A Native-Like Three-α-Helix Bundle Protein from Structure-Based Redesign: A Novel Maquette Scaffold

Jonas S. Johansson,*,†,§ Brian R. Gibney,† Jack J. Skalicky,‡ A. Joshua Wand,‡ and P. Leslie Dutton†

Contribution from the Johnson Research Foundation, Department of Biochemistry and Biophysics, and the Department of Anesthesia, University of Pennsylvania, Philadelphia, Pennsylvania 19104, and Departments of Chemistry, Biological Sciences, and Biophysical Sciences and Center for Structural Biology, State University of New York, Buffalo, New York 14260

Received October 9, 1997

Abstract: A uniquely structured 65 amino acid helix−loop−helix′−loop−helix″ three-α-helix bundle, α_3-1, was designed and chemically synthesized, using the crystallographically characterized three stranded coiled coil "Coil-Ser", as a starting point. The circular dichroism spectrum of α_3-1 has a typical α-helical signature, with a $[\theta]_{222} = -22\,600$ deg·cm^2·dmol^{-1}, indicating a 80.5% α-helical content. Sedimentation equilibrium analytical ultracentrifugation revealed that α_3-1 is monomeric in solution. Consistent with the design parameters, the fluorescence emission maximum of the unique hydrophobic core tryptophan residue occurs at 324 nm. The evaluated ΔG^{H_2O} based on reversible guanidine hydrochloride denaturation is -4.6 ± 0.3 kcal·mol^{-1} ($m = 2.2 \pm 0.2$ kcal·mol^{-1}·M^{-1}) as measured by CD spectroscopy. The amide−aromatic region of the ^{1}H-NMR spectrum of α_3-1 illustrates excellent chemical shift dispersion and resolution. All 35 expected methyl correlations are accounted for in the ^{13}C-HSQC spectrum, providing stringent evidence for the existence of a native-like hydrophobic core. The monomeric nature of α_3-1 should facilitate NMR structural studies and kinetic protein folding analysis of the current design, and on future variants with engineered binding sites. The utility of this single-chain three-α-helix bundle framework for expanding the range of biochemical cofactors bound in maquettes is being explored.

Introduction

The *de novo* design of proteins is a powerful tool for exploring fundamental issues of protein folding, structure, and thermodynamics, as well as metal and cofactor binding.[1] Pioneering studies have shown that the design and synthesis of amphiphilic α-helices based upon the heptad repeat of amino acids observed in natural coiled coils is readily achieved, and that the helices can be designed to self-associate into bundles composed of from two to six members.[2] The ability to engineer uniquely structured hydrophobic cores has been a more formidable challenge, although recent success with four-α-helix bundles illustrates promise.[3] These studies indicate that the hydrophobic cores of α-helical bundles may be designed to assume native-like configurations by incorporating β-branched or aromatic amino acids at hydrophobic **a** or **d** positions of the heptad repeat, alternating with the conformationally more flexible leucines, which display two low-energy rotamers on an α-helical framework,[4] occupying the adjacent **a** or **d** positions.

Like the more studied four-α-helix bundles, three-α-helix bundles are a common native folding motif, being found, for example, in soluble proteins such as the structural protein spectrin[5] and the extramembraneous portion of the *Staphylococcus aureus* protein A.[6] The three-α-helix bundle motif is also found in the transmembrane spanning portion of membrane proteins such as the influenza virus hemagglutinin.[7] Herein, we report on the design and chemical synthesis of a monomeric three-α-helix bundle protein, designated α_3-1, composed of three different covalently linked α-helical segments (α−loop−α′−loop−α″), which folds into a unique conformation in solution as determined by NMR spectroscopy. This design represents

* Corresponding author. Mailing address: 780A Dulles, Hospital of the University of Pennsylvania, 3400 Spruce St., Philadelphia, PA 19104. Phone: (215) 349-5472. Fax: (215) 349-5078. E-mail: johansso @mail.med.upenn.edu.

† Department of Biochemistry and Biophysics, University of Pennsylvania.

§ Department of Anesthesia, University of Pennsylvania.

‡ State University of New York.

(1) (a) O'Neil, K. T.; DeGrado, W. F. *Science* **1990**, *250,* 646−651. (b) Regan, L. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22,* 257-281. (c) 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. (d) Struthers, M. D.; Cheng, R. P.; Imperiali, B. *Science* **1996**, *271,* 342−345. (e) Beasley, J. R.; Hecht, M. H. *J. Biol. Chem.* **1997**, *272,* 2031−2034. (f) Gibney, B. R.; Rabanal, F.; Dutton, P. L. *Curr. Opin. Chem. Biol.* **1997**, *1,* 537−542.

(2) (a) Hodges, R. S.; Saund, A. K.; Chong, P. C. S.; St-Pierre, S. A.; Reid, R. E. *J. Biol. Chem.* **1981**, *256,* 1214−1224. (b) Regan, L.; DeGrado, W. F. *Science* **1988**, *241,* 976−978. (c) Hill, C. P.; Anderson, D. H.; Wesson, L.; DeGrado, W. F.; Eisenberg, D. *Science* **1990**, *249,* 543−546. (d) Harbury, P. B.; Zhang, T.; Kim, P. S.; Alber, T. *Science* **1993**, *262,* 1401−1407.

(3) (a) Raleigh, D. P.; Betz, S. F.; DeGrado, W. F. *J. Am. Chem. Soc.* **1995**, *117,* 7558−7559. (b) Dolphin, G. T.; Brive, L.; Johansson, G.; Baltzer, L. *J. Am. Chem. Soc.* **1996**, *118,* 11297−11298. (c) Munson, M.; Balasubramanian, S.; Fleming, K. G.; Nagi, A. D.; O'Brien, R.; Sturtevant, J. M.; Regan, L. *Protein Sci.* **1996**, *5,* 1584−1593. (d) Gibney, B. R.; Rabanal, F.; Skalicky, J. J.; Wand, A. J.; Dutton, P. L. *J. Am. Chem. Soc.* **1997**, *119,* 2323−2324. (e) Jiang, X.; Bishop, E. J.; Farid, R. S. *J. Am. Chem. Soc.* **1997**, *119,* 838−839. (f) Roy, S.; Ratnaswamy, G.; Boice, J. A.; Fairman, R.; McLendon, G.; Hecht, M. H. *J. Am. Chem. Soc.* **1997**, *119,* 5302−5306. (g) Betz, S. F.; Liebman, P. A.; DeGrado, W. F. *Biochemistry* **1997**, *36,* 2450−2458.

(4) McGregor, M. J.; Islam, S. A.; Sternberg, M. J. *J. Mol. Biol.* **1987**, *198,* 295−310.

(5) Yan, Y.; Winograd, E.; Viel, A.; Cronin, T.; Harrison, S. C.; Branton, D. *Science* **1993**, *262,* 2027−2030.

(6) Starovasnik, M. A.; Skelton, N. J.; O'Connell, M. P.; Kelley, R. F.; Reilly, D.; Fairbrother, W. J. *Biochemistry* **1996**, *35,* 15558−15569.

(7) Bullough, P. A.; Hughson, F. M.; Skehel, J. J.; Wiley, D. C. *Nature* **1994**, *371,* 37−43.

S0002-7863(97)03538-5 CCC: $15.00
Published on Web 04/09/1998

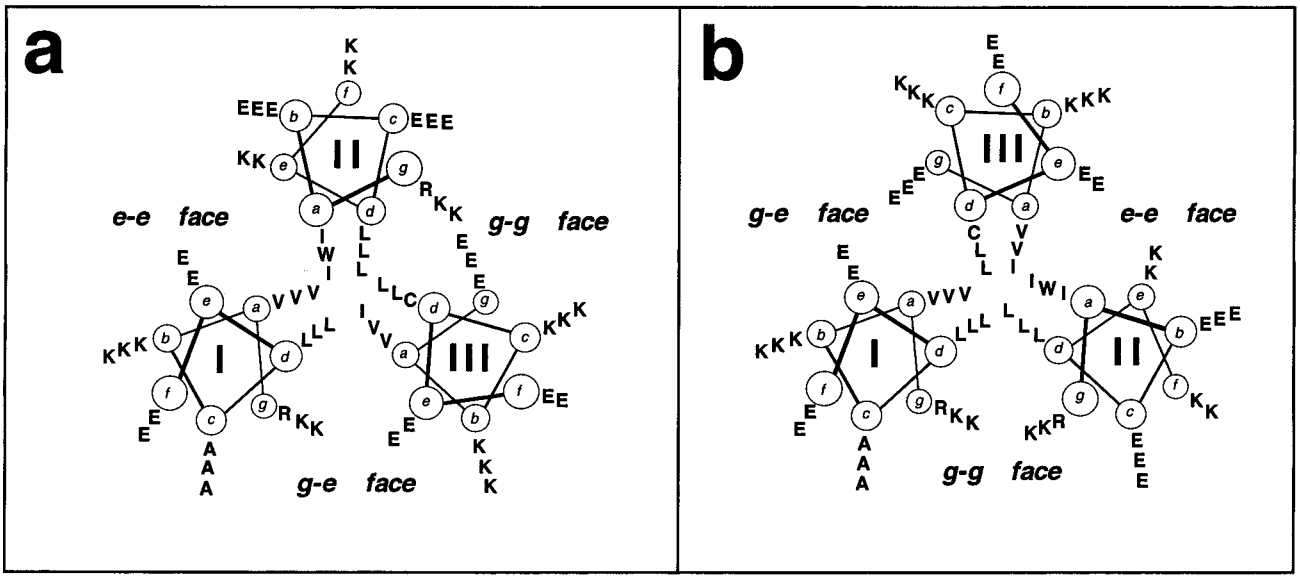

Figure 1. (a) Helical wheel diagram of α_3-1. Helix II is antiparallel to helices I and III, while helices I and III are oriented parallel to each other. Small italic letters indicate the heptad positions. Large uppercase letters are the amino acid assignments. The *e*−*e*, *g*−*g*, and *g*−*e* faces are designated using the terminology of Betz and DeGrado.[11] (b) Alternate arrangement of the three α-helices showing unfavorable electrostatic interactions at the *g*−*e* and *g*−*g* faces.

a novel protein architecture for the construction of maquettes, which are simplified functional versions of biological macromolecules.[8]

Experimental Section

Materials. Fmoc-protected amino acid perfluorophenyl esters were purchased from PerSeptive Biosystems (Framingham, MA) with the exception of Fmoc-L-Arg(Pmc)-OPfp, which was obtained from Bachem (King of Prussia, PA). NMR grade 2,2,2-trifluoroethanol was from Aldrich Chemical Co. (Milwaukee, WI). All other chemicals were reagent grade.

Protein Design. A 65 amino acid helix−loop−helix′−loop−helix″ protein (α_3-1, Figure 1) was derived from a structure-based redesign using the crystallographically characterized three stranded coiled coil, "Coil-Ser" (PDB accession code 1COS), as a backbone template for molecular modeling.[9] Since a single-chain version of coiled serine (>90 amino acids) represents a significant synthetic challenge,[10] we reduced the length of the helices from four heptads with eight hydrophobic core packing layers to two and a half heptads. The antiparallel arrangement of the three 19 amino acid helices provides six packing layers composed of either two **a** and one **d** or one **a** and two **d** heptad positions. Our hydrophobic core redesign specified that (1) each helix contain a unique sequence, (2) each heptad **a** position contain either a β-branched or an aromatic amino acid to help restrict the conformational flexibility of the adjacent leucine residues,[3bd,11] and (3) half of the layers adhere to a "small versus large" hydrophobic packing scheme observed in several native-like designed α-helical bundles.[12] Additionally, a single tryptophan was placed at the central **a** position of helix II as a spectroscopic probe and a cysteine was located at a **d** position on helix III as a unique site for future covalent modification.

Negative design[13] based on potential electrostatic interactions between glutamates and lysines/arginines at heptad **e** and **g** positions was utilized to construct the hydrophilic interfaces of the three helices. A clockwise global topology (Figure 1a) was designed to result in potentially favorable salt bridges among helix I**e**−helix II**e**, helix II**g**−helix III**g**, and helix I**g**−helix III**e**; the counterclockwise fold (Figure 1b) potentially results in unfavorable lysine/arginine−lysine/arginine interactions (helix I**g**−helix II**g**) and glutamate−glutamate interactions (helix III**g**−helix I**e**). Consequently, neutral alanine residues were placed at the **c** positions on helix I to avoid unfavorable electrostatic interactions with the adjacent heptad **f** or **g** positions. The three helices are joined by two $(glycine)_4$ linkers, chosen for the ability of glycine to adopt a wide range of backbone dihedral angles, spanning from heptad **d** to **g** positions. Finally, the model of α_3-1 was minimized using the Sybyl Software package (Tripos Associates, St. Louis, MO) on a Silicon Graphics Indigo^2 workstation, and visually inspected to ensure that no large packing defects were present in the hydrophobic core.

Protein Synthesis and Purification. The 65 amino acid protein, Ac-R·VKALEEK·VKALEEK·VKAL-GGGG-R·IEELKKK·WEE-LKKK·IEEL-GGGG-E·VKKCEEE·VKKLEEE·IKKL-$CONH_2$, was synthesized on a continuous flow Milligen 9050 solid phase synthesizer using single 60 min extended coupling cycles employing the fluorenylmethoxycarbonyl/*tert*-butyl protection strategy with NovaSyn PR-500 resin (0.33 mmol/g loading) on a 0.2 mmol scale. Pentafluorophenyl ester/1-hydroxybenzotriazole activation chemistry was employed for all amino acids. Following chain assembly, the N-terminus was manually acetylated, followed by washing with *N*,*N*-dimethylformamide and dichloromethane before cleavage from the resin with simultaneous side chain deprotection using 90:8:2 (v/v/v) trifluoroacetic acid (TFA)−ethanedithiol−water for 2 h. The crude peptide was precipitated and washed with cold ether, dissolved in water (0.1% v/v TFA), lyophilized, and purified to homogeneity by reversed phase C_{18} HPLC using an aqueous−acetonitrile gradient containing 0.1% (v/v) TFA. The molecular mass of analytically pure α_3-1 was confirmed with matrix-assisted laser desorption mass spectrometry. Protein concentration was determined using tryptophan absorbance,[14] taking $\epsilon_{280} = 5700\ M^{-1}\cdot cm^{-1}$.

Circular Dichroism Spectroscopy. Spectra were recorded with an Aviv Model 62 DS spectropolarimeter using 2 mm path length rectangular quartz cells. The cell holder was temperature controlled at 25.0 ± 0.1 °C. The buffer was 10 mM potassium phosphate at pH 7.0. The bandwidth was 1.00 nm, with a scan step of 0.5 nm and an average scan time of 3.0 s. Protein concentration was 27 μM. The α-helical content was calculated as described, using $[\theta]_{222} = 32\,000$ deg·cm^2·dmol^{-1} as 100% α-helix.[15]

(8) 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.; Dutton, P. L. *Nature* **1994**, *368*, 425−432.

(9) Lovejoy, B.; Choe, S.; Cascio, D.; McRorie, D. K.; DeGrado, W. F., Eisenberg, D. *Science* **1993**, *259*, 1288−1293.

(10) Synthetic methods for routine construction of large proteins are becoming available. See the use of natural chemical ligation in Hackeng, T. M.; Mounier, C. M.; Bon, C.; Dawson, P. E.; Griffin, J. H.; Kent, S. B. H. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 7845−7850.

(11) Betz, S. F.; DeGrado, W. F. *Biochemistry* **1996**, *35*, 6955−6962.

(12) (a) Munson, M.; O'Brien, R.; Sturtevant, J. M.; Regan, L. *Protein Sci.* **1994**, *3*, 2015−2022. (b) Ogihara, N. L.; Weiss, M. S.; DeGrado, W. F.; Eisenberg, D. *Protein Sci.* **1997**, *6*, 80−88.

(13) Hecht, M. H.; Richardson, J. S.; Richardson, D. C.; Ogden, R. C. *Science* **1990**, *249*, 884−891.

(14) Edelhoch, H. *Biochemistry* **1967**, 6, 1948−1954.

Sedimentation Equilibrium Ultracentrifugation. Sedimentation equilibrium analysis of 37 μM α_3-1 was performed on a Beckman XLI analytical ultracentrifuge operating at 30 000 rpm using both absorption and interference optics. The partial specific volume ($\bar{v}$) for the peptide was calculated from the residue-weighted average of the amino acid sequence using the method of Cohn and Edsall.[16] The density, ρ, of the solvent buffer was 1.0017 g·mL^{-1} (10 mM potassium phosphate, 100 mM KCl, pH 7.0) as measured using a Mettler Paar DMA60 density meter. The radial distribution absorbance scan data were fit to a single exponential using Igor Pro (Wavemetrics, Lake Oswego, OR). The buoyant molecular weight, M_b, was converted to the average molecular weight of the molecular species in solution, M_r, with the following relationship:

$$M_b = M_r(1 - \bar{v}\rho)$$

Fluorescence Spectroscopy. Steady-state intrinsic tryptophan fluorescence measurements on α_3-1 (18 μM) were carried out with an ISS K2 multifrequency cross-correlation phase and modulation spectrofluorometer. The quartz cell had a path length of 10 mm. The cell holder was thermostatically-controlled at 25.0 ± 0.1 °C. Excitation and emission slit widths were both 4 nm. A 295 nm cut-on filter was placed on the emission side. The buffer used for the fluorescence experimentation was 130 mM NaCl, 20 mM sodium phosphate, pH 7.0.

Denaturation Studies. Chemical denaturation of α_3-1 (15 μM protein, 10 mM potassium phosphate buffer, pH 7.0, 25 °C) was followed using circular dichroism spectroscopy, monitoring the ellipticity at 222 nm ($[\theta]_{222}$). The measured $[\theta]_{222}$ as a function of the added denaturant concentration was fit to an equation[17] describing the unfolding of a monomeric protein, using a nonlinear least-squares routine.

NMR Spectroscopy. Data were recorded on a Varian Inova spectrometer operating at a 600 MHz proton frequency. Spectra were acquired on a 1.0 mM α_3-1 sample prepared in 20 mM sodium phosphate buffer, pH 6.80, 50 mM KCl, 0.05 mM sodium azide, and 92% H_2O−8% D_2O at 308 K. NOESY[18] data were aquired with 1024 complex points in the directly detected dimension and 256 complex points in the indirect dimension. A spectral window of 7200 Hz was used for both dimensions. Natural abundance ^{13}C-HSQC data[19] were acquired with 1024 complex points in the directly detected proton dimension and 80 complex points in the carbon dimension. The carbon carrier was centered at 20 ppm, and acquired with a 4800 spectral window. A total of 256 transients were signal averaged for each free induction decay. Quadrature detection was obtained with States-TPPI[20] for the two-dimensional NMR experiments. Data were processed using Felix95.0 software (Molecular Simulations Inc., San Diego).

Results

Far-UV circular dichroism spectroscopy was performed in order to characterize the secondary structure of the designed protein α_3-1. The circular dichroism (CD) spectrum of α_3-1 (Figure 2) revealed a typical α-helical signature with a $[\theta]_{222}$ = −22 600 deg·cm^2·dmol^{-1}, indicating an 80.5% α-helical content, assuming that 57 of the total 65 amino acids make up the three α-helical segments, with the remaining eight glycines forming the two loops. Adding the α-helix stabilizing solvent 2,2,2-trifluoroethanol[21] (TFE) resulted in only a minor increase in the ellipticity at 222 nm (−24 400 deg·cm^2·dmol^{-1} in 50% TFE, corresponding to a 86.9% α-helical content), indicating that the helical potential of α_3-1 is almost fully achieved under benign conditions.

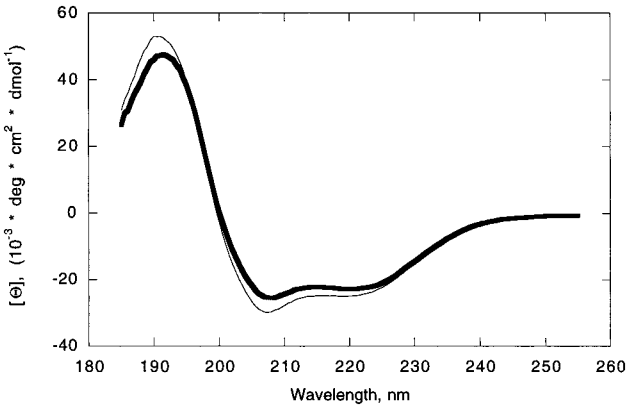

Figure 2. Far-UV CD spectrum of α_3-1 in benign buffer (thick line), and in the presence of 50% TFE (thin line).

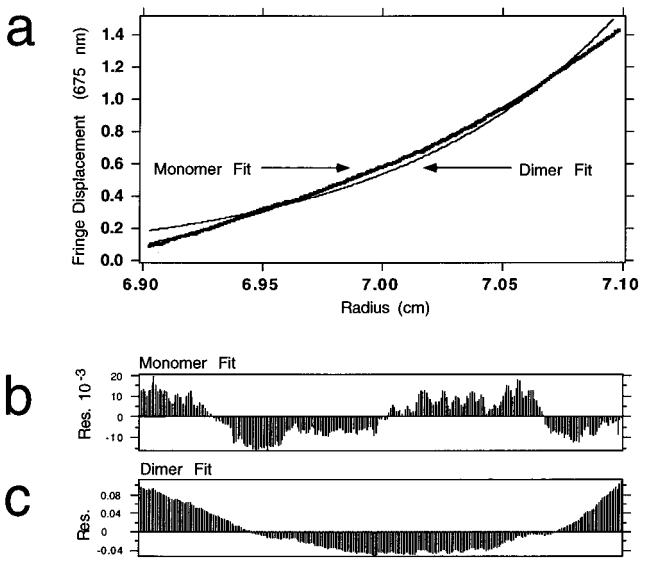

Figure 3. (a) Equilibrium analytical ultracentrifugation run for 37 μM α_3-1 (10 mM potassium phosphate, 100 mM KC1, pH 7.0) showing fringe displacement at 675 nm *versus* radius for the raw data and the superimposed monomer fit. The dimer fit is also shown for comparison. Residuals for the (b) monomer and (c) dimer fits of the data. Note the different scales for the residuals.

Analytical equilibrium sedimentation ultracentrifugation[16] was carried out in order to define the aggregation state of α_3-1, and demonstrates that the protein exists as a monomer in solution (Figure 3). A monomer model fit to the data is clearly superior to a dimer model fit as shown in Figure 3a. Furthermore, Figure 3b,c shows that the residuals for the monomer model fit display more random deviations around zero, while those for the dimer model fit have a more systematic distribution. Furthermore, the residuals for the monomer model fit are approximately 7-fold smaller than the residuals of the dimer model fit.

The position of the fluorescence emission maximum of tryptophan displays a blue shift in low dielectric environments,[22] allowing the location of this aromatic residue to be probed. Consistent with the design parameters, the fluorescence emission maximum (λ_{em}) of the unique tryptophan residue (W32) occurs at 324 nm (λ_{ex} = 280 nm), indicative of a solvent-protected indole ring at an **a** position in the hydrophobic core of the three-α-helix bundle (Figure 4).

(15) Lau, S. Y. M.; Taneja, A. K.; Hodges, R. S. *J. Biol. Chem.* **1984**, *259*, 13253−13261.

(16) Cohn, E. J.; Edsall, J. T. *Proteins, Amino Acids and Peptides as Ions and Dipolar Ions;* Reinhold Publishing Corp.: New York, 1943; pp 370−377 .

(17) Pace, C. N. *Methods Enzymol.* **1986**, *131*, 266−280.

(18) Macura, S.; Huang, Y.; Suter, D.; Ernst, R. R. *J. Magn. Reson.* **1981**, *43*, 259−281.

(19) (a) Muller, L. *J. Am. Chem. Soc.* **1979**, *101*, 4481−4484. (b) Bodenhausen, G.; Ruben, D. J. *Chem. Phys. Lett.* **1980**, *69*, 185−189.

(20) Marion, D.; Ikura, M.; Tschudin, R.; Bax, A. *J. Magn. Reson.* **1989**, *85*, 393−399.

(21) (a) Nelson, J. W.; Kallenbach, N. R. *Biochemistry* **1989**, *28*, 5256−5261. (b) Segawa, S-I.; Fukuno, T.; Fujiwara, K.; Noda, Y. *Biopolymers* **1991,** *31*, 497−509.

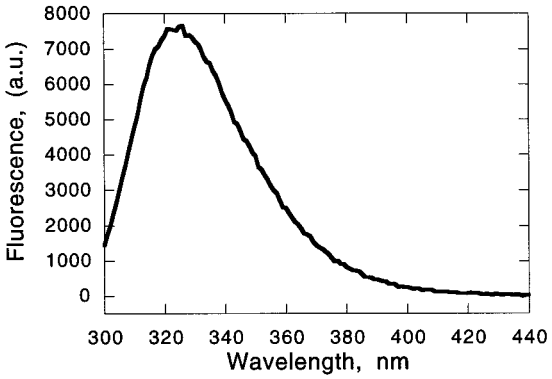

Figure 4. Fluorescence spectrum of W32 in α_3-1.

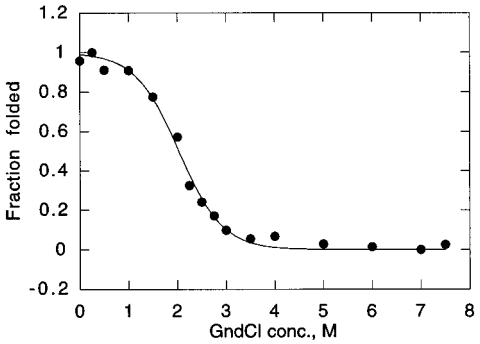

Figure 5. Reversible guanidine hydrochloride denaturation of α_3-1 as followed by CD spectroscopy at 222 nm (15 μM protein, 10 mM potassium phosphate buffer, pH 7.0).

The stability of the folded form of α_3-1 was determined in order to allow a comparison with previous α-helical bundle designs and also with small natural proteins. The global stability of α_3-1 was evaluated using isothermal guanidine hydrochloride denaturation, measuring the reversible loss of $[\theta]_{222}$ by CD spectroscopy (Figure 5). A two-state model[17] yields an apparent free energy of stabilization, $\Delta G^{H_2O} = -4.6 \pm 0.3$ kcal·mol^{-1}, with a midpoint of unfolding at 2.1 M guanidine hydrochloride and a slope of unfolding (m) of 2.2 ± 0.2 kcal·mol^{-1}·M^{-1}. The steepness of the guanidine hydrochloride denaturation curve indicates a highly cooperative unfolding event characteristic of small natural proteins.[23]

NMR analysis of α_3-1 was performed to determine the proton line widths and ranges of chemical shift dispersions used to qualitatively assess the secondary structure and the extent of structural singularity in solution. The amide-aromatic region of the one-dimensional ^{1}H-NMR spectrum of α_3-1 (Figure 6) illustrates a set of resonances having narrow line widths and excellent chemical shift dispersion. The amide proton shift dispersion is ∼2 ppm (see Figure 6) compared to the ∼0.5 ppm spread for all amino acids in the random coil conformation.[24] Likewise most H^α proton shifts are relatively upfield shifted to 4.0−4.3 ppm consistent with α-helical secondary structure (see Figure 7). All of the expected 35 methyl correlations are accounted for in the ^{13}C-HSQC spectrum, providing stringent evidence for the existence of a native-like hydrophobic core in solution (Figure 8). The full number of methyl correlations are counted assuming that the three very intense methyl correlations in Figure 8 are each the sum of two overlapping correlations. The three alanine $C^\beta H_3$, three isoleucine $C^\delta H_3$, and three

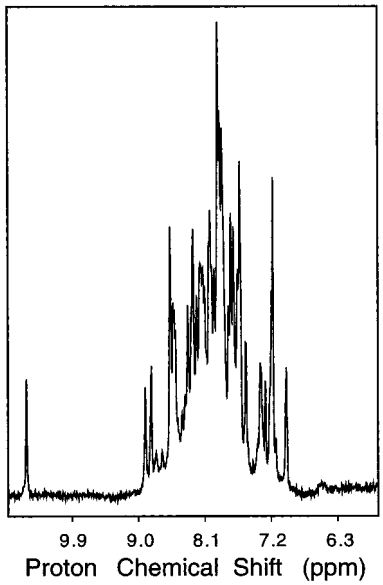

Figure 6. One-dimensional ^{1}H-NMR spectrum of the α_3-1 amide−aromatic proton region.

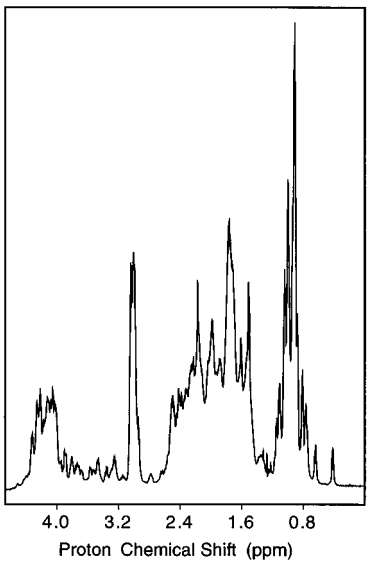

Figure 7. One-dimensional ^{1}H-NMR spectrum of the aliphatic proton region.

isoleucine $C^{\gamma 2}H_3$ methyl correlations are resolved and easily identified on the basis of their chemical shifts.[24] Likewise, the 10 valine $C^\gamma H_3$ and the 16 leucine $C^\delta H_3$ correlations are clustered in the lower half of the HSQC spectrum (see Figure 8). The effects of differential methyl group packing and the resulting range of magnetic environments are clearly evident with the three isoleucine $C^\delta H_3$ methyls that differ by 2.0 ppm carbon and 0.25 ppm proton chemical shift, and the three isoleucine $C^{\gamma 2}H_3$ methyls that differ by 1.6 ppm carbon and 0.11 ppm proton chemical shift. Likewise, the three alanine $C^\beta H_3$ methyls are clearly in different magnetic environments. The valine $C^\gamma H_3$ and leucine $C^\delta H_3$ methyls uniformly span the carbon chemical shift range of 21.0−26.7 ppm in contrast to the more restricted random coil chemical shift range of 20.3−21.1 ppm for valine, and 23.3−24.9 ppm for leucine. This spectral evidence supports the presence of a single highly populated conformer with placement of the methyl groups in slightly different magnetic environments, resulting in a spread of chemical shifts. The amide−amide region of the NOESY spectrum (Figure 9) shows the presence of many strong H^N−H^N NOEs consistent with α-helical secondary structure. Several

(22) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy;* Plenum Press: New York, 1983; p 354.

(23) Myers, J. K.; Pace, C. N.; Scholtz, J. M. *Protein Sci.* **1995,** *4,* 2138−2148.

(24) Wishart, D. S.; Bigam, C. G.; Holm, A.; Hodges, R. S.; Sykes, B. D. *J. Biomol. NMR* **1995**, *5,* 67−81.

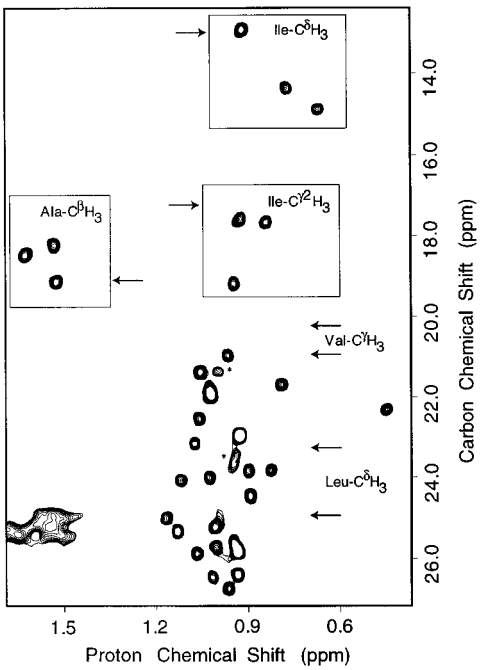

Figure 8. Expansion of the α_3-1 ^{13}C-HSQC spectrum showing the methyl region. Isoleucine C$^{\delta}$H$_3$ and C$^{\gamma 2}$H$_3$ and alanine C$^{\beta}$H$_3$ correlations are boxed, and the carbon random coil values are indicated with arrows. The random coil carbon chemical shift ranges for the prochiral valine C$^{\gamma}$H$_3$ and leucine C$^{\delta}$H$_3$ methyl resonances are shown with arrows. Intense cross-peaks at [proton, carbon] shifts of [1.02, 21.9], [0.93, 23.0], and [0.94, 25.8] are considered degenerate peaks. Asterisks indicate weak correlations that are not attributed to methyls from the major conformer and may represent strong correlations from a minor conformer *versus* leucine C$^{\gamma}$H methine correlations.

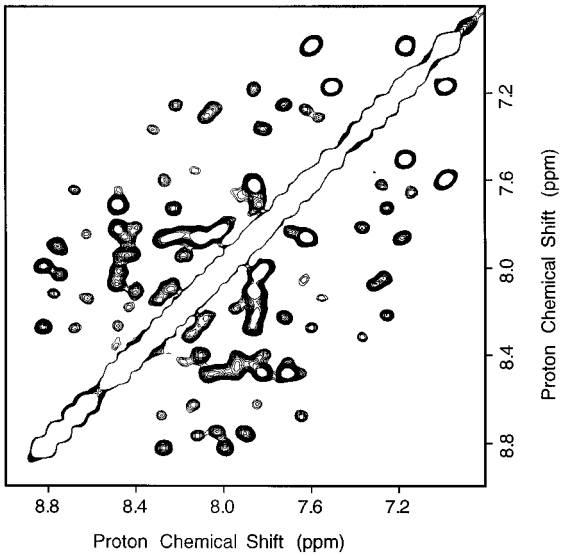

Figure 9. Expansion of the α_3-1 2D-NOESY spectrum showing the amide−amide region.

NOESY correlations are observed between the tryptophan side chain H$^{\epsilon 1}$ proton and aliphatic protons at methyl proton and methylene proton shifts consistent with a buried tryptophan side chain (not shown).

Discussion

The design, synthesis, and initial characterization of a chemically synthesized single-chain 65 amino acid protein, α_3-1, that folds into a native-like three-α-helix bundle structure in solution is described. Of note is that the current three-α-helix bundle design is composed of three structurally distinct α-helices that are covalently linked. The covalent linking facilitates the incorporation of structurally heterogeneous α-helices, and there is therefore considerably greater asymmetry in the resulting hydrophobic core of α_3-1 than in previous homotrimeric $(\alpha)_3$ designs,[9,12b,25] which should be of importance for the future incorporation of specific binding sites for small ligands, cofactors, and metals.

High-resolution solid-state X-ray crystal structures of two homotrimeric ($(\alpha)_3$) three-α-helix bundles, GCN4-pII and Coil-V_aL_d, have recently been reported.[12b,25] Solution-state sedimentation equilibrium analysis also indicates that Coil-V_aL_d aggregates as a homotrimer.[26] The principle difference between both GCN4-pII and Coil-V_aL_d and the current α_3-1 design is that the former two three-α-helix bundles are homotrimeric, while α_3-1 is monomeric. Furthermore, both GCN4-pII and Coil-V_aL_d form parallel three-α-helix bundles, while α_3-1 folds in an antiparallel conformation. The antiparallel helix topology of α_3-1 is a consequence of covalently joining the three α-helical sequences with a pair of glycine linkers.

The current design, α_3-1, has a total of six different hydrophobic core layers consisting of VLV, LIC, VLV, LWL, VLI, and LIL, in both the clockwise and the anticlockwise orientations of the helices shown in Figure 1. These six hydrophobic core layers are composed of either two **a** and one **d** or one **a** and two **d** heptad positions, because of the antiparallel orientation of the helices. In contrast, the hydrophobic core of Coil-V_aL_d consists of four layers of **d** position leucines (LLL), alternating with four layers of **a** position valines (VVV), for a total of eight hydrophobic layers.[12b] The GCN4-pII homotrimeric three-α-helix bundle also has eight hydrophobic core layers,[25] for a total of twelve **d** and twelve **a** positions, all composed of isoleucines (III).

The antiparallel orientation of α_3-1 results in *e*−*e*, *g*−*g*, and *g*−*e* interhelical interfaces according to the terminology of Betz and DeGrado.[11] All three helices in α_3-1 have two **e** positions, which in helices I and III are occupied by glutamates, while helix II has two lysines at these positions. Each helix in α_3-1 has three **g** positions, which in helices I and II are occupied by an arginine followed by two lysines, while helix III has three **g** position glutamates. While the three interfaces in α_3-1 were designed to allow favorable electrostatic interactions between neighboring helices as described,[27] the actual assignment of residues had to be altered from that used in these earlier studies due to the antiparallel orientation of the helices in α_3-1. It should be pointed out that the crystal structure of Coil-Ser had unfavorable electrostatic interactions along the entire length of two of the three hydrophilic interfaces,[9] suggesting that the side chains are flexible enough to avoid each other and that this arrangement is not necessarily destabilizing. The alignment of the α-helices in the anticlockwise arrangement as shown in Figure 1b is therefore not inconceivable, and the precise topology of the helices in α_3-1 will require further structural studies.

In contrast, the parallel orientation of both the GCN4-pII and the Coil-V_aL_d three-α-helix bundles results in three identical interhelical *g*−*e* interfaces.[12b,25] The four **e** positions in GCN4-pII are occupied by two glutamates, a leucine and a lysine, while the five **g** positions are occupied by an arginine, two lysines, a glutamate, and a leucine.[25] There are therefore three potential salt bridges per *g*−*e* interface in GCN4-pII: arginine−glutamate

(25) Harbury, P. B.; Kim, P. S.; Alber, T. *Nature* **1994**, *71*, 80−83.
(26) Boice, J. A.; Dieckmann, G. R.; DeGrado, W. F.; Fairman, R. *Biochemistry* **1996**, *35*, 14480−14485.
(27) (a) Nautiyal, S.; Woolfson, D. N.; King, D. S.; Alber, T. *Biochemistry* **1995**, *34*, 11645−11651. (b) Lombardi, A.; Bryson, J. W.; DeGrado, W. F. *Biopolymers* **1997**, *40*, 495−504.

and two glutamate−lysine pairs. The four **e** positions in Coil-V_aL_d are occupied by glutamates, while three of the four **g** positions are lysines, with the final **g** position being a glutamate.[12b] There are consequently two observed salt bridges per *g*−*e* interface in Coil-V_aL_d, each involving a glutamate−lysine pair.

The remaining hydrophilic heptad **b**, **c**, and **f** positions in α_3-1 were assigned to either glutamates or lysines in an alternating fashion in order to avoid adjacent like charges. An exception was made at the three **c** positions on helix I, where alanines were placed in order to prevent a potentially unfavorable clustering of glutamates along the *g*−*e* interface between helices I and III (Figure 1a). GCN4-pII, a variant of a natural protein, contains a greater variety of amino acids at the hydrophilic **b**, **c**, and **f** positions; eleven different amino acids are featured,[25] compared to the three in α_3-1. In Coil-V_aL_d, all four **c** positions are occupied by alanines while the four **b** positions have two glutamates, an alanine, and a glutamine, and the three **f** positions feature a serine flanked by two lysines.[12b]

The current design, α_3-1, contains a bulky tryptophan at the central **a** position of helix II, which serves as a spectroscopic probe. The position of the fluorescence emission maximum indicates that the indole ring is well protected from the surrounding aqueous solvent in the hydrophobic core of the three-α-helix bundle. Earlier work with designed four-α-helix bundle systems has shown that tryptophan residues placed at hydrophobic heptad **a** positions are shielded from the aqueous solvent and display fluorescence emission maxima of 324 nm[28] and 329 nm.[29] The current study therefore indicates that the three-α-helix bundle motif is likewise able to form a well-defined hydrophobic environment that effectively excludes bulk water molecules. This suggests that a suitable solvent-shielded core exists in this small protein, which holds promise for the future incorporation of hydrophobic binding sites.

Compared to previous maquette scaffolds based on a four-α-helix bundle motif,[3d,30] α_3-1 is considerably less stable, consistent with its smaller size and limited hydrophobic core content. Each of the six hydrophobic core layers of α_3-1 consists of three amino acids, while the earlier four-α-helix bundle maquettes have eight hydrophobic core layers, each consisting of four amino acids. However, these four-α-helix bundles are dimeric, which makes direct conformational stability comparisons difficult. Thermodynamic studies on the monomeric 75 amino acid four-α-helix bundle M-60[3f] yielded a free energy of unfolding of ca. −4 kcal·mol^{-1}, comparable to the −4.6 kcal·mol^{-1} found in the present study for α_3-1. The larger monomeric 108 residue four-α-helix bundle DHP_1 (designed helical protein 1)[31] has a conformational stability of −9.3 kcal mol^{-1}. These values are all comparable to those determined for small monomeric natural proteins[32] such as RNase A (−6.5 kcal·mol^{-1}), ferricytochrome *c* (−8.9 kcal·mol^{-1}), and myoglobin (−7.6 kcal·mol^{-1}). The finding that α_3-1 falls at the lower end of the conformational stability range is consistent with its small size, its limited hydrophobic core content, and the small number of potential electrostatic interactions along the three α-helical interfaces. Nevertheless, the data indicate that less stable, yet conformationally unique, proteins can be achieved, using simple design rules and structure-based redesign.

The NMR spectroscopy results support a single populated conformer in solution with α-helical secondary structure. Strong H^N−H^N correlations in the NOESY spectrum and upfield-shifted H^α resonances are characteristic of this class of secondary structure. Compared to previous four-α-helix maquette scaffolds the dispersion of the amide proton and methyl carbon and proton chemical shift dispersions is larger in the four-α-helix maquettes. This is not surprising since the four-α-helix bundle contains twenty aromatic groups compared to one aromatic group per three-α-helix bundle. Nonetheless, the methyl groups of α_3-1 are mostly in unique environments as shown in the ^{13}C-HSQC which is quite demonstrative of a hydrophobic core packed to van der Waals surfaces. A number of correlations at ∼10% the intensity of the major correlations, below the contour level of Figure 8, may be attributed to a minor conformer in solution. This minor conformation may be attributed to a counterclockwise-folded α_3-1 (Figure 1b), or may represent a dimeric form of α_3-1, since the NMR studies use a 20-fold higher concentration of protein than the equilibrium sedimentation ultracentrifugation experiment. Finally, these low-intensity signals may represent leucine contributions from $C^\gamma H$ methine correlations. Apparent minor conformers have been observed in other designed proteins.[3d] Overall these NMR data indicate achievement of a well-structured α_3-1 protein in solution.

The construction of a maquette scaffold based on a single polypeptide chain containing three unique helices (helix−loop−helix′−loop−heIix″) potentially allows for the design of asymmetric cofactor and small molecule binding sites. Current maquettes based on self-associating peptide dimers (with C_2 symmetry[33]) are limited by their architecture which results in the formation of two identical binding sites *per* four-α-helix bundle[28,34] with variable topology. While conformationally specific homo- and heterotrimeric three stranded coiled coils have been described previously (on the basis of electrostatic interactions),[26,27] the resulting array of global topological isomers limits their utility in maquette design. By synthesizing a continuous polypeptide three-α-helix bundle which uses negative design to favor a single topology and displays a conformationally specific hydrophobic core, we have generated a novel maquette scaffold. Currently, the utility of this three-α-helix bundle framework to expand the range of biochemical cofactors bound in maquettes is being explored. The monomeric nature of α_3-1 should facilitate NMR structural studies and kinetic protein folding analysis of the current design, and on future variants with engineered biological cofactor binding sites.

Acknowledgment. This work was supported by a FAER Young Investigator Award and NIH Grants GM55876 to J.S.J., GM35940 to A.J.W., and GM41048 to P.L.D., with NSRA grants GM17816 to B.R.G. and GM18121 to J.J.S., and in part by the NSF [DMR91-20668]. The Protein Chemistry Laboratory of the University of Pennsylvania performed mass spectrometry. We thank Dr. Jim Lear for assistance with the ultracentrifugation experiments.

Note Added in Proof. After we submitted this manuscript, it came to our attention that Professor Degrado has a manuscript in press in *Protein Sci.* describing a different single-chain native-like three-α-helix bundle.

JA973538M

(28) Johansson, J. S.; Gibney, B. R.; Rabanal, F.; Reddy, K. S.; Dutton, P. L. *Biochemistry* **1998**, *37*, 1421−1429.
(29) Handel, T. M.; Williams, S. A.; Menyhard, D.; DeGrado, W. F. *J. Am. Chem. Soc.* **1993**, *115*, 4457−4460.
(30) Gibney, B. R.; Johansson, J. S.; Rabanal, F.; Skalicky, J. J.; Wand, A. J.; Dutton, P. L. *Biochemistry* **1997**, *36*, 2798−2806.
(31) Schafmeister, C. E.; LaPorte, S. L.; Miercke, L. J. W.; Stroud, R. M. *Nat. Struct. Biol.* **1997**, *4*, 1039−1046.
(32) Makhatadze, G. I.; Privalov, P. L. *Adv. Protein Chem.* **1995**, *47*, 307−425.
(33) Skalicky, J. J.; Bieber, R. J.; Gibney, B. R.; Rabanal, F.; Dutton, P. L.; Wand, A. J. *J. Biomol. NMR*, in press.
(34) (a) Rabanal, F.; DeGrado, W. F.; Dutton, P. L. *J. Am. Chem. Soc.* **1996**, *118*, 473−474. (b) Gibney, B. R.; Mulholland, S. E.; Rabanal, F.; Dutton, P. L. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 15041−15046.