Downloaded from symposium.cshlp.org on January 9, 2015 - Published by Cold Spring Harbor Laboratory Press

The Design of a Four-helix Bundle Protein

W.F. DeGrado, L. Regan, and S.P. Ho
E.I. du Pont de Nemours and Company, Central Research and Development Department, Wilmington, Delaware 19898

The design of proteins with predetermined structural properties is a necessary first step in the de novo design of novel enzymes and receptors. A major problem associated with the design of proteins is the high flexibility inherent in polypeptide chains. It has been estimated that a protein of 100 residues can adopt up to 10^{100} different conformations, a number that is as large as the number of atoms in the universe (Creighton 1984)! The mechanism by which a protein adopts a relatively well-defined set of conformations out of such a large number of possibilities is only beginning to be understood, and it is not possible to predict the three-dimensional structure of a protein from its amino acid sequence.

Recently, the protein folding problem has been rephrased in light of the ease with which one can now prepare proteins of virtually any amino acid sequence (Drexler 1981; Pabo 1983). If it is not possible to predict the three-dimensional structure of a protein from its sequence, might it not be possible to do just the inverse? Could one begin with a reasonable three-dimensional structure and then design a peptide sequence that would fold into this structure? We think that this should be possible if a design is chosen that is relatively simple and rich in secondary structural elements. The design of protein secondary structures has become fairly straightforward. It has long been known that certain amino acids have clear-cut preferences for adopting a given secondary structure (Chou and Fasman 1978), and more recent studies with synthetic peptides have shown the importance of hydrophobic periodicity (Eisenberg et al. 1984; DeGrado and Lear 1985) and electrostatic interactions (Shoemaker et al. 1987) in determining the stability of α helices and β sheets. Thus, it is now routinely possible to design peptides that incorporate single secondary structural units (Kaiser and Kezdy 1984; O'Neil et al. 1987). A logical next step would be to design secondary structures that can pack together to form a globular protein with a predetermined three-dimensional structure. In this paper, we will describe our progress along these lines toward the design of a synthetic four-helix bundle protein.

Our designed protein (Eisenberg et al. 1986; Ho and DeGrado 1987) is an idealized version (Fig. 1) of a class of proteins that includes myohemerythrin, apoferittin, tobacco mosaic virus coat protein, and cytochrome c' (Weber and Salemme 1980). These proteins contain a common four-helix bundle structural motif that is comprised of four helices connected by three loops. The helices in the bundle are nearly antiparallel to one another with a slight tilt of about 20°. As a result of this tilt, the helices diverge from a point of closest approach, forming a cavity that can ac-

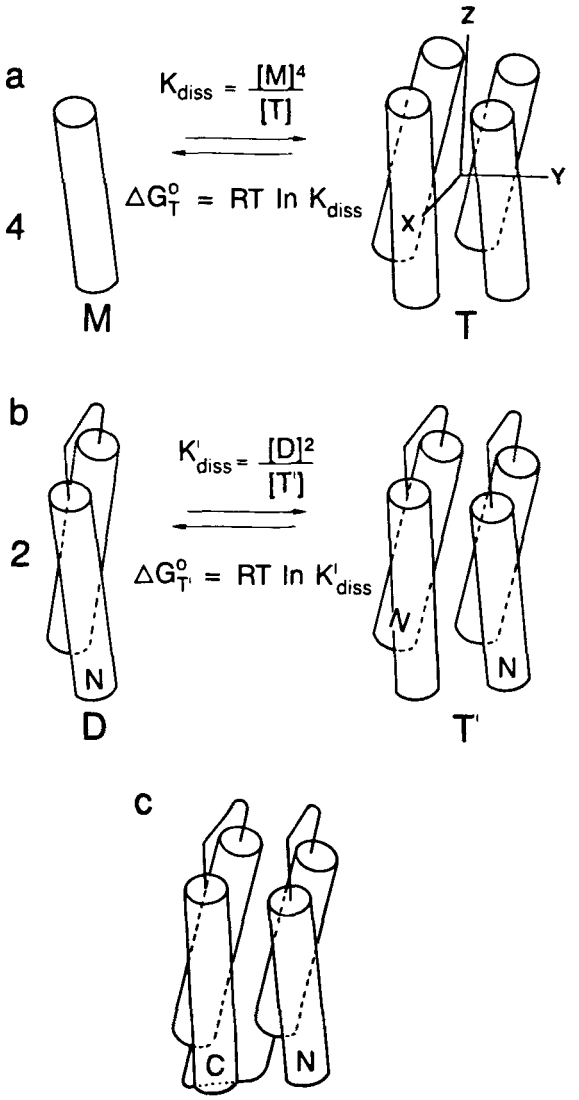

Figure 1. An incremental approach to the design of a four-helix bundle protein. (*a*) The sequence of an amphiphilic helix is designed with the aim of obtaining stable tetrameric aggregates in solution. (*b*) Once the sequence of the helix has been optimized, sequences for the loops are evaluated. Using the best helix sequence obtained in step *a*, various designed loop sequences are evaluated on the basis of the stability of the resulting dimer of helical hairpins. (*c*) The ultimate four-helix bundle is constructed from four optimized helices and three optimized loops. Dissociation constants and free energies are experimentally determined as a measure of the stability of the tetramers and dimers.

Downloaded from symposium.cshlp.org on January 9, 2015 - Published by Cold Spring Harbor Laboratory Press

commodate binuclear iron in myohemerythrin or a heme in cytochrome *c'*. The formation of cavities near the end of the bundles appears to be quite a general attribute of this folding pattern (Weber and Salemme 1980), making this motif an attractive target for understanding protein folding as well as for the eventual design of synthetic binding sites. In our approach to the design of a four-helix bundle, we decided to first make a fairly short bundle in which the helices were too short to diverge very far from one another. At a later date we intend to extend the helices from one side of the bundle to form a cavity that could be lined with appropriate side chains for binding and catalysis.

Chothia and co-workers (Chothia et al. 1977; Chothia 1984) have proposed a very simple hypothesis to account for the observed geometry of the helical packing in four-helix bundles as well as the stability of these structures. They observed a regularity in the packing of the side chains that comprise the solvent inaccessible core of four-helix bundles. Residues at positions i − 4, i, and i + 4 appear to form a ridge that packs against residues at positions k − 3, k, and k + 3 on a neighboring helix. This defines the packing angle between neighboring helices and allows tight packing of the apolar side chains, thereby driving the folding process. Electrostatic interactions between helical macrodipoles (Sheridan et al. 1982) are also thought to contribute to the stability of the structure. Finally, the loops between the helices probably serve to stabilize the structure.

In our design of a four-helix bundle protein (Eisenberg et al. 1986; Ho and DeGrado 1987), we idealized the pseudo 2,2,2 symmetry that is found in the structures of natural four-helix bundle proteins (Weber and Salemme 1980). This provided two distinct advantages over other approaches to protein design (Moser et al. 1983; B.W. Erickson et al., pers. comm.). First, the modeling was significantly simplified; rather than designing an entire protein, it was only necessary to design a single helix, which, upon application of a 2,2,2 symmetry operator, would adopt a structure with a tightly packed interior. Second, the design of the protein could be approached in the iterative, experimental manner described in Figure 1. The first step of this approach involved the evaluation of single helices that self-assemble into homotetramers. The free energy for the self-assembly process could be conveniently monitored from the monomer-to-tetramer equilibrium constant. Evaluation of several different designs allowed optimization of the helical sequence. In the next step of the project, loops were inserted between two identical helical sequences in an attempt to prepare a helical hairpin that would assemble into a four-helix bundle, as shown in Figure 1b. Finally, the entire sequence for the four-helix bundle was prepared by connecting four identical helices with three identical loops.

Methods and Experimental Procedures

The synthesis and characterization of peptides 1–4 were reported previously (Eisenberg et al. 1986; Ho and DeGrado 1987).

DNA oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer and on a Coder 300 (du Pont) DNA synthesizer. The oligonucleotides were purified by excision and elution from denaturing polyacrylamide gels (Maniatis et al. 1982) followed by C18 (Sep-Pac, Waters) chromatography. Restriction enzymes and DNA-modifying enzymes were purchased from New England Biolabs, Bethesda Research Laboratories, and Pharmacia. They were used according to the manufacturers' specifications. *Escherichia coli* protein extraction and Western blot analyses were performed as described previously (Burnette 1981; Regan et al. 1986).

Results

Design of the helical sequence. The design of the helical sequence was first accomplished using physical models (Eisenberg et al. 1986; Ho and DeGrado 1987). Models of four 16-residue helices were arranged so that their side chains could interact in a manner similar to that described by Chothia (1984) for a four-helix bundle. Leu side chains were placed at positions that project into the interior of the structure, and Glu and Lys side chains were placed at positions projecting toward the exterior of the structure. All three of these residues are known to strongly favor helix formation (Chou and Fasman 1978). In addition, glycine residues were placed at the amino and carboxyl termini of the structure to help break the helix and also to help induce the formation of a turn when the loops were added at a later stage. The resulting structure, peptide 1, was synthesized by the Merrifield solid phase method. A fragment comprising the 12 carboxy-terminal residues of peptide 1 was also isolated during the purification. This peptide (1a) was found to form tetramers in aqueous solution (Eisenberg et al. 1986). A subsequent analysis (Ho and DeGrado 1987) of a model of peptide 1 suggested that its sequence could be improved. In models, the Leu residue at position 11 appeared to be excessively exposed to solvent, and the Glu residue at position 13 was partially buried. Consequently, Leu was changed to Lys at position 11 and Glu was changed to Leu at position 13. In addition, it appeared likely that the Lys at position 2 of peptide 1 and the α-carboxylate at the carboxyl terminus might interact unfavorably with the helical macrodipole. Therefore, these were converted to a Glu and an α-carboxamide, respectively. The resulting peptide (2) should form a more stable tetramer if the model is correct.

The assembly of the above peptides into tetramers could be assessed by circular dichroism (CD) spectroscopy. In very dilute aqueous solution the peptides are monomers and show low helicity. In concentrated solutions, they form tetramers in which the peptides form helices that are stabilized by long-range interactions between the apolar side chains. These conformational changes are reflected in the CD spectra of the peptides. The concentration dependence of the spectra are extremely well described by a simple monomer-to-tetramer equilibrium (Fig. 2). Analysis of these curves

Downloaded from symposium.cshlp.org on January 9, 2015 - Published by Cold Spring Harbor Laboratory Press

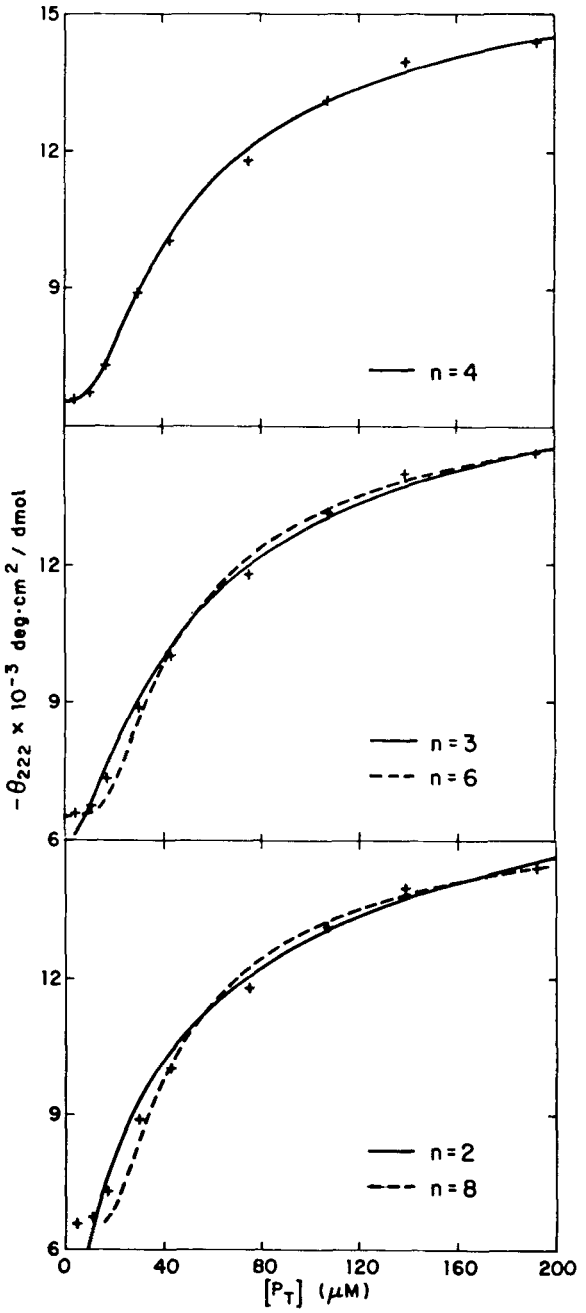

Figure 2. Concentration dependence of the ellipticity of α_1A at 222 nm. The lines are computer-generated, theoretical curves describing various monomer-n-mer equilibria. (*Top*) Monomer-tetramer equilibrium, (*middle*) monomer-trimer and monomer-hexamer equilibria, and (*bottom*) monomer-dimer and monomer-octamer equilibria. The data are optimally fit by a cooperative monomer-to-tetramer equilibrium.

provides three useful parameters: the stability of the tetramers (RT ln K_{diss}) and the approximate helical contents of the monomeric and tetrameric forms of the peptides (Table 1). Peptide 2 was found to form such stable tetramers that it was necessary to measure the monomer-to-tetramer equilibrium at various guanidine hydrochloride concentrations and then extrapolate the values of these parameters to zero guanidine concentration.

The helical content of the monomeric form depended markedly on the peptide chain length; the 12-residue peptide 1a was approximately 15% helical as a monomer, whereas the 16-residue peptides 1 and 2 were approximately 30% helical. The helical content for the tetrameric form of all three of these peptides was approximately 70%. The stabilities of the tetramers depended both on chain length and sequence and were in accord with the predictions based on computer modeling; peptide 1a was the least stable, followed by peptide 1, with peptide 2 being the most stable. The tetramer of peptide 2 was stabilized by approximately 20 kcal/mol. This large value shows that the sequence of peptide 2 is well designed for forming helical tetramers.

The design of a helical hairpin. We next attempted to design a hairpin loop that would connect two molecules of peptide 2 in a head-to-tail manner. Examination of models of tetramers of peptide 2 suggested that the ends could be joined by a single Pro if the helices were slightly unwound. Thus, peptide 3, which contains two copies of peptide 2 connected by a single Pro residue, was synthesized. Size exclusion chromatography of this peptide indicated that it was forming trimers rather than the desired dimer of helical hairpins. Operating under the assumption that the trimer formed by peptide 3 had a conformation similar to that illustrated in Figure 3, we proceeded to introduce substitutions that should destabilize the trimeric conformer with respect to the desired dimeric conformer. Two Arg residues were therefore inserted directly after the Pro with the expectation that this should electrostatically destabilize the trimer. Now, if a trimer were to form, the Arg residues from neighboring peptides would be buried together near the center of the trimer structure, thus resulting in unfavorable electrostatic interactions. Indeed, peptide 4 (Table 1) appeared to form dimers rather than trimers, as assessed by size exclusion chromatography. In addition, the concentration dependence of the CD spectra for this peptide indicated that peptide 4 was in a monomer-to-dimer equilibrium.

Peptide 4 has been substantially stabilized toward guanidine hydrochloride denaturation as compared to peptide 2. At approximately equal peptide concentrations, about twice as much guanidine hydrochloride is required to unfold peptide 4 as is required to unfold peptide 2 (Fig. 4). This success notwithstanding, a detailed analysis of the thermodynamic data for peptide 4 suggests that it might be possible to further optimize the loop sequence to provide an even more stable protein. The dimers of peptide 4 are stabilized with respect to their unfolded monomers by approximately 13 kcal/mol. Comparison of this number with that for the tetramers of peptide 2 indicates that the loop sequence could be substantially improved (Ho and DeGrado 1987).

Design of a gene encoding peptide 5. The gene for peptide 5 was constructed from eight synthetic DNA oligonucleotides. It was cloned into the vector pTM201/NS3-3 (K. Knight and R.T. Sauer, in prep.)

Downloaded from symposium.cshlp.org on January 9, 2015 - Published by Cold Spring Harbor Laboratory Press

Table 1. Amino Acid Sequences and Free Energies of Tetramerization or Dimerization of Synthetic Peptides

Peptide	Sequence	RT ln K_{diss} (kcal/mol)
1[1]	Ac-GluLeuLeuLysLysLeuLeuGluGluLeuLysGly-COOH	−11.4
2[1]	Ac-GlyLysLeuGluGluLeuLeuLysLysLeuLeuGluGluLeuLysGly-COOH	−19
3[1]	Ac-GlyGluLeuGluGluLeuLeuLysLysLeuLys GluLeuLeuLysGly-$CONH_2$	−22
4[2]	Ac-GlyGluLeuGluGluLeuLeuLysLysLeuLys GluLeuLeuLysGlyProArgArg GlyGluLeuGluGluLeuLeuLysLysLeuLys GluLeuLeuLysGly-$CONH_2$	−13

Data taken from Ho and DeGrado (1987).
[1]Monomer-tetramer equilibria.
[2]Monomer-dimer equilibrium.

at the *NcoI* and *NruI* sites, using standard techniques (Maniatis et al. 1982). This resulted in plasmid pα_4-1 in which peptide 5 is expressed from the inducible *tac* promoter (De Boer et al. 1983). (See Fig. 5.) The gene was designed such that unique restriction sites were introduced at conveniently spaced locations near the end of the helices. This should facilitate the introduction of future changes into peptide 5 by "cassette" mutagenesis (Richards 1986). Restriction analysis confirmed that all sites were indeed present in the synthetic gene.

Expression of peptide 5 in *E. coli* was detected using polyclonal antibodies generated against peptide 4. Western blot analysis (Burnette 1981) of protein extracts from strains containing the synthetic gene showed that peptide 5 was produced in *E. coli* and that its expression was inducible by the addition of isopropyl β-D-thiogalactopyranoside (Fig. 6). Further analysis revealed that the majority of peptide 5 is present in a soluble form within the bacteria. This facilitates protein purification, which is currently in progress.

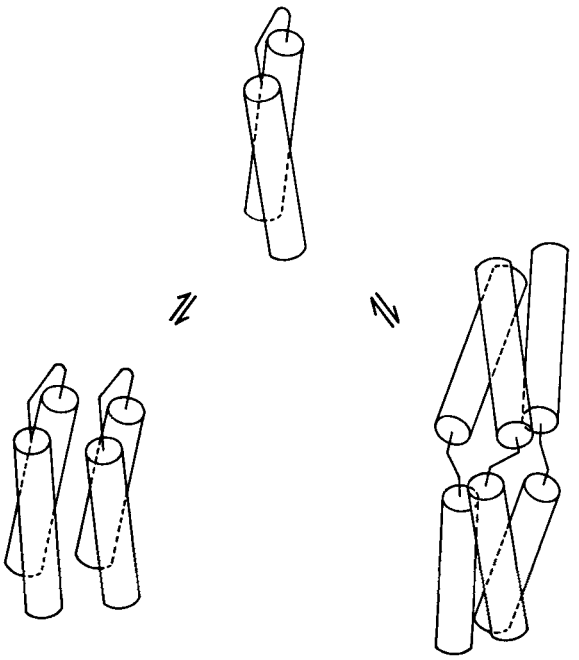

Figure 3. Diagrammatic representation of trimer formation instead of the desired dimer of helical hairpins.

Discussion

The folding of a protein is a complex process that includes helix formation, helix termination, helix packing, and loop formation. Few of these processes are very well understood, so that it seemed unlikely that a sequence that fulfilled each of these functions could be designed in a single step. Instead, it seemed more prudent to devise an incremental experimental approach, which would allow each aspect of folding to be dissected and optimized separately.

First, consider helix formation as evaluated by the helical contents of the monomeric forms of peptides 1, 1a, and 2. Helix formation by the 12-residue peptide, 1a, is quite unfavorable at room temperature, as it contains about 15% helix in dilute solution. Clearly, it is energetically unfavorable to fold this peptide into a helix, and this decreases the overall stability of the helical tetramers formed by this peptide. In contrast, peptides 1 and 2 show unusually high helical contents (approximately 30%) as monomers in water at room temperature. Few other peptides show this high a degree of monomolecular helix formation at room temperature (Shoemaker et al. 1987). Stabilizing interactions that might account for this high degree of helicity include salt bridges between Glu to Lys side chains (Eisenberg et al. 1986; Ho and DeGrado 1987) and partial dehydration of the apolar side chains in the helical conformation. Also, peptides 1 and 2 both have their amino termini blocked by an acetyl group, which

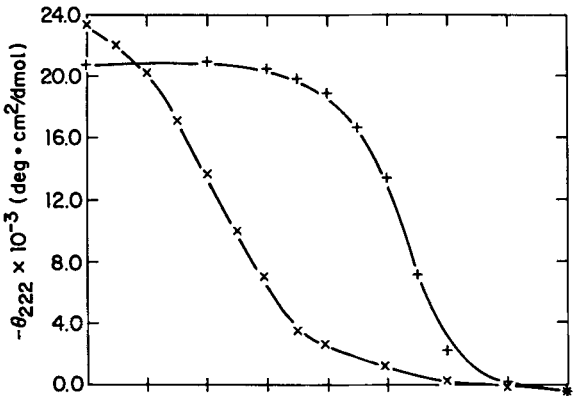

Figure 4. Guanidine denaturation curves for peptides 3 (×) and 4 (+).

Downloaded from symposium.cshlp.org on January 9, 2015 - Published by Cold Spring Harbor Laboratory Press

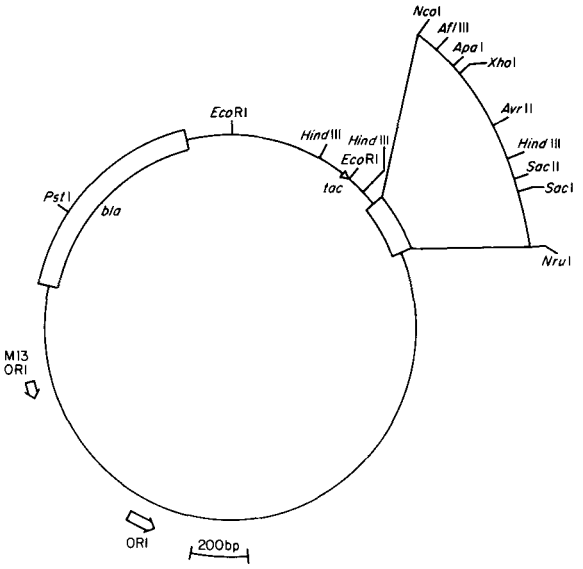

Figure 5. Restriction map of plasmid $p\alpha_4$-1.

stabilizes helix formation (Shoemaker et al. 1987). If one assumes that helix formation is an all-or-nothing process, then it can be calculated that helix formation by peptides 1 and 2 is only slightly unfavorable and on the order of a single kcal/mol. In any case, it appears that the sequences of peptides 1 and 2 have been well designed from the point of view of allowing helix formation to occur at a low energetic cost.

Peptide 2 forms more stable tetramers than peptide 1, which indicates that the packing interactions must be more favorable for peptide 2. (Both peptides have equal helicities as monomers, which suggests that helix formation is approximately energetically equivalent for both peptides.) This is in agreement with our computer modeling, which suggests that peptide 2 should form a more stable tetramer than peptide 1. The observed free energy of tetramerization of peptide 2 is consistent with dehydration of the apolar side chains being the driving force for folding. The observed free energy corresponds to −0.8 to −0.9 kcal/(mol of Leu side chains) in reasonable agreement with the value of −1.2 estimated for the transfer of a Leu side chain from water to the interior of a protein (Guy 1985).

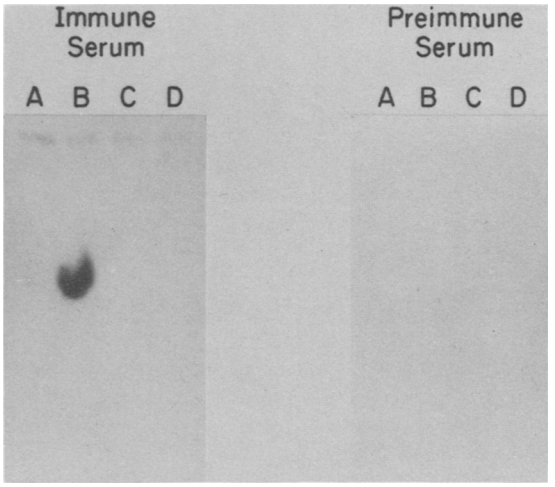

Figure 6. Western blot analysis of *E. coli* protein extracts. (Lane *A*) Extract from strain X-90 with plasmid $p\alpha_4$-1 grown in uninduced conditions. (Lane *B*) Extract from strain X-90 with plasmid $p\alpha_4$-1 grown with 5 mM isopropyl-β-D-thio-galactopyranoside. (Lane *C*) Extract from strain X-90 with the vector pTM201/NS3-3 grown in uninduced conditions. (Lane *D*) Extract from strain X-90 with the vector pTM201/NS3-3 in induced conditions. Left panel shows the filter probed with antipeptide 4 immune serum. Right panel shows the filter probed with preimmune serum.

Helix formation and loop formation appeared to be a problem in the design of peptide 3. Based on hydrodynamic measurements, it seemed likely that this peptide was forming trimers of extended rods rather than dimers of folded helical hairpins. The ease with which this problem was isolated and remedied clearly demonstrates the value of approaching the design of proteins in an iterative, experimental manner. It also illustrates an important point that must be considered in the design of new proteins. *Natural proteins have evolved not only to stabilize a given, desired fold, but also to destabilize all other possible alternatives.* A would-be designer of proteins must do no less if he or she is to succeed. One must try to consider all alternatives to the desired, designed fold and selectively destabilize the alternative folding pathways.

The hydrodynamic and thermodynamic data described herein provide excellent evidence to suggest that we have succeeded in designing proteins that are composed of four α helices. However, we cannot presently conclude that the helices are arranged into a four-helix bundle as illustrated in Figure 1. An elucidation of the helical packing geometry must await the solution of the crystal structures of peptides 1–5. Crystals of peptide 1a have been grown and diffract to high resolution (Eisenberg et al. 1986).

In the near future, we hope to purify and examine the properties of the single-chain four-helix peptide 5. If it can be shown to form a four-helix bundle, we will next try to convert it to a ligand-binding protein. By extending the helices one to two turns it should be possible to create a cavity capable of binding small, apolar compounds. The loops between the helices provide a natural location for adding catalytic groupings, including metal-binding sites. The locations of the restriction sites in the gene encoding peptide 5 should facilitate these manipulations.

ACKNOWLEDGMENTS

We thank Professor David Eisenberg for his collaboration and for many helpful discussions. We thank K. Knight and R. Sauer for a gift of plasmid pTM201/NS3-3. We thank Dan Oprian and Zelda Wasserman for helpful advice on various aspects of this work. We also thank K. Smithyman for his technical assistance throughout this work, and Carol Farber for preparing the manuscript.

Downloaded from symposium.cshlp.org on January 9, 2015 - Published by Cold Spring Harbor Laboratory Press

REFERENCES

Burnette, W.W. 1981. Western blotting: Electrophoretic transfer of proteins from sodium dodecyl sulphate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Anal. Biochem.* **112:** 195.

Chothia, C. 1984. Principles that determine the structure of proteins. *Annu. Rev. Biochem.* **53:** 537.

Chothia, C., M. Levitt, and D. Richardson. 1977. Structure of proteins: Packing of α-helices and β-pleated sheets. *Proc. Natl. Acad. Sci.* **74:** 4130.

Chou, P.Y. and G.D. Fasman. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol.* **47:** 45.

Creighton, T.E. 1984. *Proteins,* p. 161. W.H. Freeman, New York.

DeBoer, H.A., L.J. Comstock, and M. Vasser. 1983. The *tac* promoter: A functional hybrid derived from the *trp* and *lac* promoters. *Proc. Natl. Acad. Sci.* **80:** 21.

DeGrado, W.F. and J.D. Lear. 1985. Induction of peptide conformation at apolar/water interfaces: A study with peptides of defined hydrophobic periodicity. *J. Am. Chem. Soc.* **107:** 7684.

Drexler, K.E. 1981. Molecular engineering: An approach to the development of general capabilities for molecular manipulation. *Proc. Natl. Acad. Sci.* **78:** 5275.

Eisenberg, D., R.M. Weiss, and T.C. Terwilliger. 1984. The hydrophobic moment detects periodicity in protein hydrophobicity. *Proc. Natl. Acad. Sci.* **81:** 140.

Eisenberg, D., W. Wilcox, S.M. Eshita, P.M. Pryciak, S.P. Ho, and W.F. DeGrado. 1986. Design, synthesis and crystallization of a helical peptide. *Proteins Struct. Funct. Genet.* **1:** 16.

Guy, H.R. 1985. Amino acid side-chain partition energies and distribution of residues in solution. *Biophys. J.* **47:** 61.

Ho, S.P. and W.F. DeGrado. 1987. Design of a four-helix bundle: Synthesis of peptides which self-associate into a helical protein. *J. Am. Chem. Soc.* (in press).

Kaiser, E.T. and F.J. Kézdy. 1984. Amphiphilic secondary structure: Design of peptide hormones. *Science* **24:** 639.

Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Moser, R., R.M. Thomas, and B. Gutte. 1983. An artifical crystalline DDT-binding polypeptide. *FEBS Lett.* **157:** 247.

O'Neil, K.T., H.R. Wolfe, Jr., S. Erickson-Viitanen, and W.F. DeGrado. 1987. Alpha-helical periodicity reflected in the fluorescence properties of calmodulin-binding peptides. *Science* **236:** 1454.

Pabo, C.O. 1983. Molecular technology: Designing peptides and proteins. *Nature* **301:** 200.

Regan, L., J.D. Dignam, and P. Schimmel. 1986. A bacterial and silkworm enzyme share a common epitope which maps to the catalytic domain of each. *J. Biol. Chem.* **261:** 5241.

Richards, J.H. 1986. Cassette mutagenesis shows its strength. *Nature* **323:** 187.

Sheridan, R.P., R.M. Levy, and F.R. Salemme. 1982. α-Helix dipole model and electrostatic stabilization of 4-α-helical proteins. *Proc. Natl. Acad. Sci.* **79:** 4545.

Shoemaker, K.R., P.S. Kim, E.V. York, J.M. Stewart, and R.L. Baldwin. 1987. Tests for the helix dipole model for stabilization of α-helices. *Nature* **326:** 563.

Weber, P.C. and F.R. Salemme. 1980. Structural and functional diversity in 4-α-helical proteins. *Nature* **287:** 82.

Downloaded from symposium.cshlp.org on January 9, 2015 - Published by Cold Spring Harbor Laboratory Press

Cold Spring Harbor Symposia on Quantitative Biology

The Design of a Four-helix Bundle Protein

W.F. DeGrado, L. Regan and S.P. Ho

Cold Spring Harb Symp Quant Biol 1987 52: 521-526
Access the most recent version at doi:10.1101/SQB.1987.052.01.059

References This article cites 19 articles, 7 of which can be accessed free at:
http://symposium.cshlp.org/content/52/521.refs.html

Article cited in:
http://symposium.cshlp.org/content/52/521#related-urls

Email alerting service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here**

To subscribe to *Cold Spring Harbor Symposia on Quantitative Biology* go to:
http://symposium.cshlp.org/subscriptions

Copyright © 1987 Cold Spring Harbor Laboratory Press