Published in final edited form as:

Biochem Soc Trans. 2008 December ; 36(0 6): 1106–1111. doi:10.1042/BST0361106.

Controlling complexity and water penetration in functional *de novo* protein design

J.L. Ross Anderson^{*}, Ronald L. Koder[†], Christopher C. Moser^{*}, and P. Leslie Dutton^{*,1}

^{*}The Johnson Research Foundation, Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA 19104, U.S.A.

[†]Department of Physics, The City College of New York, NY 10031, U.S.A

Abstract

Natural proteins are complex, and the engineering elements that support function and catalysis are obscure. Simplified synthetic protein scaffolds offer a means to avoid such complexity, learn the underlying principles behind the assembly of function and render the modular assembly of enzymatic function a tangible reality. A key feature of such protein design is the control and exclusion of water access to the protein core to provide the low-dielectric environment that enables enzymatic function. Recent successes in *de novo* protein design have illustrated how such control can be incorporated into the design process and have paved the way for the synthesis of nascent enzymatic activity in these systems.

Keywords

abzyme; *de novo* protein design; haem; maquette; protein engineering; water penetration

Functional elements supporting catalysis

The generation of *de novo* enzymes offers the prospect of harnessing the impressive power and range of chemistry of natural enzymes in a robust package tailored to specific needs. Several protein engineering principles have been suggested as being critical for supporting catalysis [1–4]. The attempt to engineer them into natural proteins or artificial protein scaffolds presents a significant test of that understanding. These principles include the creation of a binding cavity that sequesters substrate into a hydrophobic environment and offers specific amino acid interactions to stabilize the transition state, as well as the positioning of cofactors as chemical or electron-transfer partners. Recent work *in silico* has shown the significant contribution that electrostatic interactions play in enzymatic catalysis [5]. Jencks [6] has emphasized that motion is just as important as static structures in facilitating the chemistry carried out by enzymes. There is still debate regarding the contribution of such motion in enzymatic catalysis, although recent work has shown that conformational microstates along the reaction co-ordinate occur during multi-step catalysis and that this phenomenon could drive the subsequent catalytic steps [7].

It appears to us that these engineering principles are simple and straightforward enough to consider their assembly in a relatively elementary synthetic protein scaffold. We do not subscribe to the idea that functions in proteins are necessarily highly optimized and are the rare achievement of millions of years of change and natural selection. It seems more likely that Nature selects structures and functions that are merely adequate for the organism's needs. Evidence of convergent evolution has consistently proved that one particular protein fold is not a necessity for one particular function. This functional simplicity combines with a practical understanding of how to make simple protein folds and the versatility of molecular biology and synthetic peptide chemistry to make construction of new enzymes feasible. However, certain properties of protein as a material must be considered when re-engineering natural proteins or constructing *de novo* proteins.

Darwin and Muller: complex characters

The intrinsic complexity of natural proteins presents a major challenge to delineating individual amino acid functions in natural enzymes and raises major barriers to their redesign while engineering new functions in artificial proteins. Two complementary principles (Figure 1) illustrate the roots of natural protein complexity. First, individual amino acids are naturally selected for their contribution to more than one function at a time, such as transport, folding and binding of cofactors and substrates (Figure 1A, left). This is the molecular analogue of Darwin's principle of multiple utility [8] that any one part of organism can serve multiple roles and be subject to many selective forces, and not optimized for any one force. Secondly, there is a tendency to develop an interdependency between amino acids relating to a particular function that leads to ever increasing complexity (Figure 1A, right). Muller [9] described this phenomenon in genetic systems where a change is made, typically for minor or even no selective advantage, and it then becomes essential as new changes begin to depend on the old change. The complex interdependency leads to epistatic effects observed when attempts are made to modify natural proteins [10]. Not having prior knowledge of protein fitness landscapes, it is inherently difficult to predict the outcome of multiple mutations. Additive effects of individual mutations on catalytic activity can be amplified beyond the sum of the original effects when applied simultaneously; individually negative mutations when applied together can also result in a net increase in activity compared with the wild-type enzyme.

New catalytic activity in old systems

Most work regarding the assembly of novel catalytic function has concentrated on the use of a small-molecule transition-state analogue. Catalytic antibodies, referred to as abzymes, are produced when a molecule similar to the hypothetical transition state of a given reaction is appended to a protein and exposed to the immune system of an organism [11]. The transition state of the desired reaction is thus stabilized within the abzyme's binding site and, when the actual substrate is mixed with the abzyme, catalysis occurs. This type of enzyme generation has proved effective in reproducing the basic activity, although the catalytic efficiencies are significantly lower than those exhibited by natural proteins. Another similar method entails quantum mechanical calculation of the transition state followed by exposure of this hypothetical molecule to a set of protein crystal structures *in silico* [12–14]. Once suitable

structures have been identified, further mutations to facilitate binding and catalysis are designed computationally and then tested *in vitro*. This method also suffers from significantly lower catalytic efficiencies compared with Nature, although completely novel catalytic activity has recently been incorporated into an existing protein fold [12].

Modifications like these of natural proteins suffer from four problems: (i) maintaining functional design upon changing more than a few amino acids in natural proteins is inherently difficult due to protein complexity; (ii) blindly exploring rugged protein fitness landscapes seldom leads to a successful scaling of desired peaks in this landscape; (iii) other factors, besides transition state binding, must also be introduced simultaneously to attain catalytic efficiency comparable with natural systems; and (iv) motion may play a significant role, and the protein design field has yet to address this issue.

Simplify roles for amino acids through *de novo* design

Protein complexity, however, can be kept under control in a system that has never experienced natural selection in any form: *de novo* protein maquettes ([15], and R.L. Koder, J.L.R. Anderson, L.A. Solomon, K.S. Reddy, C.C. Moser and P.L. Dutton, unpublished work) (Figure 1B). These maquettes, akin to a sculptor's simplified scale model of their work, are small robust proteins that are of the appropriate size and sequence to support both assembly into a desired framework and subsequent cofactor incorporation. Maquettes assembled in this laboratory are specifically designed to share no sequence identity with natural proteins. The design process begins by selecting a desired fold and building a generic sequence that will assemble accordingly. Amino acids are then inserted in specific positions to support cofactor binding. At this stage, the maquettes generally do not display native-like structure and can be described as molten-globule-like with rapid side-chain fluxionality, yet the overall fold is strongly maintained. To attain greater structural resolution, it is useful to iteratively adjust the sequence and test by NMR. At any stage in the process, engineering elements that are introduced purposefully or spontaneously arise can be elaborated or eliminated. Careful use of this procedure can result in a functional protein in which the reason for inclusion of every amino acid is largely understood and accounted for, and in which further changes can be thoughtfully considered and made in a tractable manner (R.L. Koder, J.L.R. Anderson, L.A. Solomon, K.S. Reddy, C.C. Moser and P.L. Dutton, unpublished work) (Figure 1B).

Our laboratory has had significant success in the design of haem-binding four- α -helix bundle maquettes (Figure 2). Regan and DeGrado [17] first demonstrated that, through binary patterning of polar and non-polar residues, soluble four- α -helix bundle proteins could be assembled from simple heptad repeat peptides [17]. In these peptides, polar residues are positioned on the protein surface and the non-polar residues are sequestered into the core, driving the protein assembly. The starting point for our design is a heptad repeat protein with three and a half repeats of the LLKKLLE sequence. Cofactor binding is achieved through replacement of two leucine residues at the interior **a**-positions with histidine residues, and loops are incorporated between sets of helices to reduce possible topomers to *syn* and *anti*. The resulting protein, designated H10H24 [15], is capable of binding four

haem moieties per bundle and reproduces UV–visible spectra typical of natural bis-histidine-ligated cytochromes *b*.

Keeping dry in water

One function of haem proteins is to bind oxygen. Protein is not essential to form a stable oxyferrous haem state, but water exclusion is. Simply dissolving imidazole-ligated haems in a non-polar plastic allows reversible formation of the oxyferrous state, whereas the same systems dissolved in water react rapidly, leading to deoxyferric haem and superoxide [18,19]. Many synthetic porphyrins when dissolved in aprotic organic solvents also form the oxyferrous state, supporting the view that isolating the binding or active site from water is key to function [20]. The catalytic function observed in other haem proteins such as the cytochromes P450 and peroxidases also relies on the exclusion and control of water [21]. If this property is common to natural proteins, then how can it be reproduced in *de novo* proteins? And how do you keep something dry in 55.6 M water?

Controlling structure to exclude water

There is evidence of water exclusion even in this first haem maquette, H10H24. The reduction potential of the haem is pH-dependent and coupled to glutamate residues [22] that rotate into the hydrophobic core upon haem binding. With two haems bound, there is a negatively co-operative electrostatic interaction between them of approx. 100 mV or 2.9 kcal · mol⁻¹ (1 kcal = 4.184 kJ) [15,22]. This electric field effect is tantalizing evidence of a relatively low dielectric environment in the interior haem environment that indicates significant shielding from water [23]. However, the presence of a low dielectric environment on average does not rule out the occasional penetration of water that may inactivate a catalytic site. This seems likely in our early bundles that display tertiary structure mobilities typical of molten globules [15]. Indeed, certain variants of H10H24 in which the histidine residue at position 24 is changed to alanine or serine [22,24] are mobile enough to flip from an all-helices parallel *syn* topology to an *anti* topology, with half of the helices pointing in the opposite direction, upon redox change of the haems. This demonstrates the ability to produce an allosterically regulated charge-activated conformational switch in these proteins [24].

The first step in reducing mobility is to introduce β -branched amino acids in key interior positions to restrict the number of possible rotamers [25,26]. The NMR [27] and crystal structure [28] of this protein, named BB L31M, showed that, although the apoprotein attained native-like structure, the interface between pairs of disulfide-linked helices is poorly formed, as there were few interfacial NOEs (nuclear Overhauser effects) in the NMR spectra. We can conclude that interface between dihelix pairs is fluxional and may even slip significantly, possibly facilitating water penetration.

Addition of specific contacts between helices provides a means to increase structural resolution in designed proteins [15,29,30]. DeGrado and colleagues showed that, for a protein with a molten-globule-like four- α -helix protein (α_4), it is possible to confer native-like structure through the introduction of Zn²⁺-binding sites into the protein [30]. The hydrophobic core of the protein is sufficient for assembly into a four- α -helix bundle, but the

binding between the metal ion and the amino acid side chains provide the specific organization that precipitates a native-like structure.

In our design process, the crystal structure of BB L31M showed that a helical rotation was necessary to position the histidine side chains in appropriate rotamers for haem ligation [28]. On the basis of simple visual modelling, the helical register was adjusted to accommodate this change. In contrast with BB L31M, the protein produced by these changes, named HP1 [31], is molten-globule-like in the apo form and attains native-like structure when haem is bound. This again illustrates the versatility of the nucleating effect of metal ions and cofactors on the protein structure. Similar effects can be seen in natural haem-containing proteins where removal of the haem can result in a large thermodynamic destabilization of the structure [32]. Although complete loss of native-like structure would be unlikely in such circumstances, specific interactions between cofactor and protein are undoubtedly strong enough to drive the holoprotein assembly to completion.

HP1, although attaining unique structure when haem is ligated, remains susceptible to water access (R.L. Koder, J.L.R. Anderson, L.A. Solomon, K.S. Reddy, C.C. Moser and P.L. Dutton, unpublished work). H/D exchange (hydrogen/deuterium exchange) of backbone amide protons occurs on the minute timescale, indicating a relatively low protection factor and large degree of solvent access, presumably through a poorly defined intermonomer interface akin to that of BB L31M [27]. Redesign of the loops connecting helices into a 'candelabra' motif and sequence diversification of surface amino acids resulted in the protein HP7 [33]. HP7 also becomes progressively structured upon haem binding and also adopts unique structures with haems other than protoporphyrin IX (Figure 3). In contrast with HP1, with two haems B bound, H/D exchange of the backbone amides of HP7 occurred on the hour timescale, indicating a significant inhibition of water access to the hydrophobic core (R.L. Koder, J.L.R. Anderson, L.A. Solomon, K.S. Reddy, C.C. Moser and P.L. Dutton, unpublished work). Concomitant with this observation of restricted water access is the acquisition of function, in this case reversible oxygen binding to the haems in aqueous solution. We believe that this is facilitated through the destabilization of the iron-histidine bond by glutamate residues imported into the core upon haem binding [33]. Burial of charge in the protein core would result in an entatic state in which the propensity for histidine-iron bond breaking is increased, thereby producing a population of penta-co-ordinate haem that can bind exogenous ligands. These glutamate residues are thought to rotate into the core in each of the precursor proteins H10H24 [22] through HP1 [31], but only become functionally important on acquisition of a watertight core and native-like structure.

Working in the membrane

Native-like structure is not the only method in which water can be effectively excluded from *de novo* proteins. The hydrophobic environment of a membrane can also provide the conditions where functional *de novo* proteins can be assembled. Lear, DeGrado and colleagues demonstrated that incorporation of multiple serine residues in specific motifs can drive the association of simple transmembrane helices into higher oligomers and that the proteins assembled were functional ion channels [34,35]. Engelman and colleagues also

demonstrated transmembrane helix association was possible through multiple serine and threonine motifs in designed peptides [36].

The design and assembly of haem-binding amphiphilic proteins has been studied in our laboratory. These proteins are modular four- α -helix bundles with their hydrophilic sequence taken from HP1 [31] and a hydrophobic trans-membrane sequence generally based on a natural protein [37]. Cofactor binding sites are included in sections of the protein relatively close to or far from bulk water and can incorporate a range of haem derivatives and chlorins. When incorporated into vesicles with two haems bound in the membrane section, the amphiphilic proteins display similar electric field effects to H10H24 [15,22], with a coupling of approx. 160 mV between adjacent haems [37]. The haems also display coupling with interior glutamate residues as is evident from their pH-dependent redox potentials. Such electric field effects are indicative of the expected hydrophobic environment of the assembled protein core in a membrane and illustrate the protection that is afforded by membrane assembly.

Conclusions

These developments show how maquettes clarify our understanding of how function can be assembled in proteins. In addressing the question of practical ways of controlling water access, we have focused on only one of many functional elements central to the assembly of enzymatic catalysis. Progress beyond our current functional proteins to true *de novo* enzymes will require the simultaneous incorporation of more engineering elements.

Acknowledgments

This work was supported by the National Institutes of Health grant GM41048.

Abbreviations used

H/D exchange	hydrogen/deuterium exchange
NOE	nuclear Overhauser effect

References

1. Jencks, WP. *Catalysis in Chemistry and Enzymology*. McGraw-Hill; New York: 1969.
2. Wolfenden R. Waterlogged molecules. *Science*. 1983; 222:1087–1093. [PubMed: 6359416]
3. Corey MJ, Corey E. On the failure of *de novo*-designed peptides as biocatalysts. *Proc Natl Acad Sci USA*. 1996; 93:11428–11434. [PubMed: 8876152]
4. Baltzer L, Nilsson J. Emerging principles of *de novo* catalyst design. *Curr Opin Biotechnol*. 2001; 12:355–360. [PubMed: 11551463]
5. Warshel A, Sharma PK, Kato M, Xiang Y, Liu H, Olsson MHM. Electrostatic basis for enzymatic catalysis. *Chem Rev*. 2006; 106:3210–3235. [PubMed: 16895325]
6. Jencks WP. Binding-energy, specificity and enzyme catalysis: Circe effect. *Adv Enzymol*. 1975; 43:219–410. [PubMed: 892]
7. Boehr DD, McElheny D, Dyson HJ, Wright PE. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science*. 2006; 313:1638–1642. [PubMed: 16973882]
8. Darwin, C. *The Origin of Species by Means of Natural Selection*. Earlton House; New York: 1872.

9. Muller HJ. The relation of recombination to mutational advance. *Mutation Res.* 1964; 1:2–9. [PubMed: 14195748]
10. Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature.* 2007; 445:383–386. [PubMed: 17251971]
11. Hilvert D. Critical analysis of antibody catalysis. *Annu Rev Biochem.* 2000; 69:751–793. [PubMed: 10966475]
12. Rothlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, et al. Kemp elimination catalysts by computational enzyme design. *Nature.* 2008; 453:190–195. [PubMed: 18354394]
13. Jiang L, Althoff EA, Clemente FR, Doyle L, Rothlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF, et al. *De novo* computational design of retro-aldol enzymes. *Science.* 2008; 319:1387–1391. [PubMed: 18323453]
14. Zhang X, DeChancie J, Gunaydin H, Chowdry AB, Clemente FR, Smith AJT, Handel TM, Houk KN. Quantum mechanical design of enzyme active sites. *J Org Chem.* 2008; 73:889–899. [PubMed: 18179229]
15. Robertson DE, Farid RS, Moser CC, Urbauer JL, Mulholland SE, Pidikiti R, Lear JD, Wand AJ, DeGrado WF, Dutton PL. Design and synthesis of multi-haem proteins. *Nature.* 1994; 368:425–432. [PubMed: 8133888]
16. Reference deleted
17. Regan L, DeGrado WF. Characterization of a helical protein designed from 1st principles. *Science.* 1988; 241:976–978. [PubMed: 3043666]
18. Wang JH, Nakahara A, Fleischer EB. Hemoglobin studies. 1 The combination of carbon monoxide with hemoglobin and related model compounds. *J Am Chem Soc.* 1958; 80:1109–1113.
19. Wang JH. Hemoglobin studies. 2 A synthetic material with hemoglobin-like property. *J Am Chem Soc.* 1958; 80:3168–3169.
20. Collman JP, Boulatov R, Sunderland CJ, Fu L. Functional analogues of cytochrome *c* oxidase, myoglobin, and hemoglobin. *Chem Rev.* 2004; 104:561–588. [PubMed: 14871135]
21. Sono M, Roach MP, Coulter ED, Dawson JH. Heme-containing oxygenases. *Chem Rev.* 1996; 96:2841–2887. [PubMed: 11848843]
22. Schifman JM, Moser CC, Kalsbeck WA, Bocian DF, Dutton PL. Functionalized *de novo* designed proteins: mechanism of proton coupling to oxidation/reduction in heme protein maquettes. *Biochemistry.* 1998; 37:16815–16827. [PubMed: 9843452]
23. Schifman JM, Gibney BR, Sharp RE, Dutton PL. Heme redox potential control in *de novo* designed four- α -helix bundle proteins. *Biochemistry.* 2000; 39:14813–14821. [PubMed: 11101297]
24. Grosset AM, Gibney BR, Rabanal F, Moser CC, Dutton PL. Proof of principle in a *de novo* designed protein maquette: an allosterically regulated, charge-activated conformational switch in a tetra- α -helix bundle. *Biochemistry.* 2001; 40:5475–5487.
25. Gibney BR, Johanson JS, Rabanal F, Salicky JJ, Wand AJ, Dutton PL. Global topology & stability and local structure & dynamics in a synthetic spin-labeled four-helix bundle protein. *Biochemistry.* 1997; 36:2798–2806. [PubMed: 9062107]
26. Gibney BR, Rabanal F, Salicky JJ, Wand AJ, Dutton PL. Iterative protein redesign. *J Am Chem Soc.* 1999; 121:4952–4960.
27. Gibney BR, Huang SS, Skalicky JJ, Fuentes EJ, Wand AJ, Dutton PL. Hydrophobic modulation of heme properties in heme protein maquettes. *Biochemistry.* 2001; 40:10550–10561. [PubMed: 11523997]
28. Huang SS, Gibney BR, Stayrook SE, Dutton PL, Lewis M. X-ray structure of a maquette scaffold. *J Mol Biol.* 2003; 326:1219–1225. [PubMed: 12589764]
29. Osterhout JJ Jr, Handel TM, Na G, Toumadje A, Long RC, Connolly PJ, Hoch JC, Johnson WC Jr, Live D, DeGrado WF. Characterization of the structural properties of α_1 B, a peptide designed to form a four-helix bundle. *J Am Chem Soc.* 1992; 114:331–337.
30. Handel TM, Williams SA, DeGrado WF. Metal ion-dependent modulation of the dynamics of a designed protein. *Science.* 1993; 261:879–885. [PubMed: 8346440]

31. Huang SS, Koder RL, Lewis M, Wand AJ, Dutton PL. The HP-1 maquette: from an apoprotein structure to a structured hemoprotein designed to promote redox-coupled proton exchange. *Proc Natl Acad Sci USA*. 2004; 101:5536–5541. [PubMed: 15056758]
32. Landfried DA, Vuletich DA, Pond MP, Lecomte JTJ. Structural and thermodynamic consequences of *b* heme binding for monomeric apoglobins and other apoproteins. *Gene*. 2007; 298:12–28. [PubMed: 17550789]
33. Koder RL, Valentine KG, Cerda J, Noy D, Smith KM, Wand AJ, Dutton PL. Nativelike structure in designed four- α -helix bundles driven by buried polar interactions. *J Am Chem Soc*. 2006; 128:14450–14451. [PubMed: 17090015]
34. Lear JD, Wasserman ZR, DeGrado WF. Synthetic amphiphilic peptide models for protein ion channels. *Science*. 1988; 240:1177–1181. [PubMed: 2453923]
35. DeGrado WF, Gratkowski H, Lear JD. How do helix–helix interaction help determine the folds of membrane proteins? Perspectives from the study of homo-oligomeric helical bundles. *Protein Sci*. 2003; 12:647–665. [PubMed: 12649422]
36. Dawson JP, Weinger JS, Engelman DM. Motifs of serine and threonine can drive association of transmembrane helices. *J Mol Biol*. 2002; 316:799–805. [PubMed: 11866532]
37. Discher BM, Noy D, Strzalka J, Ye S, Moser CC, Lear JD, Blasie JK, Dutton PL. Design of amphiphilic protein maquettes: controlling assembly, membrane insertion, and cofactor interactions. *Biochemistry*. 2005; 44:12329–12343. [PubMed: 16156646]

Figure 1. Origins of protein complexity

(A) According to Darwin's principle of multiple utility, each amino acid (dark grey circle) contributes to multiple functions (dark grey arrows and light grey circle) in a protein such as folding, cofactor binding and catalysis. Each function that a protein performs (light grey circle) is dependent on the co-operative effects of multiple amino acids (interlocking circles) leading to a type of Mullerian interdependency. (B) Through the process of assembly and testing, it is possible to create synthetic proteins where the contributions of each amino acid to individual functions are simpler, mostly understood and can be modified in a tractable manner.

Figure 2. Restricting water access in *de novo* proteins through topological redesign

Schematic representations of HP1 and HP7 where the hydrophobic interior is purple, the alternating positive and negative charge on the protein surface is shown as blue and pink, and glutamate residues are red. HP1 is oriented in an *anti*-topology with a poorly defined intermonomer interface and is thought to allow water access into the core even though native-like structure is attained upon haem ligation. The candelabra motif of HP7 prohibits water access to the core through the use of the haems and a disulfide bond to fix the intermonomer interface.

Figure 3. Haem-facilitated structuring of a *de novo* protein monitored by ^{15}N HSQC NMR
(A) Apo-HP7 is a molten-globule-like four-helix bundle. **(B)** On binding one equivalent of protoporphyrin IX, the helices containing the haem-ligating histidine residues attain unique structure, while the non-ligating helices remain disordered. Three glutamate residues (red triangles) rotate into the core upon binding. **(C)** Binding a second equivalent of protoporphyrin IX confers native-like structure to the remaining helices and another three glutamate residues enter the hydrophobic core. Reprinted with permission from [33], © 1998 American Chemical Society.