PROTEIN DESIGN

Proteins from Scratch

William F. DeGrado

Not long ago, it seemed inconceivable that proteins could be designed from scratch. Because each protein sequence has an astronomical number of potential conformations, it appeared that only an experimentalist with the evolutionary life span of Mother Nature could design a sequence capable of folding into a single, well-defined three-dimensional structure. But now, on page 82 of this issue, Dahiyat and Mayo (1) describe a new approach that makes de novo protein design as easy as running a computer program. Well almost. . .

The intellectual roots of this new work go back to the early 1980s when protein engineers first thought about designing proteins (2). At that point, the prediction of a protein's three-dimensional structure from its sequence alone seemed a difficult proposition. However, they opined that the inverse problem—designing an amino acid sequence capable of assuming a desired three-dimensional structure—would be a more tractable problem, because one could "over-engineer" the system to favor the desired folding pattern. Thus, the problem of de novo protein design reduced to two steps: selecting a desired tertiary structure and finding a sequence that would stabilize this fold. Dahiyat and Mayo have now mastered the second step with spectacular success. They have distilled the rules, insights, and paradigms gleaned from two decades of experiments (3) into a single computational algorithm that predicts an optimal sequence for a given fold. Further, when put to the test the algorithm actually predicted a sequence that folded into the desired three-dimensional structure. Thus, the rules of protein folding and computational methods for de novo design may now be sufficiently defined to allow the engineering of a variety of proteins.

Dahiyat and Mayo's program divides the interactions that stabilize protein structures

The author is in the Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104-6059, USA. E-mail: wdegrado@mail.med.upenn.edu

into three categories: interactions of side chains that are exposed to solvent, of side chains buried in the protein interior, and of parts of the protein that occupy more interfacial positions. Exposed residues contribute to stability, primarily through conformational preferences and weakly attractive, solvent-exposed polar interactions (4). The burial of hydrophobic residues in the well-packed in-
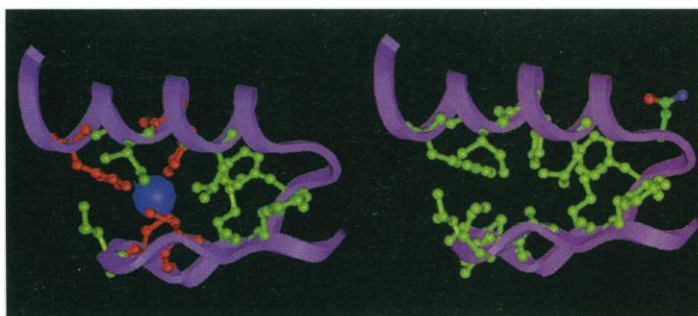
Better than the real thing. The natural zinc finger protein Zif268 (**left**) is stabilized in part by a core of hydrophobic (green) side chains and metal-chelating side chains (red). In the designed protein FSD-1 (**right**), the Zif268 core is retained but the metal-chelating His residues and one of the Cys residues of Zif268 are converted to hydrophobic Phe and Ala residues, thereby extending the hydrophobic core. The fourth metal ligand Cys[8] is converted to a Lys residue. The apolar portion of this interfacial residue shields the hydrophobic core, whereas its ammonium group is exposed to solvent. The helix is also stabilized by an N-capping interaction (19), which presumably also stabilizes the structure.

terior of a protein provides an even more powerful driving force for folding. The side chains in the interior of a protein adopt unique conformations, the prediction of which is a large combinatorial problem.

One important simplifying assumption arose from the early work of Jainin et al. (5), who showed that each individual side chain can adopt a limited number of low-energy conformations (named rotamers), reducing the number of probable conformers available to a protein. This work was subsequently extended to the design of proteins containing only the most favorable rotamers (6). Although the side chains in natural proteins deviate from ideality in a few cases (complicating the prediction of the structures of natural proteins), these deviations need not be considered in the design of idealized proteins. Thus, various algorithms have been developed to examine all possible hydrophobic residues in all possible rotameric states, to find combinations that efficiently fill the interior of a protein. A complementary ap-

proach uses genetic methods to exhaustively search for sequences capable of filling a protein core (7), and this work has been adapted for the de novo design of proteins (8).

Interfacial residues are also quite important for protein stability (9, 10). They are often amphiphilic (for example, Lys, Arg, and Tyr) and their apolar atoms can cap the hydrophobic core, while their polar groups engage in electrostatic and hydrogen-bonded interactions.

Until recently, protein designers have frequently concentrated on quantifying the energetics associated with just one of these three types of interactions (3). However, de novo design is best approached by simultaneously considering all of the side chains in the protein—unfortunately, a very high-order combinatorial problem. For instance, the volume available to the interior side chains depends on the nature and conformation of the residues at the interfacial positions and vice versa. Dahiyat and Mayo assumed that each of these three features had been adequately quantitated to provide a useful empirical energy function for protein design. Their program combines a number of feaures taken from earlier potential functions and includes a penalty for exposing hydrophobic groups to solvent. Another essential innovation included in their program is an implementation of the Dead-End Elimination theorem, to efficiently search through sequence and side chain rotamer space.

Dahiyat and Mayo's target fold is a zinc finger, a motif with a well-established history in protein structure prediction and design. In an early, prescient paper, Berg correctly inferred that this His^{2}Cys^2 Zn-binding motif must feature a β-β-α fold that would position the ligating groups in a tetrahedral array around the bound Zn(II) (11). Favorable metal ion–ligand interactions together with a small apolar core help stabilize the three-dimensional structure of this compact fold. More recently, Imperiali and co-workers have designed a peptide that folded into this motif, even in the absence of metal ions (12). The design included a D-amino acid to stabilize a type II' turn, and a large, rigid tricyclic side chain that may help consolidate the hydrophobic core. This work was particularly ex-

citing because, before their studies, it was not expected that sequences as short as 25 residues in length could fold into stable tertiary structures.

Now, Dahiyat and Mayo take these studies one step further through the design of a sequence composed of only natural amino acids that adopts the zinc finger motif. As input to their program, they introduced the coordinates of the backbone atoms from the crystal structure of the second domain of the zinc finger protein Zif268. The program then evaluated a total of 10^{62} possible side chain-rotamer combinations to find a sequence capable of stabilizing this fold without a bound metal ion. The resulting protein sequence shares a small hydrophobic core with its predecessor from Zif268. However, in the newly designed protein FSD-1 the core is enlarged through the addition of hydrophobic residues that fill the space vacated by the removal of the metal-binding site (see the figure). This increase in the size of the hydrophobic core together with the enhancements in the propensity for forming the appropriate secondary structure provide an adequate driving force for folding. The designed miniprotein actually folds into the desired structure as assessed by nuclear magnetic resonance spectroscopy, and the observed structure closely resembles the three-dimensional structure of Zif268.

Because of its small size, the protein is marginally stable. A Van't Hoff analysis of the thermal unfolding curve gives a change in the enthalpy (ΔH_{vH}) of approximately −10 kcal/mol, and indicates that the protein is about 90 to 95% folded at low temperatures (13). The small value ΔH_{vH} and the lack of strong cooperativity in the unfolding transition are expected for a native-like protein of this very small size (14). Thus, FSD-1 is the smallest protein known to be capable of folding into a unique structure without the thermodynamic assistance of disulfides, metal ions, or other subunits. This important accomplishment illustrates the impressive ability of Dahiyat and Mayo's program to design highly optimized sequences.

This new achievement caps a banner year for de novo protein design. Earlier, Regan (15) answered the challenge of changing a protein's tertiary structure by altering no more than 50% of its sequence. And although Dahiyat and Mayo have demonstrated that the stabilizing metal-binding site is not necessary in their system, Caradonna, Hellinga, and co-workers (16) have made impressive progress in automating the introduction of functional metal-binding sites into the three-dimensional structures of natural proteins. Further, other workers (17) have used less automated approaches to successfully introduce functionally and spectroscopically interesting metal-binding sites into de novo designed proteins.

To date, the most computationally intensive protein design problems have been the redesign of natural proteins of known three-dimensional structure. But the new automated approaches open the door to the de novo design of structures with entirely novel backbone conformations. It will be interesting to see if Dahiyat and Mayo's approach of designing an optimal sequence for a given fold is sufficient, or if it will be necessary also to destabilize alternate possible folds. Indeed, when using an earlier version of their algorithm to repack the interior of the coiled coil from GCN4, they had to retain the identity of a buried Asn residue from the wild-type protein. Although the inclusion of this Asn actually destabilized the desired fold, it was nevertheless essential to avoid the formation of alternate, unwanted conformers (18). The ability to ask such focused questions will reveal much about how natural proteins adopt their folded conformations while simultaneously allowing the design of entirely new polymers for applications ranging from catalysis to pharmaceuticals.

References and Notes

1. B. I. Dahiyat and S. L. Mayo, *Science* **278**, 82.
2. K. E. Drexler, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 5275 (1981); C. Pabo, *Nature* **301**, 200 (1983).
3. W. F. DeGrado, Z. R. Wasserman, J. D. Lear, *Science* **243**, 622 (1989); J. W. Bryson *et al.*, *ibid.* **270**, 935 (1995); M. H. J. Cordes, A. R. Davidson, R. T. Sauer, *Curr. Opin. Struct. Biol.* **6**, 3 (1996).
4. R. Munoz and L. Serrano, *Proteins* **20**, 301 (1994); C. A. Kim and J. M. Berg, *Nature* **362**, 267 (1993); D. L. Minor and P. S. Kim, *ibid.* **367**, 660 (1994); C. K. Smith, J. M. Withka, L. Regan, *Biochemistry* **33**, 5510 (1994).
5. J. Janin, S. Wodak, M. Levitt, B. Maigret, *J. Mol. Biol.* **125**, 37 (1978).
6. J. W. Ponder and F. M. Richards, *ibid.* **193**, 775 (1987); J. R. Desjarlais and T. M. Handel, *Protein Sci.* **4**, 2006 (1995); X. Jing, E. J. Bishop, R. S. Farid, *J. Am. Chem. Soc.* **119**, 838 (1997).
7. J. U. Bowie, J. F. Reidhaar-Olson, W. A. Lim, R. T. Sauer, *Science* **247**, 1306 (1990).
8. S. Kamtekar, J. M. Schiffer, H. Xiong, J. M. Babik, M. H. Hecht, *ibid.* **262**, 1680 (1993).
9. K. J. Lumb and P. S. Kim, *ibid.* **271**, 1137 (1996); Y. Yu, O. D. Monera, R. S. Hodges, P. L. Privalov, *J. Mol. Biol.* **255**, 367, (1996).
10. A. C. Braisted and J. A. Wells, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5688 (1996).
11. J. M. Berg, *ibid.* **85**, 99 (1988).
12. M. D. Struthers, R. P. Cheng, B. Imperiali, *Science* **271**, 342 (1996).
13. This Van't Hoff analysis of the protein is approximate because of the lack of definition of the pre- and posttransition baselines.
14. P. Alexander, S. Fahnestock, T. Lee, J. Orban, P. Bryn, *Biochemistry* **31**, 3597 (1992).
15. S. Dalal, S. Balasubramanian, L. Regan, *Nat. Struct. Biol.* **4**, 548 (1997).
16. A Pinto, H. W. Hellinga, J. P. Caradonna, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5562 (1997); C. Coldren, H. W. Hellinga, J. P. Caradonna, *ibid.*, p. 6635.
17. B. R. Gibney, S. E. Mulholland, F. Rabanal, P. L. Dutton, *ibid.* **93**, 15041 (1996); M. P. Scott, J. Biggins, *Protein Sci.* **6**, 340 (1997); P. A. Arnold, W. R. Shelton, D. R. Benson, *J. Am. Chem. Soc.* **119**, 3181 (1997); G. R. Dieckman *et al.*, *ibid.*, p. 6195.
18. P. B. Harbury, T. Zhang, P. S. Kim, T. Alber, *Science* **262**, 1401 (1993); K. J. Lumb and P. S. Kim, *Biochemistry* **34**, 8642 (1995).
19. L. G. Presta and G. D. Rose, *Science* **240**, 1632 (1988); J. S. Richardson and D. C. Richardson, *ibid.*, p. 1648.

Mathematical broth
http://math.wisc.edu/~griffeat/welcome.html

The Primordial Soup Kitchen is a site devoted to research on cellular automata (CA), collections of mathematical objects that individually are governed by simple rules but are highly complex in the aggregate. Depending on the rules, automata can evolve into ordered states from disordered beginnings. The Primordial Soup Kitchen relies on a culinary metaphor: the menu includes CAffeine (Java applications for cellular automata), Lagniappe (bite-sized descriptions of recent automata research by various researchers), and the Kitchen Sink (Web links to other resources). The chef, D. Griffeat of the University of Wisconsin, has prepared a thoroughly nutritious Web page for CA studies.

An educational voyage
http://www.reef.edu.au

The Reef Education Network (REN) Web page, developed by the University of Sydney, is a basic guide to one of the most complex ecosystems on Earth. The REN is aimed at an audience of layperson s and educators and contains a great deal of sophisticated information about reef communities. "Ask a Brain Coral" offers answers to questions about reef biology, and the Research Hut describes the people and laboratories involved in reef research. Users of the page can create their own online notebooks with links to items in the Web site.

One-stop site shop
http://www.scicentral.com

SciCentral is structured as a meta-directory of science and engineering on the Web. The site provides many links to resources in all fields of science, including daily necessities such as e-mail and phone directories and lists of college and university Web pages. Other tools are included, too, including travel guides for the frequent-flying researcher, links to media coverage of science, and grant information on the Web.

Edited by David Voss

Readers are invited to suggest excellent scientific Web sites by e-mail to editors@aaas.org