648 (1977).
52. J. F. Perdue, *Can. J. Biochem. Cell Biol.* **62**, 1237 (1984).
53. D. Maiter, L. E. Underwood, M. Maes, M. L. Davenport, J. M. Ketelslegers, *Endocrinology* **123**, 1053 (1988); K. C. Copeland, L. E. Underwood, J. J. Van Wyk, *J. Clin. Endocrinol. Metab.* **50**, 690 (1980); M. B. Grant *et al.*, *ibid.* **63**, 981 (1986).
54. J. J. Van Wyk, M. E. Svoboda, L. E. Underwood, *J. Clin. Endocrinol. Metab.* **50**, 206 (1980).
55. J. Zapf and E. R. Froesch, *Horm. Res.* **24**, 121 (1986).
56. K. Hall and V. R. Sara, in *Vitamins and Hormones* (Academic Press, New York, 1983), vol. 40, pp. 175–233; L. E. Underwood, A. J. D'Ercole, D. R. Clemmons, J. J. Van Wyk, *Clin. Endocrinol. Metab.* **15**, 59 (1986); R. C. Baxter, *Adv. Clin. Chem.* **25**, 49 (1986).
57. A. J. D'Ercole, A. D. Stiles, L. E. Underwood, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 935 (1984).
58. C. C. Orlowski and S. D. Chernausek, *Endocrinology* **123**, 44 (1988); J. Isgaard *et al.*, *ibid.* **122**, 1515 (1988); J. Isgaard, A. Nilsson, K. Vikman, O. G. P. Isaksson, *J. Endocrinol.* **120**, 107 (1988); N. L. Schlechter, S. M. Russell, E. M. Spencer, C. S. Nicoll, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7932 (1986).
59. J. Zapf, C. Hauri, M. Waldvogel, E. R. Froesch, *J. Clin. Invest.* 77, 1768 (1986).
60. A. Skottner, R. G. Clark, I. C. A. F. Robinson, L. Fryklund, *J. Endocrinol.* **112**, 123 (1987).
61. E. Schoenle, J. Zapf, C. Hauri, T. Steiner, E. R. Froesch, *Acta Endocrinol.* **108**, 167 (1985).
62. K. Hall and V. R. Sara, *Clin. Endocrinol. Metab.* **13**, 91 (1984).
63. R. C. Baxter, Z. Zaltsman, J. R. Turtle, *J. Clin. Endocrinol. Metab.* **58**, 955 (1984).
64. A. N. Corps, K. D. Brown, L. H. Rees, J. Carr, C. G. Prosser, *ibid.* **67**, 25 (1988).
65. A. Honegger and R. E. Humbel, *J. Biol. Chem.* **261**, 569 (1986).
66. L. F. Fisher and E. L. Russell, unpublished report (Toxicology Division, Lilly Research Laboratories, Eli Lilly & Co., Indianapolis, IN, 1989).
67. J. B. Terrill, unpublished report (Hazleton Laboratories of America, Inc., prepared for Monsanto Agricultural Co., St. Louis, MO, 1989).
68. T. C. White *et al.*, unpublished report MSL 8671 (Monsanto Agricultural Co., St. Louis, MO, 1989).
69. R. J. Collier *et al.*, unpublished report MSL 8531 (Monsanto Agricultural Co., St. Louis, MO, 1988).
70. D. Schams and H. Karg, unpublished Elanco report, Eli Lilly & Co., dated 23 September 1988 (Institute of Physiology, Technische Universitat Munchen, D-8050 Freising-Weihenstephan, FRG, 1988); unpublished Elanco report, Eli Lilly & Co. (Institute of Physiology, Technische Universitat Munchen, D-8050 Freising-Weihenstephan, FRG, 1988).
71. S. R. Davis, P. D. Gluckman, A. M. Bryant, unpublished Elanco report, Eli Lilly & Co. (Ruakura Agricultural Centre, Hamilton, New Zeland, 1989).
72. T. C. White *et al.*, unpublished report MSL 8633 (Monsanto Agricultural Co., St. Louis, MO, 1989).
73. M. A. Miller *et al.*, unpublished report MSL 8673 (Monsanto Agricultural Co., St. Louis, MO, 1989).
74. M. A. Miller, T. C. White, R. J. Collier, unpublished report (Monsanto Agricultural Co., St. Louis, MO, 1989).
75. We thank S. Sechen for her contributions to the manuscript.

Research Article

De Novo Design, Expression, and Characterization of Felix: A Four-Helix Bundle Protein of Native-Like Sequence

MICHAEL H. HECHT,* JANE S. RICHARDSON, DAVID C. RICHARDSON,† RICHARD C. OGDEN

The protein Felix was designed de novo to fold into an antiparallel four-helix bundle of specific topology. Its sequence of 79 amino acid residues is not homologous to any known protein sequence, but is "native-like" in that it is nonrepetitive and contains 19 of the 20 naturally occurring amino acids. Felix has been expressed from a synthetic gene cloned in *Escherichia coli*, and the protein has been purified to homogeneity. Physical characterization of the purified protein indicates that Felix (i) is monomeric in solution, (ii) is predominantly α-helical, (iii) contains a designed intramolecular disulfide bond linking the first and fourth helices, and (iv) buries its single tryptophan in an apolar environment and probably in close proximity with the disulfide bond. These physical properties rule out several alternative structures and indicate that Felix indeed folds into approximately the designed three-dimensional structure.

NEW TECHNIQUES IN MOLECULAR BIOLOGY HAVE OPENED up the potential for engineering the structural properties of proteins to desirable specifications. This possibility is being explored with mutagenesis procedures to alter the properties of existing structures and also by designing entire protein structures de novo.

De novo design represents an attempt to choose an amino acid sequence that is unrelated to any natural protein sequence, but that will fold into a desired three-dimensional structure. The principles and details of protein folding are not well enough understood to ensure the success of such attempts; nevertheless, we and others are tackling some of the simpler cases. The reason for taking such a drastic and uncertain step is that minor variants of natural proteins cannot be used to test determinants that control major topological differences in structure because of the obstacles presented by all those features of the protein that have been evolutionarily selected to fit the native structure. We believe there are important unsettled fundamental questions, and therefore are attempting de novo design

M. H. Hecht was at the Department of Biochemistry, Duke University, Durham, NC 27710. J. S. Richardson and D. C. Richardson are with the Department of Biochemistry, Duke University, Durham, NC 27710. R. C. Ogden is at Agouron Pharmaceuticals, 11025 North Torrey Pines Road, La Jolla, CA 92037.

*Present address: Department of Chemistry, Princeton University, Princeton, NJ 08544.
†To whom correspondence should be addressed.

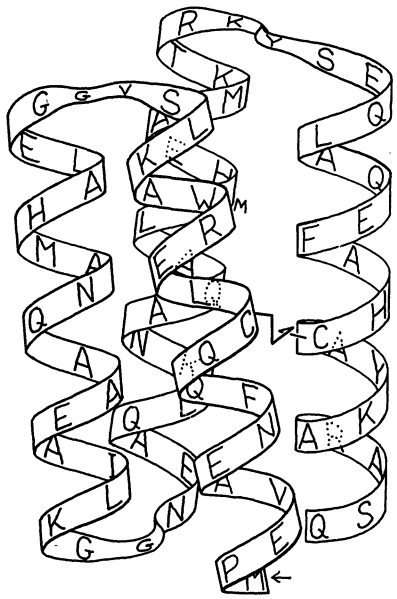

Fig. 1. Ribbon drawing of the sequence and proposed three-dimensional structure of Felix, with the disulfide indicated. The chain turns left at the top of the first helix, making this a "left-turning", up-and-down, four-helix bundle. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

as a direct experimental strategy for answering them.

De novo design has so far been tried only a few times, but by means of quite diverse strategies (*1*). Gutte synthesized a 24-residue β-sheet peptide that binds the insecticide DDT (*2*). Eisenberg, DeGrado, and colleagues designed a modular bundle of four identical α-helices, each composed of all leucine residues on the inner face and all glutamic acid and lysine residues on the outer (*3*). After some redesign, the helices were linked by identical turns and the entire sequence was expressed in vivo; the resulting protein, called "α-4", is very stable and circular dichroism (CD) indicates a high helix content (*4, 5*). DeGrado, Wasserman, and Lear have also designed membrane channel peptides with repeating sequences of leucines and serines; they show conductance properties much like those of natural ion channels (*6*). Mutter and colleagues have circumvented the protein folding problem by making "template-assisted synthetic proteins" (TASP's) in which several amphiphilic α helices or β strands are covalently linked to an artificial template (*7*).

Our own strategy for protein design is to use natural structural motifs and to design sequences that are native-like in pattern and composition, are locally nonrepetitive, and are not "homologous" to any known protein. Two of these designed proteins have been synthesized and partially characterized. The first, betabellin (*8*), was designed to form a β sandwich consisting of two identical four-stranded antiparallel β sheets. Over several cycles of peptide synthesis and redesign, it has been improved in yield, in spectroscopic properties, and especially in solubility. It shows spectra classic of β structure by both CD and laser Raman, and forms the designed cross-sheet disulfide (*1, 9*). The other is Felix (for *four helix*), a 79-residue sequence designed to fold into an up-and-down four-helix bundle (*1, 10*). In this article, we describe the design, expression in vivo, purification, and physical characterization of Felix.

Choice of a tertiary-structure motif. To increase the chances of the design succeeding, the structural motif chosen should be fairly small and simple. At the same time, if our goal is to study and control tertiary structure, the motif should be large enough to have a true tertiary structure, comprising multiple elements of secondary structure that interact with each other around a sizable, self-contained hydrophobic core. The simplest and also the most common α-helical tertiary structure is the up-and-down four-helix bundle (*11, 12*). Examples with coordinate files available from the Brookhaven Data Bank (*13*) include myohemerythrin, hemerythrin, cytochrome b562, cytochrome b5, cytochrome c′, uteroglobin, and tobacco mosaic virus protein (*14*). Felix is intended as a simple prototypical four-helix bundle (Fig. 1). Criteria for size were that the helices each be about five turns long and the connections be as short as practical; this requires a sequence of close to 80 residues.

Even with a pure up-and-down topology where each helix is connected to its nearest neighbor by a short connection, two variants are possible: If the amino-terminal helix goes up the bundle, then the second helix can be either to its right or to its left (as viewed from the outside). We call the first version a "right-turning" bundle and the second a "left-turning". A right-turning bundle contains a right turn, a left turn, and then a right turn; whereas the turns in a left-turning bundle are left, right, and left. Although right-turning bundles are much more common, left-turning bundles do occur, for example, cytochrome b5, ROP protein (*15*), and interleukin (*16*). We chose to design Felix as a left-turning bundle (Fig. 1) since its success would then be a stronger argument for control of the design. The two possible bundle topologies can be distinguished from one another by formation of an intramolecular rather than an intermolecular disulfide bond in our particular sequence, as explained below.

General sequence criteria. The sequence of Felix was designed by students and faculty in a graduate seminar in 1985 (*17*); it was designed only for three-dimensional structure, and not for function. The sequence was to be "native-like", but not homologous to any known protein, and should have naturally occurring amino acids and an NH_2-terminal methionine to allow expression in vivo. Since our aim is to understand the determinants of tertiary structure in natural proteins, we chose to work with linear, nonrepeating sequences. We even avoided duplicate tripeptides to facilitate future nuclear magnetic resonance (NMR) assignments. In addition to these general principles, the design incorporated many specific, structurally based constraints, as described below.

Four different helices were designed. Since each helix was to be at the surface of a soluble protein and in contact with two other helices, approximately half of its surface would be hydrophobic and half hydrophilic. The Felix helices are therefore amphiphilic; however, they have only moderate hydrophobic moments similar to those of helices in soluble, globular proteins rather than the extreme moments typical of helices that bind to the surface of membranes (*18*). This was accomplished with the use of approximately one-quarter charged residues, one-quarter uncharged hydrophilic residues, one-quarter large hydrophobics residues, and one-quarter alanine (Fig. 1, or 5).

Use of helix-preference statistics. In choosing the specific amino acids for the helices, preference was of course given to strong helix-formers. As a check that the helices would be predicted as helices, each proposed version of the sequence was analyzed by the secondary-structure prediction programs of Chou and Fasman and of Garnier *et al.* (*19*). Later the final consensus sequence was subjected to other prediction programs (*20*). In all cases the helices scored well, and the turns between helices were almost always strong and in the right places. The poorest-predicting amino acids in Felix are the two half-cystines of the disulfide.

At each specific location along a helix we used residues that are most frequently found at that particular location in natural helices. We identified these helical propensities by statistically analyzing the frequency at which each amino acid occurs at each particular location. The individual positions were named as follows:

N″N′ N-cap N1 N2 N3 N4 N5 .. middle .. C5 C4 C3 C2 C1 C-cap C′C″

(N1 through C1: helix)

At the time of the initial design, statistics were collected for ~100 helices from known protein structures and the residues were

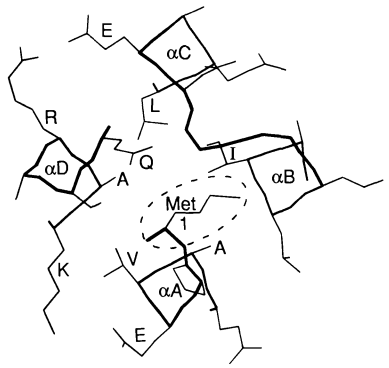

Fig. 2. A portion of a computer model of the designed three-dimensional structure of Felix. The four-helix cluster is viewed end-on, with Met 1 circled. Side chains are also labeled which make hydrophobic contacts or helix-helix salt links. *Methods*: We have built three such models so far, each with slightly different helix packing geometry and connection conformation. The modeling was done with SYBYL, from Tripos Associates, running on an Evans & Sutherland PS390 and a MicroVax II at the Duke Cancer Center Shared Resource for Macromolecular Graphics, with the help of Thomas P. Quinn.

grouped in sets of 2 or 3 (*21*). Later (*22*) the sample was increased to 215 helices and the preferences were made specific to each individual position. Those later statistics were not available at the time of the Felix design, and they would have dictated different choices in a few cases. Also not available then was the analysis of hydrogen-bond groupings at helix ends (*23*). The helix termini in Felix are fair but not outstanding by that criterion.

The N-cap residues (those that are half in and half out of a helix at its NH_2-terminus) in Felix are Met^1, Gly^{22}, Asn^{40}, and Ser^{61}. Glycine, asparagine, and serine are three of the most favored N-cap residues in natural helices, whereas methionine is found there only occasionally. Nonetheless, Met^1 was included in the sequence because it is required for expression in vivo, and indeed natural proteins also often use weaker caps at chain termini. The four C-caps in Felix are Ser^{19}, Gly^{38}, Arg^{58}, and Gln^{79}. Gly is the most favored C-cap residue, whereas the others are only moderately good C-caps. Gln^{79} is at the chain terminus, while the other two (Ser^{19} and Arg^{58}) were chosen to form long-range interactions between the first and third interhelical turns.

In the N1 positions (which immediately follows the N-cap) there are three Glu and one Pro residues. These are the most frequently observed at this position in the helices of natural proteins (Glu is three times more common than Asp at N1). In the C1 position, Lys^{37} and Ser^{78} are excellent, while Leu^{18} and Thr^{57} are only moderate. Thr^{57} is a very reasonable compromise because it makes favorable tertiary interactions in the model, whereas Leu^{18} was probably a mistake and will be changed in future versions of Felix.

Through the middle sections of the helices, those residues with low preference values of 0.3 to 0.6 (Pro, Gly, and Ser) were avoided; those with moderate values of 0.7 to 1.0 (for example, Thr, Asn, Tyr, and Cys) were included only occasionally; whereas those with values well above 1.0 (for example, Ala and Gln) were used extensively. The large hydrophobics were included in order of average preference values for their favored positions N4, C4, and C3 (Leu>Met>Phe>Val,Ile), but were located in the interior of the Felix model. Although tryptophan has a high preference value for the middle turns of helices, it was used only once (Trp^{15}) because its large ring is difficult to fit in the model and also because we wanted an unambiguous fluorescent probe.

Side-chain interactions. As a means of ensuring solubility, the 79 amino-acid sequence of Felix contained 23 charges including the termini and three histidines. Almost all of the charges were positioned so that they can make salt links either between neighboring helices (for example, Glu^3–Lys^{74}, Arg^{14}–Glu^{67}, Glu^{23}–Arg^{58}), or between positions i and $i + 4$ (*24*) within a helix (for example, Glu^{13}–Lys^{17}, Glu^{33}–Lys^{37}, and Glu^{41}–Lys^{45}). The charges are clustered at the helix ends: N′ to N6 and C6 to C′. Each end has three charges on average, with a net negative charge near each NH_2-terminus and a net positive charge near each COOH-terminus for favorable interaction with the helix dipole (*25*).

The helix-helix connections not only influence where the helices end, but also how the bundle packs together. They were made as short as is frequently observed in natural proteins, with either one or two residues between C-cap and N-cap. The right-turning αB-αC connection was modeled on the turns between helices 1 and 2 in myohemerythrin and hemerythrin. The two left-turning connections are the longer ones, in accordance with Efimov's observations that right-turning connections can be shorter than left-turning connections (*26*). The αA-αB and αC-αD turns were designed to form hydrogen bonds (Glu^{23}–Arg^{58}) and hydrophobic interactions with one another and also to interact with the helices that they connect (Tyr^{60} tucks down into the hydrophobic core, analogous to many observed connections). Since the αB–αC turn is only a single glycine, most of the long-range interactions at that end of the bundle are provided by the chain termini.

The internal packing of the entire four-helix bundle is presumably one of the most critical factors for a successful design. However, it is also one of the most difficult to model. We first set the backbone conformations by positioning the helices at distances (9 to 10 Å) and angles (+10° to +15°) commonly found among natural four-helix bundles (*12*). Side-chain conformations were then modeled with one of the χ_1 values preferred for that residue in natural α helices (*27*); those angle distributions are surprisingly narrow (*28*). The large hydrophobic side chains (Met, Phe, Leu, Trp) were positioned to pack against one another, and also to contact the main chain of an adjacent helix. This is in analogy to what we have observed with small-probe dot surfaces to analyze the packing of natural helix bundles (*27*). At areas of close helix-helix contact, as observed in real proteins, long side chains were turned out to the side of the helix pairs and did not pass between the helices.

Space-filling models of Felix were constructed, and the sequence was then modified to remove lumps or fill holes. This is still easier to do with physical models than on the computer. A rough computer

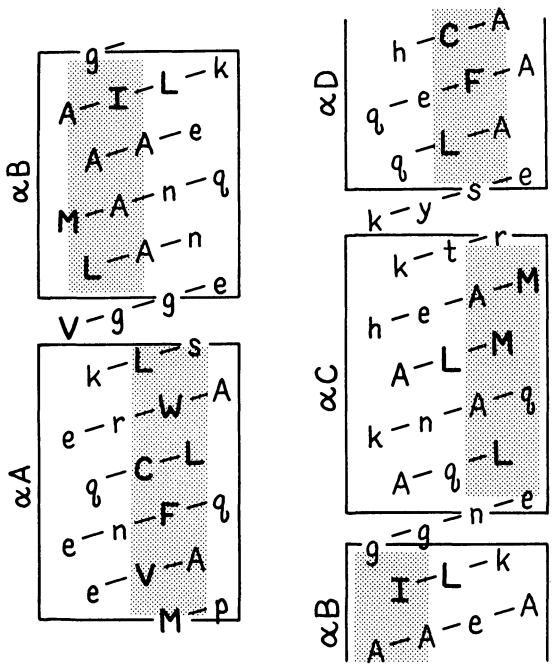

Fig. 3. A helical net with 3.5-residue pitch, continuous through the helices of Felix. The hydrophobic stripe on each helix is indicated, showing how they are offset across the designed helix connection to prevent formation of a single coiled-coil helix.

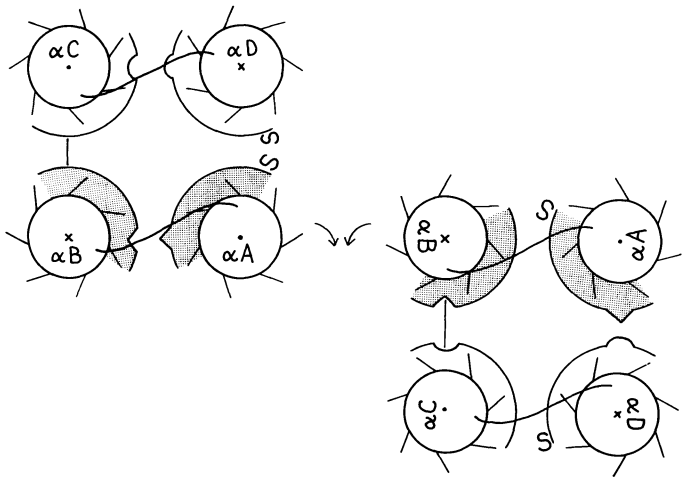

Fig. 4. The transformation between a left-turning and a right-turning helix bundle topology, with the αA-αB helix pair as a reference. Helices are viewed end-on, with a central dot for one coming up out of the page and an X for one going away. Some of the pinwheeling Cα–Cβ vectors are indicated. The hydrophobic half of each helix surface (stippled) contains two possible sites for contact with another helix. On the left, helices αA and αB interact with one pair of contacts (shown as triangles), leaving the other pair of contacts (shown smooth) available on their top surface for interacting with helices C and D; this produces a left-turning bundle topology. On the right, the two reference helices have rotated inward like meshed gears, bringing the smooth pair of contacts together and now leaving an open hydrophobic face on their lower surface against which helices C and D pack; this arrangement produces a right-turning topology.

model was built with 31 of our designated side-chain interactions as distance constraints for a dynamics calculation with CHARMM (*29*). It showed that the original αA-αB connection could not be formed with favorable ϕ,ψ angles, necessitating the addition of one residue and some sequence change in that turn. Recently, we used SYBYL to build three additional computer models (see Fig. 2), which reproduce the designed conformation and interactions fairly well. Model coordinates are available as Brookhaven Data Bank file 1FLX.

This design process resulted in a sequence whose composition and pattern compared well with natural helical proteins (except, perhaps, for an unusually high content of Gln). The sequence was checked against the PIR sequence data base (*30*), and no significant homology was found. No tripeptide was repeated, and 19 of the 20 natural amino acids are represented. An initial sequence of Felix included all 20 amino acids, but the single Asp N-cap was changed when the connection was redesigned.

Negative design. In designing (or predicting) a protein structure, it is not sufficient to show that the given sequence is compatible with a particular structure; we must also ensure that it is less compatible with alternative structures. We think of this process of designing against alternative structures as "negative design". Another intrinsic part of the negative design process is including features that allow the alternatives to be distinguished experimentally.

For secondary structure, we might expect that using only strong helix formers and good termination sequences would at least prevent formation of β strands. However, it has been shown that synthetic copolymers composed only of lysine and leucine (which have high helical propensities) form α-helices when their sequence is random, but form β-structure when the sequence contains alternating hydrophobic and hydrophilic residues (*31*). The sequence of Felix avoids such alternation, and thus β structure would not be expected. Circular dichroism (CD) spectroscopy provides a simple experimental test for such competing secondary structures.

An alternative tertiary structure for Felix might be produced if the helices continued through their designated breaks, giving a hairpin of two double-length helices or even a single very long helix. Since the helices in Felix have broad hydrophobic faces, such molecules would probably aggregate into some sort of coiled coil. We attempted to design against this possibility by two strategies. (i) We used strong helix-breaking sequences in the proposed turns. (ii) We made sure that if the helices, nevertheless, tried to continue through the intended turn sequences—for example, as in (*4*)—then the hydrophobic stripes on the individual helices would not line up with each other along the putative long helix (Fig. 3). Such an offset of hydrophobic stripes is seen for helical hairpins in the natural proteins and may, coupled with loop conformation, help to determine the length of α-α connections. For Felix, then, the hydrophobic stripes were offset by a one-third to one-half helical turn across each connection as shown on the helical net in Fig. 3, except for αC-αD, whose offset is only marginal. If extended or oligomeric "wrong" structures did form, they could be detected experimentally by an increase in the apparent molecular mass in size-exclusion chromatography.

Not all antiparallel four-helix bundles have an up-and-down topology; growth hormone (*32*) and ferretin (*33*) contain long crossover connections. We designed against such structures with the use of well-defined helix termini and short interhelical connections.

Another alternative tertiary structure would be a right-turning rather than a left-turning helix bundle (Fig. 4). Our major strategy for avoiding the alternative topology was to arrange the hydrophobic faces of the four helices so that they would be complementary to one another and pack tightly in the desired left-turning topology, but would fit badly in the right-turning arrangement. Since there may also be some influence of connection lengths on preferred topology (*26*), Felix was given the long, short, long connection pattern that should work best in a left-turning bundle.

The two bundle topologies can be distinguished experimentally by the formation of an intramolecular disulfide bond. Disulfides linking the centers of α-helices are highly constrained and rarely observed; they are found only in crambin (1CRN) (*34*) and phospholipase A_2 (3BP2), (*35*) among natural proteins, and in the disulfide dimer link designed into λ repressor (*36*). Only one position in the Felix model has the correct geometry, namely, connecting the centers of helices A and D, by means of residues 11 and 71. These side chains would point toward one another in the desired left-turning topology, but would be widely separated and point outward in the alternative structure (Fig. 4). Thus, ready formation of the Cys^{11}–Cys^{71} intramolecular disulfide bond rather than SS-linked multimers would provide experimental evidence for the left-turning topology.

Gene design and expression. The final amino acid sequence of Felix is shown in Fig. 5, along with the synthetic gene designed to direct the expression of Felix in *Escherichia coli*. To encourage high levels of expression, the gene was composed primarily of codons used by *E. coli* to express its most abundant proteins (*37*). Each strand can be divided into six moderate-length pieces which fit together with specific nine-base overlapping ends. Eleven different restriction sites were included in the DNA sequence in order to facilitate the future construction of sequence variants by cassette replacement.

The Felix gene was synthesized in 12 overlapping pieces, each between 34 and 48 nucleotides long, with the use of cyanoethyl phosphoramidite chemistry. The oligonucleotides were ligated together, cloned into M13mp8, and sequenced (Fig. 5).

Our initial attempts to express Felix in *E. coli* were not successful. Although the synthetic gene was efficiently transcribed after being subcloned in several different expression vectors, Felix protein could not be detected and apparently underwent proteolysis in vivo. Using various expression systems, we tried to circumvent proteolysis. Our

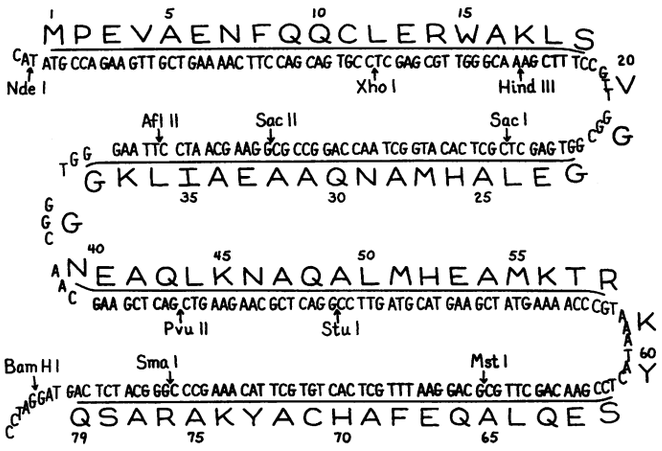

Fig. 5. The amino acid sequence of Felix and DNA sequence of its gene. The restriction sites are indicated. Oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer with cyanoethyl phosphoramidite chemistry. The oligonucleotides were purified by high-performance liquid chromatography (HPLC), phosphorylated on their 5′ termini, annealed, mixed together, ligated, and cloned into pBR322. Clones were screened by restriction analysis and by hybridization with the original oligonucleotides, which had been labeled with ^{32}P. The correct sequence was verified by sequencing both strands of DNA.

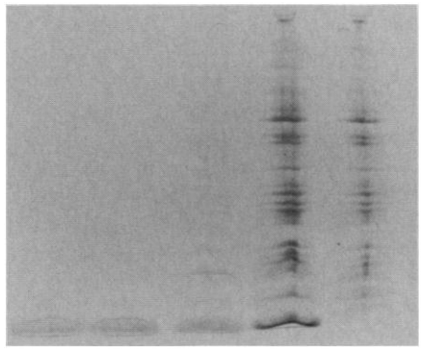

Fig. 6. SDS gel electrophoresis showing the expression and purification of Felix. Lanes 1 and 2 (at right) show whole cell lysates before and 3 hours after IPTG induction respectively. Lane 3 shows insoluble material after cell lysis, lane 4 shows Felix after purification by size exclusion chromatography in 6.0 M urea, and lane 5 shows the same material after dialysis into native buffer. *E. coli* strain BL21(DE3)(pLysS) (*38*) harboring plasmid pET3a-Felix was grown in rich medium in the presence of chloramphenicol (25 μg/ml) and ampicillin (100 μg/ml). When the culture reached an absorbance of 1.0 at 600 nm, IPTG was added to 150 μg/ml, and the culture was grown for 3 more hours. Cells were centrifuged, washed in 50 mM tris-HCl, pH 8, 200 mM NaCl, 2 mM EDTA, and lysed in a French Pressure cell. The lysate was centrifuged and Felix was found almost exclusively in the pellet (lane 3). The pellet was suspended in 10 mM tris-HCl, pH 8.0, 2 mM EDTA, 5 percent sucrose, and layered over a sucrose step gradient. After centrifugation overnight at 70,000*g*, the purified inclusion bodies were recovered from a layer that sedimented between 50 and 55 percent sucrose. The inclusion bodies were then solubilized in 50 mM tris-HCl, pH 8.0, 200 mM Glycine, 1 mM EDTA, 6.0 mM urea, 50 mM DTT. This material was placed onto a Superose 12 (Pharmacia) size exclusion column and chromatographed in 50 mM tris-HCl, pH 8.0, 200 mM Glycine, 1 mM EDTA, 6.0 M urea, 5 mM DTT. Pure Felix was recovered from this column (lane 4). Urea was removed by dialysis into 10 mM tris-HCl, pH 8.0, 1 mM DTT, 0.1 mM EDTA; the protein remained in solution (lane 5).

aim was either to overwhelm cellular proteases by using strong *Escherichia coli* promoters, to remove Felix from intracellular proteases by secretion, or to stabilize Felix by synthesizing it as part of a fusion protein. Even in these systems, purification of intact Felix was difficult.

The T7 system (*38*) made it possible to purify large quantities of Felix. In this system, Felix was cloned downstream of a phage T7 promotor, while the T7 RNA polymerase (which recognizes this promoter uniquely) was expressed under *lac* control. When the T7

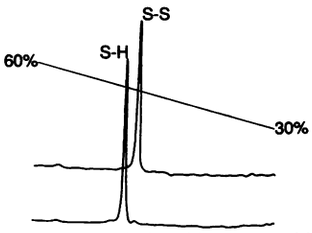

Fig. 7. Analytical reversed-phase chromatographs showing the conversion of Felix S-H to Felix S-S, which contains the Cys^{11}-Cys^{71} intramolecular disulfide bond. Felix S-H was maintained in the reduced state by an excess of DTT. This material, at a protein concentration of less than 5 μM, was dialyzed into a buffer containing 10 mM tris-HCl, pH 8.0, 50 mM NaCl. After dialysis, oxidized glutathione was added to 0.2 mM, and disulfide exchange was allowed to proceed overnight at 2°C. The pH was then adjusted to 5.3 by the addition of 50 mM sodium acetate, and Felix S-S was concentrated with the simultaneous removal of glutathione by binding it to an S-Sepharose Fast Flow (Pharmacia) cation exchange resin, and eluting it with 50 mM sodium phosphate, pH 7.6, 100 mM NaCl. The resulting material was analyzed by reversed-phase chromatography on a PEP/RPC column (Pharmacia) in which solvent A is 0.2 percent trifluoroacetic acid (TFA) in water, solvent B is 0.2 percent TFA in acetonitrile, and a gradient is run from 30 percent to 60 percent B. Felix S-S eluted at 46 percent B, while Felix S-H eluted at 48 percent B.

RNA polymerase was induced with isopropyl-β-D-thiogalactopyranoside (IPTG), Felix accumulated as a major component of total cell protein (Fig. 6). Under these conditions, Felix was sequestered in insoluble inclusion bodies, and presumably for this reason it escaped intracellular proteolysis.

Purification and disulfide bond formation. Purification was facilitated by the fact that Felix was already more than 50 percent pure in the insoluble material recovered from cell lysis (Fig. 6, lane 3). Inclusion bodies were separated from cell debris by centrifugation of the suspended material through a sucrose gradient. The purified inclusion bodies were then solubilized in 6.0 M urea in the presence of excess DTT. Felix was then purified by size exclusion chromatography in the presence of urea and DTT. The purified protein was allowed to fold by dialyzing it from urea into a native buffer. Following dialysis, Felix remained in solution, suggesting that it had folded into a globular structure in which the hydrophobic side chains are buried and thereby prevented from causing intermolecular aggregation and precipitation.

The purified protein was examined for amino acid composition and by 11 steps of automated NH_2-terminal sequence analysis. These showed the expected composition and sequence for Felix, with the exception that the NH_2-terminal methionine had been removed quantitatively in vivo. We had not expected cleavage of the Met^1-Pro^2 peptide bond, and had designed the Met side chain to provide stabilizing interactions, which are now missing. There is no evidence of either sequence heterogeneity or impurities.

In designing Felix, we included an intramolecular disulfide bond connecting Cys^{11} with Cys^{71} for four reasons. (i) This covalent link should stabilize the folded structure of the protein; (ii) the disulfide would provide a quick and simple probe of tertiary structure by showing that two regions distant in sequence are in fact close to one another; (iii) if such a highly constrained SS can form without lowering helix content, then the local structure near it is likely to be as modeled; (iv) as described earlier, the disulfide can distinguish between a left-turning and a right-turning helix bundle by forming either monomers or aggregates.

The Cys^{11}–Cys^{71} disulfide bond was formed in the purified protein by disulfide exchange with oxidized glutathione. Formation of the disulfide-containing species (Felix S-S) was detected by the appearance of a new peak in reversed-phase chromatography (Fig. 7) and its stoichiometric reconversion to the starting material by the addition of DTT. In that the same peak was observed after alternative methods of oxidation (including ferricyanide oxidation and Cu^{2+}-catalyzed air oxidation), it does not represent material containing a mixed disulfide between Felix and glutathione.

Monomeric state. Formation of a disulfide bond in Felix might have yielded either the desired Cys^{11}–Cys^{71} intramolecular bond or one of various intermolecular disulfide bonds, which would produce dimers or higher order oligomers. Although multimeric forms can be detected, even at high protein concentrations they are minor peaks and are completely absent in Fig. 7, where Felix S-S is a single species. The monomeric state of Felix S-S in native buffer was demonstrated by size exclusion chromatography with compact globular proteins as calibration standards (see Fig. 8). By this technique, the apparent molecular mass of Felix (9600 daltons) is in close agreement with the molecular mass of a monomer calculated from the amino acid sequence (8700 daltons). Since covalent multimers, aggregated multimers, or extended conformations would all show significantly larger apparent molecular mass, we conclude that Felix is a monomer in solution, and furthermore, that it folds into a compact globular structure that contains the desired intramolecular disulfide bond.

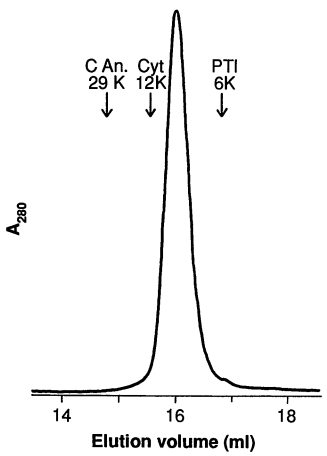

Fig. 8. Size exclusion chromatography indicating that Felix is a monomer under nondenaturing conditions. Felix S-S, at an initial concentration of 0.2 mg/ml, was chromatographed on a prepacked Superose-12 gel filtration column (Pharmacia) in 50 mM sodium phosphate, pH 7.6, 100 mM NaCl. Molecular size markers are carbonic anhydrase (C An), cytochrome c (Cyt), and bovine pancreatic trypsin inhibitor (PTI).

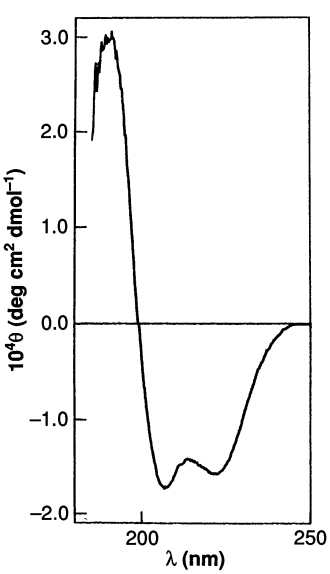

Fig. 9. Circular dichroism spectrum of Felix S-S. Felix S-S was dissolved at 0.09 mg/ml (~10 μM) in 50 mM sodium phosphate, pH 7.6 in a cuvette with a 1-mm path length. Spectra were measured at 2°C with a Jobin Yvon Mark 5 spectropolarimeter interfaced with an Apple computer. The data were collected in triplicate from 260 nm to 185 nm with a step size of 0.2 nm, and a response time of 2 seconds.

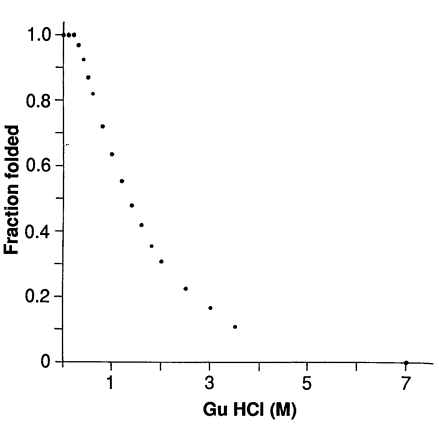

Fig. 10. Denaturation of Felix S-S. The fraction of protein in the native state (calculated as the ratio between the observed mean residue ellipticity at 222 nm at a given concentration of denaturant relative to the observed value in the absence of denaturant) versus the concentration of Gu-HCl. The procedures were as described for Fig. 9, except that the denaturation was performed at 20°C.

Spectroscopic probes of structure. The secondary structure of proteins is readily probed by CD. This technique is particularly diagnostic for α-helical proteins, which have spectra that include a maximum at 190 nm, a crossover at 200 nm, and negative minima at 208 nm and 222 nm. The CD spectrum of Felix S-S (Fig. 9) displays these α-helical features. The magnitudes of the characteristic minima suggest that Felix contains 50 to 65 percent α-helical secondary structure (*39, 40*). That is less than would be predicted by our model, but is similar to or only slightly smaller than the fraction observed for natural four-helix bundles such as growth hormone (*41*) or cytochrome b_{562} (*42*) or for the designed bundle α-4 (*5*). The helix content of Felix is slightly increased by the formation of the Cys^{11}–Cys^{71} disulfide bond (compared to data that we have obtained for Felix SH), thus indicating that this bond was able to form without significant distortion of the α-helices.

The stability of Felix S-S toward denaturation by guanidine hydrochloride (Gu-HCl) was measured by monitoring the mean residue ellipticity at 222 nm as a function of denaturant concentration (Fig. 10). The midpoint of the unfolding transition was approximately 1.3 M Gu-HCl.

A single tryptophan was included in the Felix sequence as a convenient spectroscopic probe of structure, since Trp fluorescence is sensitive to the environment of the indole side chain (*43*). The fluorescence maxima for Trp in folded proteins (where they are buried in apolar environments) are shifted to shorter wavelengths relative to the same proteins that have been denatured by guanidine hydrochloride. Comparison of the fluorescence spectra of Felix S−S in native buffer and in 7.0 M Gu-HCl shows that the emission maximum in the folded protein is suitably blue-shifted (Fig. 11), thus indicating the Trp^{15} is buried in the folded structure of Felix. The emission maximum for the reduced form of the protein (Felix S-H) does not show a similar blue shift, indicating that in the absence of the disulfide bond Trp^{15} is accessible to solvent.

The intensity of the fluorescence in the folded protein was diminished significantly relative to the unfolded state (Fig. 11). This is in sharp contrast to most proteins, where the fluorescence intensity of Trp is severalfold greater in the folded form. Thus it appears that the fluorescence of Trp^{15} in Felix is being quenched. The most likely candidate for an apolar quencher is the disulfide bond linking Cys^{11} with Cys^{71}. Since fluorescence quenching by disulfide bonds requires close contact (*44*), these results support the designed model of Felix in which the i to $i + 4$ separation between Cys^{11} and Trp^{15} places these side chains on the same side of an α-helix.

Assessment of protein design. In summary, we have described the design, expression, purification, and initial physical characterization of Felix. Our results indicate that (i) Felix is a compact monomer in solution; (ii) Felix has a predominantly α-helical secondary structure; (iii) the tertiary structure of Felix accommodates an intramolecular disulfide bond that brings together regions of the chain distant from one another in the primary sequence; this bond is in a highly constrained location between the centers of two α-helices, but nevertheless forms with no loss of helicity; and (iv) the single tryptophan is buried in an apolar environment and is probably close to the disulfide bond.

Taken together, the above results suggest that the structure of Felix in solution is a compact, left-turning, four-helix bundle.

Two other α-helical tertiary structures would be compatible with

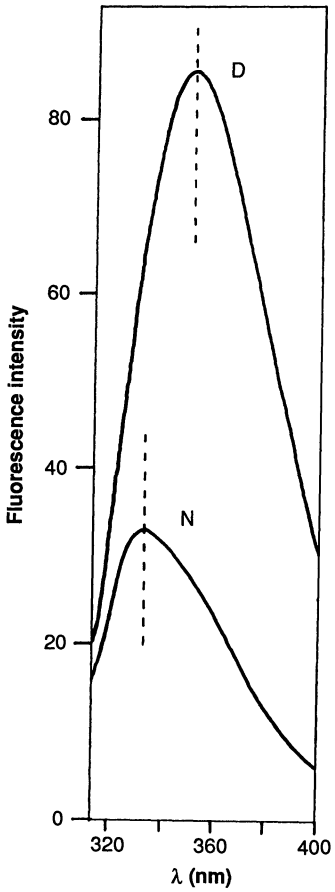

Fig. 11. Fluorescence spectra of Felix in native (N) and denaturing (D) buffers, showing the blue shift and lower intensity for the folded protein. Felix S-S was dissolved at 0.06 mg/ml in 50 mM sodium phosphate, pH 7.6 in the presence or absence of 7.0 M Gu-HCl. Spectra were measured at 18°C on a Shimadzu RF-540 spectrofluorophotometer. The excitation wavelength was 295 nm.

the observed CD spectrum of Felix but can be ruled out by other considerations. A coiled-coil hairpin (two long helices rather than four short ones) would expose a long hydrophobic face and would be expected to dimerize; however, Felix is monomeric and compact. A four-helix bundle with a right-turning rather than the designed left-turning topology would position Cys^{11} and Cys^{71} pointing away from one another; however, in Felix these residues form the designed intramolecular disulfide bond. Therefore, by these multiple criteria, the Felix sequence has folded into approximately the structure designed for it.

More detailed information is needed to see how well the helix packing geometry, helix end locations, and loop conformations match the designed model. It is already clear that Felix is at the low end of stabilities and cooperativities observed for natural proteins. Although Felix maintains a well-folded α-helical structure under native conditions, measurements of its CD spectrum as a function of Gu-HCl concentration show a rather broad denaturation curve with a midpoint of approximately 1.3 M Gu-HCl (Fig. 10). An extrapolation of ΔG_{obs} as a function of Gu-HCl concentration (*45*) suggests that the free energy stabilizing the folded structure in the absence of denaturant is approximately 0.8 kcal/mol; natural proteins typically are stable by 5 to 15 kcal/mol (*46*). The marginal stability of Felix suggests that some of the interior side chains may not have only one position and may actually be fairly mobile, as has been hypothesized for the "molten globule" state (*47*).

Remarkably, Felix and the other design projects surveyed above all show evidence of producing something close to the designed structure in spite of the use of quite different methodologies and a wide variety of philosophical approaches to choice of sequence. Comparing the various design projects leads to several general conclusions. First, using all leucine for helix contacts is a remarkably effective way to produce very stable structures, as shown by α_4 (*4, 5*) and by recent "leucine zipper" peptides (*48*). It will be very interesting to determine why natural proteins typically use a less-stabilizing mix of hydrophobic residues. Second, helices are probably easier to design than β sheets. Felix has achieved on the first try the same degree of success that betabellin, our designed β-sandwich protein, achieved after nine syntheses incorporating four major changes (*1, 8, 9*). This difference is probably due to the greater modularity of helical structures and the solubility problems inherent in designing β sheets.

In all these projects, the full three-dimensional structures will doubtless reveal some surprises, and the most reliable and specific new information will be obtained only after several iterative cycles of design, structure determination, and redesign. Nonetheless, our present highly encouraging results have already demonstrated that the design of proteins de novo is indeed a feasible enterprise.

REFERENCES AND NOTES

1. J. S. Richardson and D. C. Richardson, *Trends Biochem. Sci.* **14**, 304 (1989).
2. R. Moser, R. M. Thomas, B. and Gutte, *FEBS Lett.* **157**, 247 (1983); R. Moser *et al.*, *Protein Eng.* **1**, 339 (1987).
3. D. Eisenberg *et al.*, *Proteins* **1**, 16 (1986).
4. S. P. Ho and W. F. DeGrado, *J. Am. Chem. Soc.* **109**, 6751 (1987).
5. L. Regan and W. F. DeGrado, *Science* **241**, 976 (1988).
6. J. D. Lear, Z. R. Wasserman, W. F. DeGrado, *Science* **240**, 1177 (1988).
7. M. Mutter, *Trends Biochem. Sci.* **13**, 260 (1988); M. Mutter *et al.*, *Proteins* **5**, 13 (1989).
8. C. G. Unson, B. W. Erickson, D. C. Richardson, J. S. Richardson, *Fed. Proc.* **43**, 1837 (1984); J. S. Richardson, D. C. Richardson, B. W. Erickson, *Biophys. J.* **45**, 25a (1984).
9. J. S. Richardson and D. C. Richardson, in *Protein Engineering*, D. L. Oxender and C. F. Fox, Eds. (Liss, New York, 1987), pp. 149–163; S. B. Daniels, R. W. Williams, J. S. Richardson, D. C. Richardson, B. W. Erickson, *FASEB J.* **2**, A1543 (1988).
10. M. H. Hecht, D. C. Richardson, J. S. Richardson, R. C. Ogden, *J. Cell. Biochem.* **13A**, 86 (1989).
11. P. Argos, M. Rossmann, J. Johnson, *Biochem. Biophys. Res. Commun.* **75**, 83 (1977); J. S. Richardson, *Adv. Protein Chem.* **34**, 167 (1981).
12. P. C. Weber *et al.*, *Nature* **286**, 302 (1980).
13. F. Bernstein *et al.*, *J. Mol. Biol.* **112**, 535 (1977).
14. Myohemerythrin (file 2MHR): W. A. Hendrickson and K. B. Ward, *J. Biol. Chem.* **252**, 3012 (1977); hemerythrin (file 1HMQ): R. E. Stenkamp, L. C. Sieker, L. H. Jensen, J. E. McQueen, Jr., *Biochemistry* **17**, 2499 (1978); cytochrome b562 (file 156B): F. S. Mathews, P. H. Bethge, E. W. Czerwinski, *J. Biol. Chem.* **254**, 1699 (1979); cytochrome b5 (file 3B5C): F. Mathews, P. Argos, M. Levine, *Cold Spring Harbor Symp. Quant. Biol.* **36**, 387 (1972); cytochrome c′ (file 2CCY): B. Finzel, P. Weber, K. Hardman, F. R. Salemme, *J. Mol. Biol.* **186**, 627 (1985); uteroglobin (file 1UTG): I. Morize *et al.*, *J. Mol. Biol.* **194**, 725 (1987); tobacco mosaic virus protein (file 2TMV): K. Namba, R. Pattanayek, G. Stubbs, *J. Mol. Biol.* **208**, 307 (1989).
15. D. Banner, M. Kokkinidis, D. Tsernoglou, *J. Mol. Biol.* **196**, 97 (1987).
16. B. J. Brandhuber, T. Boone, W. C. Kenney, D. B. McKay, *Science* **238**, 1707 (1987).
17. Biochemistry 222 at Duke University, spring 1985, was taught by D.C.R., J.S.R. and R.C.O. Students were Paul Garner, Ali Khorrami, Diana Pitterle, Kim Schandel, and Susan Slezinska Gardlik.
18. D. Eisenberg, R. M. Weiss, T. C. Terwilliger, *Nature* **299**, 371 (1985); D. Eisenberg, R. M. Weiss, T. C. Terwilliger, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 140 (1984).
19. P. Y. Chou, and G. D. Fasman, *Adv. Enzymol.* **47**, 45 (1978); J. Garnier, D. J. Osguthorpe, B. Robson, *J. Mol. Biol.* **120**, 97 (1978).
20. We thank A. V. Finkelstein and M. Karplus for testing our sequence with their programs [O. B. Ptitsyn and A. V. Finkelstein, *Biopolymers* **22**, 15 (1983); L. H. Holley and M. Karplus, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 152 (1989)].
21. J. S. Richardson, P. Argos, D. Kneller, D. Osguthorpe, M. Scharf, *EMBL BIOcomputing Technical Document* **1**, 71 (1986).
22. J. S. Richardson and D. C. Richardson, *Science* **240**, 1648 (1988).
23. L. G. Presta and G. D. Rose, *Science* **240**, 1632 (1988).
24. S. Marquesee and R. L. Baldwin, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8898 (1987).
25. K. R. Shoemaker, P. S. Kim, E. J. York, J. M. Stewart, R. L. Baldwin, *Nature* **326**, 563 (1987).
26. A. V. Efimov, *Mol. Biol.* **16**, 271 (1982).
27. J. S. Richardson and D. C. Richardson, in *Prediction of Protein Structure and Protein Conformation*, G. D. Fasman, Ed. (Plenum Press, New York, 1989), pp. 1–98.
28. J. W. Ponder and Richards, *J. Mol. Biol.* **193**, 775 (1987).
29. Done for us by Bernard Brooks, at NIH.
30. Protein Identification Resource, National Biomedical Research Foundation, Georgetown University Medical Center, Washington, D.C. 20007.
31. A. Brack and G. Spach, *J. Amer. Chem. Soc.* **103**, 6319 (1981).
32. S. Abdel-Meguid *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6434 (1987).
33. D. Rice, G. Ford, J. White, J. Smith, P. Harrison, *Adv. Inorg. Biochem.* **5**, 39

(1983).
34. W. Hendrickson and M. Teeter, *Nature* **290**, 107 (1981).
35. B. W. Dijkstra *et al.*, *Biochemistry* **23**, 2759 (1984).
36. R. T. Sauer *et al.*, *ibid.* **25**, 5992 (1986).
37. H. A. de Boer and R. A. Kastelein, in *Maximizing Gene Expression*. W. Reznikoff and L. Gold, Eds. (Butterworth, Stoneham, Mass., 1986), pp. 225–286.
38. F. W. Studier and B. A. Moffatt, *J. Mol. Biol.* **189**, 113 (1986); A. H. Rosenberg *et al.*, *Gene* **56**, 125 (1987).
39. N. Greenfield and G. D. Fasman, *Biochemistry* **8**, 4108 (1969).
40. S. Brahms and J. Brahms, *J. Mol. Biol.* **138**, 149 (1980); W. C. Johnson, Jr., *Annu. Rev. Biophys. Biophys. Chem.* **17**, 145 (1988).
41. L. A. Holladay, R. G. Hammonds, Jr., D. Puett, *Biochemistry* **13**, 1653 (1974).
42. P. A. Bullock and Y. P. Myer, *ibid.* **17**, 3084 (1978).
43. Y. P. Lakowicz, *Principles of Fluorescence Spectroscopy* (Plenum Press, New York, 1983).
44. R. H. Cowgill, *Biochim. Biophys. Acta* **207**, 556 (1970); D. E. Isenman, D. Lancet, I. Pecht, *Biochemistry* **18**, 3327 (1979).
45. C. N. Pace, *Methods Enzymol.* **131**, 266 (1986).
46. C. N. Pace, *CRC Crit. Rev. of Biochem.* **1**, 1 (1975).
47. M. Ohgushi and A. Wada, *FEBS Lett.* **164**, 21 (1983); O. B. Ptitsyn, *J. Protein Chem.* 6, 273 (1987); K. Kuwajima, *Proteins* 6, 87 (1989).
48. W. H. Landschulz, P. F. Johnson, S. L. McKnight, *Science* **240**, 1759 (1988); E. K. O'Shea, R. Rutkowski, P. S. Kim, *Science* **243**, 538 (1989).
49. Supported by the General Medical Sciences of NIH; by the NSF; by Burroughs Wellcome through the Life Sciences Foundation; and by the MacArthur Foundation. We thank L. Musick, T. Quinn, K. Gernert, and T. Vedvick for assistance with specific experimental techniques; N. Tweedy and B. Studier for discussions; J. Swan for scientific editing; C. Usher for help with the manuscript; and the 1985 Biochemistry 222 class at Duke, whose hard work and enthusiasm helped launch the Felix project.

12 December 1989; accepted 18 June 1990

"But on the bright side, we're* much *less likely to have heart disease."