Protein Design, a Minimalist Approach

WILLIAM F. DEGRADO, ZELDA R. WASSERMAN, JAMES D. LEAR

The question of how the amino acid sequence of a protein specifies its three-dimensional structure remains to be answered. Proteins are so large and complex that it is difficult to discern the features in their sequences that contribute to their structural stability and function. One approach to this problem is de novo design of model proteins, much simpler than their natural counterparts, yet containing sufficient information in their sequences to specify a given function (for example, folding in aqueous solution, folding in membranes, or formation of ion channels). Designed proteins provide simple model systems for understanding protein structure and function.

UNDERSTANDING THE RELATIONS BETWEEN AMINO ACID sequence and protein structure and protein function is a daunting prospect because of the complexity of protein structures. Given the 20 commonly occurring amino acids, there are 20^{100} possible permutations for a 100-residue protein. Furthermore, each protein of this length could potentially have about 10^{100} energetically reasonable conformations (*1*). Nevertheless, under the proper solution conditions, natural proteins adopt virtually unique, sequence-determined conformations. How this comes about is only beginning to be unraveled. Proteins with fewer than 10% of the same residues at comparable positions in the protein chain can have remarkably similar structures (*2*). This degeneracy of the sequence-structure "code" suggests that the forces determining protein structure are fundamental ones in that they remain constant within large variations of amino acid sequence. Although the specification and quantitation of these forces have not yet progressed to the ultimate goal of allowing prediction of the three-dimensional structures of natural proteins from their amino acid sequences, we feel that enough is known to begin the de novo design of model proteins.

Design and synthesis of model compounds has long been a useful technique for the study of small molecules. This classical approach can now be applied to large molecules such as proteins (*3*). For example, the use of site-directed mutagenesis (*4*) to create variants of proteins is now a standard technique for determining which residues in a protein are essential for folding or function. De novo protein design is a complementary approach; it starts with elementary principles and attempts to design model proteins from scratch (*5*). The fundamental advantage of this approach is that it critically tests our understanding of protein structure. The success of the designs depends on how correctly the underlying principles have been fathomed. This approach also allows the construction of model systems that should be easier than natural proteins to study and understand. Finally, this approach should ultimately allow the construction of novel proteins and polymers whose structures and properties are unprecedented in nature.

De novo design (*5*) has been successfully applied to the study of peptides (*6, 7*). Peptides differ from proteins in their size as well as their conformational properties. Linear peptides (less than about 50 residues) consisting of the 20 commonly occurring amino acids generally populate large ensembles of conformational states in water and adopt well-defined conformations only when bound to metal ions (*8*), surfaces (*7*), and receptors (*6*). Thus a major goal of peptide design is to decrease the flexibility of peptides through the introduction of conformational constraints such as macrocycles (*6, 9*). For peptides larger than about 15 residues, this approach is less feasible because of the large number of potential conformations that need to be considered, but the approach pioneered by Kaiser, Kézdy, and their co-workers (*7*) can produce impressive results. Based on a hypothetical structure for the receptor-bound conformation of a peptide, analogs are designed that have minimal sequence homology to the parent, but that nonetheless embody the conformational and physical characteristics thought to be responsible for activity. This approach has been used to show that amphiphilic helices are important for the properties of lipoproteins, peptide toxins, and peptide hormones (*7*).

We have adopted a similar approach to the design of proteins. Proteins can be considered as a number of covalently connected peptide subdomains that, although individually highly flexible and disordered, cooperatively assemble into a well-defined three-dimensional structure. Although individually each peptide subdomain may not be well ordered, the overall folding process would be favorable and driven largely by long-range interactions among peptide segments (*10*). Thus protein design involves specifying a set of connected peptides capable of assuming conformations that are complementary to one another. This need to design a specific molecular recognition surface from multiple, flexible segments makes protein design a challenging endeavor.

A few preliminary attempts to design functional proteins based on stereochemical packing considerations met with some encouraging successes [reviewed in (*5*), (*11*), and (*12*)], although their physical properties were not extensively investigated. Recently, a previously designed, 24-residue DDT-binding peptide that has been proposed to adopt an antiparallel β-sheet conformation has been cloned and expressed in bacteria (*13*). Other recent design attempts have been based on idealized naturally occurring folding motifs. Several years ago, a design for a β-barrel protein consisting of two sheets synthesized simultaneously from a cross-linker was reported (*14*). The implementation of this design has proven to be difficult, although recently the protein has been made as two separate, identical halves that can be combined to form a disulfide-bonded pair (*15*). Some encouraging preliminary studies on the design and synthesis of β-α-β supersecondary structures have recently been reported (*16*). A synthetic protein designed to be a four-helix cluster with a composition of amino acids that could occur naturally has

The authors are with E. I. du Pont de Nemours & Company, Central Research and Development Department, Experimental Station, Wilmington, DE 19898.

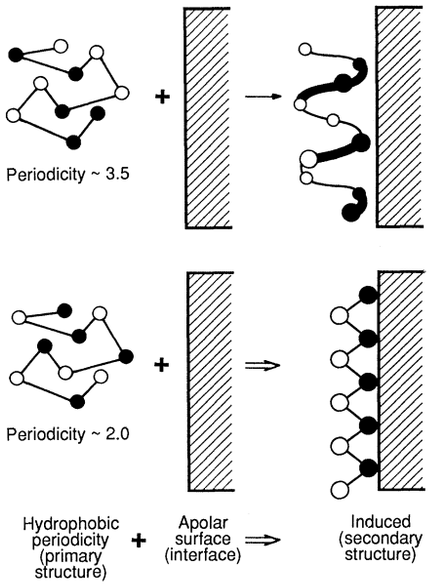


Fig. 1. How an apolar-water interface might influence the conformation of a peptide. Closed and open circles represent apolar and polar residues, respectively. In homogeneous aqueous solution (**left**) the peptides lack ordered conformations, but assume well-defined conformations at apolar-water interfaces that allow them to minimize their free energies of solvation, α helix (**top**) and β sheet (**bottom**). [Redrawn from (*39*)]

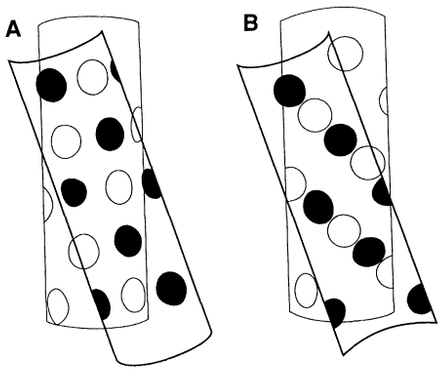


Fig. 2. (**A**) and (**B**) illustrate "knobs into holes" (*42*) and "ridges into grooves" (*47*) side chain–side chain packing arrangements. The helices are depicted as cylinders of approximately 10 Å radius, and the positions of the side chains are denoted as open and closed circles for the bottom and top helix, respectively. The upper part of the top helix has been cut away to allow inspection of the packing.

been cloned and expressed both alone and as a fusion protein with dihydrofolate reductase, but neither product is as yet well characterized (*15*).

We have concentrated on the design of α-helical proteins with the simplest conceivable sequence consistent with function [for example, folding in aqueous solution (*17–20*), folding in phospholipid membranes (*21*), or formation of ion-conductive channels (*21*)]. The design of model structures with minimal complexity simplifies the sequences and the interpretation of the results. We have focused on α-helical proteins for several reasons: α helices are structurally simple, much is known about their stability in solution (*22–25*), and they are internally hydrogen-bonded, forming autonomous folding units. In contrast, β strands require hydrogen bonds to other strands that are distant in sequence.

Principles for the Design of α-Helical Proteins

α-Helical secondary structures are stabilized by interatomic interactions that may be categorized according to the distance between interacting atoms in the sequence of the protein. We define the short-range interactions as those between atoms within a given amino acid residue, medium-range as those between atoms within or proximal to a given secondary, and long-range as those between atoms close in space but distant in sequence.

Short-range interactions account for different amino acids having different conformational preferences. Both statistical [for example, Chou and Fasman (*26*)] and experimental [for example, H. A. Scheraga's "host-guest" (*27*)] methods show that residues such as Glu, Ala, and Met tend to stabilize helices, whereas residues such as Gly and Pro are destabilizing. Conformational preferences can arise from both entropic (*28*) and enthalpic (*29*) effects. However, these intrinsic preferences are not sufficient to account for the stability of helices in globular proteins. Analysis of the free-energy requirements for helix initiation and propagation (*30*) indicates that peptides of 10 to 20 residues should show little helix formation in water (*23*). Although this is generally true for amino acid homopolymers, sequence-specific, monomeric helix formation in water has been observed for peptides as short as 13 residues (*22–25, 31*). Medium-range interactions appear to account for the additional stabilization in these cases.

Studies (*22–25*) on peptides derived from the NH_2-terminal helix of ribonuclease A (the C-peptide) have helped specify some of these medium-range interactions. The α-helical geometry allows the formation of a variety of helix-stabilizing side chain–side chain and side chain–main chain interactions. In an α helix, the positive and negative ends of the amide group dipole point toward the helix NH_2-terminus and COOH-terminus, respectively, giving rise to a significant macrodipole (*32*). Appropriately charged residues near the ends of a helix can favorably interact with the helical dipole and stabilize helix formation (*23–25, 32*). Hydrogen bonds between side chains and the terminal helical N–H and C=O groups may also stabilize helix formation (*33*). Some medium-range side chain–side chain interactions implicated as being important (*25, 31, 34*) include electrostatic interactions, hydrogen bonding, and the perpendicular stacking of aromatic residues (*35*). Although the stabilizing influence of each of these interactions is probably small in comparison with the unfavorable entropy associated with helix formation, they work in concert to make the free energy associated with C-peptide helix formation only slightly unfavorable (*22*) at room temperature. Thus the energetic cost of helix formation is small, and C-peptide binds with high affinity and in a helical conformation to S-protein, a large fragment of the protein from which C-peptide was excised (*36*). The driving force for helix formation and complexation arises from long-range interactions between C-peptide and S-protein (*36*).

Several types of long-range interactions can act cooperatively to stabilize secondary structures. In particular, the importance of hydrophobic interactions can be recognized in the amino acid sequences of individual secondary structures (*37, 38*). Hydrophobic residues often repeat every three to four residues in α helices and every two residues in a β strand (Fig. 1). The role of hydrophobic interactions in determining secondary structures was studied for a series of peptides containing only Leu and Lys in their sequences, but with the residues arranged in different ways (*39*). The peptides tended to have random conformations in very dilute solution, but formed secondary structures matching their hydrophobic periods when they self-associated or bound to the air-water interface (Fig. 1). These results indicate that hydrophobic interactions can play a dominant role in secondary structure formation, and may be used as a central driving force for inducing secondary structure formation in designed proteins.

The next level of complexity in the analysis of helical proteins is specifying the factors determining the packing of helices within a protein. The pairwise geometric arrangement of helices in proteins such as four-helix bundles (*40*) is determined by the packing of side chains protruding from adjacent helices. Side chains in the interiors of proteins are invariably closely packed (*41*). Close packing constrains the interhelical angles by allowing only those interhelical geometries that promote effective interdigitation of side chains (Fig. 2). With an interhelical crossover angle of about 20° the side chains efficiently pack in a "knobs into holes" manner (*42*) (Fig. 2). At this

angle, pairs of straight α helices interact favorably for about six helical turns before they begin to diverge (Fig. 3). In fibrous proteins such as tropomyosin, which has helices interacting over 80 turns, close packing is maintained by coiling of the right-handed helices about one another with a slight left-handed superhelical twist (Fig. 3). This arrangement produces an integrally periodic interaction pattern that repeats every seven residues. The coiled-coil arrangment of α helices has been confirmed in the crystal structures of a number of proteins (*43–45*) and is widespread (*46*).

Analysis of side chain packing interactions determining the interhelical geometries of shorter and less regular helices found in globular proteins has been investigated by means of geometrical methods (*47*) and by the examination of computer or physical models of α helices (*48–51*). Side chains on interacting helices pack favorably when the helical axes are inclined by about −80°, −60°, or 20°, which are the more frequently observed crossover angles in protein structures. However, the distribution of angles around each maximum is broad, and it is difficult to distinguish among different types of idealized packings proposed (*47–50*) to give these preferred helical crossover angles. Residues as small as Gly and as large as Trp are accommodated at the interfaces between interacting helices and can produce large deviations from idealized crossover angles, generating structural diversity in proteins. So although the simple packing diagrams described by others (*42, 47, 49*) (Fig. 2) show a regular, lattice-like structure, the interhelical interaction patterns found in globular proteins correspond to highly flawed lattices in which the order is difficult or impossible to discern. Nevertheless, idealized packing diagrams (Fig. 2) provide conceptual frameworks for the analysis of protein structures and starting points for protein design.

The Design of α-Helical Coiled Coils, from Tropomyosin Models to Ion Channels

Hodges and co-workers (*52*) have designed models for the two-stranded coiled-coil protein, tropomyosin, that shows the potential of the minimalist approach for testing structural hypotheses. Analyses of the sequence of tropomyosin indicated that this protein contains a heptapeptide homology unit that is repeated 40 times (*53*). Polymers of a heptapeptide paradigm for the repeated sequence (Fig. 4) formed parallel, two-stranded, α-helical coiled coils in aqueous solution, and their thermal stability exceeded that of tropomyosin. Close packing of the Leu side chains (Fig. 4) probably determines this exceptional stability. Interestingly, a number of gene-regulatory proteins also contain a segment with a periodic distribution of Leu residues repeating every seventh position (*54*). These segments (named "leucine zippers" by the authors) may form two-stranded coiled coils that mediate dimerization of the DNA-binding domains of these proteins.

The α-helical coiled coil has also been proposed to be an important structural feature in many transmembrane proteins (*46, 55, 56*), and amino acid sequence analysis and electron microscopy (*57*) of ion channel proteins have provided encouraging support for models involving bundles of transmembrane helices. Channel proteins, including the nicotinic acetylcholine receptor (*58*), the sodium channel (*59*), and the potassium channel (*60*), are believed to consist of four or five homologous subunits arrayed about a rotational symmetry axis perpendicular to the bilayer. Each subunit is believed to consist of a bundle of hydrophobic transmembrane helices, some of which also contain a few polar residues that might associate to form a low-energy pathway for conduction of ions (*61–63*). Based on these analyses, we have designed several peptide models for protein ion channels.

The recently published structure (*45*) of the water-soluble, anti-

Table 1. Amino acid sequences and free energies of tetramerization of synthetic peptides. Data are from (*18*) and 1*M* is the standard state.

Peptide	Sequence*	$RT \ln K_d$† (kcal/mol)
α_1	Ac-ELLKKLLEELKG-COOH	−11.4
α_1A	Ac-GKLEELLKKLLEELKG-COOH	−19
α_1B	AcGELEELLKKLKELLKG-$CONH_2$	−22

*Abbreviations for the amino acid residues are: E, Glu; G, Gly; K, Lys; and L, Leu. †*R*, gas constant; *T*, absolute temperature; and K_d, dissociation constant.

parallel four-stranded coiled coil, ColE1 Rop protein, considerably aided our analysis and design. An idealized packing arrangement for a typical heptapeptide repeat of an antiparallel four-stranded coiled coil (parallel coiled coils are fundamentally similar) is shown in Fig. 5. In Rop, close packing of the apolar interior of the structure is achieved by alternating large (Leu or Ile) and small (Ala, Thr, or Cys) residues at positions d and a, respectively. Positions e and g also contribute to structural stability and are often largely apolar or amphiphilic, whereas the remaining residues at positions b, c, and f are generally hydrophilic and provide a polar surface for the water-soluble structure.

We have designed an "inside-out" version of Rop in which the positons of the hydrophobic and hydrophilic residues have been reversed (*21*), providing an apolar exterior and a polar interior (Fig. 6). To function as an ion channel, the interior was designed to leave room for ions and water molecules rather than to be closely packed as in Rop. To accomplish this, we chose Ser, the smallest of the commonly occurring polar amino acids, for positions a and d. Leu, an apolar residue that packs well at helix-helix interfaces, was chosen for the remaining positions of the heptad repeat. The basic heptad was then repeated three times to give a 21-residue model helix, H_2N-(Leu-Ser-Leu-Leu-Leu-Ser-Leu)$_3$-$CONH_2$ [(LSLLLSL)$_3$], just long enough to span the hydrophobic portion of the phospholipid bilayer. In models for a tetramer of this helix there is sufficient room for isolated water molecules between the layers of Ser residues, although there is insufficient room for a hydrated alkali metal ion to traverse the length of the structure. Indeed, (LSLLLSL)$_3$ forms efficient proton channels that are impervious to lithium or other alkali metal cations (*21*). The conductance states and lifetimes of the (LSLLLSL)$_3$ channels show a degree of homodispersity similar to that found in natural ion channel proteins, and very different from membrane-perturbing peptide toxins such as melittin (*64*), which give rise to erratic increases in membrane conductance (*65*).

We also extensively modeled another peptide, which we had designed based on maximizing lateral amphiphilicity (*21*). When this peptide, H_2N-(Leu-Ser-Ser-Leu-Leu-Ser-Leu)$_3$-$CONH_2$ [(LSSLLSL)$_3$], was modeled into the structure of a four-stranded coiled coil, only two Ser side chains per heptad repeat could be accommodated in the interior of the structure, and one Ser occupied a position at the helical interface (Fig. 7), disrupting the otherwise regular packing of the Leu side chains. However, in a six-stranded coiled coil (*55, 56*) (Fig. 7), all three of the Ser side chains could be better accommodated in the central channel, which was large enough (8 Å diameter) to accept hydrated alkali metal cations. Indeed, this peptide formed channels (*21*) that conducted protons, alkali metal cations, and organic cations of less than 8 Å diameter.

Considerable theoretical, physical, and experimental studies will be required to test the validity of the models in Fig. 6 and 7. One aspect of the models is almost certainly correct; circular dichroism (CD) spectroscopy and infrared spectroscopy show that the peptides adopt helical conformation in phospholipid membranes (*66*). The experimental determination of the precise number and orientation of the helices in the channels may be approached by analyzing

the kinetics and thermodynamics of channel formation. Current-time tracings of membranes with concentrations of peptide low enough to contain a small number of channel molecules (*21*) show the abrupt, quantized increases and decreases in conductance characteristic of the single molecular events associated with natural ion channel proteins (*67*). This quantized conductance behavior has classically been analyzed in terms of models in which the channels are spontaneously forming and breaking down, with the rate constants for these events depending on the transmembrane potential and the peptide concentration (*67–69*). Significant channel formation occurs only when the side of the bilayer opposite peptide addition is positive, and the frequency increases exponentially (about *e*-fold/10 mV) with increasing voltage, suggesting that the helices might adopt two different orientations in membranes. In the surface-bound state the helices would lie parallel to the membrane with the Ser side chains projecting toward water and the Leu side chains projecting toward the apolar portions of the phospholipid membranes. In the perpendicularly inserted state the peptide helices transverse the phospholipid bilayer and associate to form channels. In this model, the voltage dependence of channel formation would indicate that, in the absence of an applied transmembrane voltage, the peptides strongly prefer the surface-bound state, and hence channel formation is infrequent. However, a transmembrane potential can electrostatically stabilize the perpendicular orientation of the peptides if they insert with the negative end of their helical dipoles oriented toward the positive side of the bilayer.

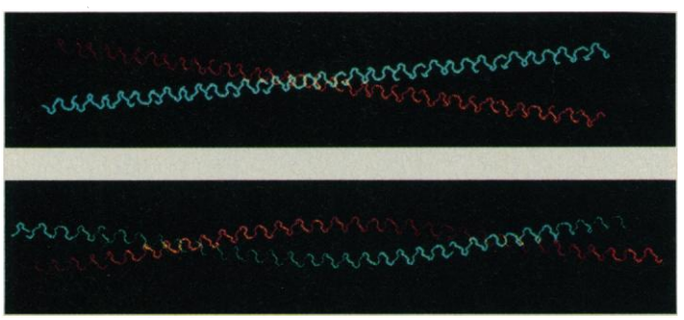


Fig. 3. α Helices packed with a 20° crossover angle diverge for straight (**top**) but not supercoiled (**bottom**) helices.

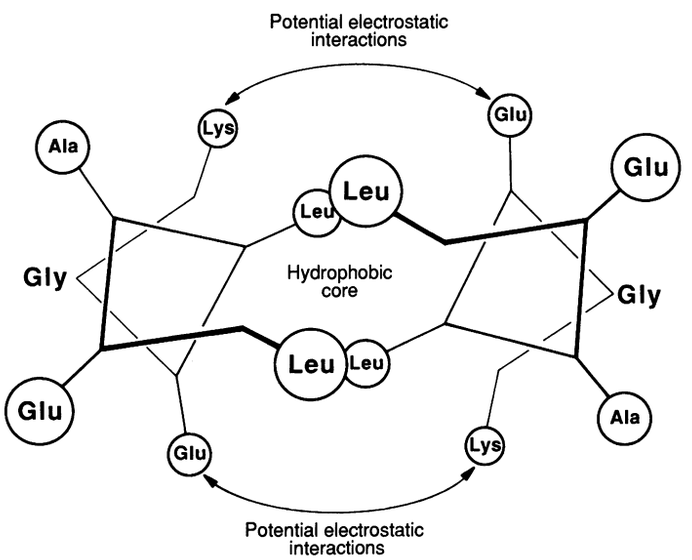
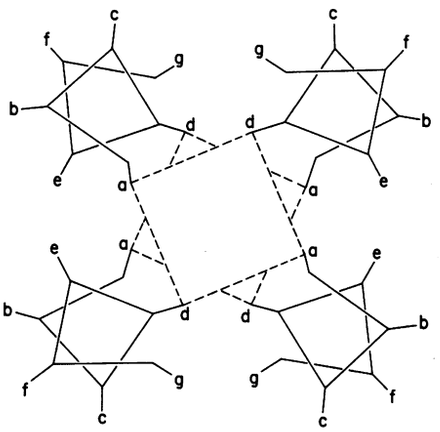


Fig. 4. A single heptapeptide repeat of an idealized model for two-stranded coiled-coil proteins (*53*).

Fig. 5. An idealized packing diagram of the basic heptapeptide repeat for an antiparallel four-stranded coiled coil. [Redrawn from (*45*)]



Similar peptide concentration and voltage dependences of channel formation have been observed for the α-helical peptide antibiotic, alamethicin, although the analysis is even more complex, because this peptide forms "bursts" of channel activity with up to nine different conductance states (*68, 69*). Channels occur more frequently when the side of the bilayer opposite to peptide addition is at a negative potential, a fact interpreted as preferential insertion of the very hydrophobic, acetylated NH_2-terminus of that peptide. Both $(LSSLLSL)_3$ and $(LSLLLSL)_3$, which contain a free, ionizable NH_2-terminus and a blocked, carboxamide COOH-terminus, show the opposite behavior.

The number of helices in the ion channels is classically estimated from the peptide concentration dependence of the "macroscopic" conductance measured in bilayers containing a large number of channels (*68, 69*). However, the interpretation of such experiments is made extremely difficult by the number of equilibria that need to be considered (monomer-to-aggregate in the solution, membrane-bound, and inserted states) and other experimental difficulties. For instance, the macroscopic conductance is between second and eleventh order in bulk alamethicin concentration depending on the lipid used to form membranes (*69*). We have instead initially focused on molecular modeling to determine reasonable packing schemes that could account for the conductance properties of the peptide channels (*21*). A variety of models for trimeric, tetrameric, pentameric, and hexameric helical bundles were constructed from straight α helices, and then energy was minimized. The resulting structures with the most uniform side chain packing and lowest energies were found to be approximate left-handed supercoils with rotational symmetry and knobs-into-holes packing. Based on the known cation size–selectivities of the LSLLLSL and LSSLLSL channels, the tetramer and hexamer, respectively, appear most reasonable, although an extremely well packed three-stranded coiled-coil arrangement could be constructed for $(LSLLLSL)_3$, and $(LSSLLSL)_3$ appeared to be capable of forming coiled coils of more than six strands. Further experimental studies will be required to definitively determine the aggregation state.

Design of a Water-Soluble Four-Helix Protein

The four-helix bundle (*40, 51*) conformation is a common folding motif found in the structures of such functionally diverse proteins as myohemerythrin, cytochrome c′, and TMV coat protein (Fig. 8). Like Rop, they are composed of four nearly antiparallel α helices packed with interhelical angles of about 20°. However, in Rop the helices maintain a constant interhelical distance throughout the length of the bundle by virtue of their being left-handedly super-

coiled, whereas in myohemerythrin and cytochrome c′ the helices gradually diverge from a point near one end of the bundle (*40*) (Fig. 8). Divergence of the helices is functionally significant; it gives rise to a cavity that accommodates binuclear iron in myohemerythrin and hemes in the cytochromes.

The divergent four-helix bundle motif is an excellent target for molecular engineering because of its structural simplicity and natural tendency to produce cavities that accommodate prosthetic groups. An important first question to ask is what causes the helices to diverge in some four-helix proteins but to supercoil in others? The previously discussed periodic distribution of large and small apolar residues that occurs in the highly symmetric Rop protein, but not in the divergent bundles, is important for forming a four-stranded coiled coil. Other factors should become apparent when the crystal structures of other four-helix bundle proteins such as growth hormone (*70*) have been refined to high resolution.

We have applied the techniques of de novo design to probe the structural requirements necessary for formation of the four-helical bundle fold (*17–20*). Our initial target, α_4 (Fig. 8), is an idealized, shortened version of the bundles found in natural proteins such as myohemerythrin. The helices in α_4 are three to four turns in length, a size that appears minimally sufficient to provide adequate long-range interactions to drive the folding process. Although this is not sufficiently long to allow the helices to diverge significantly, subsequent designs could extend the helices to create cavities useful for constructing binding sites for small molecules. In the initial design, we idealized the approximate symmetry characteristic of natural four-helix bundle proteins (Fig. 8). This simplifies the design and modeling process considerably; rather than designing an entire protein it is only necessary to create a single helical sequence that, upon application of the appropriate symmetry operator, will give rise to a closely packed, four-helix bundle.

The helical sequence was initially designed in collaboration with Eisenberg and co-workers with the use of physical models and computer graphics (*17*). Physical models of four-polyalanine α helices were arranged in an approximate 222 symmetrical array, with an interhelical distance of approximately 10 Å and a left-handed tilt of about 20° to provide a favorable interhelical packing geometry. Side chains were then added to stabilize the structure. Good interior packing could be accomplished with Leu as the only interior residue. Glu and Lys residues were chosen as charged, helix-stabilizing residues for the solvent-accessible exterior of the protein and were generally spaced a helical turn apart to help stabilize helix formation by electrostatic interactions.

To test the assumption that helix-helix interactions could indeed provide the primary driving force for folding, we synthesized peptides with the sequence of the designed helices and evaluated the thermodynamics associated with their self-assembly into α-helical tetramers from the peptide concentration dependence of the CD spectrum (*17, 18*). As monomers in very dilute, aqueous solution, the peptides are predominantly nonhelical, because they lack the stabilizing long-range interactions necessary for folding, whereas at higher peptide concentrations they self-associate into predominantly α-helical tetramers. The monomer-to-tetramer equilibrium is highly cooperative, indicating that tetramers are formed preferentially over other aggregates. The free energy associated with the self-assembly process provided an experimental parameter for evaluation of alternate helical sequences (Table 1). The best helical sequence, α_1B, combines several stabilizing features of earlier designs (*17–19, 39*), including a chain length of 16 residues (a shorter peptide formed a much less stable tetramer), and the potential for numerous medium-range electrostatic and long-range hydrophobic interactions.

The next stage in the design involved linking two adjacent α_1B helices in an antiparallel arrangement (*18*). Examination of a physical model of the α_1B tetramer suggested that this could be accomplished by inserting a single Pro residue between the NH_2- and COOH-terminal Gly residues of this peptide. However, the resulting peptide formed an unforeseen trimeric aggregate rather than the desired dimer of helical hairpins. This result illustrates an important point concerning the protein-folding problem; the sequences of natural proteins have evolved not only to stabilize a desirable three-dimensional structure, but to destabilize all other possible alternatives. A successfully designed protein must do the same. The inappropriate aggregation was alleviated by introducing two Arg residues after the Pro residue in the link sequence to provide additional flexibility for loop formation and to electrostatically destabilize a likely conformation of the trimer (*18, 19*). Finally, a gene directing the synthesis of the complete α_4 protein (Fig. 9) with four α_1B helices connected by three Pro-Arg-Arg links was designed and expressed in *Escherichia coli* (*19, 20*).

The complete α_4 protein is the first example of a designed protein that has been shown to adopt a folded, globular conformation in aqueous solution (*20*). Circular dichroism spectroscopy indicates

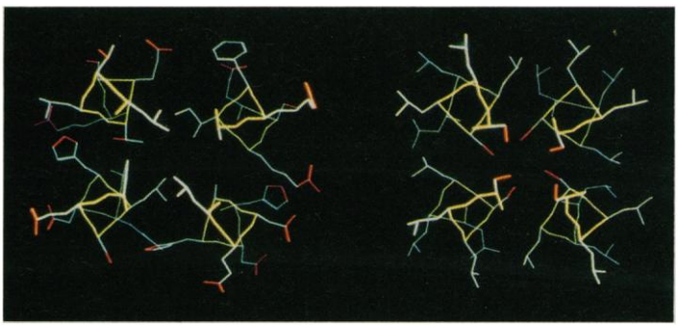


Fig. 6. Membrane-soluble (**left**) and water-soluble (**right**) four-stranded coiled-coil structures. In each case a single heptad is viewed looking down the central axis of the superhelix. The backbone atoms are shown in yellow, polar side chain functionality (alcohols, imidazole, amines, and carboxylates) in red, and apolar side chain functionality (C–C bonds) in white. The left structure is from the crystal structure of Rop (*45*), and the right structure is the model for the $(LSLLLSL)_3$ channels generated as described (*21*). As supported by the experimental data (*21*), the $(LSLLLSL)_3$ structure is a parallel bundle, whereas the Rop structure is an antiparallel bundle. The packing interactions are fundamentally similar for both orientations.

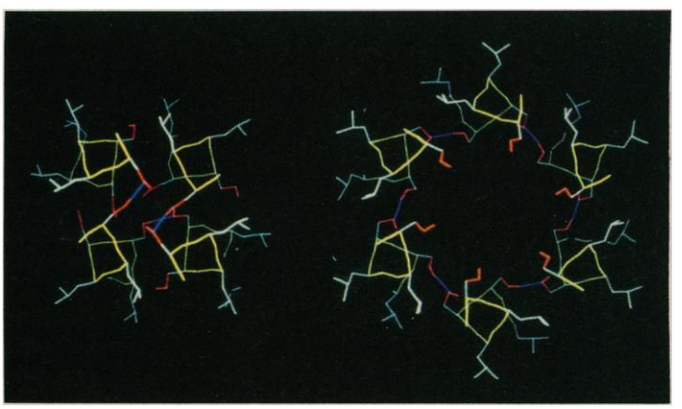


Fig. 7. Axial views of parallel, four- and six-stranded coiled-coil structures for the peptide $(LSSLLSL)_3$. The coloring of functional groups is as in Fig. 6, and hydrogen bonds between Ser hydroxyls are shown in blue. Note that in the four-stranded structure one Ser per heptad repeat lies at the otherwise well-packed apolar interface between neighboring helices [compare with the tetramer of $(LSLLLSL)_3$ shown in Fig. 6]. In the hexamer, all of the Ser side chains project more directly toward the interior of the structure, and are able to hydrogen bond to one another, as well as to the solvent in the central cavity. The models were generated as described (*21*).

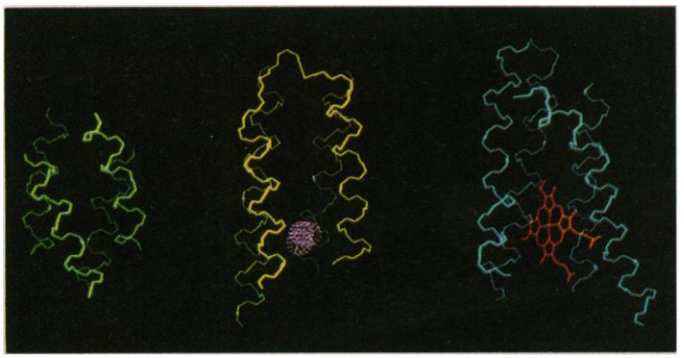


Fig. 8. The crystal structures of myohemerythrin (*74*) (**middle**) and cytochrome c′ (*75*) (**left**), and the intended folding pattern for α_4 (**right**). Only the helical portions of the proteins are shown. The van der Waals surface associated with the binuclear iron and bound oxygen atoms are indicated with dots.

Helix	-Gly-Glu-Leu-Glu-Glu-Leu-Leu-Lys-Lys-Leu-Lys-Glu-Leu-Leu-Lys-Gly-
Loop	-Pro-Arg-Arg-
α_1B	Ac-Helix-$CONH_2$
α_2	Ac-Helix-Loop-Helix-$CONH_2$
α_4	Met-Helix-Loop-Helix-Loop-Helix-Loop-Helix-COOH

Fig. 9. The amino acid sequences of the designed peptides α_1B, α_2, and α_4. Ac, NH_2-terminal acetyl; $CONH_2$, COOH-terminal carboxamide.

that it is predominantly α-helical, and size-exclusion chromatography indicates that it is compactly folded and monomeric. Furthermore, the folded conformation of the protein is extraordinarily stable, particularly when compared to other small, natural proteins. It displays a highly cooperative guanidine hydrochloride (GuHCl)–induced unfolding transition between approximately 6 and 7*M* GuHCl. Most other small natural proteins show unfolding transitions of similar cooperativity, but the midpoints for the transitions generally occur between 1 and 3*M* GuHCl. The increased stability of the α_4 protein is in keeping with its idealized design; its helices are composed of more helix-stabilizing residues and contain more stabilizing electrostatic interactions than the helices in naturally occurring proteins. Also, the interior of the protein is more regular and perfectly hydrophobic than in natural proteins. The extreme stability of α_4 indicates that it should be a suitable framework for the construction of binding and catalytic sites.

The protein, α_4, can reasonably be assumed to contain four tightly packed helices, based on its physical properties and the incremental approach with which it was designed. However, it remains to be determined whether the helices are arranged in antiparallel bundles as envisioned in Fig. 8. A disulfide bond between the links at the top of the bundle would constrain the orientation of the helices and may stabilize the folded conformation if the hypothetical structure is correct. To test this, a peptide composed of two α_1B sequences connected by a Cys-Arg-Arg link was synthesized (*71*). The disulfide-bonded homodimer of this peptide is indeed highly stable toward GuHCl-induced denaturation, with a midpoint at slightly higher GuHCl concentration than for α_4. Furthermore, modeling indicated that it should be considerably less stabilizing to move the disulfide one residue toward the COOH-terminus. A covalent dimer with two α_1B sequences connected by a Pro-Cys-Arg link is less stable toward reduction and denaturation than the dimer with the Cys-Arg-Arg link. It will be of interest to determine the effect of introducing the corresponding mutations into the original, full-length α_4 protein. Ultimately, the structure of α_4 should be determined by x-ray crystallography. Crystals of unlinked, designed helices suitable for x-ray diffraction studies have been grown by Eisenberg and co-workers (*17*) and structural studies are under way.

Implications

The proteins described in this article embody many protein structural principles that have emerged over the last 20 to 30 years. The similarity of the properties of these model proteins to natural proteins indicates that it is possible to design simple proteins from scratch and that the study of such models should enrich our understanding of natural proteins. For instance, it has been proposed that cation-selective (*21*) channel proteins have negatively charged side chains that aid in charge discrimination (*62*). However, the $(LSSLLSL)_3$ peptide that lacks negatively charged groups is, in fact, cation-selective, which indicates that dipolar effects involving formally neutral side chains can produce impressive selectivity. Thus the minimalist design approach is capable of producing fundamental insights that are useful in the understanding of natural proteins.

Does the symmetry of our designed proteins limit their appropriateness as models for less symmetrical natural proteins? Various degrees of symmetry are apparent in the structures of natural proteins. The individual subunits of multisubunit proteins (*50*) are often similar or identical and are regularly arrayed in proteins such as the acetylcholine receptor, hemoglobin, and viral coat proteins. Even the individual domains of many small globular proteins such as calcium-binding proteins (*72*) show significant approximate symmetry. For instance, intestinal calcium-binding protein contains two homologous symmetry-related domains that evolved from a single primordial gene (*72, 73*). Although the sequences are currently nonidentical, when the gene duplicated they must have been identical, and the three-dimensional structure of the two-domain protein must have been more symmetrical. As discussed by McLachlan (*73*), tandem gene duplication and subsequent divergence is a general mechanism that has been used to evolve function in a variety of proteins including those containing zinc fingers, iron-sulfur clusters, and β barrels. In a similar vein, the availability of a clone for the α_4 protein should allow us to introduce asymmetric changes to one or more of the helices and loops as we attempt to evolve our structural protein into a functional protein.

Finally, in the work described herein we have restricted our studies to peptides composed of the 20 commonly occurring amino acids, because our primary interest was to explore some of the principles underlying the structures and properties of natural proteins. However, because designed proteins can be small enough to be approached by chemical methods, the palette of potential design elements is virtually unlimited. In subsequent designs, it should be possible to incorporate a variety of chemical functionalities to achieve desired functional or structural properties.

Note added in proof: Mutter has recently described a promising approach to the design of simple, symmetrical proteins using polypeptide templates (*77*).

A recent analysis of Rop suggests that its helices achieve the overall effect of supercoiling by bending at discreet junctions rather than bending smoothly throughout the length of the bundle as in an idealized four-stranded coiled coil (*78*). It will be interesting to determine whether this result is general, both to long "bent" helices in globular proteins as well as to the very long helices in coiled coils such as tropomyosin.

REFERENCES AND NOTES

1. T. Creighton, *Proteins* (Freeman, San Francisco, 1984), p. 135.
2. A. M. Lesk and C. Chothia, *J. Mol. Biol.* **160**, 325 (1982); *ibid.* **136**, 225 (1980);

R. F. Doolittle, *Science* **214**, 149 (1981).
3. D. L. Oxender and C. F. Fox, Eds., *Protein Engineering* (Liss, New York, 1987).
4. J. R. Knowles, *Science* **236**, 1252 (1987); G. K. Ackers and F. R. Smith, *Annu. Rev. Biochem.* **54**, 597 (1985).
5. W. F. DeGrado, *Adv. Protein Chem.*, in press.
6. G. D. Rose, L. M. Gierasch, J. A. Smith, *ibid.* **37**, 1 (1985).
7. E. T. Kaiser and F. J. Kézdy, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1137 (1983); *Science* **223**, 249 (1984); S. Erickson-Viitanen, K. T. O'Neil, W. F. DeGrado, in (*3*), pp. 201–211; J. W. Taylor and E. T. Kaiser, *Pharmacol. Rev.* **38**, 291 (1986).
8. J. Gariepy, B. D. Sykes, R. S. Hodges, *Biochemistry* **22**, 1765 (1983); R. E. Reid, *ibid.* **26**, 6070 (1987).
9. V. J. Hruby, *Life Sci.* **31**, 189 (1982); ——— and M. E. Hadley, in *Design and Synthesis of Organic Molecules Based on Molecular Recognition*, G. van Binst, Ed. (Springer-Verlag, Berlin, 1984), p. 269; P. W. Schiller, *Peptides (New York)* **6**, 219 (1984).
10. H. Taniuchi and C. B. Anfinson, *J. Biol. Chem.* **244**, 3864 (1969); I. Kato and C. B. Anfinson, *ibid.* **244**, 1004 (1967).
11. R. Moser *et al.*, *Angew. Chem. Int. Ed. Engl.* **24**, 719 (1985).
12. M. Mutter, *ibid.*, p. 639.
13. R. Moser *et al.*, *Protein Eng.* **1**, 339 (1987).
14. J. S. Richardson, D. C. Richardson, B. W. Erickson, *Biophys. J.* **45**, 25A (1984).
15. J. S. Richardson and D. C. Richardson, in (*3*), pp. 149–164; J. S. Richardson, personal communication.
16. M. Mutter, K.-H. Altmann, T. Vorherr, *Z. Naturforsch. Teil B* **41**, 1315 (1986).
17. D. Eisenberg *et al.*, *Proteins* **1**, 16 (1986).
18. S. P. Ho and W. F. DeGrado, *J. Am. Chem. Soc.* **109**, 6751 (1987).
19. W. F. DeGrado, L. J. Regan, S. P. Ho, *Cold Spring Harbor Symp. Quan. Biol.* **52**, 521 (1987).
20. L. Regan and W. F. DeGrado, *Science* **241**, 976 (1988).
21. J. D. Lear, Z. R. Wasserman, W. F. DeGrado, *ibid.* **240**, 1177 (1988).
22. A. Bierzynski, P. S. Kim, R. L. Baldwin, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2470 (1982).
23. K. R. Shoemaker *et al.*, *ibid.* **82**, 2349 (1985).
24. K. R. Shoemaker *et al.*, *Nature* **326**, 563 (1987).
25. K. R. Shoemaker *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **52**, 391 (1987).
26. P. Y. Chou and G. D. Fasman, *Adv. Enzymol.* **47**, 45 (1978).
27. M. Sueki *et al.*, *Macromolecules* **17**, 148 (1984).
28. L. Piela, G. Nemethy, H. A. Scheraga, *ibid.* **26**, 1273 (1987); M. J. McGregor, S. A. Islam, M. J. E. Sternberg, *J. Mol. Biol.* **198**, 295 (1987); J. Janin, S. Wodak, M. Levitt, B. Maigret, *ibid.* **125**, 357 (1978).
29. M. Vásquez, G. Némethy, H. A. Scheraga, *Macromolecules* **16**, 1043 (1983).
30. H. Scheraga, *Pure Appl. Chem.* **50**, 315 (1978).
31. S. Marqusee and R. L. Baldwin, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8898 (1987).
32. W. G. J. Hol, L. M. Halie, C. Sander, *Nature* **294**, 532 (1981); A. Wada, *Adv. Biophys.* **9**, 1 (1977); D. E. Blagdon and M. Goodman, *Biopolymers* **14**, 241 (1975).
33. L. G. Presta and G. D. Rose, *Science* **240**, 1632 (1988); J. S. Richardson and D. C. Richardson, *ibid*, p. 1648.
34. M. Sundaralingam, Y. C. Sekharudu, N. Yathindra, V. Ravichandran, *Proteins* **2**, 64 (1987).
35. T. Blundell, J. Singh, J. Thornton, S. K. Burley, G. A. Petsco, *Science* **234**, 1005 (1986).
36. F. M. Richards and H. W. Wyckoff, in *The Enzymes*, P. D. Boyer, Ed. (Academic Press, New York, 1971), vol. 4, pp. 647–806; F. M. Finn and K. Hofmann, *Acc. Chem. Res.* **6**, 169 (1973); T. Kanmera, G. A. Homanderg, A. Komoriya, I. M. Chaiken, *Int. J. Peptide Protein Res.* **21**, 74 (1983); A. Komoriya and I. M. Chaiken, *J. Biol. Chem.* **257**, 2599 (1982); H. C. Taylor, A. Komoriya, I. M. Chaiken, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 6423 (1985).
37. M. I. Kanehisa and T. Y. Tsong, *Biopolymers* **19**, 1617 (1980).
38. D. Eisenberg, R. M. Weiss, T. C. Terwilliger, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 140 (1984).
39. W. F. DeGrado and J. D. Lear, *J. Am. Chem. Soc.* **107**, 7684 (1985).
40. P. C. Weber and F. R. Salemme, *Nature* **287**, 82 (1980).
41. F. M. Richards, *Annu. Rev. Biophys. Bioeng.* **6**, 151 (1977).
42. F. H. C. Crick, *Acta Cryst.* **6**, 689 (1953).
43. G. N. Phillips, Jr., J. P. Fillers, C. Cohen, *J. Mol. Biol.* **192**, 111 (1986).
44. I. A. Wilson, J. J. Skehel, D. C. Wiley, *Nature* **289**, 366 (1981).
45. D. W. Banner, M. Kokkinidis, D. Tsernoglou, *J. Mol. Biol.* **196**, 657 (1987).
46. C. Cohen and D. A. D. Parry, *Trends Biochem. Sci.* **11**, 245 (1986).
47. C. Chothia, *Annu. Rev. Biochem.* **53**, 537 (1984).
48. K.-C. Chou, G. Némethy, H. A. Scheraga, *J. Am. Chem. Soc.* **106**, 3161 (1984).
49. T. J. Richmond and F. M. Richards, *J. Mol. Biol.* **119**, 537 (1978).
50. J. Richardson, *Adv. Prot. Chem.* **34**, 167 (1981).
51. D. H. Ohlendorf, B. C. Finzel, P. C. Weber, F. R. Salemme, in (*3*), pp. 165–174.
52. J. A. Talbot and R. S. Hodges, *Acc. Chem. Res.* **15**, 224 (1982); S. Y. M. Lau, A. K. Taneja, R. S. Hodges, *J. Biol. Chem.* **259**, 13253 (1984).
53. A. D. McLachlan and M. Stewart, *J. Mol. Biol.* **98**, 293 (1975); R. S. Hodges, J. Sodek, L. B. Smilie, L. Jurasek, *Cold Spring Harbor Symp. Quant. Biol.* **37**, 299 (1972); A. D. McLachlan and J. Karn, *J. Mol. Biol.* **164**, 605 (1983).
54. W. H. Landshulz, P. F. Johnson, S. L. McKnight, *Science* **240**, 1759 (1988).
55. M. Inouye, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 2396 (1974).
56. A. K. Dunker and D. A. Marvin, *J. Theor. Biol.* **72**, 9 (1978).
57. A. Brisson and P. N. T. Unwin, *Nature* **315**, 474 (1985).
58. M. Noda *et al.*, *ibid.* **305**, 818 (1983); J. Boulter *et al.*, *ibid.* **319**, 368 (1986).
59. M. Noda *et al.*, *ibid.* **312**, 121 (1984); L. Salkoff *et al.*, *Science* **237**, 744 (1987).
60. D. M. Papazian *et al.*, *Science* **237**, 749 (1987); B. L. Tempel *et al.*, *ibid.*, p. 770.
61. H. R. Guy and F. Hucho, *Trends Neurosci.* **10**, 318 (1987); J. Finer-Moore and R. Stroud, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 155 (1984).
62. R. E. Greenblatt, Y. Blatt, M. Montal, *FEBS Lett.* **193**, 125 (1985); H. R. Guy, *Biophys. J.* **45**, 249 (1984); E. M. Kosower, *FEBS Lett.* **182**, 234 (1985).
63. S. Oiki, W. Danho, M. Montal, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2393 (1988).
64. M. T. Tosteson and D. C. Tosteson, *Biophys. J.* **36**, 109 (1981).
65. M. T. Tosteson *et al.*, *Biochemistry* **26**, 6627 (1987).
66. W. F. DeGrado and M. Rafalski, unpublished results.
67. B. Hille, *Ionic Channels of Excitable Membranes* (Sinauer, Sunderland, MA, 1984).
68. R. Latorre and O. Alvarez, *Physiol. Rev.* **61**, 77 (1981).
69. J. E. Hall, I. Vodyanov, T. M. Balasubramanian, G. R. Marshall, *Biophys. J.* **45**, 233 (1984).
70. S. S. Abdel-Meguid *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6434 (1987).
71. L. Regan and W. F DeGrado, unpublished observations.
72. D. M. E. Szebenyi and K. Moffat, *J. Biol. Chem.* **261**, 8761 (1986); R. H. Kretsinger, *Cold Spring Harbor Symp. Quant. Biol.* **52**, 499 (1987).
73. A. D. McLachlan, *Cold Spring Harbor Symp. Quant. Biol.* **52**, 411 (1987).
74. S. Sheriff, W. A. Hendrickson, J. L. Smith, *J. Mol. Biol.* **197**, 273 (1987).
75. B. C. Finzel, P. C. Weber, K. D. Hardmann, F. R. Salemme, *ibid.* **186**, 627 (1985).
76. D. M. E. Szebenyi, S. K. Obendort, K. Moffat, *Nature* **294**, 327 (1981).
77. M. Mutter, *Trends Biochem. Sci.* **13**, 260 (1988).
78. C. Chothia, unpublished results.
79. This article is dedicated to the memory of Professor E. Thomas Kaiser and Irving S. Sigal. We thank our many co-workers who have contributed significantly to the work described herein: G. Murray III, R. Wilk, J. Kallal, S. P. Ho, L. Regan, D. Eisenberg, F. R. Salemme, R. Hilmer, P. Weber, and S. Brenner.