Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 8729–8730, September 1994

Commentary

De novo design of β-sheet proteins

Michael H. Hecht

Department of Chemistry, Princeton University, Princeton, NJ 08544-1009

Why Design Novel Proteins?

Nature has evolved an enormous number of different proteins. They fold into a variety of structures and carry out a bewildering diversity of functions. With so many natural proteins to observe and manipulate, what can be gained by the design and characterization of novel proteins?

There are two main motivations for pursuing protein design. The first is a practical one: Current efforts to design simple protein structures represent the essential first steps toward the eventual design of novel macromolecules "made to order" for solving important chemical or biochemical problems.

The second motivation for protein design is to test our understanding of basic principles. A complete understanding of any structure or phenomenon is ultimately demonstrated by an ability to design a similar structure or phenomenon from first principles. Current attempts at *de novo* protein design might be likened to initial attempts at designing a suspension bridge. As a first step, one studies the detailed structures of known examples. Next one might study stability—for a bridge, the chaotropic agent would be wind or water and for a protein it might be urea or high temperature. Finally, one might engineer alterations and observe their effects on structure and stability. For a preexisting suspension bridge, this might be the addition of a lower deck; for proteins it might involve modification of an active site residue by site-directed mutagenesis. However, even after observing many bridges, testing their stabilities in floods and high winds, and engineering a variety of structural alterations, one cannot claim to fully understand suspension bridges without building such a bridge *de novo*, crossing the river upon it, and demonstrating that one's feet remain dry. By analogy, our understanding of natural proteins—their folding pathways, thermodynamic stabilities, and catalytic properties—is ultimately tested by our ability to design novel proteins with predetermined structures and properties.

Two recent papers, one in these *Proceedings* (1) and the other in *Protein Science* (2), demonstrate that the ability to design novel proteins has reached another milestone.

α-Helical Versus β-Sheet Designs

The production and characterization of *de novo* proteins is a relatively new endeavor. Although designing proteins "from scratch" has been an appealing prospect for some time (3, 4), the actual production of amino acid sequences longer than ≈50 residues was not practical until quite recently. However, over the past several years, substantial advances in both peptide synthesis and molecular biology have made it possible to produce virtually any sequence the would-be designer chooses. So now, the question has become: What to choose? The majority of initial attempts at *de novo* design chose simple α-helical motifs, most notably, four-helix bundles (5–11). In contrast to these earlier works, the authors of the two recent papers (1, 2) chose the more difficult task of designing β-sheet proteins.

α-helical proteins are easier to design from both theoretical and practical perspectives. An individual α-helix can exist in isolation (12, 13), and thus α-helices can be considered independent structural modules to be used as building blocks for protein design. In contrast, an isolated β-strand is not stable, rendering β-structure less modular and inherently more difficult to design. This contrast between α- and β-structures stems from the fundamental difference in the hydrogen bonding patterns of the two types of secondary structures (14, 15). In the α-helix, backbone hydrogen bonding is intrasegmental. It connects the C=O of residue i to the N—H of residue $i + 4$. Thus the α-helix can be relatively self-contained. It can satisfy most of its backbone hydrogen bonding needs without help from a partner. The situation for β-strands is quite different. The C=O and N—H groups in β-strands are hydrogen bonded to N—H and C=O groups on neighboring strands. Thus the β-strand by its very nature is more gregarious.

This difference has significant practical implications for protein design. Since α-helices satisfy most of their backbone hydrogen bonding within their modular structures, the design of an α-helical protein can focus on the hydrophobic contacts between the nonpolar faces of several amphiphilic α-helices. If the design is successful, the molecule will adopt a unique fold, with the number of helices dictated by the specific contacts between them. Open-ended and uncontrolled oligomerization into high-order oligomers typically is not a serious problem.

The situation is quite different for β-structures. Both the formation of backbone hydrogen bonds and the burial of hydrophobic surface area require a β-strand to interact with neighboring elements of structure. Furthermore, different neighbors satisfy different needs. For a β-strand going into the page (see Fig. 1), the neighboring strands on its right and left sides can form hydrogen bonds to the backbone N—H and C=O groups. Side chains will point up and down, enabling hydrophobic residues to interact with neighbors above or below the original strand. Thus, the β-strand can form favorable interactions with neighbors in four directions (left, right, up, and down). This neighborliness has significant practical implications: β-strands are so gregarious that they frequently aggregate and precipitate out of solution (16, 17).

This tendency to aggregate makes β-structures more difficult to design than α-structures. It is not sufficient merely to design favorable interactions between several β-strands; indeed, assiduous design of extensive interstrand contacts may drive β-strands to aggregate and precipitate. For a design to succeed, it must ensure that interactions with additional "unwanted" β-strands are unfavorable. In particular, for the design of a β-sandwich protein such as betadoublet (1) or betabellin 14D (2), one must ensure that the intended edge strand of a β-sheet indeed forms an edge and does not reach over the edge to form further interactions leading to extensive aggregation. Thus, *de novo* β-sandwiches, far more than 4-helix bundles, require attention to "negative design" (8). It is not sufficient to design for one particular structure; it is equally important to design against competing alternative structures. Failure to restrict the innate gregariousness of β-strands can result in promiscuous and unwanted interactions, ultimately leading to irreversible downfall into an amorphous precipitate.

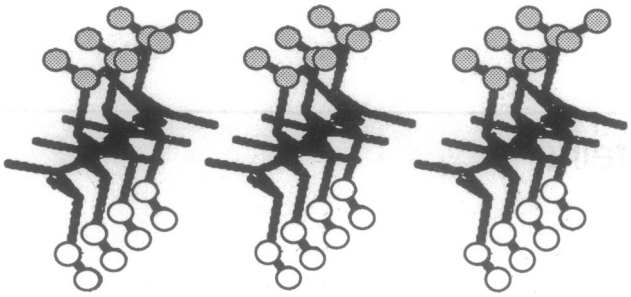

FIG. 1. Three parallel strands of a β-sheet. The N—H and C=O groups on the central β-strand can hydrogen bond to the neighboring β-strands on the left and right sides. Within each β-strand, side chains alternate pointing up and down. Hydrophobic interactions involving shaded side chains would occur above the plane of the β-sheet and hydrophobic interactions involving unshaded side chains would occur below the plane.

There have been several earlier attempts to design β-sheet proteins (17–20), and both betadoublet (1) and betabellin 14D (2) are direct descendants of the early betabellins designed by the Richardson and Erickson groups (17, 18). However, previous attempts to design native-like β-sheet proteins typically led to insoluble aggregates, or at best, marginally soluble material. The two new proteins reported by Quinn *et al.* (1) and by Yan and Erickson (2) represent a significant advance. Both proteins are soluble at 10 mg/ml in aqueous buffers (1, 2). This high solubility has enabled the study of both proteins by NMR spectroscopy. This would have been unthinkable for the aggregating and insoluble progenitors of these *de novo* β-sheet proteins (17, 18).

Native Proteins and Novel Proteins

The two novel β-sheet proteins were produced by two very different technologies. Quinn *et al.* (1) made a synthetic gene and expressed betadoublet in *Escherichia coli* as part of a fusion protein. This biological approach facilitates high-level production, rapid purification, and straightforward mutagenesis. In contrast, Yan and Erickson (2) synthesized betabellin 14D by solid-phase methods. This approach is inherently nonbiological and facilitated the incorporation of 12 nonnative D-amino acids in the 64-residue protein.

If *de novo* design is the ultimate test of our understanding of a natural system, then we must judge the success of a design by its ability to recapitulate the properties of the natural system. How closely do the novel proteins betadoublet (1) and betabellin 14D resemble the native proteins found in living systems?

Natural water-soluble proteins form compact globular structures with a discreet oligomeric state. They contain hydrophobic interiors, polar exteriors, and an abundance of secondary structure. Both betadoublet and betabellin 14D appear to have captured these properties quite well.

More subtle features of natural proteins are revealed by detailed structural and thermodynamic studies. Recapitulation of these subtleties is currently the major challenge of *de novo* protein design (21, 22). Natural proteins typically form unique structures with well-ordered hydrophobic cores. They bury their nonpolar side chains quite effectively and, therefore, do not bind the hydrophobic dye 1-analinonaphthalene-8-sulfonate (ANS). Furthermore, native proteins typically give rise to NMR spectra with well-dispersed chemical shifts, long-range nuclear Overhauser effects (i.e., between residues distant in primary sequence), and amide hydrogens that are protected from exchange with solvent. Additionally, the thermal denaturation profiles of natural proteins show cooperative two-state transitions, indicating that the well-ordered native structure unfolds in a concerted fashion into a disordered unfolded state.

Betadoubet and betabellin 14D have recapitulated many of these properties. Although neither structure has been shown to be fully native-like by NMR and ANS-binding, both proteins demonstrate temperature melting profiles reminiscent of natural proteins. Since cooperative thermal denaturation in the absence of chaotropes has been quite difficult to attain even for designed 4-helix bundles (5, 8, 22–24), the melting behavior of these two β-sheet proteins is very encouraging. Perhaps the field of protein design is one step closer to getting across the river—with dry feet—on a bridge designed *de novo*.

1. Quinn, T. P., Tweedy, N. B., Williams, R. W., Richardson, J. S. & Richardson, D. C. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 8747–8751.
2. Yan, Y. & Erickson, B. W. (1994) *Protein Sci.* **3**, 1069–1073.
3. Gutte, B., Daumigen, M. & Wittschieber, E. (1979) *Nature (London)* **281**, 650–655.
4. Drexler, K. E. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 5275–5278.
5. Regan, L. & DeGrado, W. F. (1988) *Science* **241**, 976–978.
6. Sasaki, T. & Kaiser, E. T. (1988) *J. Am. Chem. Soc.* **111**, 380–381.
7. Hahn, K. W., Klis, W. A. & Stewart, J. M. (1990) *Science* **248**, 1544–1547.
8. Hecht, M. H., Richardson, J. S., Richardson, D. C. & Ogden, R. C. (1990) *Science* **249**, 884–891.
9. Ghadiri, M. R., Soares, C. & Choi, C. (1992) *J. Am. Chem. Soc.* **114**, 4000–4002.
10. Mutter, M., Tuchscherer, G. G., Miller, C., Altmann, K.-H., Carey, R. I., Wyss, D. F., Labhardt, A. M. & Rivier, J. E. (1992) *J. Am. Chem. Soc.* **114**, 1463–1470.
11. Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M. & Hecht, M. H. (1993) *Science* **262**, 1680–1685.
12. Brown, J. E. & Klee, W. A. (1971) *Biochemistry* **10**, 470–476.
13. Kim, P. S. & Baldwin, R. L. (1984) *Nature (London)* **307**, 329–334.
14. Pauling, L., Corey, R. B. & Branson, H. R. (1951) *Proc. Natl. Acad. Sci. USA* **37**, 205–211.
15. Pauling, L. & Corey, R. B. (1951) *Proc. Natl. Acad. Sci. USA* **37**, 729–740.
16. Osterman, D. G. & Kaiser, E. T. (1985) *J. Cell. Biochem.* **29**, 57–72.
17. Richardson, J. S. & Richardson, D. C. (1987) in *Protein Engineering*, eds. Oxander, D. L. & Fox, C. F. (Liss, New York), pp. 149–163.
18. Unson, C. G., Erickson, B. W., Richardson, D. C. & Richardson, J. S. (1984) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **4**, 1837 (abstr.).
19. Mutter, M., Altmann, E., Altmann, K.-H., Hersperger, R., Koziej, P., Nebel, K., Tuchscherer, G., Vuilleumier, S., Gremllich, H.-U. & Müller, K. (1988) *Helv. Chim. Acta* **71**, 835–847.
20. Pessi, A., Bianchi, E., Crameri, A., Venturini, S., Tramontano, A. & Sollazzo, M. (1993) *Nature (London)* **362**, 367–369.
21. Richardson, J. S., Richardson, D. C., Tweedy, N. B., Gernert, K. M., Quinn, T. P., Hecht, M. H., Erickson, B. W., Yan, Y., McClain, R. D., Donlan, M. E. & Surles, M. C. (1992) *Biophys. J.* **63**, 1186–1209.
22. Betz, S. F., Raleigh, D. P. & Degrado, W. F. (1993) *Curr. Opin. Struct. Biol.* **3**, 601–610.
23. Raleigh, D. P. & Degrado, W. F. (1992) *J. Am. Chem. Soc.* **114**, 10079–10081.
24. Handel, T. M., Williams, S. A. & DeGrado, W. F. (1993) *Science* **261**, 879–885.