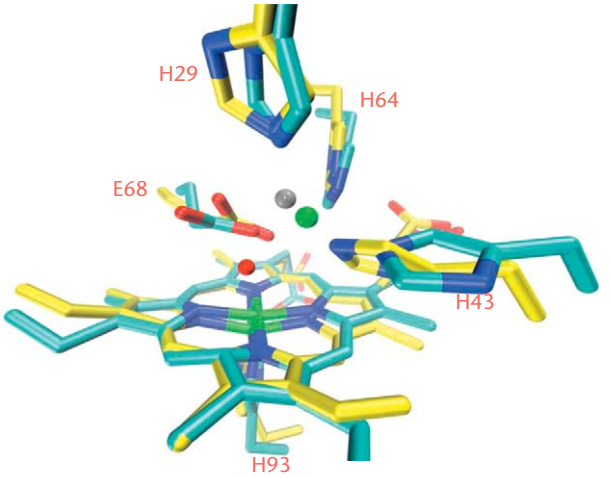

Figure 2 The rational redesign of myoglobin to a nitric oxide reductase: overlay of the 1.72 Å resolution crystal structure (blue) of a designed Fe_B-Mb centre and the computer model (yellow). The designed protein was found to mimic the structure and function of native nitric oxide reductase. Reproduced with permission from Yeung *et al.* (2009).

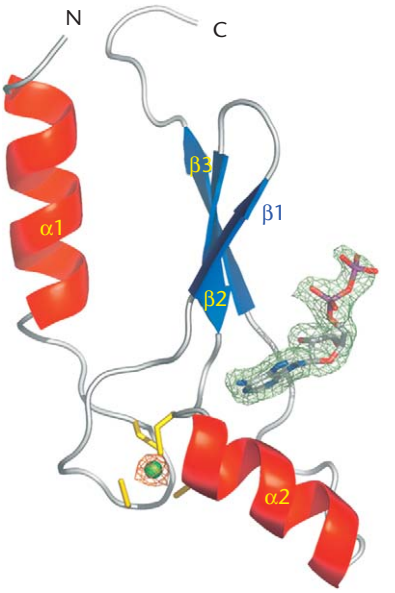

Figure 3 The first protein crystal structure derived from function-directed *in vitro* evolution. It has a Zn(II)-binding as well as an ADP-binding site. PDB ID=1UW1. Reproduced with permission from Lo Surdo *et al.* (2004).

(…PNPPNNP…) or β sheet (…PNPN…). Pools of such alternating amino acids can be generated starting at the deoxyribonucleic acid (DNA) level using semirandomly synthesised DNA sequences and further varied using cassette mutagenesis, error-prone polymerase chain reaction (PCR) or gene shuffling techniques (Jackel *et al.*, 2008; Tafelmeyer and Johnsson, 2004). The challenge is then to create a selection scheme to select for the characteristic of interest (Bloom and Arnold, 2009; Ventura and Serrano, 2004). One method for accomplishing this is through yeast, ribosome, phage or messenger ribonucleic acid (mRNA) display techniques (Jestin and Pecorari, 2007; Nizak *et al.*, 2005). In phage display, the protein of interest is covalently attached to the phage coat proteins and can be identified through the phage DNA (Cirino and Frei, 2009). Similarly, RNA display allows the linking of the protein of interest with its mRNA (Frankel *et al.*, 2003; Liu *et al.*, 2000). Assays are then performed to separate the displayed proteins, often by affinities to ligands or by susceptibility to protease cleavage. In addition, monoclonal antibodies can be isolated with protein structures that tightly bind metal ions or other target molecules (He *et al.*, 2004; Hilvert, 2000). **See also**: Catalytic Antibodies

These combinatorial methods have been used successfully to design both the protein scaffolds and the active sites of proteins. For example, dozens of α helical and β sheet proteins have been designed using the binary patterning method (Bradley *et al.*, 2007). Using directed evolution techniques, proteins with inverted stereochemistry, improved stability, improved substrate binding, novel cofactor binding or novel activities have been obtained (**Figure 3**) (Bloom and Arnold, 2009; Jestin and Pecorari, 2007). Results from these studies often reveal structural elements that would not necessarily have been obvious to rational designers.

Designing Artificial Proteins with Unnatural Amino Acids, Backbone Linkages or Cofactors

As mentioned in the section on Introduction, the main goals of protein design and engineering are to test our knowledge and to apply the designed proteins for practical applications. A true test of our understanding is whether one can not only create a protein that is a replica of the naturally occurring proteins but also translate that knowledge into designing novel proteins with unnatural moieties or with unprecedented efficiency, selectivity or even new activities. The latter proteins may be better suited for biochemical and biotechnological applications (Lu, 2006a; Lu *et al.*, 2009a).

There are a number of methods for incorporating unnatural moieties, such as unnatural amino acids, backbone linkages or cofactors, into proteins (Beatty and Tirrell, 2009; Cheng, 2004; Wang *et al.*, 2006). These methods include protein expression in auxotrophic organisms, chemical modification of cysteine residues, *in vitro* and *in vivo* protein translation and total synthesis using a solid-state peptide synthesiser. In addition, peptide ligation methods have been developed to assemble synthetic peptides containing unnatural moieties into proteins (Merkel *et al.*, 2009). A variation of this method, called expressed protein ligation, is a semisynthetic technique that couples a bacterially expressed peptide containing the majority of the protein sequence with a synthetic peptide containing the unnatural amino acids (Clark *et al.*, 2009; Pratt and Muir, 2007). The technique of choice for

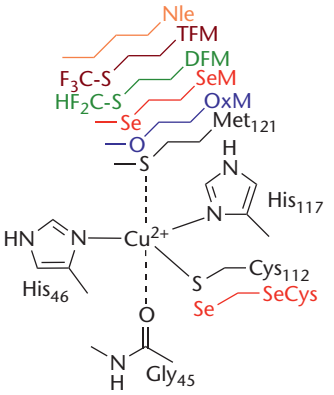

Figure 4 The semisynthetic technique of expressed protein ligation was used to vary the active site Met_{121} and Cys_{112} residues in the blue copper protein azurin with iso-structural amino acids to probe the electronic contributions of the ligands without altering steric parameters. Reproduced with permission from Garner *et al.* (2006).

incorporating artificial moieties depends on the system of study. If the addition of a single synthetic amino acid is desired, an auxotrophic expression host could be used and the growth medium supplemented with the desired amino acid, such as selenomethionine, to replace methionine. On the contrary, more complicated amino acid changes may require the total synthesis of a protein from scratch. This, however, is challenging, and for proteins larger than ~ 60 amino acids, semisynthetic methods such as expressed protein ligation might be the preferred technique.

Many examples of non-natural elements successfully incorporated into proteins are known (Lu, 2005; Taylor and Imperiali, 2009). These include proteins with amino acid analogues (**Figure 4**), D-amino acids (instead of the L-amino acids that occur in nature) and fluorescent dyes. Furthermore, in contrast to the α-amino acid backbone linkages in the native proteins, β- and γ-amino acids linkages have been introduced into designed proteins and have been shown to support formation of interesting and unusual helical and sheet structures (Koyack and Cheng, 2006). Finally, covalent attachments of both metal-containing (such as copper phenanthroline, Fe-EDTA, Mn-SALEN or organometallic compounds) and nonmetal containing (such as pyridoxamine) cofactors into proteins have resulted in new proteins with unprecedented activities (Abe *et al.*, 2009; Heinisch and Ward, 2010; Lu, 2005).

Designing Functional Proteins

The ultimate goal of protein design is to design functional proteins. As outlined earlier, we are learning more about protein structure and improving the combinatorial synthesis of protein libraries that has allowed multiple successes in designing specific protein structures. The next level of complexity, however, involves learning the specific structural factors that dictate protein function. The field of protein design is now focused on learning these factors and advancing towards the synthesis of functional proteins. A few recent examples of successfully engineered functional proteins have been realised (Abe *et al.*, 2009; Koder and Dutton, 2006; Lu *et al.*, 2009b; Mandell and Kortemme, 2009). These include examples of proteins with designed functional active sites as well as proteins with designed protein–protein or protein–macromolecule interactions.

In the design of functional proteins, a stable scaffold is typically chosen and rational design as well as combinatorial variation is applied to the putative active site to enhance or create novel functions. Often, many rounds and combinations of rational design and combinatorial selections are used to fine-tune different aspects of the function, from substrate binding to active site polarity and electronics. For example, the design of an oxygen-binding haemoglobin mimic marked an interesting success in the field (Koder *et al.*, 2009). The researchers began with a stable *de novo*-designed 4 α-helical bundle and through several rounds of rational design added two haem centres followed by modifications that were intended to stabilise the hydrophobic core, minimise dynamical movements and exclude water. The oxygen-binding affinities of the final protein matched natural globins and even improved on their discrimination of carbon monoxide binding. In many other studies, binding and catalytic functions, such as hydroxylation, hydrolysis and a Kemp elimination catalyst (**Figure 5**), have been designed (Bloom and Arnold, 2009; Jestin and Vichier-Guerre, 2005; Kaplan and DeGrado,

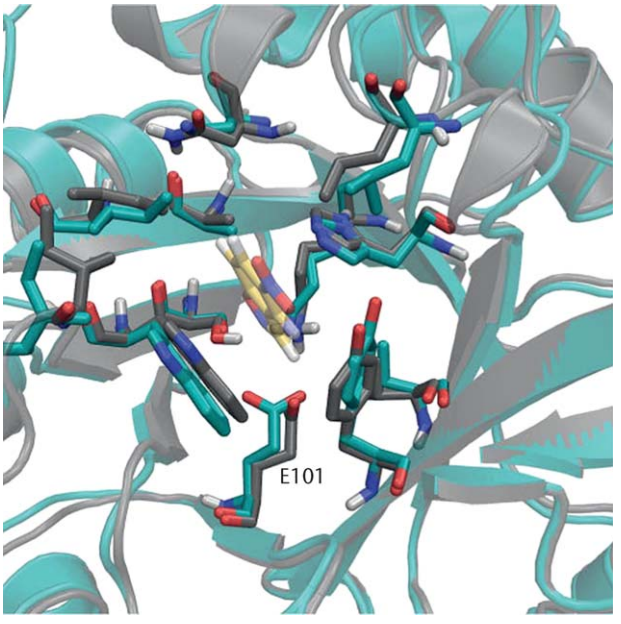

Figure 5 Overlay of crystal structure (cyan) and computationally designed structure (grey) of the active site for a prototypical Kemp elimination catalyst (removes an hydrogen ion from carbon). The final protein was found to accelerate the reaction rate by 10^5. Further rounds of directed evolution increased the rate parameters by 200 times, nicely demonstrating the combined use of *de novo* design followed by combinatorial directed evolution techniques. Reproduced with permission from Röthlisberger *et al.* (2008).

2004; Lu, 2006b; Lu *et al*., 2009b). Further rounds of directed evolution offer the exciting possibility of improving any given enzyme design given an appropriate selection method.

In addition to designing individual proteins, the design of specific protein–protein or protein–macromolecule interactions is being pursued by designers, as such interactions play important functional roles in biology. Protein–protein interactions were fostered between a protein and non-native partner, which structurally resembled the native partner, after redesigning their interfacial contacts (Mandell and Kortemme, 2009). Furthermore, a monomeric 56-amino acid domain of streptococcal protein G was dimerised with itself (Huang *et al*., 2007). Computational analysis identified 24 residues in a putative dimer interface. Following 8- and 12-fold mutations in the interface, a stable heterodimer was isolated. In addition, a haem-containing helical bundle was designed with both lipophilic and lipophobic sections assembled stably in phospholipid vesicles as well as stable protein–DNA complexes (Koder and Dutton, 2006; Mandell and Kortemme, 2009). Finally, because it is difficult to design specific interactions for two proteins with few interfacial contacts, Tezcan and coworkers have demonstrated the design of a metal-binding site at the interface of two proteins to overcome the entropic barrier for their interactions. Upon obtaining crystal structures of such metal-containing multiple domain proteins, they went on to use a variant of the RosettaDesign algorithms to optimise side-chain rotamer conformations. With as few as six mutations in cytochrome cb_{562}, the proteins formed dimers at micromolar concentration in the absence of any metal ions (Salgado *et al*., 2010). We are getting closer to attaining the dream of protein designers: designing novel proteins and catalysts for any given function of interest. **See also**: Protein–Ligand Interactions: General Description; Protein–protein Interactions: Identification

Summary and Conclusion

We have described different approaches to protein structure design and engineering and the progress made in each area. It is difficult to judge which approach is the best as the choice is highly dependent on the goal of the research and the protein systems involved. As structural knowledge accrues, the rational design of proteins can be performed more readily. However, when the knowledge available is limited, combinatorial methods can be implemented with more success. Recently, the rational design approach is becoming more popular because enhanced computer capabilities can allow more sequence space to be sampled than many combinatorial methods allow. However, many successes rely on both techniques. Rational methods may be best used to create an initial structure, whereas combinatorial selection can be better at optimising specific structural elements within the protein. The advantages, as well as drawbacks, of the rational design and combinatorial approaches complement one another.

The area of protein design is an exciting and fast-growing field. It is advancing from the design of stable protein structures to the recent designs of functional proteins (Lu *et al*., 2009b). The possible selectivity, specificity and high catalytic efficiency of these biologically based molecules make them extremely attractive targets for catalysts and materials (Lu, 2006a; Lu *et al*., 2009a). Proteins' environmentally friendly nature is another important advantageous property. With advances made in protein synthesis, both *in vivo* and *in vitro*, and in protein structure determination, through nuclear magnetic resonance (NMR) and X-ray crystallography, the everyday design of novel functional proteins is becoming a reality.

References

Abe S, Ueno T and Watanabe Y (2009) Artificial metalloproteins exploiting vacant space: preparation, structures, and functions. *Topics in Organometallic Chemistry* **25**: 25–43.

Beatty KE and Tirrell DA (2009) Noncanonical amino acids in protein science and engineering. *Nucleic Acids and Molecular Biology* **22**: 127–153.

Bloom JD and Arnold FH (2009) In the light of directed evolution: pathways of adaptive protein evolution. *Proceedings of National Academy of Sciences of the USA* **106**: 9995–10000.

Bradley LH, Wei Y, Thumfort P, Wurth C and Hecht MH (2007) Protein design by binary patterning of polar and nonpolar amino acids. In: *Methods in Molecular Biology*, vol. 352, pp. 155–166. Totowa, NJ: Humana Press Inc.

Calhoun JR, Nastri F, Maglio O *et al*. (2005) Artificial diiron proteins: from structure to function. *Biopolymers* **80**: 264–278.

Cheng RP (2004) Beyond *de novo* protein design – *de novo* design of non-natural folded oligomers. *Current Opinion in Structural Biology* **14**: 512–520.

Cirino PC and Frei CS (2009) Combinatorial enzyme engineering. In: Park SJ and Cochran JR (eds) *Protein Engineering and Design*, pp. 131–152. Boca Raton, FL: Taylor and Francis.

Clark KM, van der Donk WA and Lu Y (2009) Expressed protein ligation for metalloprotein design and engineering. *Methods in Enzymology* **462**: 97–115.

Das R and Baker D (2008) Macromolecular modeling with rosetta. *Annual Review of Biochemistry* **77**: 363–382.

DeGrado WF, Summa CM, Pavone V, Nastri F and Lombardi A (1999) *De novo* design and structural characterization of proteins and metalloproteins. *Annual Review of Biochemistry* **68**: 779–819.

Frankel A, Li S, Starck SR and Roberts RW (2003) Unnatural RNA display libraries. *Current Opinion in Structural Biology* **13**: 506–512.

Garner DK, Vaughan MD, Hwang HJ *et al*. (2006) Reduction potential tuning of the blue copper center in *Pseudomonas aeruginosa* azurin by the axial methionine as probed by unnatural amino acids. *Journal of the American Chemical Society* **128**: 15608–15617.

Ghosh D and Pecoraro VL (2005) Probing metal-protein interactions using a *de novo* design approach. *Current Opinionin Chemical Biology* **9**: 97–103.

He M, Cooley N, Jackson A and Taussig MJ (2004) Production of human single-chain antibodies by ribosome display. *Methods in Molecular Biology* **248**: 177–189.

Heinisch T and Ward TR (2010) Design strategies for the creation of artificial metalloenzymes. *Current Opinionin Chemical Biology* **14**: 184–199.

Hilvert D (2000) Critical analysis of antibody catalysis. *Annual Review of Biochemistry* **69**: 751–793.

Huang P-S, Love JJ and Mayo SL (2007) A de novo designed protein-protein interface. *Protein Science* **16**: 2770–2774.

Jackel C, Kast P and Hilvert D (2008) Protein design by directed evolution. *Annual Review of Biophysics* **37**: 153–173.

Jestin JL and Pecorari F (2007) Protein design. In: Hecht S and Huc I (eds) *Foldamers: Structure, Properties, and Applications*, pp. 267–289. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co, KGaA.

Jestin J-L and Vichier-Guerre S (2005) How to broaden enzyme substrate specificity: strategies, implications and applications. *Research in Microbiology* **156**: 961–966.

Kang S-G and Saven JG (2007) Computational protein design: structure, function and combinatorial diversity. *Current Opinion in Chemical Biology* **11**: 329–334.

Kaplan J and DeGrado WF (2004) *De novo* design of catalytic proteins. *Proceedings of the National Academy of Sciences of the USA* **101**: 11566–11570.

Koder RL, Anderson JLR, Solomon LA *et al.* (2009) Design and engineering of an O_2 transport protein. *Nature* **458**: 305–309.

Koder RL and Dutton PL (2006) Intelligent design: the *de novo* engineering of proteins with specified functions. *Dalton Transactions* **1**: 3045–3051.

Koyack MJ and Cheng RP (2006) Design and synthesis of beta-peptides with biological activity. In: *Methods in Molecular Biology*, vol. 340, pp 95–109. Totowa, NJ: Humana Press Inc.

Kuhlman B, Dantas G, Ireton GC *et al.* (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**: 1364–1368.

Liu R, Barrick JE, Szostak JW and Roberts RW (2000) Optimized synthesis of rna-protein fusions for *in vitro* protein selection. *Methods in Enzymology* **318**: 268–293.

Lo Surdo P, Walsh MA and Sollazzo M (2004) A novel ADP- and zinc-binding fold from function-directed *in vitro* evolution. *Nature Structural & Molecular Biology* **11**: 382–383.

Lu Y (2005) Design and engineering of metalloproteins containing unnatural amino acids or non-native metal-containing cofactors. *Current Opinion in Chemical Biology* **9**: 118–126.

Lu Y (2006a) Biosynthetic inorganic chemistry. *Angewante Chemie International Edition* **45**: 5588–5601.

Lu Y (2006b) Metalloprotein and metallo-DNA/RNAzyme design: current approaches, success measures, and future challenges. *Inorganic Chemistry* **45**: 9930–9940.

Lu Y, Garner DK and Zhang J-l (2009a) Artificial metalloproteins: design and engineering. *Wiley Encyclopedia of Chemical Biology* **1**: 124–133.

Lu Y, Yeung N, Sieracki N and Marshall NM (2009b) Design of functional metalloproteins. *Nature* **460**: 855–862.

Mandell DJ and Kortemme T (2009) Computer-aided design of functional protein interactions. *Nature Chemical Biology* **5**: 797–807.

Marshall NM, Garner DK, Wilson TD *et al.* (2009) Rationally tuning the reduction potential of a single cupredoxin beyond the natural range. *Nature* **462**: 113–116.

Merkel L, Moroder L and Budisa N (2009) Chemical protein engineering: synthetic and semisynthetic peptides and proteins. *Nucleic Acids and Molecular Biology* **22**: 29–64.

Nanda V and Koder RL (2010) Designing artificial enzymes by intuition and computation. *Nature Chemistry* **2**: 15–24.

Nizak C, Moutel S, Goud B and Perez F (2005) Selection and application of recombinant antibodies as sensors of rab protein conformation. *Methods in Enzymology* **403**: 135–153.

Pratt MR and Muir TW (2007) Expressed protein ligation. In: Schreiber SL, Kapoor TM and Wess G (eds) *Chemical Biology: From Small Molecules to Systems Biology and Drug Design*, vol. 2, pp. 537–566. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co, KGaA.

Röthlisberger D, Khersonsky O, Wollacott AM *et al.* (2008) Kemp elimination catalysts by computational enzyme design. *Nature* **453**: 190–195.

Salgado EN, Ambroggio XI, Brodin JD *et al.* (2010) Metal templated design of protein interfaces. *Proceedings of the National Academy of Sciences of the USA* **107**: 1827–1832.

Slusky JS, Yin H and DeGrado WF (2009) Understanding membrane proteins. How to design inhibitors of transmembrane protein-protein interactions. *Nucleic Acids and Molecular Biology* **22**: 315–337.

Suarez M and Jaramillo A (2009) Challenges in the computational design of proteins. *Journal of the Royal Society, Interface/The Royal Society* **6**: S477–S491.

Tafelmeyer P and Johnsson K (2004) Directed molecular evolution of proteins. In: Schmuck C and Wennemers H (eds) *Highlights in Bioorganic Chemistry: Methods and Applications*, pp. 341–343. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co, KGaA.

Taylor EV and Imperiali B (2009) Native chemical ligation: semisynthesis of post-translationally modified proteins and biological probes. *Nucleic Acids and Molecular Biology* **22**: 65–96.

Ventura S and Serrano L (2004) Designing proteins from the inside out. *Proteins* **56**: 1–10.

Vizcarra CL and Mayo SL (2005) Electrostatics in computational protein design. *Current Opinion in Chemical Biology* **9**: 622–626.

Wang C, Vernon R, Lange O, Tyka M and Baker D (2010) Prediction of structures of zinc-binding proteins through explicit modeling of metal coordination geometry. *Protein Science* **19**: 494–506.

Wang L, Xie J and Schultz PG (2006) Expanding the genetic code. *Annual Review of Biophysics and Biomolecular Structure* **35**: 225–249.

Wei Y and Hecht MH (2004) Enzyme-like proteins from an unselected library of designed amino acid sequences. *Protein Engineering, Design & Selection* **17**: 67–75.

Ye S, Strzalka JW, Discher BM *et al.* (2004) Amphiphilic 4-helix bundles designed for biomolecular materials applications. *Langmuir* **20**: 5897–5904.

Yeung N, Lin Y-W, Gao Y-G *et al.* (2009) Rational design of a structural and functional nitric oxide reductase. *Nature* **462**: 1079–1082.

Further Reading

Arndt KM and Müller KM (eds) (2007) Protein engineering protocols. In: *Methods in Molecular Biology*, vol. 352, 328 pp. Totowa, NJ: Humana Press Inc.

Arnold FH and Georgiou G (eds) (2003) Directed enzyme evolution: screening and selection methods. In: *Methods in Molecular Biology*, vol. 230, 370 pp. Totowa, NJ: Humana Press Inc.

Brakmann S and Schwienhorst A (eds) (2004) *Evolutionary Methods in Biotechnology: Clever Tricks for Directed Evolution*, 227 pp. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co, KGaA.

Camarero JA (2005) Polypeptide chemical ligation tools in protein engineering. *Protein & Peptide Letters* **12**: 721–799.

Farinas ET (2006) Directed evolution approaches for protein engineering. *Combinatorial Chemistry & High Throughput Screening* **9**: 235–329.

Guerois R and de la Paz ML (eds) (2006) *Protein Design: Methods and Applications*, 299 pp. Totowa, NJ: Humana Press Inc.

Köhrer C and RajBhandary UL (2009) Protein engineering. *Nucleic acids and Molecular Biology* **22**: 205–229.

Lutz S and Bornscheuer UT (2009) *Protein Engineering Handbook*, vols. 1 & 2. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co, KGaA, p. 1015.

Park SJ and Cochran JR (2009) *Protein Engineering and Design*, 416 pp. Boca Raton, FL: Taylor and Francis.

Robertson DE and Noel JP (eds) (2004) Protein Engineering. In: *Methods in Enzymology*, vol. 388, 456 pp. London, UK: Elsevier Academic Press.