Acta Chemica Scandinavica, 1996: 50: 688–696
Printed in UK – all rights reserved

ACTA
CHEMICA SCANDINAVICA
ISSN 0904–213X

Review Article

Expression of *de novo* Designed α-Helical Bundles†

S. F. Betz,[a] J. W. Bryson,[a] M. C. Passador,[b] R. J. Brown,[a] K. T. O'Neil[a,b] and W. F. DeGrado*[,b]

[a]The DuPont Merck Pharmaceutical Company, PO Box 80328, Wilmington, DE 19880–0328, USA and [b]Johnson Research Foundation, Department of Biochemistry and Biophysics University of Pennsylvania, Philadelphia PA 19104, USA

Betz, S. F., Bryson, J. W., Passador, M. C., Brown, R. J., O'Neil, K. T. and DeGrado, W. F. 1996. Expression of *de novo* Designed α-Helical Bundles. – Acta Chem. Scand. 50: 688–696. © Acta Chemica Scandinavica 1996.

The successful design of proteins requires careful consideration of the multiplicity of forces that stabilize their three-dimensional structures including hydrophobic interactions, hydrogen-bonding, electrostatics and weakly polar interactions. Early attempts to design proteins relied too heavily on hydrophobic interactions to provide stability, resulting in structures with dynamic properties.

Addition of more specific interactions to these initial designs gives rise to proteins with more native-like properties. This manuscript describes the design of native-like three- and four-helix bundles, and their cloning and expression of these proteins.

How proteins fold into their well-defined tertiary structures is an important but unsolved question. Native protein folds are defined by a large number of weak interactions including hydrophobic interactions,[1] hydrogen-bonding,[2,3] electrostatic[4,5] and weakly polar interactions.[6] A full understanding of protein folding will be achieved only by deciphering how these interactions interrelate to produce the specific folds seen in natural proteins.

De novo protein design provides one approach to the elucidation of some of these forces.[7–9] In this approach, one attempts to design a sequence capable of forming a pre-chosen three-dimensional structure. The success or failure of the design then provides a critical test of the underlying concepts and principles guiding the work. Through repeated cycles of design and rigorous experimental characterization of the products it should thus be possible to test and refine our understanding of protein structure and function.

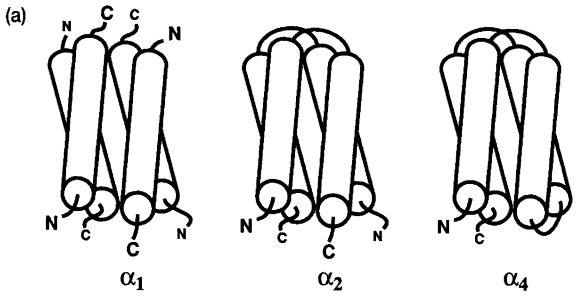
The design of amino acid sequences that adopt a desired three-dimensional fold has been of keen interest over the past decade and a great deal of progress has been made recently.[7–11] In this paper, we review the principles that have emerged from our recent design of proteins, and then discuss the design, cloning and expression of several proteins that have native-like characteristics.

Results and discussion

Non-covalent self-assembly of novel proteins. In our first attempts to design proteins we focussed on four-helix bundles.[12,13] This fold is an attractive target because it has a high degree of pseudo-symmetry and also is found in a large variety of functionally diverse proteins including cytochromes, receptors, hormones and structural proteins.[14] The partial symmetry of the structures suggested that they could be prepared in an iterative fashion[12,13] in which we first designed single helix peptides (designated α_1 peptides) that self-assembled into tetramers, followed by helix–loop–helix peptides (α_2 peptides) that dimerized to form a four-helix bundle, and ultimately single-chain bundles (α_4 proteins) consisting of four helices connected by three loops[15] [Fig. 1(a)]. This iterative strategy allowed us to optimize separately the construction of each element of structure required for the formation of a four-helix bundle: helix formation, helix/helix association and loop formation. A second advantage of this strategy is that it allowed evaluation of initial designs with α_1 or α_2 peptides, which are easily synthesized. In contrast, the preparation of an α_4 protein is more time-consuming, and often required construction

† Contribution presented at the Nobel Symposium on *Catalytic Asymmetric Synthesis*, September 3–7, 1995 at Tammsvik, Bro, Sweden.

* To whom correspondence should be addressed.



(b) **Progression of α_1 family of designed four-helical bundle proteins.**

α_1A: G K L E E L L K K L L E E L K G

α_1B: G E L E E L L K K L K E L L K G

α_2B: G E L E E L L K K L K E L L K G - P R R - G E L E E L L K K L K E L L K G

α_4: G E L E E L L K K L K E L L K G - P R R - G E L E E L L K K L K E L L K G - P R R
-G E L E E L L K K L K E L L K G - P R R - G E L E E L L K K L K E L L K G

α_2C: G E V E E L L K K F K E L W K G - P R R - G E I E E L F K K F K E L I K G

α_2D: G E V E E L E K K F K E L W K G - P R R - G E I E E L H K K F H E L I K G

α_4D: G E V E E L E K K F K E L W K G - P R R - G E I E E L H K K F H E L I K G - P R R
-G E V E E L E K K F K E L W K G - P R R - G E I E E L H K K F H E L I K G

α_4E: G E V E E L E K K F K E L W K G - P R R - G E I E E L H K K F H E L I K N M -
G E V E E L E K K F K E L W K G - P R R - G E I E E L H K K F H E L I K N M

Fig. 1. Incremental design of four-helix bundle proteins: (a) representation of the progression from α_1 (tetramer) to α_2 (dimer) to α_4 (single chain) proteins. Helices are displayed as cylinders, free helix termini are indicated by N or C; (b) sequence progression of the α_1, α_2 and α_4 family of four helix bundle proteins. The sequences are listed to illustrate the hierarchic, iterative design from simple to more complex molecules. The sequence of α_aD is different from earlier reports (9,11) which contain a typographical error.

and expression of a synthetic gene encoding the protein of interest.

Subsequent to the initial reports of the self-assembly of helical bundles from isolated helices a large number of bundles have been prepared and structurally characterized. Examples of assemblies of helical bundles include parallel dimers,[16] trimers,[17] and tetramers,[18] antiparallel dimers,[19,20] trimers,[21] and tetramers,[22] as well as pentamers[23] and hexamers[24] of undefined topology. In addition, the design of dimeric β-sheet proteins has recently been described.[25,26]

A major motivation for creating model proteins is to provide a vehicle for understanding the structures and folding of natural proteins, so it is instructive to consider whether the symmetry of self-assembled helical bundles limits their appropriateness as models for less symmetrical proteins. In fact, many proteins have evolved by a process involving gene duplication, which gives rise to residual symmetry in their structures.[27] The $(\beta\text{–}\alpha)_8$ barrels and calcium-binding proteins (Fig. 2), zinc fingers, and iron sulfur cluster proteins are representative examples of proteins with internal symmetry.[27] A particularly good example is intestinal calcium-binding protein,[28] which contains two homologous Ca^{2+}-binding domains (EF hands), that evolved from a single primordial gene.[29] 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.

It is interesting to note that EF-hands currently exist in a variety of different proteins with disparate functions.[29] In a similar manner we feel that it should be possible to evolve a variety of different functions into the structures of helical bundles.

Hydrophobicity is a strong driving force determining protein structures. Hydrophobicity has long been postulated to be the primary driving force influencing the structure of proteins,[30] but this simple view has been questioned.[31,32] Only through the construction of model proteins with minimial complexity has it been possible to determine conclusively the contribution of this force to folding. In early work,[33] the role of hydrophobic interactions in determining secondary structures was studied through the synthesis of a series of peptides containining only Leu (hydrophobic) and Lys (hydrophilic) in their sequences, but with the residues arranged in different ways. The Leu residues in these peptides were arranged with different periodicities so that the

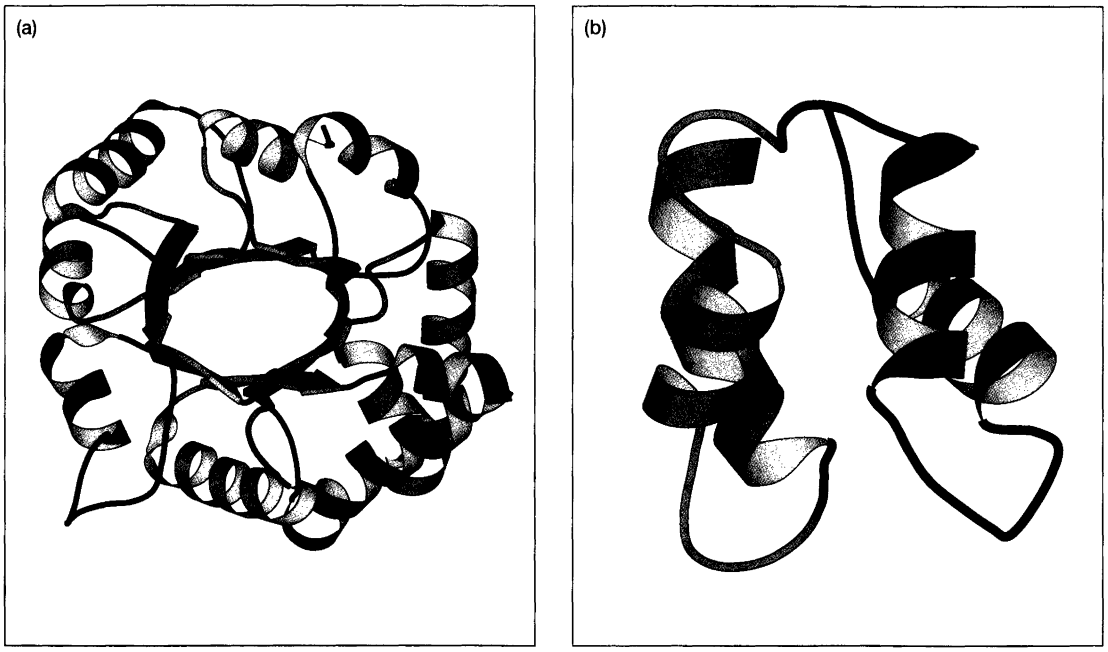


Fig. 2. MOLSCRIPT[60] diagrams of naturally occurring proteins with residual internal symmetry: (a) yeast triose phosphate isomerase (1ypi[61]); (b) calf intestinal calcium-binding protein (4icb[62]). Adjacent regions of internal symmetry are distinguished by alternating light and dark shading.

hydrophobes would line up along one face of either an α-helix or a β-sheet.

In aqueous solution, the peptides self-assembled into protein-like aggregates with secondary structures matching their hydrophobic repeats. These findings are particularly significant because the sequences were chosen such that stabilizing electrostatic interactions or side chain–side chain hydrogen bonding could not occur. Thus, hydrophobic interactions alone were shown to be sufficient to specify secondary structure. In fact, recent experiments have demonstrated that hydrophobic periodicity overrides inherent amino acid secondary structure propensities in the determination of the overall secondary structure in self-associating peptides.[34]

In subsequent work electrostatic interactions were also included in the design, leading to the design of a series of four-helix bundle proteins (α_4).[12,13,15] This protein was initially designed in collaboration with David Eisenberg using physical models and computer graphics.[12] Four α-helices were arranged in an approximate 222 symmetrical array with the helices approximately 10 Å apart and a left-handed tilt of 20° to provide favorable interhelical packing. Side chains were next added to stabilize the structure. Good interior packing could be accomplished with Leu as the only interior residue, and Glu and Lys residues were added at interfacial or exposed positions to provide electrostatic stabilization and water solubility. The α-helical peptides resulting from this design were then synthesized and their abilities to assemble into α-helical tetramers were characterized by CD, NMR, and sedimentation equilibrium.[12,13,35,36] The peptide that associated with the most favorable free energy of association was designated α_1B [Fig. 1(b)]. Four copies of this helical peptide were connected with three Pro-Arg-Arg loops, resulting in a highly stable four-helical bundle protein, α_4.[15] It contains the correct secondary structure as assessed by CD spectroscopy, and it is monomeric and compact. Further, it is far more resistant to denaturation (e.g., by guanidine hydrochloride) than most natural proteins.[15]

Hecht and coworkers[37] introduced a genetic, combinatorial approach to evaluate the role of hydrophobic interactions in stabilizing the structure of four-helix bundles. They prepared an immense library of partially random proteins with the same hydrophobic repeat as α_4 in which: (*i*), its Leu side chains were substituted randomly with Val, Leu, Phe or Ile; (*ii*), its Glu and Lys were substituted with Asn, Asp, Gln, Glu, His, or Lys; (*iii*), the positions of Gly, Pro and hydrophilic residues in the interhelical loops were essentially unchanged. Most of the resulting proteins were well expressed in *E. coli*, suggesting that they folded into conformations that protected them from rapid proteolysis.

Three of the proteins (protein 'B', 'F' and '86') were purified and shown to be helical and compactly folded in aqueous solution.[37]

The above work along with numerous other studies of synthetic proteins[8,9,38] has shown that hydrophobic interactions provide a very strong driving force for folding, and to a first approximation, proteins can be successfully designed by considering an exceptionally simple code — hydrophobics in, hydrophilics out. However, careful characterization of the conformations of derivatives of α_4 (and other similarly designed proteins[39–41]) indicated

that it did not have as well-defined a structure as native proteins. The interior amino acid side chains of α_4 were more mobile than in native proteins, indicating that α_4 shares some characteristics with the molten globule,[42,43] a non-native state of proteins with dynamically averaging conformations.

Specificity in protein design, a native-like four-helix bundle. The design of proteins that specifically bind substrates or catalyze reactions with high efficiencies requires the design of sequences that assume unique native-like structures.

This requires the introduction of more specific interactions than were included in the initial design of α_4. We have explored several strategies towards this goal including the introduction of metal-binding sites, improving the side chain packing interactions, and decreasing the hydrophobic content of the protein.

About a third of known natural proteins bind metal ions, and in many cases the bound ions are essential for maintaining the structural integrity and minimizing the dynamic behavior of the protein chain.[44] An extreme example is the zinc finger motif, which is unstructured in the absence of metal ions, and only folds upon complexation with metal ions.[45,46] A number of different metal-binding sites have been engineered into α_4, and the resulting proteins carefully characterized.[47–49] In each case, the mutant proteins appeared to fold and bind the metal in the predicted manner. Furthermore, NMR spectroscopy showed that the side chains in the vicinity of the metal-binding sites were well ordered, but side chains that were more distal to the binding sites appeared less well ordered.[47,48] Other experiments including the binding of hydrophobic dyes and the absence of well defined thermal unfolding transitions suggest that the proteins still exhibited some molten globule character.[47,48]

A second strategy involved changing the Leu residues in the interior of α_2B and α_4 to a different set of residues that would fit together with more geometric complementarity, providing a structure that more closely resembled a three-dimensional jig-saw puzzle. Single or double changes in α_2B resulted in structures with long-range NOEs consistent with the designed antiparallel structures,[50] but these proteins still retained many of the hallmarks of molten globules. We therefore used computer modeling to guide the choice of seven simultaneous changes within the hydrophobic core of α_2B.[51] The resulting dimeric protein [α_2C, Fig. 1(b)] undergoes a temperature-dependent transition from a native-like state to a molten globule-like state with a transition midpoint near room temperature. However, a fully cooperative transition from native-like to fully unfolded, as is generally seen for small natural proteins, was not observed.

We next attempted to introduce a metal-binding site into the repacked α_2C dimeric protein to provide a fully native structrure. Surprisingly, the resulting dimeric four-helix bundle, α_2D, was native-like, even in the absence of metal ions.[11] It no longer bound the hydrophobic dye, ANS, and its NMR spectrum was well resolved. It had intense bands in its near-UV CD spectrum, indicating that its aromatic groups were held in relatively rigid asymmetric environments. Further, this protein's near- and far-UV CD signals (indicative of the tertiary structure and secondary structure, respectively) are lost in a single, cooperative transition with values of $\Delta H°$ and ΔC_p similar to those of natural proteins of its size. These data strongly suggest that α_2D has a well packed hydrophobic core.

Subsequent synthesis of α_2D analogues[11] showed that the improvements were a consequence of changing two hydrophobic residues to hydrophilic residues (originally intended to bind metal ions). Modeling suggested that these substitutions in α_2D serve to destabilize one of two energetically degenerate conformations available to α_2C and earlier versions of this dimeric protein. Because the identities of these hydrophilic residues are not essential for native-like behavior, it is likely that they do not participate in specific stabilizing interactions. Instead they might destabilize conformations that would be accessible in their absence. Examination of an idealized model of the original protein shows that several Leu residues at interfacial positions are partially solvent-exposed (Fig. 3, top). Rotation of two of the helices

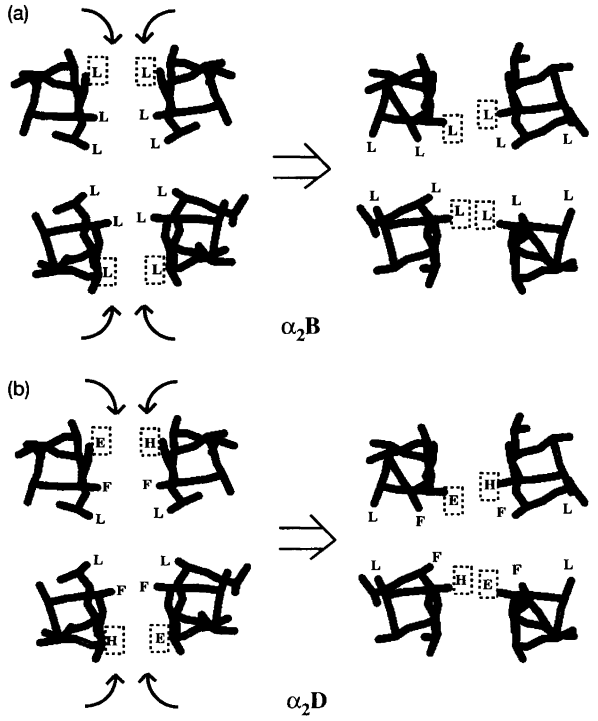


Fig. 3. Axial view of the central portion (seven residue slice) of the working model of the α_2B (top) and α_2D dimers (bottom panel). Residues that are hydrophilic in α_2D but hydrophobic in α_2B are boxed. A counter-rotation of the helices (indicated by the curved arrows) results in burial of the interfacial positions. This alternate packing should be highly energetically unfavorable for α_2D, but approximately isoenergetic for α_2B.

would produce a nearly equivalent conformation in which the interfacial residues are now buried, while other previously buried positions become partially exposed. If the interfacial residues are polar, this rotation requires unfavorable desolvation of these side chains (Fig. 3, bottom). Thus, the hydrophobic to hydrophilic substitutions in α_2D decrease the conformational degeneracy of the folded state. This finding illustrates an important principle in protein engineering which has been coined 'negative design' by Jane Richardson.[52] The sequences of natural proteins have evolved not only to stabilize a desirable three-dimensional structure but also to destabilize all other possible alternatives. A successfully designed protein must do the same.

The solution structure of α_2D is being investigated using NMR spectroscopy. Several key NOEs have been identified that indicate that the protein has an antiparallel structure as in the design. However, complete assignments will require uniform ^{15}N labeling, which requires the synthesis of a gene encoding this protein. This paper will describe the cloning of a gene encoding a protein, α_4D, which consists of two tandemly repeated α_2D sequences.

A native-like three-helix bundle protein. While much work has been published on the design of four-helix bundles, considerably less has been accomplished in the area of the design of three-helix bundle proteins despite their occurrence in such structures as protein A.[53] We have recently designed a three-helix bundle protein based on our published coordinates of an α-helical peptide that forms three-stranded coiled coils.

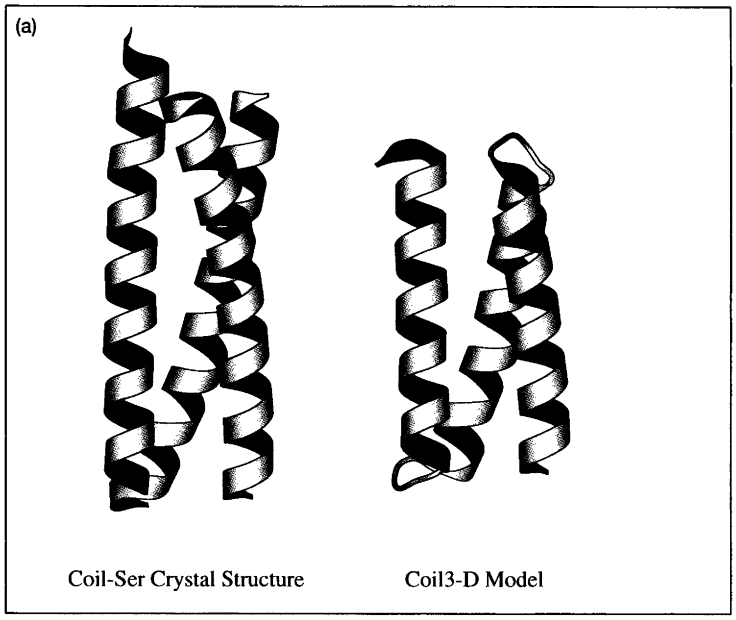
The α-helical coiled coil represents a structure of intermediate complexity bridging the gap between simple monomeric helices and native proteins. In pioneering work,[54] Hodges and coworkers designed polyheptapeptides based on the repeating sequence ($Leu_aGlu_bAla_cLeu_dGlu_eGly_fLys_g$), and showed that these peptides formed highly stable aggregates of helices. The Leu residues hydrophobically stabilize the bundle, while Glu and Lys residues at positions 'e' and 'g' were included to stabilize the structure through interhelical electrostatic interactions that are only possible in a parallel orientation.

We built on Hodges' original polyheptapeptides to prepare a 29-residue peptide, 'coil-Ser', as a model for studying helix stability.[55] The peptide contained Leu residues at 'a' and 'd' positions, Glu and Lys at 'e' and 'g', respectively, and a 'host' site on the solvent-exposed face of the helices for determining the helix propensities of various amino acid 'guests'. Surprisingly, the crystal structure of this peptide[21] shows a trimer rather than the designed dimeric structure. As in the design, it consisted of a parallel dimer of α-helices, but a third helix docks against this pair in an antiparallel manner. The Leu residues cluster in the interior of the structure in a manner consistent with the design, and there are favorable salt-bridges along one face of the parallel pair of helices. However, the electrostatics are unfavorable at the remaining helix–helix interfaces indicating that the driving force for the antiparallel orientation of the third helix may be related to efficient packing of the hydrophobic Leu residues.[21] In solution, the peptide exists in a relatively non-cooperative monomer/dimer/trimer equilibrium[56] (Hill coefficient = 2.4). Similarly a version of Hodges' original peptide, TM-43, is trimeric under a variety of conditions.[56]

We have used the crystallographic coordinates of the coil-Ser trimer as a starting point for the design of a monomeric version of this three-helix bundle. Using computer graphics we inserted loops between the helices, which were shortened to 21 residues — a length more typical of natural three-helix bundle proteins.[53] This protein, Coil3-D (Fig. 4) was chemically synthesized and we encountered no difficulties in purifying it in adequate quantities to allow characterization. Features that were considered in this design included capping interactions near the ends of the helices and the rearrangement of the charged residues to provide electrostatically stabilizing interactions along each helix–helix interface. The resulting protein has been examined by a multiplicity of techniques[8] used to differentiate native proteins from 'molten globules' and it shows most of the properties of a native protein. It has a cooperative guanidine unfolding curve, is monomeric in solution (as determined by sedimentation equilibrium), its amides exchange very slowly with solvent deuterons, and it does not bind ANS. However, thermal denaturation experiments determine that ΔC_p for the unfolding transition (8 cal mol^{-1} K^{-1} res^{-1}) is slightly smaller than expected for a native protein (10–15 cal mol^{-1} K^{-1} res^{-1}). Also, it has not been possible to solve the structure of this protein directly by NMR spectroscopy because its spectrum is too congested in the aliphatic region due to the large number of Leu residues.

In order to overcome these problems in Coil3-D, the all-Leu hydrophobic core was replaced with a more diverse selection of hydrophobic side chains, intended to pack in a more specific manner. This was accomplished using a computer program that exhaustively tries all possible low energy combinations of hydrophobic amino acid side chains in their preferred rotomeric states to determine which combination fills space most efficiently.[57,58] Additionally, the interfacial charged residues were redesigned to favor one of two possible topologies. The sequence of the resulting protein is given in Fig. 5(c).

Cloning and expression of designed helical bundle proteins. The overlap in the aliphatic region of ^{1}H NMR spectra of α_2D requires the incorporation of heteronuclei for aid in resonance assignment, and hence structural determination. For expression purposes we felt that a 35-residue peptide would be difficult to express and purify as the gene product might be rapidly degraded. For that reason, and to increase the ratio of desired product to expressed protein, the decision was made to express a tandem



(b) **Progression of Coil-Ser family of three-helical bundle proteins.**

```
CoilSer:     E WEALEKK LAALESK LQALEEKK LEALEHG

Coil3-D:     E WEALEKK LAALESK LQALGG-
         NPDE-WAALKKE LAALKSE LQALKGKG-
           NPE-WEALEKK LAALESK LQALEHG

Coil3-G:     S WAEFKER LAAIKSR LQALGG-
         SEAE-LAAFEKE IAAFESE LQAYKGKG-
           NPE-VEALRKE AAAIRSE LQAYRHN
```

Fig. 4. Design of three-helix bundle proteins: (a) MOLSCRIPT[60] diagrams of the crystal structure of Coil-Ser (1cos[21]) and the model structure of Coil3-D; (b) Sequence progression of the Coil-Ser family of three helix bundle proteins.

repeat of α_2D. The initial target, α_4D, contained a Pro-Arg-Arg loop connecting two α_2D units, analogous to the extension of α_2B to α_4.

We also wished to create a version of α_4D that could be cleaved to yield a fully labeled derivative of α_2D.

Because both repeats need to be identical, Met was designed as part of the loop between the dimer units. The presence of Met allows the use cyanogen bromide (CNBr) to cleave the desired protein from a poly-His affinity tag, and simultaneously cause the release of two identical homodimers.[59] To facilitate cloning, the *C*-terminal Gly in each α_2D unit was mutated to Asn-Met. A gene encoding α_4E was prepared from α_4D by two rounds of PCR [Figs. 5(a) and 5(b)]. Protein expression under a T7 Promotor was optimized in both rich and minimal media, and the product purified to homogeneity using Ni^{2+}-affinity chromatography. The expressed protein has a mass of 12472 as determined by ESI-MS in reasonable agreement with the predicted mass of 12476. After CNBr cleavage the major product was purified by preperative HPLC and has a mass 4408 consistant with 'α_2E' in which the *C*-terminal Met has been converted into homoserine lactone.

The gene construction (described below) of Coil3-G [see Fig. 5(c)] is complete and expression optimization trials are underway.

Materials and methods

Gene construction general procedures. The peptide sequence of each designed protein was translated into DNA using the program Pincers v1.1. The final DNA sequence was determined by favoring the most common codons found in *E. coli*, while optimizing the nucleotide sequence for heterogeneity. The sequences were adjusted to facilitate cloning, and final gene sequences are displayed in Fig. 5.

The α_4D and Coil3-G sequences were divided into sections of roughly 60–75 base pairs (bp) with 9–12 base overhangs. The resulting oligodeoxyribonucleotides were synthesized in house or obtained from Ana-Gen Technologies (San Francisco, CA). Each oligo was puri-

(a) α_4D Protein and Gene Sequences

```
Nco 1                                -20
  Met Gly His His His His His His His His His Ser Ser Gly His Ile Glu Gly
C ATG GGC CAT CAT CAT CAT CAT CAT CAT CAT CAC AGC AGC GGC CAT ATC GAA GGT
      CCG GTA GTA GTA GTA GTA GTA GTA GTA GTA GTG TCG TCG CCG GTA TAG CTT CCA

-10              BamH 1               1                              10
Arg His Met Leu Glu Asp Pro Gly Arg Met Gly Glu Val Glu Glu Leu Glu Lys Lys Phe
CGT CAT ATG CTC GAG|GAT CCC GGT CGT ATG GGC GAA GTT GAA GAA CTG GAG AAA AAA TTC
GCA GTA TAC GAG CTC CTA GGG CCA GCA TAC CCG CTT CAA CTT CTT|GAC CTC TTT TTT AAG

                                 20                              30
Lys Glu Leu Trp Lys Gly Pro Arg Arg Gly Glu Ile Glu Glu Leu His Lys Lys Phe His
AAA GAG CTG TGG AAA GGC CCG CGT CGT GGT GAA ATC GAG GAG CTT CAC AAA AAG TTT CAC
TTT CTC GAC ACC TTT CCG GGC GCA GCA CCA CTT TAG CTC CTC GAA GTG TTT TTC AAA|GTG

                                 40 Sap 1                        50
Glu Leu Ile Lys Gly Pro Arg Arg Gly Glu Val Glu Glu Leu Glu Lys Lys Phe Lys Glu
GAA CTT ATC AAG GGT CCA CGC CGC GGC GAA GTA GAA GAG CTT GAA AAG AAG TTC AAG GAG
CTT GAA TAG TTC CCA GGT GCG GCG CCG CTT CAT CTT CTC GAA CTT TTC TTC AAG TTC CTC

                                 60                              70
Leu Trp Lys Gly Pro Arg Arg Gly Glu Ile Glu Glu Leu His Lys Lys Phe His Glu Leu
TTG TGG AAA GGT CCT CGT|CGC GGC GAG ATT GAG GAA TTG CAT AAG AAA TTT CAT GAA CTC
AAC ACC|TTT CCA GGA GCA GCG CCG CTC TAA CTC CTT AAC GTA TTC TTT AAA GTA CTT GAG

        73          Hind III
Ile Lys Gly
ATT AAG GGC TAA TAA GCA
TAA TTC CCG ATT ATT CGT TCG A
```

(b) α_4E Protein and Gene Sequences

```
Nco 1                                -20
  Met Gly His His His His His His His His His Ser Ser Gly His Ile Glu Gly
C ATG GGC CAT CAT CAT CAT CAT CAT CAT CAT CAC AGC AGC GGC CAT ATC GAA GGT
      CCG GTA GTA GTA GTA GTA GTA GTA GTA GTA GTG TCG TCG CCG GTA TAG CTT CCA

-10                              1                              10
Arg His Met Leu Glu Asp Pro Gly Arg Met Gly Glu Val Glu Glu Leu Glu Lys Lys Phe
CGT CAT ATG CTC GAG GAT CCC GGT CGT ATG GGC GAA GTT GAA GAA CTG GAG AAA AAA TTC
GCA GTA TAC GAG CTC CTA GGG CCA GCA TAC CCG CTT CAA CTT CTT GAC CTC TTT TTT AAG

                                 20                              30
Lys Glu Leu Trp Lys Gly Pro Arg Arg Gly Glu Ile Glu Glu Leu His Lys Lys Phe His
AAA GAG CTG TGG AAA GGC CCG CGT CGT GGT GAA ATC GAG GAG CTT CAC AAA AAG TTT CAC
TTT CTC GAC ACC TTT CCG GGC GCA GCA CCA CTT TAG CTC CTC GAA GTG TTT TTC AAA GTG

                                 40                              50
Glu Leu Ile Lys Asn Met Gly Glu Val Glu Glu Leu Glu Lys Lys Phe Lys Glu Leu Trp
GAA CTT ATC AAG AAC ATG GGG GAA GTA GAA GAG CTT GAA AAG AAG TTC AAG GAG TTG TGG
CTT GAA TAG TTC TTG TAC CCC CTT CAT CTT CTC GAA CTT TTC TTC AAG TTC CTC AAC ACC

                                 60                              70
Lys Gly Pro Arg Arg Gly Glu Ile Glu Glu Leu His Lys Lys Phe His Glu Leu Ile Lys
AAA GGT CCT CGT CGC GGC GAG ATT GAG GAA TTG CAT AAG AAA TTT CAT GAA CTC ATT AAG
TTT CCA GGA GCA GCG CCG CTC TAA CTC CTT AAC GTA TTC TTT AAA GTA CTT GAG TAA TTC

                75          Sal 1
Asn Met Gly Gly Gly
AAC ATG GGC GGT GGG TAG TGA G
TTG TAC CCG CCA CCC ATC ACT CAG CT
```

(c) Coil3-G Protein and Gene Sequences

```
Nco 1            -10                                  1
  Met Gly His His His His His His His His His His Gly Met Gly Ser Trp Ala Glu
C ATG GGC CAT CAT CAT CAT CAT CAT CAT CAT CAC GGC ATG GGG TCT TGG GCT GAA
      CCG GTA GTA GTA GTA GTA GTA GTA GTA GTG CCG TAC|CCC AGA ACC CGA CTT

             10                                  20
Phe Lys Glu Arg Leu Ala Ala Ile Lys Glu Arg Leu Gln Ala Leu Gly Gly Ser Glu Ala
TTC AAA GAA CGT CTG GCT GCT ATC AAA GAA CGT CTG CAG|GCT CTG GGT GGT TCT GAA GCT
AAG TTT CTT GCA GAC CGA CGA TAG TTT CTT GCA GAC GTC CGA GAC CCA|CCA AGA CTT CGA

             30                                  40
Glu Leu Ala Ala Phe Glu Lys Glu Ile Ala Ala Phe Glu Ser Glu Leu Gln Ala Tyr Lys
GAA CTG GCT GCT TTC GAA AAA GAA ATC GCT GCT|TTC GAA TCT GAA CTG CAA GCT TAC AAA
CTT GAC CGA CGA AAG CTT TTT CTT TAG CGA CGA AAG CTT AGA|CTT GAC GTT CGA ATG TTT

             50                                  60
Gly Lys Gly Asn Pro Glu Val Glu Ala Leu Arg Lys Glu Ala Ala Ala Ile Arg Asp Glu
GGT AAA GGT AAC CCG GAA GTT GAA GCT|CTG CGT AAA GAA GCG GCC GCT ATC CGT GAC GAA
CCA TTT CCA TTG GGC CTT CAA CTT CGA GAC GCA TTT CTT|CGC CGG CGA TAG GCA CTG CTT

             70          BamH 1
Leu Gln Ala Tyr Arg His Asn
CTG CAG GCA TAC CGT CAC AAC TAG
GAC GTC CGT ATG GCA GTG TTG ATC CTA G
```

fied by electrophoresis on a 10% polyacrylamide urea gel. Appropriate bands were visualized, excised and the oligos eluted overnight in 200 µl of 3 M NaOAc. The oligos were ethanol precipitated at −20 °C, and resuspended in 20 µl of distilled, deionized water (ddH_2O).

Oligos were kinased with T4 polynucleotide kinase (all molecular biology materials were purchased from New England Biolabs, unless otherwise noted). Genes were assembled stepwise. The genes were divided approximately in half and annealing reaction mixtures prepared by combining the oligos required to generate the desired gene fragment. The annealing reaction mixtures contained 40 µg ml^{-1} of each oligo in 1X T4 DNA ligase reaction buffer (50 mM TrisCl, pH 7.8, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, 25 µg μl^{-1} BSA), heated to 70 °C for 10 min, slow-cooled to room temperature in a large volume water bath, and placed on ice. T4 DNA ligase (400 U) was added to each reaction mixture and the mixtures were incubated for >4 h at 16 °C. The reactions were loaded onto 3% NuSieve-GTG agarose gel (FMC Corp) and bands of the appropriate size excised and purified using a Qiaex Gel Extraction Kit (Qiagen).

Each partial construct (ca. 5 µg each gene fragment) was brought up to 100 µl ligase buffer, heated to 55 °C, slow-cooled to room temperature, and placed on ice. Ligase (20 U) was added and the mixture incubated for >4 h at 16 °C. As before, bands of the appropriate size were excised and purified. Genes were then amplified by the polymerase chain reaction (PCR) using 1 pmol of the 5′ and 3′ terminal oligos in 20 µl for 30 cycles. Typically, three-temperature PCR was performed varying the amount of template at 94 °C for 45′, 64 °C for 30′, 72 °C for 20′ with a Perkin Elmer Gene Amp 9600. Alternatively, the partial constructs can be mixed and the entire gene amplified by PCR using the 5′ and 3′ primers. The completed gene segments were purified as above.

Cloning. pET16b (Novagen) and insert were digested with appropriate restriction endonucleases (Fig. 5) and purified by electrophoresis on either a 1% SeaPlaque, or 3% Nu-sieve GTG in TBE. Visualized bands were excised and the DNA recovered as described for gene constructions. Cut plasmid was then treated with 1 U of shrimp alkaline phosphatase (USB) and 1 mM rATP for 1 h at 37 °C to prevent self-ligation. Ligations with varying ratios of insert to template were performed at 16 °C for >4 h. Five to 10 µl of each reaction were transformed into either DH5α (Gibco BRL, genotype: F^-

Fig. 5. Synthetic gene sequences of designed helical bundle proteins. The translated protein sequence is shown in bold above the gene sequence. Restriction sites used for cloning are indicated. Sequences of individual oligodeoxyribonucleotides used in gene assembly are indicated in the figure with solid lines: (a) α_4D; (b) α_4E; (c) Coil3-G.

F80d*lac*ZΔM15 D(*lac*ZYA-*arg*F)U169 *deo*R *rec*A1 *end*A1 *pho*A *hsd*R17(r_K^-, m_K^+) *sup*E44 l^- *thi*-1 *gyr*A96 *rel*A1), or XL-1-Blue (Stratagene, genotype: *rec*A1 *end*A1 *gyr*A96 *thi*-1 *hsd*R17 *sup*E44 rel*A1* lac [F′ *pro*AB, *lac*^{q}ZΔM15, *Tn*10 (tet^r)]), on LB plates with 100 μg ml^{-1} ampicillin at 37 °C. Positive clones were identified by diagnostic restriction digests, and confirmed by DNA sequence analysis.

Cloning of α_4E. From the α_4D background, PCR was used to introduce a methionine at position 36 [as well as other changes, see Fig. 1(b)] using Nco I-Sap I restrictions sites. The PCR product was isolated and ligated into α_4D/pET-16b cut with the same enzymes. Positive clones from this ligation were used as a template in a second PCR reaction to introduce methionine at position 72 in Fig. 1(b). This material was digested with Nco I and Sal I and the insert ligated into pET-16b digested with Nco I and Xho I to yield the final α4E gene displayed in Fig. 5(b).

Expression. Plasmids were transformed into BL21(DE3) [Novagen, genotype: F^- *omp*R *hsd*SB(r_B^-,m_B^-) *gal dcm* (DE3)] and plated on LB-Amp. Overnight cultures were grown in 2–5 ml LB-Amp, and used to inoculate one-liter cultures of LB-Amp. Large cultures were grown with vigorous shaking until OD_{600} exceeded 0.6 (approximately 2–3 h). Protein expression was induced by addition of isopropyl β-D-thiogalactopyranoside (IPTG, Sigma) to a final concentration of 1 mM. Samples were taken every hour after induction and analyzed by SDS-PAGE. Generally, optimum expression was achieved 2.5–3.0 h after induction. The cultures were centrifuged, and cell pellets frozen at −70 °C.

Protein purification. Cell pellets were thawed at room temperature and resuspended in (20 ml per 500 ml culture) 1X Binding Buffer (5 mM imidazole, 0.5 M NaCl, 40 mM TrisCl, pH 7.9), with 1 mM pefabloc (Boehringer Mannheim) and 250 μg ml^{-1} leupeptin (Sigma). The cell suspension was passed once through an SLM-Aminco french press using a 40 kpsi high-pressure cell. The lysate was centrifuged at 15 krpm for 20 min. All samples were kept on ice. A 2.5 ml His-Bind resin (Novagen) column was charged with 0.5 M $NiSO_4$, then equilibrated with 1X binding buffer. Supernatant from the lysate centrifugation was loaded onto the column at a flow rate of approximately 0.5 ml min^{-1}. The column was washed with binding buffer, and eluted with a gradient of increasing imidazole concentration (10 mM to 0.5 M). The eluent was analyzed by polyacrylamide gel electrophoresis. Excess imidazole was removed by dialysis against ddH_2O, and the desired product lyophilized. Analytical HPLC was performed to assess the purity of the dialyzed protein.

Cyanogen bromide cleavage. A small amount (1–10 mg) of CNBr was dissolved in 10 ml 70% formic acid and added to the lyophilized protein. The reaction was performed in the dark with gentle rocking for 12–18 h and was stopped by addition of 10 volumes (100 ml) ddH_2O. The diluted reaction mixture was lyophilized and the products purified by preparative HPLC.

Acknowledgements. We would like to thank Ron Hoess for advice and helpful discussions, Marvin Kendall for in-house oligodeoxyribonucleotide synthesis, and Rich McKay for ESI-MS analysis.

References

1. Makhatadze, G. I. and Privalov, P. L. *Adv. Protein Chem. 47* (1988) 191.
2. Stickle, D. F., Presta, L. G., Dill, K. A. and Rose, G. D. *J. Mol. Biol. 226* (1992) 1143.
3. McDonald, I. K. and Thornton, J. M. *J. Mol. Biol. 238* (1994) 777.
4. Yang, A.-S. and Honig, B. *Curr. Opin. Struct. Biol. 2* (1992) 40.
5. Honig, B. and Nicholls, A. *Science 268* (1995) 1144.
6. Burley, S. K. and Petsko, G. A. *Adv. Protein Chem. 39* (1988) 125.
7. DeGrado, W. F., Wasserman, Z. R. and Lear, J. D. *Science 243* (1989) 622.
8. Betz, S. F., Raleigh, D. P. and DeGrado, W. F. *Curr. Opin. Struct. Biol. 3* (1993) 601.
9. Bryson, J. W., Betz, S. F., Lu, H. S., Suich, D. J., Zhou, H. X., O'Neil, K. T. and DeGrado, W. F. *Science 270* (1995) 935.
10. Robertson, D. E., Farid, R. S., Moser, C. C., Urbauer, J. L., Mulholland, S. E., Pidikiti, R., Lear, J. D., Wand, A. J., DeGrado, W. F. and Dutton, P. L. *Nature (London) 368* (1994) 425.
11. Raleigh, D. P., Betz, S. F. and DeGrado, W. F. *J. Am. Chem. Soc. 117* (1995) 7558.
12. Eisenberg, D., Wilcox, W., Eshita, S. M., Pryciak, P. M., Ho, S. P. and DeGrado, W. F. *Proteins: Struct. Funct. Genet. 1* (1986) 16.
13. Ho, S. P. and DeGrado, W. F. *J. Am. Chem. Soc. 109* (1987) 6751.
14. Harris, N. L., Presnell, S. R. and Cohen, F. E. *J. Mol. Biol. 236* (1994) 1356.
15. Regan, L. and DeGrado, W. F. *Science 241* (1988) 976.
16. Hodges, R. S. *Curr. Biol. 2* (1992) 122.
17. Harbury, P. B., Kim, P. S. and Alber, T. *Nature (London) 371* (1994) 80.
18. Harbury, P. B., Zhang, T., Kim, P. S. and Alber, T. *Science 262* (1993) 1401.
19. Fezoui, Y., Weaver, D. L. and Osterhout, J. J. *Proc. Natl. Acad. Sci. USA 91* (1994) 3675.
20. Kuroda, Y., Nakai, T. and Ohkubo, T. *J. Mol. Biol. 236* (1994) 862.
21. Lovejoy, B., Choe, S., Cascio, D., McRorie, D. K., DeGrado, W. F. and Eisenberg, D. *Science 259* (1993) 1288.
22. Schafmeister, C. E., Miercke, L. J. W. and Stroud, R. M. *Science 262* (1993) 734.
23. Lutgring, R. and Chmielewski, J. *J. Am. Chem. Soc. 116* (1994) 6451.
24. Chin, T.-M., Berndt, K. D. and Yang, N. C. *J. Am. Chem. Soc. 114* (1992) 2279.
25. Yan, Y. and Erickson, B. W. *Protein Sci. 3* (1994) 1069.
26. Quinn, T. P., Tweedy, N. B., Williams, R. W., Richardson, J. S. and Richardson, D. C. *Proc. Natl. Acad. Sci. USA 91* (1994) 8747.
27. McLachlan, A. D. *Cold Spring Harbor Symposia on Quantitative Biology LII* (1987) 411.

28. Szebenyi, D. M. E. and Moffat, K. *J. Biol. Chem. 261* (1986) 8761.
29. Kretsinger, R. H. *Cold Spring Harbor Symposia on Quantitative Biology LII* (1987) 499.
30. Kauzmann, W. *Adv. Protein Chem. 14* (1959) 1.
31. Makhatadze, G. I. and Privalov, P. L. *J. Mol. Biol. 232* (1993) 639.
32. Privalov, P. L. and Makhatadze, G. I. *J. Mol. Biol. 232* (1993) 660.
33. DeGrado, W. F. and Lear, J. D. *J. Am. Chem. Soc 107* (1985) 7684.
34. Xiong, H., Buckwalter, B. L., Shieh, H.-M. and Hecht, M. H. *Proc. Natl. Acad. Sci. USA 92* (1995) 6349.
35. Ciesla, D. J., Gilbert, D. E. and Feigon, J. *J. Am. Chem. Soc. 113* (1991) 3957.
36. Osterhout, J. J., Handel, T., Na, G., Toumadje, A., Long, R. C., Connolly, P. J., Hoch, J. C., Johnson, W. C., Live, D. and DeGrado, W. F. *J. Am. Chem. Soc.* (1992) 331.
37. Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M. and Hecht, M. H. *Science 262* (1993) 1680.
38. Betz, S. F., Bryson, J. W. and DeGrado, W. F. *Curr. Opin. Struct. Biol. 5* (1995) 457.
39. Kaumaya, P. T. P., Berndt, K. D., Heldorn, D. B., Trewhella, J., Kezdy, F. J. and Goldberg, E. *Biochemistry 29* (1990) 13.
40. Morii, H., Ichimura, K. and Uedaira, H. *Proteins 11* (1991) 133.
41. Nishino, N., Mihara, H., Tanaka, Y. and Fujimoto, T. *Tetrahedron Lett. 33* (1992) 5767.
42. Kuwajima, K. *Proteins 6* (1989) 87.
43. Ptitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E. and Razgulyaer, O. L. *FEBS Lett 262* (1990) 20.
44. da Silva, J. J. R. F. and William, R. J. P. *The Biological Chemistry of the Elements*, Clarendon Press, Oxford, England 1993.
45. Parraga, G., Horvath, S., Hood, L., Young, E. T. and Klevit, R. E. *Proc. Natl. Acad. Sci. USA 87* (1990) 137.
46. Mau, T., Baleja, J. D. and Wagner, G. *Protein Sci. 1* (1992) 1403.
47. Handel, T. and DeGrado, W. F. *J. Am. Chem. Soc. 112* (1990) 6710.
48. Handel, T. M., Williams, S. A. and DeGrado, W. F. *Science 261* (1993) 879.
49. Regan, L. and Clarke, N. D. *Biochemistry 29* (1990) 10878.
50. DeGrado, W. F., Raleigh, D. P. and Handel, T. *Unpublished data.*
51. Raleigh, D. P. and DeGrado, W. F. *J. Am. Chem. Soc. 114* (1992) 10079.
52. Richardson, J., Richardson, D. C., Tweedy, N. B., Gernert, K. M., Quinn, T. P., Hecht, M. H., Erickson, B. W., Yan, Y., McClain, D. M. E. and Surles, M. C. *Biophys. J. 63* (1992) 1186.
53. Cedergren, L., Andersson, R., Jansson, B., Uhlén, M. and Nilsson, B. *Protein Eng. 6* (1993) 441.
54. Lau, S. Y. M., Taneja, A. K. and Hodges, R. S. *J. Biol. Chem. 259* (1984) 13253.
55. O'Neil, K. T. and DeGrado, W. F. *Science 250* (1990) 646.
56. Betz, S., Fairman, R., O'Neil, K., Lear, J. and DeGrado, W. *Phil. Trans. R. Soc. London, Ser. B 348* (1995) 81.
57. Ponder, J. W. and Richards, F. M. *J. Mol. Biol. 193* (1987) 775.
58. Desjarlais, J. R. and Handel, T. M. *Protein Sci. 4* (1995) 2005.
59. Kuliopulos, A. and Walsh, C. T. *J. Am. Chem. Soc. 116* (1994) 4599.
60. Kraulis, P. J. *J. Appl. Crystallogr. 24* (1991) 946.
61. Lolis, E., Alber, T., Davenport, R. C., Rose, D., Hartman, F. C. and Petsko, G. A. *Biochemistry 29* (1990) 6609.
62. Svensson, L. A., Thulin, E. and Forsen, S. *J. Mol. Biol. 223* (1992) 601.

Received September 5, 1995.