Protein Engineering vol 2 no.4 pp.249–250, 1988

Commentary

Passing the first milestone in protein design

When the wave of recombinant DNA technology led to much speculation about the future design of proteins with novel functions, pessimists pointed out that we still could not predict the folding of a protein from its amino acid sequence. Despite the pessimists, a few groups set out to design and synthesize a protein *de novo*.

A recent publication by Regan and DeGrado (1988) announced the passing of the first milestone in the protein design journey. The milestone can be defined as the *de novo* design and synthesis of a polypeptide chain of significant length, which does not contain non-natural components, and which folds into a stable structure in aqueous solution. The protein designed by Regan and DeGrado (1988) is based on a true *de novo* design, is 74 residues in length (i.e. comparable to small single-domain proteins found in nature), is composed of only natural amino acids (in fact, it was produced in *E.coli* by expression of a synthetic gene), and there is unequivocal evidence that the protein folds as a monomer in physiological conditions.

The protein was purified to apparent homogeneity from lysed cells, the first 25 residues were sequenced using Edman degradation, and amino acid analysis confirmed the expected composition. The evidence for folding is based on (i) a circular dichroism (CD) spectrum that indicates significant α-helix content (α-helices can be inferred from CD spectra with much more confidence than any other secondary structure) and (ii) loss of the CD signal at 222 nm (i.e. a helical band) in a cooperative manner on addition of GuHCl. A remarkable aspect of the GuHCl denaturation is that the protein is extremely stable: the midpoint of the denaturation curve occurs at 6.5 M GuHCl.

It is important to note that DeGrado and his co-workers did not travel the first mile in one giant leap, but rather they took a systematic, stepwise approach to the problem. The goal was to design a 4-helix bundle protein, which is a common folding motif found in nature (Weber and Salemme, 1980).

In the first step (Eisenberg *et al.*, 1986), a 16-residue peptide was designed to form a tetrameric α-helical bundle. The design of the bundle was simplified by using identical sequences for each of the four helices. Basically, a sequence was sought which, upon application of a 222 symmetry operator, gave a structure with the characteristics of natural 4-helix bundles (close packing of hydrophobic sidechains in the interior, helix crossing angles of $\sim 18°$ and polar residues exposed on the surface). Such a sequence was designed using leucine as the only hydrophobic residue and lysine or glutamic acid as polar residues. A glycine residue (i.e. a helix breaker) was added at each end of the peptide in anticipation of adding a hairpin loop between the helices in future work. CD spectra of the peptide indicated that it folds into a helical conformation, and gel filtration experiments showed that the peptide associates into a tetrameric structure (Ho and DeGrado, 1987).

In the second step (Ho and DeGrado, 1987) a loop was designed to connect two helices in an anti-parallel orientation. In addition, a few changes in the amino acid sequence were made based on the observation that charges at the ends of the helix affect helix stability (Shoemaker *et al.*, 1987), presumably through interactions with the α-helix macro-dipole. At first, a single proline residue was placed between the two glycines at each end of the helices to give a Gly-Pro-Gly linker. When this peptide was studied, however, it was found by gel filtration to form a trimer (i.e. containing a total of six original units) instead of the desired dimer or dimers (Ho and DeGrado, 1987). A lesson from this result is that natural proteins have evolved to avoid forming wrong structures, in addition to forming correct ones (DeGrado *et al.*, 1987). The desired dimer was obtained (DeGrado *et al.*, 1987; Ho and DeGrado, 1987) using a linker of Gly-Pro-Arg-Arg-Gly, which presumably caused electrostatic destabilization of the trimeric structure.

The third step, reported in the publication by Regan and DeGrado (1988), involved adding a third linker between the two dimers. The sequence Pro-Arg-Arg was used again, and this time the polypeptide was produced in *E.coli* with a synthetic gene which contains unique restriction sites for future use. Based on the helical CD spectrum, the monomeric nature of the folded protein and the results of previous steps described above, a working assumption (Regan and DeGrado, 1988) made for future work is that the protein contains four closely packed α-helical segments. Structural studies are now needed to test this assumption and to establish the packing geometry of the helices.

What have we learned from this work? First, the skeptics will now have to admit that it is possible to design proteins *de novo*, even though we do not yet understand protein folding. The designed protein is admittedly simple, but it is also much more stable than most natural proteins, which suggests that there is considerable room for error in future work. Second, the advantages of taking a stepwise approach to the problem are now obvious, and they are part of the elegance of the work. Starting with principles learned from the design of amphiphilic peptide hormone analogues (Kaiser and Kezdy, 1984) and from studies of model coiled-coils based on the tropomyosin heptad repeat (Lau *et al.*, 1984), the first step taken by DeGrado and co-workers was relatively straightforward. Each subsequent step was kept small; nevertheless, mistakes were made. The stepwise approach allowed these mistakes to be detected and corrected before moving on. Moreover, one mistake provided an appreciation for the ability of natural proteins to avoid folding into alternative conformations. Third, these results emphasize the conclusion, based on studies of mutant proteins (review; Alber, 1988), that proteins have not evolved for maximal stability. The folded form of the synthetic protein is ~ 22 kcal/mole more stable than the unfolded form (Regan and DeGrado, 1988). Stabilities of this magnitude are virtually unheard of with natural proteins.

The extraordinary stability of the designed protein also emphasizes the large driving force provided by the hydrophobic effect in protein folding. The fact that natural proteins have not taken full advantage of the hydrophobic effect for stability probably reflects selection for other functions: avoiding wrong structures in folding, a need for protein turnover *in vivo*, and perhaps most important, the need for flexibility to optimize function. It will be interesting to see how far one can get with function (e.g. catalysis) in designed proteins, without needing to give up something in terms of overall stability.

As far as function is concerned, the protein designed by DeGrado and co-workers can now be used as a template for adding functional residues. The 4-helix bundle is particularly attractive because divergence of the helices at the end of the

bundle is known to create functionally important cavities in natural proteins. Designed proteins that bind ligands and/or carry out chemical reactions do not seem too far away.

The field of protein design is here to stay. The design of β-sheet-containing proteins that fold in aqueous solution appears to be a tough problem for the future, although a new version of betabellin (Richardson and Richardson, 1987) is much more soluble than older versions and may contain β-structure as judged by CD and laser Raman spectroscopy (J. Richardson, personal communication). In addition, synthetic templates upon which both α- and β-structures can be assembled have been reported (Mutter, 1988), and an artificial DTT-binding peptide designed to contain β-sheets was found to contain activity in aqueous/organic mixed cosolvents (Moser *et al.*, 1987). Finally, the exciting prospect of designing membrane proteins appears to be feasible. Toward this goal, DeGrado and co-workers have again taken a stepwise approach with α-helices: the results of the first step suggest that ion-specific membrane channel proteins can be designed *de novo* (Lear *et al.*, 1988).

References

Alber,T. (1988) In Fasman,G.D. (ed.), *Prediction of Protein Structure and the Principles of Protein Conformation*. Plenum, New York. In press.

DeGrado,W.F., Regan,L. and Ho,S.P. (1987) *Cold Spring Harbor Symp. Quant. Biol.*, **52**, 521–526.

Eisenberg,D., Wilcox,W., Eshita,S.M., Pryciak,P.M., Ho,S.P. and DeGrado,W.F. (1986) *Proteins*, **1**, 16–22.

Ho,S.P. and DeGrado,W.F. (1987) *J. Am. Chem. Soc.*, **109**, 6751–6758.

Kaiser,E.T. and Kezdy,F.J. (1984) *Science*, **223**, 249–255.

Lau,S.Y.M., Taneja,A.K. and Hodges,R.S. (1984) *J. Biol. Chem.*, **259**, 13253–13261.

Lear,J.D., Wasserman,Z.R. and DeGrado,W.F. (1988) *Science*, **240**, 1177–1181.

Moser,R., Frey,S., Munger,K., Hehlgans,T., Klauser,S., Langen,H., Winnacker,E., Metz,R. and Gutte,B. (1987) *Protein Eng.*, **1**, 339–343.

Mutter,M. (1988) In Marshall,G.R. (ed.), *Proceedings of the 10th American Peptide Symposium*. ESCOM, Leiden, pp. 349–353.

Shoemaker,K.R., Kim,P.S., York,E.J., Stewart,J.M. and Baldwin,R.L. (1987) *Nature*, **326**, 563–567.

Regan,L. and DeGrado,W.F. (1988) *Science*, **241**, 976–978.

Richardson,J.S. and Richardson,D.C. (1987) In Oxender,D.L. and Fox,F. (eds), *Protein Engineering*. Alan R. Liss, New York, pp. 149–163.

Weber,P.C. and Salemme,F.R. (1980) *Nature*, **287**, 82–84.

Peter S. Kim
Whitehead Institute for Biomedical Research
Nine Cambridge Center
Cambridge, MA 02142
and
Department of Biology
Massachusetts Institute of Technology
Cambridge, MA 02139, USA