Solution Structure of α_2D, a Nativelike de Novo Designed Protein

R. Blake Hill and William F. DeGrado*

Contribution from The Johnson Research Foundation, Department of Biochemistry & Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6059

Received September 25, 1997

Abstract: De novo protein design provides an attractive means for testing and refining the principles governing the stability and tertiary structure of proteins. We describe the NMR solution structure of a 35-residue peptide designed to form a helix−loop−helix that dimerizes into a four-helix bundle. Structures were calculated on the basis of 834 NMR-derived restraints including 140 long-range NOEs. With 24 restraints per residue, the structure is well determined (0.28 Å RMSD for backbone residues 3-33) and includes many features of the design yet adopts a novel topology that was unexpected. The forces that caused this peptide to adopt this unique fold are discussed.

Introduction

How hydrophobic, electrostatic, hydrogen-bonding, and van derWaals forces contribute to the thermodynamic stability and conformational specificity of a protein is being investigated by a variety of approaches.[1−10] Protein folding requires a thermodynamically favorable driving force, which arises largely through the sequestering of apolar groups from solvent into the interior of a protein. To adopt a unique fold, a large energy gap must exist between the native fold and any other conformations;[11−13] otherwise the result is an ensemble of conformers such as in the molten globule state. However, the relative contributions of these noncovalent forces that specify a protein's structure remain undetermined because no single component dominates the free energy gap. For instance, many side chains involved in hydrogen bonding can be deleted from some proteins with retention of activity.[14,15] Further, the uniquely packed hydrophobic cores of proteins can be replaced with a single flexible amino acid such as Met,[16] or with a designed or random collection of hydrophobic amino acids to produce proteins that are functional, although they often show reduced stability.[17−19] Thus, the study of variants of natural proteins has failed to provide a unifying picture of the balance of forces required for specific stabilization of a unique fold.

De novo protein design is the engineering of a protein from first principles without reference to any known particular protein structure or consensus sequence. It provides a powerful tool for addressing the question of conformational specificity, because it allows one to test the interplay of the aforementioned forces. If the intended design is achieved, then the principles guiding the design may be valid. Studies of designed peptides have demonstrated how hydrophobic interactions drive the condensation of a protein chain into a globular structure rich in secondary structure,[20,21] and how electrostatic forces[22−24] and conformational preferences specifically stabilize α-helices[25−28] and β-sheets.[29−33] However, the majority of the studies of designed proteins have used low-resolution structural information which inherently limits the evaluation of the forces involved in protein stability and specificity. Therefore, high-resolution structural data are crucial to the understanding of designed protein structure and have only recently become available for a limited number of cases.[34−39]

* To whom correspondence should be addressed. Phone: (215) 898-4590. Fax: (215) 573-7229. E-mail: wdegrado@mail.med.upenn.edu.

(1) Bai, Y.; Sosnick, T. R.; Mayne, L.; Englander, S. W. *Science* **1995**, *269*, 192−197.
(2) Baldwin, R. *L. J. Biomol. NMR* **1995**, *5*, 103−109.
(3) Creamer, T. P.; Rose, G. D. *Protein Sci.* **1995**, *4*, 1305−14.
(4) Dill, K. A.; Chan, H. S. *Nat. Struct. Biol.* **1997**, *4*, 10−19.
(5) Dobson, C. M. *Nat. Struct. Biol.* **1995**, *2*, 513−517.
(6) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144−9.
(7) Roder, H.; Colón, W. *Curr. Opin. Struct. Biol.* **1997**, *7*, 15−28.
(8) Stickle, D. F.; Presta, L. G.; Dill, K. A.; Rose, G. D. *J. Mol. Biol.* **1992**, *226*, 1143−1159.
(9) Yang, A.-S.; Honig, B. *Curr. Opin. Struct. Biol.* **1992**, *2*, 40−45.
(10) Richardson, J. S.; Richardson, D. C. *Science* **1988**, *240*, 1648−1652.
(11) Onuchic, J. N.; Socci, N. D.; Luthey-Schulten, Z.; Wolynes, P. G. *Folding Des.* **1996**, *1*, 441−50.
(12) Mirny, L. A.; Shakhnovich, E. I. *J. Mol. Biol.* **1996**, *264*, 1164−79.
(13) Lazaridis, T.; Archontis, G.; Karplus, M. *Adv. Protein Chem.* **1995**, *47*, 231−306.
(14) Fersht, A. R.; Serrano, L. *Curr. Opin. Struct. Biol.* **1993**, *3*, 75−83.
(15) Matthews, B. W. *Adv. Protein Chem.* **1990**, *46*, 249−278.
(16) Gassner, N. C.; Baase, W. A.; Matthews, B. W. *Proc. Nat. Acad. Sci. U.S.A.* **1996**, *93*, 12155−8.
(17) Desjarlais, J. R.; Handel, T. M. *Protein Sci.* **1995**, *4*, 2006−2018.
(18) Cordes, M. H.; Davidson, A. R.; Sauer, R. T. *Curr. Opin. Struct. Biol.* **1996**, *6*, 3−10.
(19) Axe, D. D.; Foster, N. W.; Fersht, A. R. *Proc. Nat. Acad. Sci. U.S.A.* **1996**, *93*, 5590−4.
(20) DeGrado, W. F.; Lear, J. D. *J. Am. Chem. Soc.* **1985**, *107*, 7684.
(21) Kamtekar, S.; Schiffer, J. M.; Xiong, H.; Babik, J. M.; Hecht, M. H. *Science* **1993**, *262*, 1680−1685.
(22) Armstrong, K. M.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11337−11340.
(23) Scholtz, J. M.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *J. Am. Chem. Soc.* **1991**, *113*, 5102−5104.
(24) Shalongo, W.; Stellwagen, E. *Protein Sci.* **1995**, *4*, 1161−6.
(25) Padmanabhan, S.; Marqusee, S.; Ridgeway, T.; Laue, T. M.; Baldwin, R. L. *Nature* **1990**, *344*.
(26) O'Neil, K. T.; DeGrado, W. F. *Science* **1990**, *250*, 646−651.
(27) Creamer, T. P.; Rose, G. D. *Proteins* **1994**, *19*, 85−97.
(28) Lyu, P. C.; Liff, M. I.; Marky, L. A.; Kallenbach, N. R. *Science* **1990**, *250*, 669−673.
(29) Otzen, D. E.; Fersht, A. R. *Biochemistry* **1995**, *34*, 5718−24.
(30) Minor, D. L.; Kim, P. S. *Nature* **1994**, *371*, 264−267.
(31) Smith, C. K.; Withka, J. M.; Regan, L. *Biochemistry* **1994**, *33*, 5510−7.
(32) Smith, C. K.; Regan, L. *Science* **1995**, *270*, 980−2.
(33) Kim, C. A.; Berg, J. M. *Nature* **1993**, *362*, 267−270.

S0002-7863(97)03364-7 CCC: $15.00
Published on Web 02/03/1998

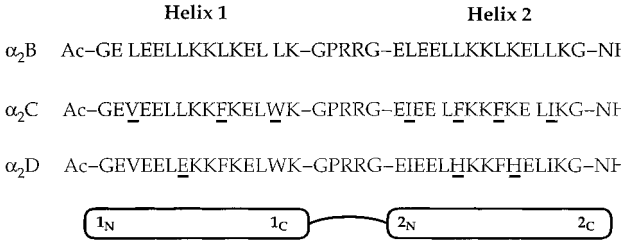

Figure 1. Sequences of the α_2 family of peptides. α_2B was the first α_2-like design and contained two identical helices with Leu residues as the sole hydrophobic amino acid. Seven residues of α_2B were changed to give α_2C (underlined), which showed behavior intermediate between a nativelike and a molten globule protein. Three additional changes to the sequence of α_2C resulted in a nativelike dimer, α_2D (underlined). Note that the changes made at each step decrease the sequence homology between helix 1 and helix 2. In each sequence, the N-terminus is acetylated and the C-terminus is amidated. The sequence of α_2D differs from that reported previously[45,47,49] by an inversion of residues 11 and 30. The previously reported sequence contained a typographical error.

Here we report the solution structure of α_2D, a de novo designed four-helix bundle comprised of a dimer of helix−loop−helix peptides. Four-helix bundles are frequently observed in natural proteins.[40] Covalently associated dimers of helix−loop−helix motifs figure largely in the structures of metalloproteins such as calmodulin[41,42] and bacterioferritin[43] while noncovalently associated dimers of helix−loop−helix peptides appear in DNA-binding proteins and RNA-binding proteins, such as Rop.[44] α_2D is the third generation of a series of peptides designed to test the hierarchical requirements for stabilizing the native structure of proteins.[45] The initial peptide, α_2B, contained a hydrophobic core comprised solely of Leu residues (Figure 1).[46] The next generation, α_2C, contained a hydrophobic core that included a more diverse set of residues to better reflect the complementarity of packing found in the core of native proteins.[47] These peptides were found to form thermodynamically stable dimers with dynamically averaging structures indicative of an energetically degenerate ensemble of interconverting conformers. To specify a nativelike conformer of α_2C, a metal binding site[48] was engineered into two of the four helix−helix interfaces by replacing three residues: Leu7Glu, Phe26His, and Lys30His (Figure 1). This peptide, α_2D, associated into dimers that show all the thermodynamic and spectral properties expected for a nativelike four-helix bundle, even in the absence of metal ions.[49] The determination of this structure allows the examination of the forces that drive conformational specificity in protein folding.

Experimental Section

Sample Preparation. The peptide was synthesized using 9-fluorenylmethyloxycarbonyl-protected amino acids on a Milligen 9050 synthesizer, and purified by preparative reversed phase high-performance liquid chromatography using a VYDAC C18 column. The molecular weight of the peptide was confirmed using electrospray mass spectrometry (observed mass 4289.1, calculated mass 4288.92). The dimeric aggregation state of α_2D ($K_d = 3\ \mu$M) at 2.0 mM was confirmed by analytical ultracentrifugation as described previously.[49] Samples for NMR were at 2.0 mM monomer concentration in 50 mM d_{11}-Tris-(hydroxymethyl)methylamine, pH 7.3. To aid in structure determination, a sample was prepared with ^{13}C-labeled δ-leucines. The molecular weight and extent of incorporation of the ^{13}C-label were verified by electrospray mass spectrometry.

NMR Spectroscopy. Proton assignments were obtained using a combination of TOCSY,[50] DQF-COSY,[51] and 200 ms NOESY[52] homonuclear 2D experiments. NOE crosspeak volumes for structure calculations were determined from a 2D homonuclear NOESY with a 100 ms mixing time and a 3D ^{13}C-separated NOESY experiment with a 100 ms mixing time.[53] The sample for the heteronuclear experiment was labeled at the δ-methyl groups of Leu6, Leu13, Leu25, and Leu32. These crosspeak volumes were converted into four classes of distance restraints: strong (0−2.7 Å), medium (2.7−3.3 Å), weak (3.3−4.0 Å), and very weak (4.0−6.0 Å) on the basis of standard methods of calibration.[54] Dihedral angle restraints were determined by extraction of $^3J_{\alpha N}$ coupling constant data from a DQF-COSY spectrum using the method of Kim and Prestegard[55] and by measurement of $^3J_{\alpha\beta}$ coupling constant data from an E-COSY spectrum obtained in D_2O.[56] All experiments were performed on a Bruker AMX-600 at 25 °C and pH 7.3 with the ^{1}H carrier positioned on the HDO resonance (4.75 ppm relative to TSP at 0.0 ppm). The acquisition times were identical for all experiments: 254 ms in t_2 and 127 ms in t_1, with an identical dwell time of 124 μs. The data were processed using the AZARA suite of programs provided by Wayne Boucher and the Department of Biochemistry, University of Cambridge (http://www-.bio.cam.ac.uk/ftp/pub/azara). Processed data were analyzed with the aid of ANSIG.[57]

Structure Calculations. The program X-PLOR[58] was used to incorporate NMR-derived distance and torsion angle restraints to generate structural models of α_2D. The search protocol used dynamical simulated annealing[59] for 48 ps at a temperature of 2000 K and slow cooling to 50 K. To limit the closest approach of nonbonded atoms, no attractive van der Waals or electrostatic terms were used. Sixty structures were generated from an extended chain starting structure on the basis of 834 NMR-derived restraints including 140 long-range distance restraints from NOEs, 30 ϕ and 14 χ_1 torsion angle restraints.

(34) Fezoui, Y.; Connolly, P. J.; Osterhout, J. J. *Protein Sci.* **1997**, *6*, 1869−1877.
(35) Ilyina, E.; Roongta, V.; Mayo, K. H. *Biochemistry* **1997**, *36*, 5245−5250.
(36) Schafmeister, C. E.; Miercke, L. J. W.; Stroud, R. M. *Science* **1993**, *262*, 734−738.
(37) Struthers, M. D.; Cheng, R. P.; Imperiali, B. *Science* **1996**, *271*, 342−345.
(38) Ogihara, N. L.; Weiss, M. S.; DeGrado, W. F.; Eisenberg, D. *Protein Sci.* **1997**, *6*, 80−8.
(39) Lovejoy, B.; Choe, S.; Cascio, D.; McRorie, D. K.; DeGrado, W. F.; Eisenberg, D. *Science* **1993**, *259*, 1288−1293.
(40) Harris, N. L.; Presnell, S. R.; Cohen, F. E. *J. Mol. Biology* **1994**, *236*, 1356−1368.
(41) Babu, Y.; Bugg, C.; Cook, W. *J.Mol. Biol.* **1988**, *204*, 191.
(42) Finn, B. E.; Forsen, S. *Structure* **1995**, *3*, 7−11.
(43) Frolow, F.; Kalb, A. J.; Yariv, J. *Nat. Struct. Biol.* **1994**, *1*, 453−460.
(44) Banner, D.; Kokkinidis, M.; Tsernoglou, D. *J. Mol. Biol.* **1987**, *196*, 657.
(45) Bryson, J. W.; Betz, S. F.; Lu, H. S.; Suich, D. J.; Zhou, H. X.; O'Neil, K. T.; DeGrado, W. F. *Science* **1995**, *270*, 935−941.
(46) Ho, S. P.; DeGrado, W. F. *J. Am. Chem. Soc.* **1987**, *109*, 6751−6758.
(47) Raleigh, D. P.; DeGrado, W. F. *J. Am. Chem. Soc.* **1992**, *114*, 10079−10081.
(48) Handel, T.; DeGrado, W. F. *J. Am. Chem. Soc.* **1990**, *112*, 6710−6711.
(49) Raleigh, D. P.; Betz, S. F.; DeGrado, W. F. *J. Am. Chem. Soc.* **1995**, *117*, 7558−7559.
(50) Bax, A.; Davis, D. G. *J. Magn. Reson.* **1985**, *65*, 355−360.
(51) Rance, M.; Sorensen, O. W.; Bodenhausen, G.; Wagner, G.; Ernst, R. R.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1983**, *117*, 479−485.
(52) Anil-Kumar; Ernst, R. R.; Wüthrich, K. W. *Biochem. Biophys. Res. Commun.* **1980**, *95*, 1−6.
(53) Fesik, S. W.; Zuiderweg, E. R. *Q. Rev. Biophys.* **1990**, *23*, 97−131.
(54) Wüthrich, K. *NMR of Proteins and Nucleic Acids;* Wiley: New York, 1986.
(55) Kim, Y.; Prestegard, J. H. *Proteins* **1990**, *8*, 377−385.
(56) Griesinger, C.; Sorensen, O. W.; Ernst, R. R. *J. Am. Chem. Soc.* **1985**, *107*, 6394.
(57) Kraulis, P. J. *J. Magn. Reson.* **1989**, *84*, 627−633.
(58) Brünger, A. T. *X-PLOR, Version 3.1;* Yale University Press: New Haven, CT, 1993.
(59) Nilges, M.; Clore, G. M.; Gronenborn, A. M. *FEBS Lett.* **1988**, *239*, 129−136.

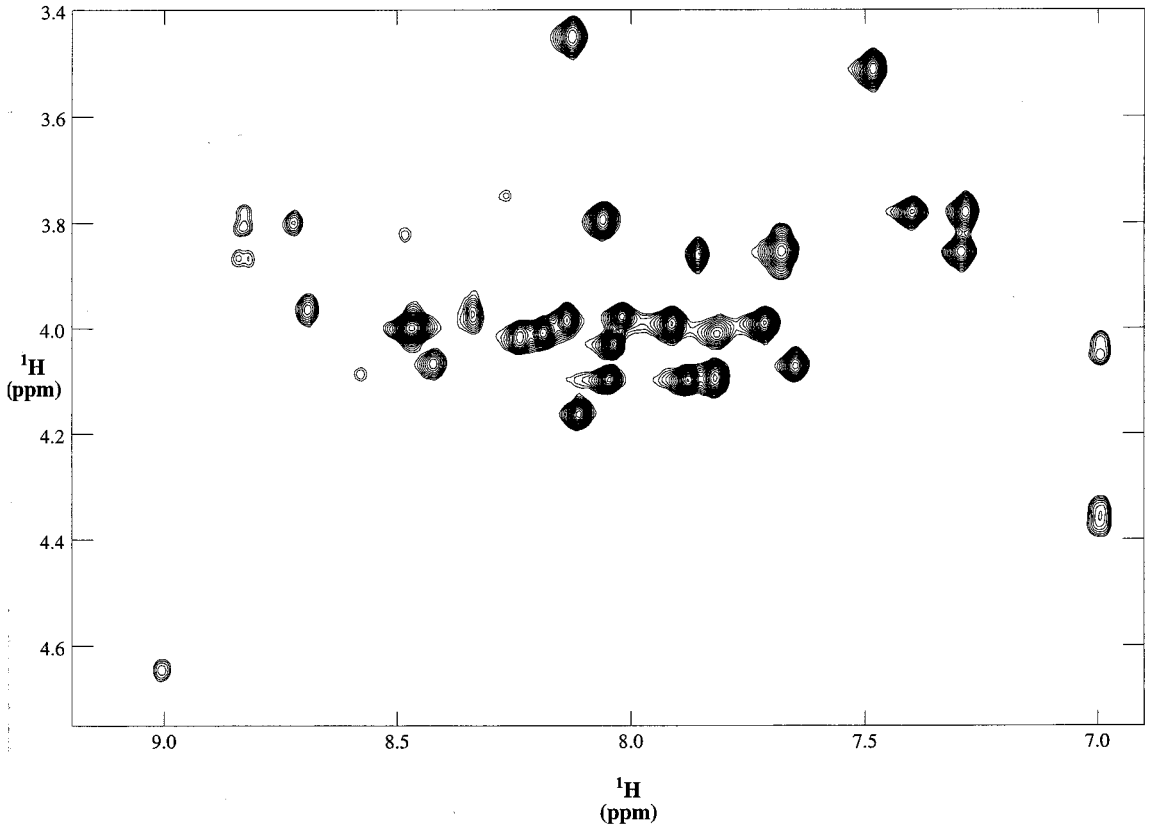

Figure 2. Fingerprint region of a ^{1}H−^{1}H TOCSY spectrum of 2 mM α_2D at 600 MHz, 50 mM Tris, pH 7.3, 25 °C. The TOCSY (or HOHAHA) experiment shows intraresidue, through-bond correlations. This portion of the spectrum shows correlations that occur between the H^N and the H^α of the same residue. The observation of a single crosspeak for each residue (two for Gly) supports the conclusion that α_2D exists as a symmetric homodimer under these conditions. Complete resonance assignments are given in the Supporting Information.

Fifty-nine percent of the long-range NOEs were determined to arise from intermonomer correlations by computational methods.[60,61] Specifically, 60 structures were calculated initially assuming α_2D to exist as a monomer. Those NOE restraints that were violated by more than 5 Å in all structures were not used in subsequent monomer calculations. When the monomer calculations converged to a single species that had no NOE violations greater than 0.3 Å, the NOEs that had been disregarded were considered candidates for intermonomer correlations in subsequent dimer calculations. Dimer structures were then calculated as described above with these potential intermonomer NOEs slowly introduced. In this manner, 83 intermonomer NOEs (59%) that occur between 20 residues were identified (see the Supporting Information). The other NOEs were treated using a weighting function developed for symmetric homodimers[62] that allows each NOE to arise from either intra- or intermonomer correlations. Symmetry terms were employed for C^α carbons only. This iterative approach has been used previously in the successful determination of several homodimeric proteins.[60,61,63−66]

Results and Discussion

Structure Determination. The structure of α_2D at pH 7.3 in aqueous solution was solved by standard two-dimensional NMR methods[54] supplemented by ^{13}C-edited NOESY experiments. Figure 2 illustrates the fingerprint region of a typical TOCSY spectrum; the spectrum is reasonably well dispersed for a helical protein, and a single set of resonances is observed for each residue, suggesting that the peptide adopts a single conformation on the NMR time scale.

The NMR data used in the structure calculations are summarized in Figure 3. The helices within each monomer span residues 3−14 (helix 1) and 20−34 (helix 2), which are within one residue of the positions expected in the design. Sixty structures were calculated, and all converged to the same topology shown in Figure 4, where an overlay of the 16 best structures is depicted without (Figure 4A) and with (Figure 4B) the hydrophobic side chains. With 24 restraints per residue, the structure is well determined with a backbone RMSD of 0.21 Å for residues 3−33 of the monomer and 0.28 Å for residues 3−33 of the dimer (Table 1).

Comparison of the Overall Fold with the Designed Fold. The structure of α_2D shows many of the features of the intended design, but differs in many respects from the expected structure. At the time that we initiated structural studies, two topologies were anticipated as possible folds for α_2D, Figure 5A,B. The first places the loops at the same side of the bundle (syn topology, Figure 5A), as in the initial design of α_2B,[46] which actually has two possible syn topologies with either right-handed or left-handed loops.[67] Two anti topologies are also possible in which the loops lie on opposite sides of the bundle (Figure 5B) with either right-turning or left-turning loops.[67]

(60) Clore, G. M.; Appella, E.; Yamada, M.; Matsushima, K.; Gronenborn, A. M. *Biochemistry* **1990**, *29*, 1689−1696.
(61) Kay, L. A.; Forman-Kay, J. D.; McCubbin, W. D.; Kay, C. M. *Biochemistry* **1991**, *30*, 4323−4333.
(62) Nilges, M. *Proteins* **1993**, *17*, 297−309.
(63) MacKenzie, K. R.; Prestegard, J. H.; Engeleman, D. M. *Science* **1997**, *276*, 131−133.
(64) Breg, J. N.; van Opheusden, J. H.; Burgering, M. J.; Boelens, R.; Kaptein, R. *Nature* **1990**, *346*, 586−589.
(65) Burgering, M. J. M.; Boelens, R.; Gilbert, D.; JN, B.; Knight, K.; Sauer, R.; Kaptein, R. *Biochemistry* **1994**, *33*, 15036−15045.
(66) Rico, M.; Jimenez, M. A.; González, C.; De Filippis, V.; Fontana, A. *Biochemistry* **1994**, *33*, 14834−14847.
(67) Betz, S. F.; Liebman, P. A.; DeGrado, W. F. *Biochemistry* **1997**, *36*, 2450−2458.

Figure 3. Summary of NMR data for α_2D at 50 mM d_{11}-Tris, pH 7.3, 25 °C. NOE, torsion angle, and chemical shift data identify the secondary structure of α_2D. ϕ angle information is summarized by the three-bond H^α-H^N coupling constant ($^3J_{\alpha N}$) directly below the sequence. Filled circles are for $^3J_{\alpha N}$ values less than 6 Hz and are consistent with an α-helix, whereas open boxes are for $^3J_{\alpha N}$ values greater than 6 Hz and indicate an open structure. NOE correlations are indicated by a line drawn between the two residues. The width of the line is proportional to the intensity of the NOE. Several connectivities indicating α-helical configurations are observed, including sequential amide proton correlations (d_{NN}) and NOEs between the H^α of residue i and H^N of residues $i + 3$ ($d_{\alpha N}(i, i + 3)$) and $i + 4$. Chemical shift indices[83] were derived from the chemical shifts of α_2D; for H^α resonances, a score of +1 is predictive of a β-strand, whereas a score of −1 predicts a region of an α-helix; for the C^α resonances (obtained from a natural abundance HMQC at 500 MHz), the scoring is reversed. Resonances with statistically insignificant secondary shifts received scores of 0. The secondary structure is indicated at the bottom and correlates with that found from structure calculations.

The experimentally determined conformation of α_2D features a 2-fold symmetrical pair of helix−loop−helix motifs with the loops on opposite sides of the bundle, as in the Rop family of dimeric anti four-helix bundles.[68] However, rather than connecting neighboring helices as in the anti conformation pictured in Figure 5B, the loops reach across the bundle to connect diagonally opposed helices, Figure 5C. Near the loop, the helices within a monomer are in van der Waals contact, but they diverge from this point, forming a U-like shape. Two helix−loop−helix monomeric units dock in a "bisecting U" motif, with extensive interactions between the two monomers. Indeed, 59% of the long-range NOEs in α_2D are intermonomer.

The fold of α_2D is most similar to the anti topology pictured in Figure 5B, in which the loops lie on opposite sides of the bundle. Both of these topologies have in common two antiparallel, intermolecular interfaces that occur between helix 1 and helix 1′ and between helix 2 and helix 2′. A major consequence of the diagonal crossover loops in α_2D, however, is that helices 1 and 2′ (and helices 1′ and 2) pack together in a parallel rather than an antiparallel fashion.

Given that the bisecting U motif had not previously been described, we examined a variety of structures for this motif. Indeed, this fold forms the dimerization domain of *E. coli* FIS, a DNA-binding protein[69] (Figure 6A). Topological correlates of this motif also occur in the intramolecular folded structures of various other proteins. For example, the conserved core of the ubiquitous immunoglobulin fold[70] contains a β-strand−loop−B-strand motif in which the b and c strands interact with the e and f strands in a manner that is topologically similar to the bisecting U motif of α_2D (Figure 6B). Further, examination of the recently published structures of the Fas death domain[71] (Figure 6C) and the δ subunit of the *E. coli* ATP synthase[72] (Figure 6D) reveals that the central four helices of these proteins interact in a topologically equivalent manner, with the loops crossing over either side of the bundle. Thus, though the bisecting U has not previously been recognized as a discrete motif, it is a fundamental element of the structures of both β-strand and α-helical proteins.

Determinants of the Fold of α_2D. The structure of a four-helix bundle may be analyzed in terms of the four helix−helix interfaces lying between the four individual helices, shown schematically in Figure 5. In the syn and the anti folds, each helix−helix interface is antiparallel, while the bisecting U motif contains two antiparallel and two parallel helical interfaces. It is interesting to explore the possible folds of α_2D in the context of its predecessor, α_2B. Because both of the helices of α_2B are identical and contain Leu as the sole hydrophobic residue, this dimer should have similar helix−helix interfaces in the syn, anti, and bisecting U motifs. Consequently, one would expect a similar free energy of folding for each of the possible

(68) Banner, D. W.; Kokkinidis, M.; Tsernoglou, D. *J. Mol. Biol.* **1987**, *196*, 657−675.
(69) Kostrewa, D.; Granzin, J.; Koch, C.; Choe, H.-W.; Raghunathan, S.; Wolf, W.; Labahn, J.; Kahmann, R.; Saenger, W. *Nature* **1991**, *349*, 178−180.
(70) Bork, P.; Holm, L.; Sander, C. *J. Mol. Biol.* **1994**, *242*, 309−20.
(71) Huang, B.; Eberstadt, M.; Olejniczak, E.; Meadows, R.; Fesik, S. *Nature* **1996**, *384*, 638−641.
(72) Wilkens, S.; Dunn, S.; Chandler, J.; Dahlquist, F.; Capaldi, R. *Nat. Struct. Biol.* **1997**, *4*, 198−901.

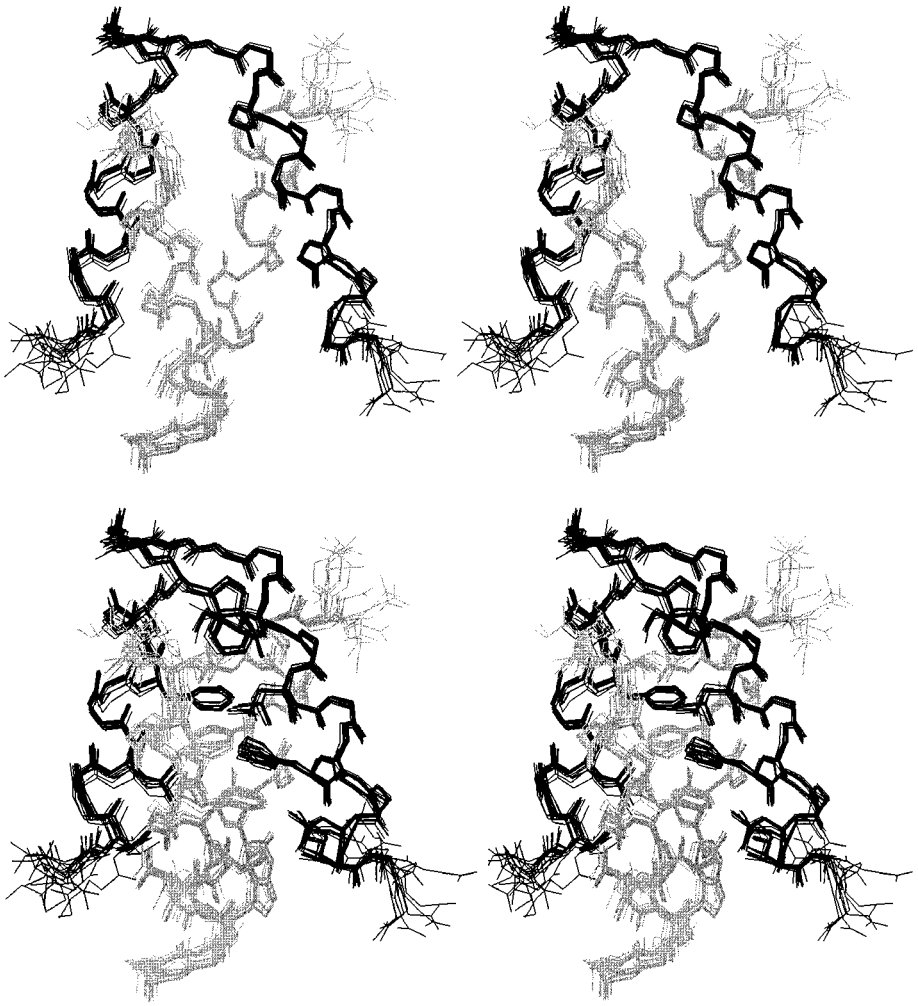

Figure 4. Stereoview of a superposition from the 16 best structures of (A, top) the backbone residues and (B, bottom) the backbone plus hydrophobic core residues of α_2D. The backbone RMSD for residues 3−33 of the monomer is 0.21 Å and for residues 3−33 of the dimer is 0.28 Å. The RMSD for all non-hydrogen atoms from the hydrophobic (internal) residues is 0.5 Å.

topologies, perhaps giving rise to the observed dynamic behavior of α_2B.[48] By contrast, the sequence asymmetry introduced in the design of α_2D removes the degeneracy of the interactions at each helix−helix interface, resulting in a large enough energy difference between each of the possible folds to define a unique conformation. Thus, in contrast to the equivalent interactions between helices for α_2B, each helix−helix interaction in the observed structure of α_2D is unique and stabilized by a distinct combination of amino acid side chains, as described below.

In the α_2D dimer, helix 1 packs in an antiparallel manner against its symmetry-related partner, helix 1′, with a crossing angle of 22°, close to the ideal crossing angle expected for an antiparallel four-helix bundle.[73] The interface contains a series of well-packed, interdigitated Leu side chains (Figure 7A). Interfacial Glu and Lys residues from neighboring helices interact in a geometrically and electrostatically complementary manner and shield the interior Leu residues from contact with the solvent. The residues contained in this Leu-rich interface are retained from the original design of α_2B. The favorable interactions observed at this interface may account in part for the extreme stability of α_2B ($\Delta G = -13$ kcal/mol).[46] As discussed above, however, these interactions may be formed in a number of different topologies of α_2B and are insufficient to define a unique conformation.

The second antiparallel interface lies between helices 2 and 2′, and is comprised of interactions between His and Ile residues on symmetry-related helices. The crossing angle of these two helices is 35°, somewhat larger than the ideal helix crossing angles for antiparallel-docked helices, but well within the range observed for natural four-helix bundle proteins.[40,74] Two pairs of His residues in the second helix at positions 26 and 30 interact in an intermonomer, H-bonded aromatic cluster (Figure 7B). These two His residues lack homologous residues on the first helix, and hence may provide considerable conformational specificity to the observed fold. Indeed a peptide in which His26 was changed to Leu shows properties similar to those of the molten globule state.[49] It is also interesting to note that in the four-helix bundle protein, Rop, aromatic His and Phe residues on equivalent monomers also interact across the 2-fold symmetry axis.

Thus, the two antiparallel interfaces of α_2D are quite similar to the antiparallel helical interfaces in the intended design. Further, the observed interactions could be accommodated within an anti-type four-helix bundle with conventional crossings between neighboring helices at each of the four antiparallel interfaces. It is therefore important to ask what causes the loop to adopt a diagonal crossover connection. The answer may lie, in part, in the interactions formed by the residues within the

(73) Gonzalez, L., Jr.; Plecs, J. J.; Alber, T. *Nat. Struct. Biol.* **1996**, *3*, 510−5.

(74) Presnell, S. R.; Cohen, F. E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 6592−6596.

Table 1. Structural Statistics of α_2D

structural statistics	⟨SA⟩[a]
rms deviations from experimental distance restraints (Å)[b]	
all (834)	0.028 ± 0.003
rms deviations from exptl dihedral restraints (deg)[c]	0.6 ± 0.2
deviations from idalized covalent geometry	
bonds (Å)	0.0030 ± 0.0001
angles (deg)	0.56 ± 0.01
impropers (deg)	0.46 ± 0.01
coordinate precision (Å)[d]	
backbone atoms	0.28 ± 0.09
backbone atoms of one monomer	0.21 ± 0.08
all non-hydrogen atoms	1.1 ± 0.2
all non-hydrogen atoms of internal residues[e]	0.5 ± 0.1

[a] ⟨SA⟩ is the emsemble of the final 16 simulated annealing structures. These structures were used to obtain a mean structure by averaging the individual structures best fitted to each other. [b] The NOE restraints employed a square well potential using a force constant of 50 kcal/mol and comprised of 171 interresidue sequential ($|i - j| = 1$), 147 interresidue short-range ($1 < |i - j| < 5$), 140 interresidue long-range ($|i - j| > 5$), 334 intraresidue, and 42 H-bond correlations. The structures exhibit no NOE violations greater than 0.4 Å. Two distance restraints were used for each backbone H-bond: $r_{NH-O} = 1.7-2.5$ Å and $r_{N-O} = 2.4-3.3$ Å. [c] ϕ angles were restrained to $-60° \pm 40°$ for 19 residues whose $^3J_{\alpha N}$ coupling constant was less than 6 Hz and to $-120° \pm 40°$ for 8 residues whose $^3J_{\alpha N}$ coupling constant was greater than 8 Hz. Force constants of 200 kcal/(mol·rad^2) were applied for all torsion restraints. The structures exhibit no dihedral violations greater than 5°. [d] The values reported for coordinate precision are the average atomic rms differences between the individual simulated annealing structures and the mean structure for residues 3-33. The first and last two residues are disordered and were not included in the analysis. [e] Residues of the hydrophobic core: Val3, Leu6, Phe10, Leu13, Trp14, Ile22, Leu25, Phe29, Leu32, Ile33.

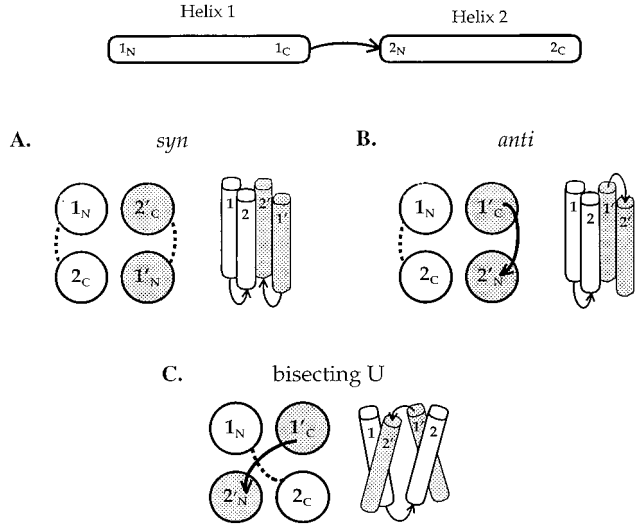

Figure 5. Possible topologies of α_2D with loops on the same (syn) or opposite (anti) sides of the bundle. The bisecting U is the actual topology of α_2D. Right-handed turn loops are depicted, but left-handed turn loops are also possible, giving rise to six distinct four-helix bundle topologies.

loops. The crossover connecting loops in β-barrels often contain hydrophobic residues that cap the hydrophobic cores of the structure as the loop traverses the top of the barrel. In a similar manner, the loop of α_2D drapes diagonally across the bundle, with the pyrrolidinyl ring of Pro and the alkyl portions of the Arg side chain engaging in van der Waals interactions with the apolar residues comprising the core of the α_2D dimer (Figure 7C).

The diagonal crossover connecting loop also dictates a parallel interface between helices 1 and 2′ (and equivalently between helices 1′ and 2), providing a second probable reason for the

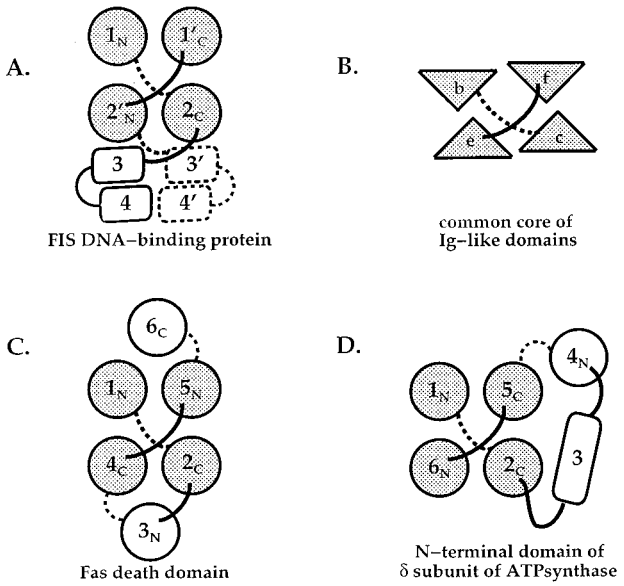

Figure 6. Topology diagrams of the bisecting U motif in the immunoglobulin fold and three helical proteins. (A) The dimerization domain of the factor for inversion stimulation (FIS) protein[69] is formed by the bisecting U. (B) β-Strands are indicated by triangles following the convention of Woolfsen et al.[84] Each triangle represents a strand with its apex depicting its orientation: pointing up, or down, indicates the strand nuns toward, or away, from the observer, respectively. The b, c and e, f strands form the conserved core of all Ig-like domains. β-Strands may be added to each end of these sheets in various proteins. A similar bisecting U topology is found in helical bundles as highlighted in (C) the Fas death domain[71] and (D) the N-terminal domain of the δ-subunit from *E. coli* ATP synthase.[72]

observed stereochemistry of the loops. These helices cross with a 19° angle (Figure 7D) close to the value found in parallel helical bundles.[75] The aromatic interactions at these interfaces, between residues Trp14 and Phe10 of one helix and Phe29 of the adjacent helix, may provide considerable stability. Specifically, the edge-to-face aromatic interaction of the Trp and Phe is very favorable.[76−79] This Trp is also the last residue in the helix and, together with the above-mentioned loop residues, serves to cap the hydrophobic core of α_2D. The face-to-face stacking between Phe10 and Phe29′ provides another example of the excellent geometric complementarity within the packing of α_2D; $\pi-\pi$ interactions may provide an additional driving force for this interaction.[77,78,80]

We have built anti models of α_2D, similar to that depicted schematically in Figure 5B with four antiparallel interfaces. It is possible to reproduce the good packing observed at the Leu-rich and His-rich interfaces in these models. However, the antiparallel interfaces between helices 1 and 2 are not as well packed in these models as the parallel interface observed in the NMR structure of α_2D. Thus, the geometric complementarity of the aromatic side chains at the parallel interface of α_2D must also contribute to the conformational specificity of this protein.

(75) Gonzalez, L., Jr.; Woolfson, D. N.; Alber, T. *Nat. Struct. Biol.* **1996**, *3*, 1011−8.
(76) Armstrong, K. M.; Fairman, R.; Baldwin, R. L. *J. Mol. Biol.* **1993**, *230*, 284−291.
(77) Hunter, C. A.; Sanders, J. K. M. *J. Am. Chem. Soc.* **1990**, *112*, 5525−5534.
(78) Hunter, C. A.; Singh, J.; Thornton, J. M. *J. Mol. Biol.* **1991**, *218*, 837−846.
(79) Burley, S. K.; Petsko, G. A. *Adv. Protein Chem.* **1988**, *39*, 125−189.
(80) Jorgensen, W. J.; Severance, D. L. *J. Am. Chem. Soc.* **1990**, *112*, 4768−4774.

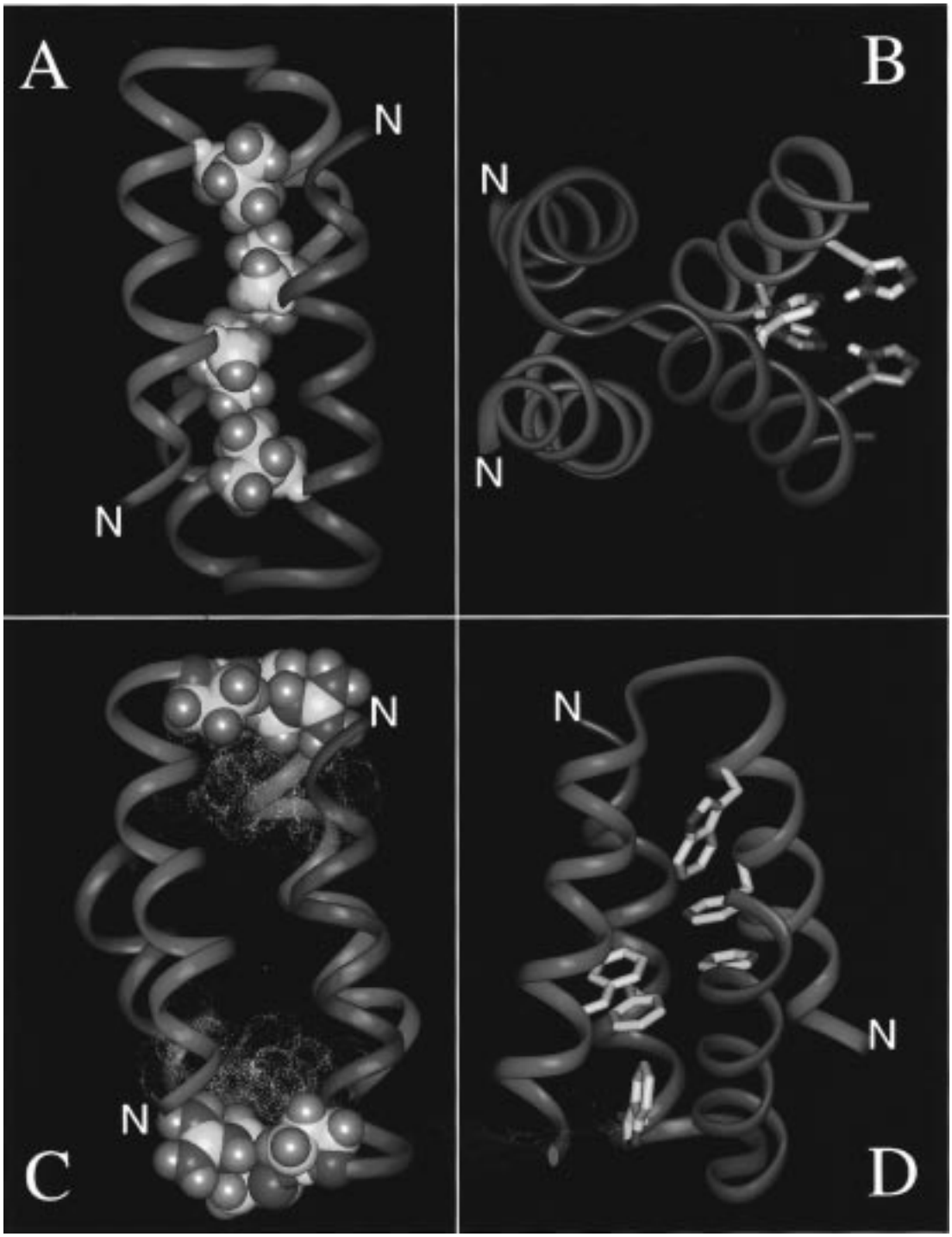

Figure 7. Four different views of the lowest energy structure of the homodimer α_2D. (A) Antiparallel interface between helix l and helix 1′ of α_2D with CPK surfaces shown for only the Leu6 and Leu13 side chains of each monomer. The radii used in generating the figure are identical to those used in X-PLOR for calculating the structures. The C^{α} atom is intentionally depicted smaller for clarity. (B) Aromatic cluster found in the antiparallel interface between helix 2 and helix 2′. The δ1 hydrogen of His30 of helix 2′ appears hydrogen bonded to the ϵ-nitrogen of His26 of helix 2. A similar interaction occurs between the symmetry-related histidines. This arrangement allows the δ1 hydrogens of each His26 to point into the hydrophobic core, leaving the unprotonated ϵ-nitrogens of each His30 available for hydrogen bonding with solvent. (C) CPK surfaces are used to depict loop residues (Pro17, Arg18, and Arg19) that cap the hydrophobic core, which is partly shown using a stippled van der Waals surface for residues Val3, Leu13, and Ile22. (D) Aromatic interactions observed between Trp14 and Phe10 of helix 1 and Phe29 of helix 2′ in the two identical parallel interfaces of α_2D that are related by C_2 symmetry. The figure was created using INSIGHTII (Biosym Technologies, San Diego).

Conclusion

This work provides the first high-resolution solution structure of a de novo designed protein of this size and complexity. This dimer consists of a total of 70 residues, indicating that it could theoretically adopt on the order of approximately 10^{70} different main chain conformers assuming 10 possible main conformers for each residue.[81] The finding that this protein adopts a single, well-defined structure in solution represents an important step in the engineering of proteins. Further, the designed fold was similar to the expected fold with respect to the lengths and

(81) Creighton, T. E. *Proteins: Structures and Molecular Properties*, 2nd ed.; W. H. Freeman and Co.: New York, 1993.

positions of the helices, and the presence of two antiparallel interfaces. However, other aspects of the structure were unanticipated. To successfully engineer a desired fold requires both positive and negative design elements; one must destabilize the undesired folds as well as stabilize the desired fold.[46,82] At the time that α_2D was designed, the bisecting U had not yet been recognized as a recurring folding motif. Our recognition of this motif has enriched the understanding of the structural repertoire of natural proteins.

Note Added in Proof. Since the completion of this work, Baltzer and co-workers have reported the NMR characterization of an aromatic cluster in a designed helix−loop−helix dimer, GTD-43.[85] Although a high-resolution structure of this dimer is not yet available, their data clearly show that the aromatic residues restrict the dynamics of GTD-43 in a manner similar to that of the aromatic groups in α_2D.

(82) Betz, S. F.; DeGrado, W. F. *Biochemistry* **1996**, *35*, 6955−62.
(83) Wishart, D. S.; Sykes, B. D. *J. Biomol. NMR* **1994**, *4*, 171−180.
(84) Woolfsen, D. N.; Evans, P. A.; Hutchinson, E. G.; Thornton, J. M. *Protein Eng.* **1993**, *6*, 561−470.
(85) Brive, L.; Dolphin, G. T.; Baltzer, L. *J. Am. Chem. Soc.* **1997**, *119*, 8598−8607.

Acknowledgment. We are indebted to Peter J. Domaille of the DuPont Merck Pharmaceutical Co. for NMR instrument time and computational assistance. We thank Gregg R. Dieckmann and Stephen F. Betz for assistance with graphics and peptide synthesis, Kevin R. MacKenzie for assistance in collecting the natural abundance ^{13}C-HMQC and graphics, James D. Lear for assistance with equilibrium centrifugation data collection and analysis, Carl H. Schwartz for ESMS analysis, Joel P. Schneider for assistance with peptide synthesis, and Stephen F. Betz for a critical reading of the manuscript. The above individuals plus James W. Bryson and Scott T. R. Walsh also contributed helpful discussions. This work was supported by NIH postdoctoral NRSA (to R.B.H.) 1 F32 GM18491-01 and NIH Grant GM54616-0 (to W.F.D.).

Supporting Information Available: A table containing the chemical shift assignments for α_2D, a table of the intermonomer NOEs determined, and a Ramachandran plot of α_2D (4 pages). See any current masthead page for ordering and Internet access instructions.

JA9733649