Solution Structure of a Designed Four-α-Helix Bundle Maquette Scaffold

ARTICLE *in* JOURNAL OF THE AMERICAN CHEMICAL SOCIETY · MAY 1999

Impact Factor: 12.11 · DOI: 10.1021/ja983309f

6 AUTHORS, INCLUDING:

Brian R Gibney
City University of New York - Brooklyn College
78 PUBLICATIONS **2,927** CITATIONS
SEE PROFILE

Ramona J Bieber Urbauer
University of Georgia
39 PUBLICATIONS **963** CITATIONS
SEE PROFILE

Andrew Joshua Wand
University of Pennsylvania
193 PUBLICATIONS **8,724** CITATIONS
SEE PROFILE

Solution Structure of a Designed Four-α-Helix Bundle Maquette Scaffold

Jack J. Skalicky,† Brian R. Gibney, Francesc Rabanal,‡ Ramona J. Bieber Urbauer, P. Leslie Dutton, and A. Joshua Wand*

Contribution from the Johnson Research Foundation, Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Received September 15, 1998. Revised Manuscript Received March 11, 1999

Abstract: The solution structure of a de novo designed disulfide-bridged two-α-helix peptide that self-assembles to form a 2-fold symmetric four-α-helix bundle protein $(\alpha'\text{-SS-}\alpha')_2$ has been solved by NMR spectroscopy. The 33-residue peptide, (α′-SH), that is the basic building block of the bundle has been recombinantly expressed. The three-dimensional structure of the asymmetric unit of the bundle has been determined using interproton distance restraints derived from the nuclear Overhauser effect (NOE), covalent torsion angle restraints derived from three bond scalar coupling constants, and longer range angular restraints derived from residual dipolar couplings. The covalent α′-SS-α′ unit forms a pair of parallel α-helices that use heptad a-, d-, e-, and g-side chains to form a hydrophobic core extending the length of the molecule. The distribution of polar and nonpolar side chains on the surface of α′-SS-α′ structure is asymmetric. The hydrophilic face is comprised of glutamate and lysine side chains, while the opposite face is comprised of leucine, isoleucine, phenylalanine, tryptophan, and neutral histidine side chains. Equilibrium sedimentation analysis, size-exclusion chromatography, pulsed field gradient translation diffusion measurements, and a rotational correlation time derived from ^{15}N NMR relaxation studies all indicate that the covalent α′-SS-α′ unit forms a noncovalent dimer, $(\alpha'\text{-SS-}\alpha')_2$, in solution. The structure confirms many expected design features and illuminates an apparent dichotomy of structure where the helical interface of the disulfide bridged two-α-helix peptide appears nativelike while the adjacent, noncovalent interface shows non-nativelike behavior. Available evidence indicates the four α-helix bundle can adopt either an anti or syn topology. The structure is discussed with respect to the potential origins of conformational specificity and nativelike protein structure.

Introduction

The characterization of numerous de novo designed α-helical and coiled coil proteins has revealed the crucial features required for stabilizing this ubiquitous structural element.[1−3] Side chain α-helix-forming propensity and their contribution to helix thermodynamics is well understood.[4,5] The heptad repeat and its role in helix supercoiling,[6] the packing of aromatic and aliphatic side chains in a hydrophobic core,[7−9] and the placement of polar and charged side chains to both promote water solubility and to stabilize or destabilize intra- or intermolecular electrostatic interactions have also been extensively investigated.[1,10] This array of information has been used to design and structurally characterize both all-parallel and anti-parallel α-helix-containing proteins. Nevertheless, many of the early de novo helical proteins displayed biophysical properties that are unlike those of natural proteins and lack a variety of features indicative of structural specificity.[2] Introducing structural specificity remains a major design challenge although significant advances using metal ions that drive local specificity[11] and computer algorithms to optimally pack protein cores[12−14] have been reported. Recently, substitution of aromatic[9] and β-branched side chains for the heptad a- and d-positions have also been successfully employed to introduce structural specificity.[15] The latter approach has led to several de novo designed proteins that adopt single solution conformations that may provide clues to the origins of structural specificity in this family of proteins.[15,16]

* To whom correspondence should be addressed.
† Present address: Department of Chemistry, State University of New York at Buffalo, Buffalo, NY 14260.
‡ Present address: Department de Quimica Organica, Universitat de Barcelona, Marti I Franques, 1-11, 080028 Barcelona.

(1) DeGrado, W. F.; Wasserman, Z. R.; Lear, J. D. *Science* **1989**, *243*, 622−8.
(2) Betz, S. F.; Bryson, J. W.; DeGrado, W. F. *Curr. Opin. Struct. Biol.* **1995**, *5*, 457−63.
(3) Fezoui, Y.; Weaver, D. L.; Osterhout, J. J. *Protein Sci.* **1995**, *4735*, 286−295.
(4) O'Neil, K. T.; DeGrado, W. F. *Science* **1990**, *250*, 646−51.
(5) Padmanabhan, S.; Marqusee, S.; Ridgeway, T.; Laue, T. M.; Baldwin, R. L. *Nature* **1990**, *344*, 268−70.
(6) Crick, F. H. C. *Acta Crystallogr.* **1953**, *6*, 689−697.
(7) Hodges, R. S.; Zhou, N. E.; Kay, C. M.; Semchuk, P. D. *Pept. Res.* **1990**, *3*, 123−37.
(8) Zhou, N. E.; Kay, C. M.; Hodges, R. S. *Biochemistry* **1992**, *31*, 5739−46.
(9) Dolphin, G. T.; Brive, L.; Johansson, G.; Baltzer, L. *J. Am. Chem. Soc.* **1996**, *118*, 11297−11298.
(10) Zhou, N. E.; Kay, C. M.; Hodges, R. S. *J. Mol. Biol.* **1994**, *237*, 500−512.
(11) Handel, T. M.; Williams, S. A.; DeGrado, W. F. *Science* **1993**, *261*, 879−85.
(12) Desjarlais, J. R.; Handel, T. M. *Protein Sci.* **1995**, *4*, 2006−18.
(13) Dahiyat, B. I.; Mayo, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10172−7.
(14) Hellinga, H. W. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10015−7.
(15) Gibney, B. R.; Rabanal, F.; Skalicky, J. J.; Wand, A. J.; Dutton, P. L. *J. Am. Chem. Soc.* **1997**, *119*, 2323−2324.

10.1021/ja983309f CCC: $18.00
Published on Web 05/06/1999

Very few solution structures of de novo designed proteins have been described, presumably due to the lack of structural specificity observed in many designed proteins.[17−19] Here a high-resolution model, determined by NMR-based methods, is reported for a disulfide bridged di-α-helical peptide structure. This structure provides a reference point for interpreting the remarkable single amino acid switch for conformational specificity[15] and sets the stage for structure-based redesign of the $(\alpha'\text{-SS-}\alpha')_2$. Last, this structure provides a platform to introduce biological function within the context of a "nativelike" protein structure.

Experimental Section

Design Strategy. The prototype protein $(\alpha\text{-SS-}\alpha)_2$ (formerly called H10H24) was previously designed[20,21] on the basis of general features of the minimalist α_2 designed protein[1,22,23] and the natural cytochrome bc_1 complex.[24] Each constituent peptide of the disulfide-bridged di-α helical monomer is identical and comprises approximately 3.5 heptad repeats containing variations on the prototype $(Leu_a\text{-}Glu_b\text{-}Glu_c\text{-}Leu_d\text{-}Leu_e\text{-}Lys_f\text{-}Lys_g)_n$ heptad and has an N-terminal Cys-$(Gly)_3$ linker. Heptad modifications based on cytochrome bc_1 complex include the substitution of histidines for Leu_a at His^{10} and His^{24} (see sequence below). These histidine side chains are required for heme binding and are spaced two heptads, or 14 residues, apart as observed in the cytochrome bc_1 complex (His^{97} and His^{111}; *Rhodobacter capsulatus* sequence). Phe^{17} was substituted for the central Leu_a and corresponds to Phe^{104} in the cytochrome bc_1 complex sequence and serves to separate the heme binding sites in the complex. Additionally, Arg^{27} was substituted for a Leu_d and corresponds to Arg^{94} in the cytochrome bc_1 complex which modulates the redox potential of the $His^{24}/H^{24'}$ bis-ligated heme. Finally, the minimalist heptad has only g-position lysines, whereas α and α′ have Leu^9, Lys^{16}, Leu^{23}, and Lys^{30}. The prototype $(\alpha\text{-SS-}\alpha)_2$ displays characteristics of a folded helical protein of non-nativelike structure. The product of an iterative redesign protocol, the double helix variant of this sequence designated $(\alpha'\text{-SS-}\alpha')_2$, previously named [H10H24-L6I,L13F], substitutes β-branched and aromatic hydrophobic side chains, Ile^6 and Phe^{13}, for the d-position leucines of the prototype. These substitutions are responsible for transforming the apoprotein to a nativelike and uniquely structured protein that is the subject of this

prototype design

(α) CGGGE**L**^{6}WKL·**H**^{10}EE**L**^{13}LKK·**F**^{17}EELLKL·**H**^{24}EE**R**^{27} LKK·L

double variant

(α′) GSCGGGE**I**^{6}WKL·**H**^{10}EE**F**^{13}LKK·**F**^{17}EELLKL·**H**^{24}EE**R**^{27} LKK·L

Recombinant Protein Expression and Purification. A synthetic α′ protein gene was cloned into a modified pET-32b (Novagen, Inc.) vector that contains the isopropyl β-D-thiogalactopyranoside (IPTG)-inducible T7 promotor and a thioredoxin fusion protein. To limit the N-terminal extension to a Gly-Ser dipeptide (see α′ sequence above) a *Bam*HI restriction site was introduced within the thrombin cleavage site of the vector. The gene encoding the 33 amino acid α′ protein was assembled from synthetic DNA oligonucleotides. Termini of the assembled double-stranded DNA were digested with *Bam*HI and *Xho*I and then directionally ligated into linearized vector previously digested with the same restriction enzymes. This construct was then used to transform BL21(DE3) *Escherchia coli* cells for protein overexpression.

The $(\alpha'\text{-SS-}\alpha')_2$ protein was prepared as follows. A starter culture consisting of 50 mL of standard M9 minimal media was inoculated with BL21-(DE3) *E. coli* cells containing the expression vector and grown for 12−15 h at 35 °C with continuous shaking. A flask containing 1/L of M9 media (35 °C) was then inoculated with the starter culture, and cells were grown to an A^{600} of 0.6. The cells were induced with 1 mM IPTG and grown for an additional 4−5 h. The cells were isolated by centrifugation at 4 °C, resuspended in 4−6 volumes of binding buffer (20 mM Tris pH 7.9, 500 mM NaCl, 5 mM Imidazole), centrifuged a second time and resuspended in 4−6 volumes of binding buffer, and flash frozen in liquid nitrogen. The cells were thawed and continuously stirred for 30 min with 50 μg/mL hen egg white lysozyme (Sigma Chemical) at room temperature, sonicated for 4 min on ice, and centrifuged to remove cellular debris. The supernatant was loaded on a 10-mL Ni^{2+} Sepharose (Novagen) column and extensively washed with binding buffer. The thioredoxin-α′ fusion protein was eluted with 50 mL of 20 mM Tris pH 7.9 containing 500 mM NaCl and 500 mM imidazole. The eluent ($A^{280} > 0.05$) was extensively dialyzed against thrombin cleavage buffer (20 mM Tris pH 8.4, 140 mM NaCl, 2.5 mM $CaCl_2$ and then centrifuged to remove particulates. The fusion protein was incubated with ∼50 NIH units of human thrombin (Sigma) for 12−15 h at room temperature. Quantitative cleavage of the α′ from the thioredoxin was achieved as judged by SDS gel electrophoresis. Thrombin was removed by passage over a 2-mL column of benzamidine−sepharose and then acidified with trifluoroethanol to 0.1% (v/v) (Pierce, Sequanal grade) and filtered through an 0.22-μM disposable filter immediately prior to HPLC purification. The α′ peptide was HPLC-purified by reverse phase C18 chromatography (Waters μBondapack C18) with a 10%−50% acetonitrile (0.1%TFA) linear gradient. The α′ peptide eluted at 35% acetonitrile and was flash frozen with liquid nitrogen and lyophilized to dryness. The purified peptide was the oxidized covalent molecule, α′-SS-α′, judged by different HPLC retention times for reduced and oxidized forms of the peptide. A yield of 15.0 mg/L was routinely obtained.

Uniform ^{15}N-labeled protein was prepared from cells grown in M9 minimal media containing 1.0 g/L of 99% $^{15}NH_4Cl$ (Isotec) as the sole nitrogen source. Uniform ^{15}N- and ^{13}C-enriched protein was produced from cells grown in M9 minimal media containing 1.0 g/L 99% $^{15}NH_4Cl$ and 2.0 g/L 99% [$^{13}C_6$]-glucose (Isotec) as the sole nitrogen and carbon sources. Proteins were prepared for NMR spectroscopy at concentrations of 1−2 mM in buffer containing 20 mM sodium phosphate (pH 6.60), 50 mM KCl, 0.01 mM NaN_3 dissolved in 92% H_2O and 8% D_2O.

Equilibrium Sedimentation. Equilibrium sedimentation analysis was performed on a Beckman XLI analytical ultracentrifuge operating at 30 000 rpm. Initial peptide, α′-SS-α′, concentrations were either 160 nM, 8 μM or 175 μM in 10 mM potassium phosphate, 100 mM KCl, D_2O (pH 7.0, uncorrected). The partial specific volume, ($\bar{v}$), for the peptide (0.7616 mg/L) was calculated from the residue-weighted average of the amino acid sequence using the method of Cohn and Edsall.[25] The density, ρ, of the solvent buffer was 1.1077 g mL^{-1} 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). The buoyant molecular weight, M_b, was converted to the average molecular weight of the molecular species in solution, M_r, with eq 1

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

NMR Spectroscopy. All NMR experiments were performed on Varian Inova spectrometers operating at 500, 600, or 750 MHz proton

(16) Skalicky, J. J.; Bieber, R. J.; Gibney, B. R.; Rabanal., F.; Dutton, P. L.; Wand, A. J. *J. Biomol. NMR* **1998**, *2*, 227−228.
(17) Saudek, V.; Pastore, A.; Castiglione Morelli, M. A.; Frank, R.; Gausepohl, H.; Gibson, T.; Weih, F.; Roesch, P. *Protein Eng.* **1990**, *4*, 3−10.
(18) Kuroda, Y.; Nakai, T.; Ohkubo, T. *J. Mol. Biol.* **1994**, *236*, 862−868.
(19) Hill, B. R.; DeGrado, W. F. *J. Am. Chem. Soc.* **1998**, *120*, 1138−1145.
(20) Choma, C. T.; Lear, J. D.; Nelson, M. J.; Dutton, P. L.; Robertson, D. E.; DeGrado, W. F. *J. Am. Chem. Soc.* **1994**, *116*, 856−65.
(21) 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*, 425−32.
(22) DeGrado, W. F.; Regan, L.; Ho, S. P. *Cold Spring Harbor Symp. Quant. Biol.* **1987**, *52*, 521−6.
(23) Degrado, W. F. *Adv. Protein Chem.* **1988**, *39*, 51−124.
(24) Xia, D.; Yu, C. A.; Kim, H.; Xia, J. Z.; Kachurin, A. M.; Zhang, L.; Yu, L.; Deisenhofer, J. *Science* **1997**, *277*, 60−6.
(25) Cohn, E. J.; Edsall, J. T. *Proteins, Amino acids and Peptides as Ions and Dipolar Ions*; Reinhold Publishing Corporation: New York, NY, 1965.

frequencies. Two-dimensional NOESY,[26] ROESY,[27] clean-TOCSY,[28] DQF-COSY,[29] triple quantum spectroscopy,[30] ^{13}C or ^{15}N HSQC,[31,32] ^{15}N HMBC,[33] and $H^{\beta}(C^{\beta})C^{\gamma}C^{\delta}(H^{\delta})$ and $H^{\beta}(C^{\beta})C^{\gamma}C^{\delta}C^{\epsilon}(H^{\epsilon})$[34] data were acquired in a standard manner. The ^{15}N HMBC was acquired in a mixed line shape mode using a 22-ms delay. The two-dimensional $H^{\beta}(C^{\beta})$-$C^{\gamma}C^{\delta}(H^{\delta})$ experiment was modified to frequency label H^{β} protons in the indirectly detected dimension instead of C^{β}. Three-dimensional ^1H TOCSY-relayed-CT-[^{13}C,^1H]-HMQC,[35] ^{15}N-resolved NOESY HSQC[36,37] and ^{13}C-resolved NOESY HSQC were acquired using standard pulse sequences. E-COSY based HNCA[38] and quantitative-J HNC^{γ}[39] and COC^{γ}[39] were acquired in a standard manner. Quadrature detection in the indirect dimensions was accomplished with TPPI-States.[40] Translation self-diffusion measurements were made with water-suppressed LED (longitudinal encode−decode) experiments[41−44] using a 100-ms diffusion time (Δ) and a pair of 7-ms pulsed field gradients (δ) that were applied with field strengths from 0.542 to 32.520 G/cm in 0.542 G/cm steps. The translational diffusion coefficient, D, was obtained by fitting the observed intensity decay profiles to eq 2

$$A = \exp[-(\gamma G\delta)^2(\Delta - (\delta/3))D] \quad (2)$$

where A is the signal intensity at a given gradient field strength (G) of duration δ ms and γ is the nuclear gyromagnetic ratio. Signal intensity is the summed integral for the methyl region of the ^1H NMR spectrum.

Structural Constraints. Interproton distance constraints were computed using a two-spin approximation. Cross-peak volumes involving two nonexchangeable protons were measured from a 100-ms mixing time 3D-^{13}C-edited or 2D-NOESY spectrum and scaled to the largest $H^{\beta 2}-H^{\beta 3}$ NOE cross-peak volume assuming a 1.8-Å separation. Similarly, cross-peak volumes involving one or more amide protons were measured from a 100-ms mixing time 3D-^{15}N-edited or 2D-NOESY spectra and scaled to the largest sequential $HN^{(i)}-HN^{(i+1)}$ cross-peak volume assuming a sequential interproton distance of 2.8 Å in a regular α-helix. Cross-peak volumes involving methyl protons were converted to distances using reference leucine CH_3-CH_3 or CH_3-CH interproton NOE volumes. The upper bound for each distance constraint was increased by 10%.

Torsion angle constraints were determined from J-couplings and their associated empirical Karplus relations (see Supporting Information Table S1). Constraints on the main chain ϕ torsion angle were determined from measurements of $^3J_{HN-H\alpha}$ in the HNCA using H^{α} as the passive spin.[38] Torsion angles (ϕ) were computed using the parametrized Karplus relation, $^3J_{HN-H\alpha} = 6.51\cos^2(\phi - 60) - 1.76\cos(\phi - 60) + 1.60$.[45] Aromatic χ^1 torsion angles were constrained using the $^3J_{C\gamma-C'}$ and $^3J_{C\gamma-N}$ couplings measured from the $C'C^{\gamma}$ and NC^{γ} spin−echo difference experiments.[39]

Residual dipolar couplings arising from molecular alignment in a dilute liquid crystal solution were measured from the $^1J_{NH}$ splitting[46−48] (see Supporting Information Table S1). Measurements were obtained from the difference in the $^1J_{NH}$ splitting between a proton-coupled ^{15}N HSQC spectrum acquired in the absence or presence of 4.6% (w/v) bicelles (1:2.9 mole ratio of dimyristoyl phosphotidylcholine:dihexanoyl phosphotidylcholine) in otherwise identical solution conditions of 20 mM sodium phosphate (pH 6.60, 50mM potassium chloride, 10 μM sodium azide, 92% H_2O and 8% D_2O and 39 °C). The residual deuterium quadrapole splitting for D_2O in the bicelle containing solution was 8.5 Hz at 39 °C.

Several α′-SS-α′ NH bond vectors were constrained as torsion angles using information from residual dipolar couplings. Because of the limited geometric distribution of amide vectors in the molecule, it was not possible to rigorously define the magnitude of the alignment tensor. Nevertheless, the fact that many of the NH vectors have large and uniform residual dipolar splittings (30 ± 3 Hz) indicates that they are collinear and aligned parallel with the z-axis of the alignment tensor. This view is supported by the observation of collinear α-helical NH vectors in a family of preliminary structures calculated without dipolar constraints. Assuming the largest dipolar coupling (37 Hz) represents an NH bond vector parallel to the z-axis of the molecular alignment tensor then the expected range of θ from the dipolar coupling range of 29−37 Hz is less than ∼±15° (see eq 3). This conservative view was used to define pairwise [H−N−N−H] torsion angles of 0° ± 30° (see Supporting Information Table S1), their main effect being to introduce long range order into the structural model.

The dependence of dipolar splitting on θ was assessed using eq 3[46] and assuming an axially symmetric tensor (i.e., $A_a \gg A_r$)

$$D_{PQ}(\theta,\phi) = -Sk[A_a(3\cos^2\theta - 1) + 1.5A_r(\sin^2\theta\cos 2\phi)] \quad (3)$$

$$k = \mu_o\gamma_P\gamma_Q h/16\pi^3(r^3)_{PQ}$$

where D_{PQ} is the experimentally determined residual dipolar coupling, S is the generalized order parameter for internal motion of the vector PQ, A_r and A_a are the rhombic and axial components of the diagonal tensor A, and θ and ϕ are the polar angles defining the orientation of the vector PQ in the alignment tensor coordinate system.[46]

Structure Calculations. A family of structural models consistent with the covalent geometry and NMR-derived interproton distance, torsion angle constraints, and residual dipolar coupling constraints was computed using Xplor3.1.[49] To avoid bias from the starting coordinates, 50 randomized beginning structures having an average pairwise backbone root-mean-square deviation (rmsd) of ∼13 Å were used to seed each independent structure calculation. Beginning structures contained no significant bond length, bond angle, improper, or van der Waals energy violations. Folding of the unstructured starting coordinates to a three-dimensional fold was accomplished using 7.5 ps of high-temperature molecular dynamics (MD) simulation with a 1000-fold reduced van der Waals energy weighting (compared to final van der Waals energy weighting) and a gradual increase in the NOE energy weighting. High-temperature structures were then cooled to 100 K with a gradual increase in the van der Waals energy weighting for an additional 7.5 ps of MD simulation. This protocol resulted in 100% convergence to the same fold. Twenty-five structures having the smallest NOE energy violation energies were further refined with the addition of dipolar coupling restraints and using van der Waals repulsive and attractive energy terms. The structural statistics are shown in Table 1.

(26) Macura, S.; Ernst, R. R. *Mol. Phys.* **1980**, *41*, 95−117.
(27) Bothner-By, A. A.; Stephens, R. L.; Lee, J.-M.; Warren, C. D.; Jeanloz, R. W. *J. Am. Chem. Soc.* **1984**, *106*, 811−13.
(28) Griesinger, C.; Otting, G.; Wüthrich, K.; Ernst, R. R. *J. Am. Chem. Soc.* **1988**, *110*, 7870−72.
(29) Rance, M.; Sorensen, O. W.; Bodenhausen, G.; Wagner, G.; Ernst, R. R.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1983**, *117*, 479−85.
(30) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley and Sons: New York, NY, 1986.
(31) Muller, L. *J. Am. Chem. Soc.* **1979**, *101*, 4481−84.
(32) Bodenhausen, G.; Ruben, D. J. *Chem. Phys. Lett.* **1980**, *69*, 185−89.
(33) Bax, A.; Marion, D. *J. Magn. Reson.* **1988**, *78*, 186−191.
(34) Yamazaki, T.; Forman-Kay, J. D.; Kay, L. E. *J. Am. Chem. Soc.* **1993**, *115*, 4−11055.
(35) Zerbe, O.; Szyperski, T.; Ottiger, M.; Wüthrich, K. *J. Biomol. NMR* **1996**, 99−106.
(36) Fesik, S. W.; Zuiderweg, E. R. *Q. Rev. Biophys.* **1990**, *23*, 97−131.
(37) Zhang, O.; Kay, L. E.; Olivier, J. P.; Forman-Kay, J. D. *J. Biomol. NMR* **1994**, *4*, 845−58.
(38) Weisemann, R.; Ruterjans, H.; Schwalbe, H.; Schleucher, H.; Bermel, W.; Griesinger, C. *J. Biomol. NMR* **1994**, *4*, 231−240.
(39) Hu, J.-S.; Grzesiek, S.; Bax, A. *J. Am. Chem. Soc.* **1997**, *119*, 1803−1804.
(40) Marion, D.; Ikura, M.; Tschudin, R.; Bax, A. *J. Magn. Reson.* **1989**, *85*, 393−99.
(41) Stejskal, E. O.; Tanner, J. E. *J. Chem. Phys.* **1965**, *42*, 288−292.
(42) Stilbs, P. *Progr. Nucl. Magn. Reson. Spectroscopy* **1987**, *19*, 1−45.
(43) Gibbs, S. J.; Johnson, C. S. *J. Magn. Reson.* **1991**, *93*, 395−402.
(44) Altieri, A. S.; Hinton, D. P.; Byrd, R. A. *J. Am. Chem. Soc.* **1995**, *117*, 7566−7567.
(45) Cavanagh, J.; Fairbrother, W. J.; Palmer, A. G.; Skelton, N. J. *Protein NMR Spectroscopy: Principles and Practices*; Academic Press: San Diego, CA, 1996.
(46) Bax, A.; Tjandra, N. *J. Biomol. NMR* **1997**, *10*, 289−92.
(47) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111−4.
(48) Clore, G. M.; Gronenborn, A. M. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5891−8.
(49) Brunger, A. T. *X-PLOR version 3.1: A System for X-ray Crystallography and NMR*; Yale University Press, 1992.

Table 1. Structural Statistics for $(\alpha'\text{-SS-}\alpha')_2$[a]

rms deviations from distance restraints (Å)	0.035 ± 0.005
rms deviations from torsional restraints (deg)	2.0 ± 0.5
deviations from idealized covalent geometry	
bonds (Å)	0.0044 ± 0.0003
bond angles (deg)	0.47 ± 0.04
impropers (deg)	0.32 ± 0.03
coordinate precision—average pairwise rmsd (Å)	
main chain (glu5-lys30, glu5′-lys30′)	0.54 ± 0.06
main chain plus hydrophobic core (glu5-lys30, glu5′-lys30′)	0.73 ± 0.09

[a] Statistics are based upon 25 independently refined structures. Root-mean-square (rms) deviations are computed on the basis of an average pairwise computation. Five hundred and eleven interproton distances were constrained in a square well potential using a 60 kcal/mol Å force constant. Fifty-three covalent torsion angles were constrained in a square well potential using a 250 kcal/mol rad force constant. Thirty-seven dipolar couplings were converted to 1332 pairwise H−N−N−H noncovalent torsion angle constraints and restrained to a square well potential with a 40 kcal/mol rad force constant. The hydrophobic core is defined as I6, I6′, L9, L9′, H10, H10′, F13, F13′, F17, F17′, L20, L20′, L21, L21′, H24, H24′, L28, L28′. Residues K30, K30′, L31, L31′ have low ^{15}N order parameters and are disordered in the structure family and were left out of the average pairwise calculations above and the cysteine and glycine linker residues (C1-G4, C1′-G4′) were also omitted from the rmsd calculation.

Results

The α′-SS-α′ main chain and nonaromatic side chain assignments have been previously reported.[16] The two helices in α′-SS-α′ are magnetically distinct and are referred to as Cys^1-Leu^{31} and Cys$^{1'}$-Leu$^{31'}$. The $(\alpha'\text{-SS-}\alpha')_2$ protein contains a 2-fold symmetry axis. This paper presents completion of the aromatic side chain assignments, a solution structural model for α′-SS-α′, and evidence supporting the four-α-helix bundle, $(\alpha'\text{-SS-}\alpha')_2$, quaternary structure.

Histidine Side Chain Assignments. There are four magnetically distinct histidines in α′-SS-α′ at His$^{10,10'}$ and His$^{24,24'}$(see Supporting Information Table S2 and Figure 1). The imidazole H$^{\epsilon1}$ and H$^{\delta2}$ were identified and connected to their directly attached carbons C$^{\epsilon1}$ and C$^{\delta2}$ using the constant time ^{13}C HSQC (see Figure 1) and carbon coupled ^{13}C HSQC. The C$^{\epsilon1}$−H$^{\epsilon1}$ correlations have relatively large downfield carbon chemical shifts that lack a $^1J_{CC}$ splitting in the carbon coupled ^{13}C HSQC spectrum. The C$^{\delta2}$−H$^{\delta2}$ correlations were identified by the unique situation of the C$^{\delta2}$ atom being directly bonded to a carbon and nitrogen. This results in the C$^{\delta2}$ carbons having a single $^1J_{CC}$ coupling in a carbon coupled ^{13}C HSQC or their signal will be inverted in sign relative to all other aromatic carbons (Figure 1), except for tryptophan C$^{\delta1}$, in the $1/^1J_{CC}$ constant time ^{13}C HSQC spectrum. The carbon bound H$^{\delta2}$ and H$^{\epsilon1}$ resonances were correlated with the nitrogen N$^{\epsilon2}$ and N$^{\delta1}$ resonances on each imidazole ring using an ^{15}N HMBC experiment making use of the small two-bond $^2J_{NH}$ couplings (see Figure 2). The histidine N$^{\delta1}$ resonances show a correlation to H$^{\epsilon1}$ and the N$^{\epsilon2}$ resonance shows correlations to both H$^{\epsilon1}$ and H$^{\delta2}$. The imidazole CH and NH correlations display similar intensity and line widths in the ^{13}C HSQC and ^{15}N HMBC, respectively, with the notable exception of His^{24} correlations. His^{24} shows a broad C$^{\delta2}$−H$^{\delta2}$ correlation in the ^{13}C HSQC spectrum and broad or multiple NH correlations in the ^{15}N HMBC spectrum. The N$^{\delta1}$ chemical shifts at 242−245 ppm and N$^{\epsilon2}$ chemical shifts at 165−172 ppm (see Supporting Information Table S2 and Figure 2) indicate neutral imidazoles under these solution conditions.[50,51] Furthermore, the presence of relatively strong N$^{\epsilon2}$−H$^{\epsilon1}$, N$^{\epsilon2}$−H$^{\delta2}$, and N$^{\epsilon2}$−H$^{\delta2}$ correlations and the absence of N$^{\epsilon2}$−H$^{\epsilon1}$ correlations (see Figure 2) indicates the N$^{\epsilon2}$−H$^{\epsilon2}$ neutral tautomer state is preferred for all histidines.[50,51]

Phenylalanine Side Chain Assignments. There are four magnetically distinct phenylalanines in α′-SS-α′ at Phe$^{13,13'}$ and Phe$^{17,17'}$ (see Supporting Information Table S3 and Figure 1). The four benzyl {H$^\delta$, H$^\epsilon$, H$^\zeta$} spin systems were initially identified with TOCSY, DQF-COSY, TQ, and NOESY using established methods.[30] The benzyl proton resonances were connected to their directly attached C$^\delta$, C$^\epsilon$, or C$^\zeta$ carbons with the constant time ($1/^1J_{CC}$) ^{13}C HSQC spectrum (Figure 1) and unambiguously correlated using the ^{1}H-TOCSY-CT-^{13}C-HMQC spectrum (see Figure 3).

Tryptophan Side Chain Assignments. There are two magnetically distinct tryptophans in α′-SS-α′ at Trp$^{7,7'}$ (see Supporting Information Table S4 and Figure 1). The tryptophan four-spin group consisting of H$^{\epsilon3}$, H$^{\zeta3}$, H$^{\eta2}$, and H$^{\zeta2}$ and the two-spin group of H$^{\delta1}$ and H$^{\epsilon1}$ were identified with TOCSY, DQF-COSY, TQ, and NOESY. The proton bound N$^{\epsilon1}$ nitrogen was assigned from the ^{15}N-HSQC spectrum and the proton bound carbons C$^{\epsilon3}$, C$^{\zeta3}$, C$^{\eta2}$, C$^{\zeta2}$, and C$^{\delta1}$ assigned with carbon-coupled and constant time ^{13}C-HSQC spectra. The ^{15}N-HMBC spectrum correlated the H$^{\delta1}$ and N$^{\epsilon1}$ spins via the $^2J_{NH}$ coupling confirming the two-proton spin groups (see Figure 2). The ^{1}H-TOCSY-CT-^{13}C-HMQC confirmed the four-proton spin systems and their directly attached carbons (see Figure 3). The two-proton and four-proton spin systems were then connected with the H$^{\zeta2}$−N$^{\epsilon1}$ correlation in the ^{15}N HMBC spectrum and a strong H$^{\epsilon1}$−H$^{\zeta2}$ correlation in the NOESY spectrum (data not shown).

Correlation of Aromatic Side Chain and C$^\beta$H$_2$ Resonances. The histidine-H$^{\delta2}$, phenylalanine H$^\delta$ and H$^\epsilon$, and tryptophan H$^{\delta1}$ protons were connected to their covalent bonded methylene group carbon (C$^\beta$) and proton (H$^\beta$) resonances with the (H$^\beta$)C$^\beta$(C$^\gamma$C$^\delta$)H$^\delta$ and (H$^\beta$)C$^\beta$(C$^\gamma$C$^\delta$C$^\epsilon$)H$^\epsilon$ experiments and variants thereof to label H$^\beta$ instead of C$^\beta$ (see Figure 4). The H$^{\delta2}$−H$^\beta$ correlations were essential for assigning His^{24} and His$^{24'}$ since they have C$^\beta$ shifts that differ by only 0.2 ppm but have quite different H$^\beta$ shifts (Figure 4). Likewise, the unique H$^\delta$−H$^\beta$ correlations were essential for connecting the benzyl groups to the methylene resonances since the four phenylalanine C$^\beta$ shifts differ by only 0.3 ppm, preventing sequence specific assignment based solely on the C$^\beta$−H$^\delta$ correlations.

Secondary Structure. After the resonance assignments were completed, the secondary structure of α′-SS-α′ was determined using information from secondary chemical shifts, main chain NOEs, $^3J_{HN-H\alpha}$ coupling constant values, amide hydrogen-exchange kinetics, and amide proton thermal coefficients. Upfield-shifted H$^\alpha$ and C$^\beta$ resonances and downfield-shifted C$^\alpha$ and C′ resonances, relative to the corresponding random coil resonance values, are characteristic of α-helical secondary structure.[52,53] This pattern of secondary shifts is observed for Glu^5-Lys^{30} and Glu$^{5'}$-Lys$^{30'}$ and is consistent with the designed α-helical structure previously indicated by CD and IR spectra.[15] Another key spectral property of helical structure are sequential HN$^{(i)}$−HN$^{(i+1)}$, H$^{\alpha(i)}$−HN$^{(i+3)}$, and H$^{\alpha(i)}$−H$^{\beta(i+3)}$ NOEs and these are observed from Gly^4-Leu^{31} and Gly$^{4'}$-Leu$^{31'}$, further supporting the presence of helical structure. A hallmark characteristic of α-helix secondary structure is small (2−5 Hz) $^3J_{HN-H\alpha}$

(50) Bloomberg, F.; Maurer, W.; Ruterjans, H. *J. Am. Chem. Soc.* **1977**, *25*, 5, 8149−8159.

(51) Pelton, J. G.; Torchia, D. A.; Meadow, N. D.; Roseman, S. *Protein Sci.* **1993**, *2*, 543−58.

(52) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *FEBS* **1991**, *293*, 72−80.

(53) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *Biochemistry* **1992**, *31*, 1647−51.

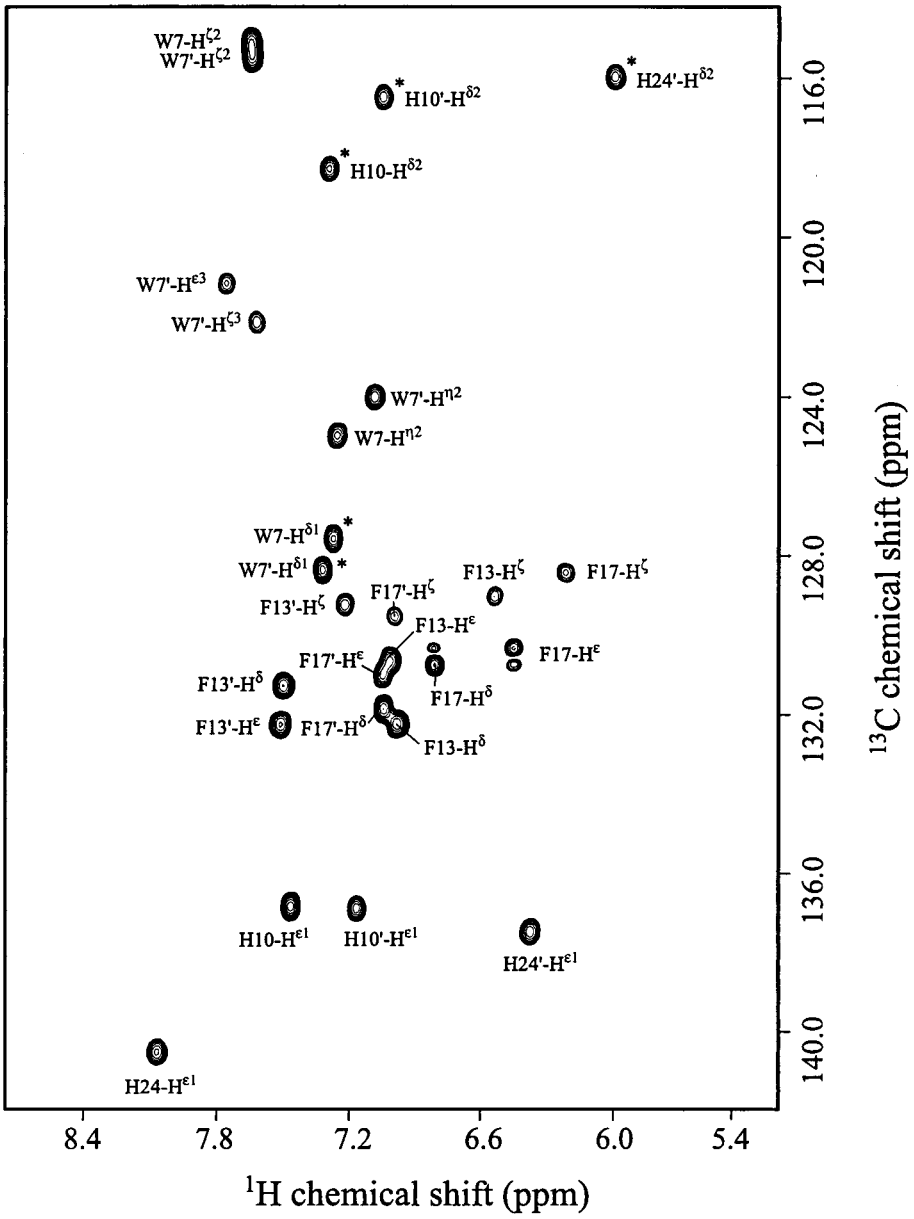

Figure 1. $1/^1J_{CC}$ constant time ^{13}C HSQC spectrum with carriers and delays optimized for the aromatic $^{13}C-^1H$ correlations. The correlations noted with asterisks are inverted in sign. Three correlations $H24C^{\delta 2}-H^{\delta 2}$, $W7C^{\epsilon 3}-H^{\epsilon 3}$, and $W7C^{\zeta 3}-H^{\zeta 3}$ are below the lowest contour level. These correlations are clearly observed in the carbon-coupled version of the ^{13}C HSQC.

couplings[54] and α′-SS-α′ displays a range from 1.9 to 5.3 Hz (see Supporting Information Table S1). The slowest amide hydrogens to exchange with bulk solvent are centrally located in the α-helices, indicating that the most protected amide hydrogens are located furthest from the ends of the helix which is consistent with α-helical secondary structure (see Table 2). Finally the amide proton thermal coefficients are small (1−5 ppb) for Glu[5]-Leu[31] and Glu[5′]-Leu[31′] and consistent with the sequestering of these amide protons from water as expected for a regular α-helical structure (data not shown). These results are fully consistent with α-helical secondary structure for Gly[4]-Leu[31] and Gly[4′]-Leu[31′].

Tertiary Structure. As designed, the α′-SS-α′ covalent unit is a composed of a pair of close packed parallel α-helices (see Figure 5 and Supporting Information Table S5 for qualitative interhelical interproton distance constraints). The structural family is well-determined, and each structure has small interproton distance and torsion angle violations with good covalent geometry (Table 1). The α-helices are of regular conformation; however, the main chain from Ile[6′]-His[10′] shows a more tightly wound helix. The ϕ and ψ torsion angles for Glu[5]-Lys[30] and Glu[5′]-Lys[30′] lie in the most favorable region for α-helix structure in a Ramachandran plot. The ϕ torsion angles are constrained for most of these residues (Supporting Information Table S1); however, no ψ torsion angle constraints are defined, and thus, ψ is being constrained to ideal helical values by a combination of interproton distances, ϕ torsion angle constraints, and planarity constraints on ω. In contrast, Cys[1]-Gly[4] and Cys[1′]-Gly[4′] ϕ and ψ torsion angles show considerable variation; however, neither ϕ nor ψ torsion angles are constrained, and no interproton constraints are defined for these residues. Left-handed supercoiling is present in each of the helices; left-handed supercoiling extends from ∼Phe[13]-Leu[31]:∼Phe[13′]-Leu[31′] and is absent from the N-terminal third of the structure.

(54) Pardi, A.; Billeter, M.; Wüthrich, K. *J. Mol. Biol.* **1984**, *180*, 741−51.

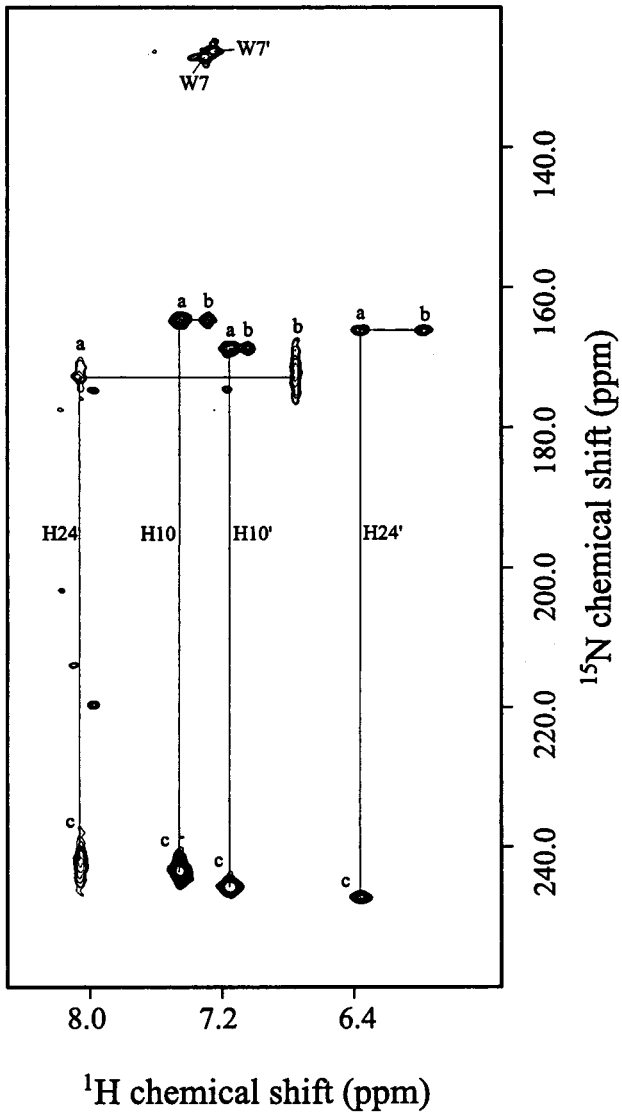

Figure 2. ^{15}N−^{1}H HMBC spectrum showing the histidine (a) $N^{\epsilon 2}$−$H^{\epsilon 1}$, (b) $N^{\epsilon 2}$−$H^{\delta 2}$, and (c) $N^{\delta 1}$−$H^{\epsilon 1}$ correlations and the tryptophan $N^{\epsilon 1}$−$H^{\delta 1}$ correlations.

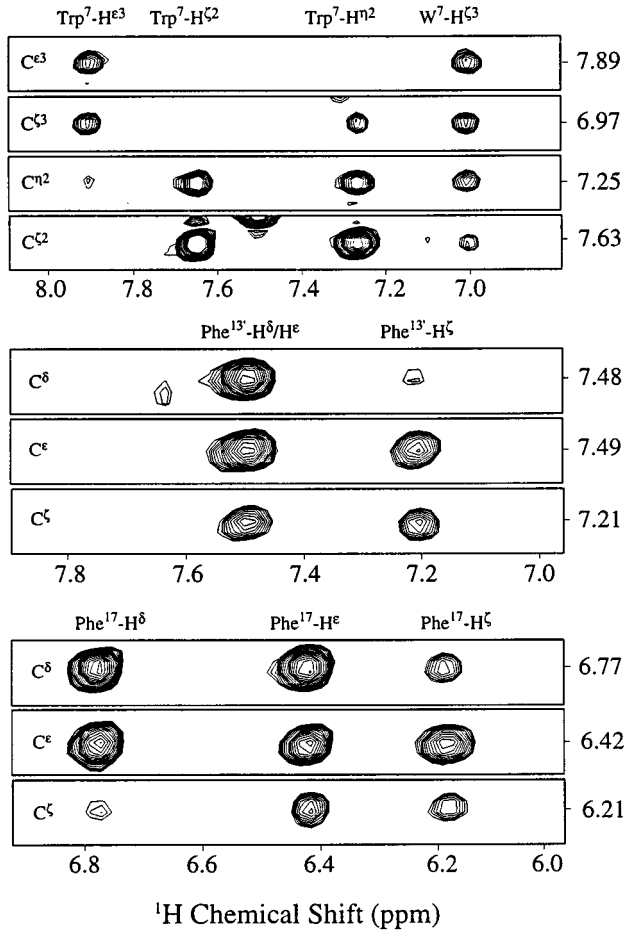

Figure 3. ^{1}H−^{1}H planes at specified carbon frequencies from the ^{1}H TOCSY-relayed-CT ^{13}C HMQC spectrum showing proton and carbon connectivity for Trp^{7} in the top group of panels, Phe$^{13'}$ in middle panels, and Phe^{17} in lower panels.

The α′-SS-α′ hydrophobic core consists of aliphatic and aromatic side chains, from heptad a-, d-, e-, and g-positions, pack at the helix−helix interface and form a spine-like core extending the length of the helices. Many of these side chains also contribute to a hydrophobic face of the isolated α′-SS-α′ covalent unit that is the putative dimerization interface forming $(\alpha'\text{-SS-}\alpha')_2$ (Figures5 and 6). The N-terminal boundary of the hydrophobic core is where the Trp$^{7'}$ indole, the Leu^{9} isopropyl, the His^{10} imidazole, and the Ile$^{6'}$ isobutyl groups pack tightly. These side chains are between a His$^{10'}$ imidazole on the bottom and a Ile^{6} isobutyl on the top. The His$^{10'}$ imidazole bridges the helix−helix interface while the His^{10} imidazole is on the hydrophobic face. The Phe^{13}/Phe$^{13'}$ benzyl groups pack with Leu$^{9'}$ and His$^{10'}$ to continue the hydrophobic core. The phenylalanine-rich region, Phe$^{13,13'}$-Phe$^{17,17'}$, shows interhelix packing of Phe^{13} with Phe$^{13'}$ and Phe^{17} with Phe$^{17'}$ and likewise intrahelix packing of Phe^{13} with Phe^{17} and Phe$^{13'}$ with Phe$^{17'}$. These benzyl groups also contribute to the hydrophobic face. Likewise, a cluster of isopropyl groups from residues Leu$^{20,20'}$ and Leu$^{21,21'}$ pack intra- and interhelix to extend the hydrophobic core. The Leu^{21,21} side chains are well positioned and adopt gauche(−) χ^1 torsion angles while the Leu$^{20,20'}$ side chains adopt gauche-(+) conformations. His^{24} imidazole packs with this isopropyl cluster and the His$^{24'}$ imidazole points into the hydrophobic face. Thus, both sets of histidine residues show one imidazole packing between the α-helices and the other pointing directly into the hydrophobic face. The isopropyl groups of Leu^{28} and Leu$^{28'}$ pack with each other as do those from Leu^{31} and Leu$^{31'}$; however, these side chains are not well-ordered in the structural family.

The distribution of the surface polar and apolar side chains is clearly asymmetric (see Figure 6), consistent with the amphiphilic di-α-helix design. The lysine and glutamate side chains are clustered on one face and edge of the covalent monomer structure, while the opposite side is comprised of hydrophobic or neutral imidazole side chains. The hydrophilic surface shows a distribution of basic and acidic residues that promote favorable charge−charge interactions as anticipated from the design.

Many unusual chemical shifts observed in α′-SS-α′ can be attributed to ring current effects. For example, Leu^{21} and Leu$^{21'}$ methyl groups are significantly upfield-shifted. These methyls are close packed with each other and are proximal to the Phe$^{17,17'}$ and His$^{24,24'}$ aromatic rings. Similarly, Leu^{28} and Leu$^{28'}$ show upfield-shifted methyl proton chemical shifts, and these groups are proximal to the His$^{24,24'}$ rings. By contrast, Leu^{23} and Leu$^{23'}$ side chains are not packed in the protein interior nor are they proximal to aromatic rings, and their methyl proton and carbon shifts are near random coil values (See Supporting Information Table S1 in ref 16). Thus, the structure is consistent with

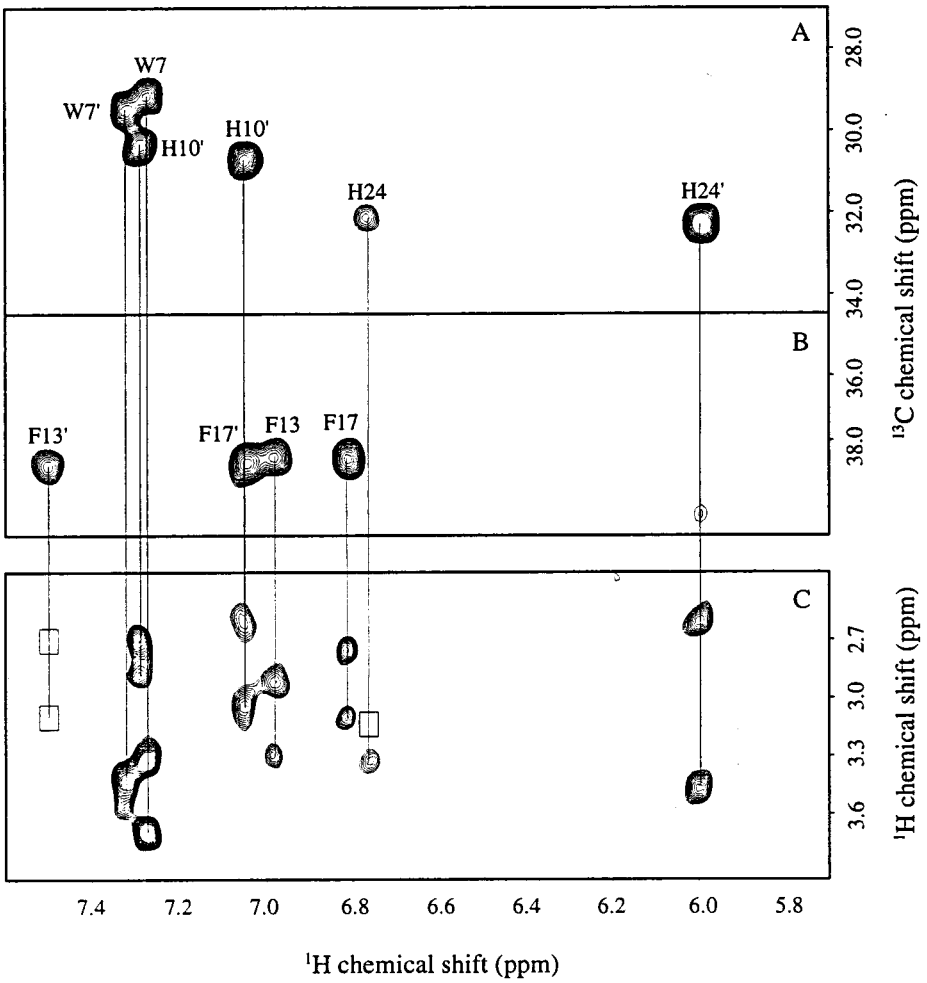

Figure 4. Composite spectrum of Cβ(CγCδ)Hδ and Hβ(CβCγCδ)Hδ spectra. Panel A shows C$^{\beta}$/H$^{\delta}$ correlations optimized for histidine and tryptophan frequencies and Panel B shows C$^{\beta}$/H$^{\delta}$ correlations optimized for the phenylalanine frequencies. Panel C shows the corresponding H$^{\beta}$/H$^{\delta}$ correlations. The open boxes are the positions of H$^{\beta}$ correlations that are not observed.

Table 2. Amide Hydrogen-Exchange Protection Factors and ΔG's for $(\alpha'\text{-SS-}\alpha')_2$[a]

residue	PF ($\times 10^{-4}$) pD 6.5	ΔG (kcal/mol)	residue	PF ($\times 10^{-4}$) pD 6.5	ΔG (kcal/mol)
Leu14	1.9	5.98	Leu14′	1.8	5.95
Lys15	0.12	4.31	Lys15′	0.15	4.44
Lys16	2.2	6.06	Lys16′	2.1	6.03
Phe17	4.9	6.54	Phe17′	5.1	6.56
Glu18	1.3	5.73	Glu18′	1.4	5.77
Glu19	0.11	4.24	Glu19′	0.11	4.24
Leu20	2.9	6.22	Leu20′	4.2	6.44
Leu21	0.35	4.98	Leu21′	4.3	6.46
Lys22	3.5	6.35	Lys22′	4.2	6.44
Leu23	0.74	5.40	Leu23′	0.90	5.51

[a] PF, protection factor (ratio of observed exchange rate/calculated exchange rate).

expected ring current effects on proximal proton and carbon resonance frequencies.

Quaternary Structure. The $(\alpha'\text{-SS-}\alpha')_2$ exists as a homodimer in solution. Equilibrium sedimentation (see Figure 7) and gel permeation experiments indicate a solution species of M_r 15.4 and 19.0, respectively. Pulsed field gradient translational diffusion measurements indicate a molecular weight consistent with $(\alpha'\text{-SS-}\alpha')_2$. Finally, preliminary ^{15}N NMR relaxation measurements provide an estimate for the molecular reorientation correlation time of 9.3 ns at 25 °C, which is consistent with the molecular weight of $(\alpha'\text{-SS-}\alpha')_2$ and not lower or higher oligomer states.

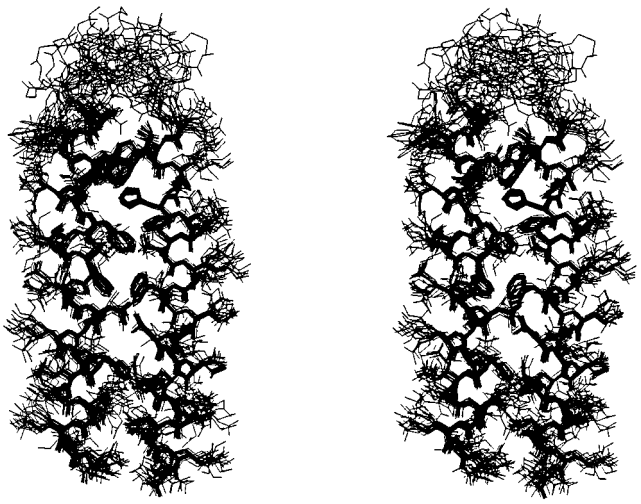

Figure 5. Stereoview showing the superposition of 25 α′-SS-α′ structures having the smallest distance and torsion angle violations. The pairwise rmsd is minimized for the Gly^{4}-Lys^{30} and Gly$^{4'}$-Lys$^{30'}$ main chain (C, O, CA, N) atoms. Amino acid residue numbers are shown. The B strand is the prime strand nomenclature of the text. Only non-hydrogen atoms are shown.

Despite strong and unequivocal evidence that the covalent dimer forms a symmetric dimer in solution, we are unable to detect nuclear Overhauser effects between protons on opposing α′-SS-α′ monomers in the four-helix bundle. Since the molecule

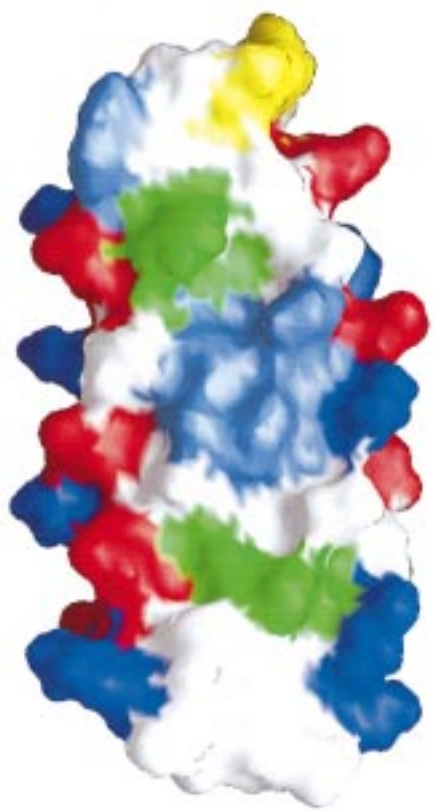

Figure 6. Representative α′-SS-α′ structure showing the hydrophobic face. Side chain colors are red for Glu, dark blue for Lys and Arg, light blue for Phe and Trp, green for His, yellow for Cys, and white for Leu, Ile, and Gly. Only non-hydrogen atoms are shown. Drawn with the program Grasp.[69]

exists as a homodimer, $(\alpha'\text{-SS-}\alpha')_2$, a labeling strategy was used to prepare populations of mixed isotopic labeled four-α-helix bundle proteins. Mixed labeled samples of [(^{15}N-α′-SS-α′)(^{13}C-α′-SS-α′)] or [(^{12}C-α′-SS-α′)(^{13}C-α′-SS-α′)] were prepared by combining equimolar amounts of each isotopically labeled molecule and allowing for dimer dissociation and association to provide a mixture of $^1/_2$ homodimers and $^1/_2$ heterodimers. This exchange has been demonstrated to occur on the seconds to minute time scale for a nitroxide spin label or coproporphyrin-containing prototype $(\alpha\text{-SS-}\alpha)_2$ protein.[55] To rule out an experimental reason for not observing the expected NOEs, a positive control using uniform ^{13}C,^{15}N-lableled $(\alpha'\text{-SS-}\alpha')_2$ demonstrated that ^{15}N−^{1}H ⇒ NOE ⇒ ^{13}C−^{1}H transfers were selected; a $^{13}C^{\alpha}$-Leu calmodulin and a $^{13}C^{\delta}H_3/^{12}C^{\gamma}H_3$-labeled ubiquitin provided positive and negative controls for demonstrating that ^{12}C−^{1}H ⇒ NOE ⇒ ^{13}C−^{1}H transfers were selected. Furthermore, NOESY experiments performed over a temperature range of 15 to 60 °C failed to reveal new intermolecular NOEs. Finally, the rotating frame nuclear Overhauser (ROESY) spectrum of $(\alpha'\text{-SS-}\alpha')_2$ is qualitatively identical to the NOESY spectrum. Thus, the absence of intermonomer NOEs (and ROEs) is strong evidence that a single interface structure is not forming in solution.

Orientation of α′-SS-α′ Covalent Units in $(\alpha'\text{-SS-}\alpha')_2$. Although the topology of the $(\alpha'\text{-SS-}\alpha')_2$ dimer could not be directly determined via observation of nuclear Overhauser effects, there is spectroscopic evidence that both syn and anti orientations can be formed in solution. Fluorescence spectroscopy of the pyrene butyric labeled $(\text{pyr-}\alpha'\text{-SS-}\alpha')_2$ indicates the

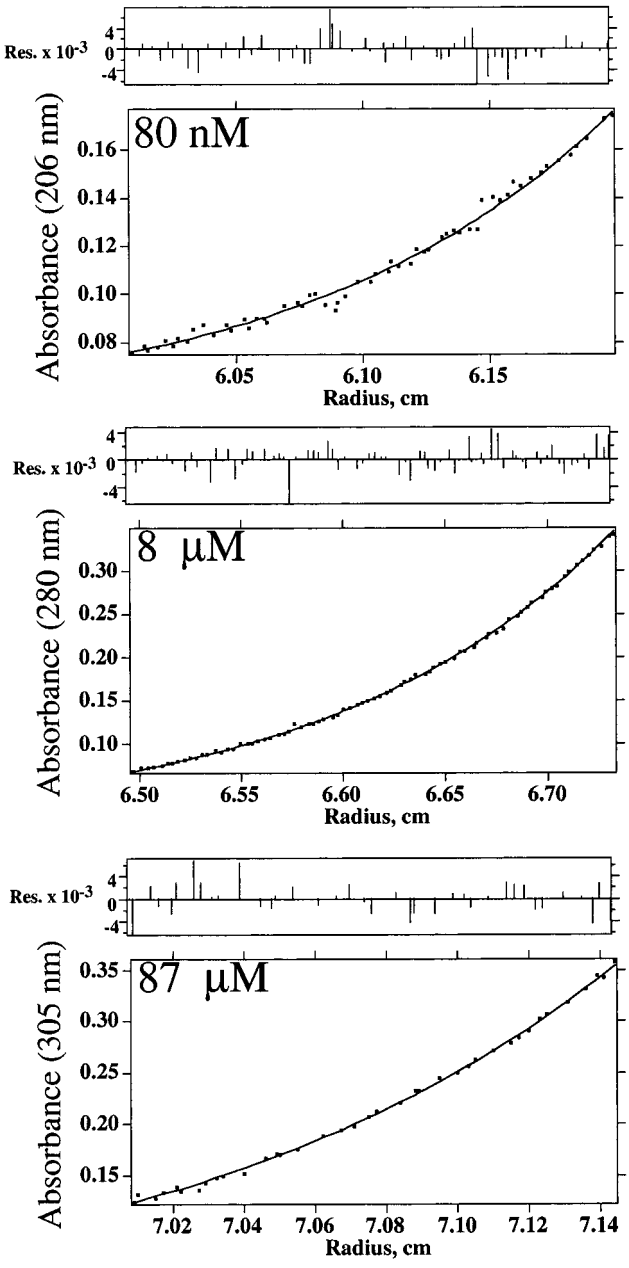

Figure 7. Sedimentation equilibrium analysis for $(\alpha'\text{-SS-}\alpha')_2$. The three panels show the results for α′-SS-α′ concentrations of 160 nM (M_r = 15.8), 8 μM (M_r = 14.9) and 175 μM (M_r = 15.5), from top to bottom, respectively. Each set of radial distribution absorbance scan data was fit to a single species of molecular weight that is consistent with a four-α-helix bundle aggregation state with the residuals shown above each panel.

pyrene groups are isolated from each other in the excited state (see Supporting Information Figure S1) consistent with a four-α-helix bundle having the probes positioned at opposite ends of the protein consistent with an anti-parallel orientation. In contrast, UV−Vis spectroscopy of the coproporphyrin-labeled $(\text{CP-}\alpha'\text{-SS-}\alpha')_2$ molecule indicates that the coproporphyrin groups are proximal to one another (see Supporting Information Figure S1) and is consistent with an all-parallel (syn) bundle. Thus, the bundle appears to adopt either syn- or anti-topology and is label-dependent (pyrene versus coproporphyrin) suggesting that there is a only a small energetic difference between syn and anti orientations. This is consistent with a nonspecific intermolecular interface.

Histidine Titration. The histidine imidazole ring ^{13}C, ^{15}N, and ^{1}H chemical shifts are exquisitely sensitive to the imidazole

(55) Gibney, B. R.; Johansson, J. S.; Rabanal, F.; Skalicky, J. J.; Wand, A. J.; Dutton, P. L. *Biochemistry* **1997**, *36*, 2798−806.

ionization state and can be used to monitor its ionization state. Thus the titration of the histidine was carried out by monitoring the pH dependence of the $H^{\epsilon 1}$ and $H^{\delta 2}$ chemical shifts and fitting these data to the Henderson−Hasselbach equation (see Supporting Information Figure S2 and Table S2)

$$\delta = \delta^{+} - (\delta^{+} - \delta^{(\text{neutral})})(10^{(\text{pH}-\text{p}Ka)})/(1 + 10^{(\text{pH}-\text{p}Ka)})$$

where $\delta^{(+)} = H^{\epsilon 1}$ or $H^{\delta 2}$ chemical shifts on the imidazolium cation; $\delta^{+} - \delta^{(\text{neutral})} = H^{\epsilon 1}$ or $H^{\delta 2}$ chemical shift difference between the imidazolium cation and a neutral imidazole. The fully protonated shifts (δ^{+}) and the ($\delta^{+} - \delta^{(\text{neutral})}$) differences could not be empirically determined, and thus the titration curves were fitted to the three parameters δ^{+}, $\delta^{+} - \delta^{(\text{neutral})}$, and the apparent p$K_a$ (see Supporting Information Figure S2). Each independent histidine titration curve was fitted to a similar apparent pK_a of ∼5.4, indicating the neutral form of the imidazole side chain is predominant under these solution conditions. Below pH 5, the $H^{\epsilon 1}$ or $H^{\delta 2}$ resonances broaden significantly, and the $^{13}C-^{1}H$ correlations are in the spectral noise.

Discussion

Determinants of Structural Specificity. Structural specificity is the degree of attainment of a unique or highly populated solution conformation. Thus, a designed protein of high structural specificity will show a single set of nuclear magnetic resonances having narrow line widths and a large dispersion of chemical shifts from their random coil values. In contrast, a protein of low structural specificity will display a multiple set of resonances and/or large line widths and a narrow range of chemical shift dispersion. Proteins that display intermediate NMR spectral properties compared to these extreme situations can be difficult to quantitatively interpret with regard to protein structural specificity. Typical of minimalist de novo designed proteins, the prototype (α-SS-α)$_2$ is a protein of low structural specificity. This protein contains a minimalist leucine-rich hydrophobic core of d-position Leu[6], Leu[13], and Leu[20] and e-position Trp[7], Leu[14], and Leu[21]. Transformation of α-SS-α to a unique structure α′-SS-α′ was based on an iterative redesign protocol where substitution of the d-position leucines on each helix with a variety of different hydrophobic amino acids resulted in significantly different structural specificity outcomes.[15] These substitutions had no effect on the helical content or preferred aggregation state of this four-α-helix bundle protein, illustrating its robust nature as a design template. The accompanying manuscript of Gibney and co-workers[15] describes in detail the structural and thermodynamic outcomes from a systematic replacement of many of the critical amino acids in this maquette design and lends considerable insight to the following discussion.

Replacement of the prototype Leu[6] with conformationally restricted β-branched amino acids, isoleucine and valine, clearly improves the conformational specificity of the proteins. Substitutions of Leu[6] with Ile[6] or Val[6] in the prototype α-SS-α result in high structural specificity with at least one minor conformer present in addition to the major conformer for the Val[6] substituted protein. Substitution of Leu[6] with Phe[6] retains the low structural specificity of the prototype.[15,56] Global structural sensitivity at this particular d-position may be explained, in part, by the observation that Ile[6] and Ile[6′] side chains are well-packed in the α′-SS-α′ solution structure. Since the leucine side chain can adopt two isoenergetic rotamers, this may result in multiple local and consequently global conformations. Restricting side chain rotamer conformations with β-branched isoleucine or valine is likely the source of conformational specificity. In contrast, the bulky phenylalanine side chain is apparently too large to be accommodated at the Leu[6] position.

Clearly, the above results implicate the packing environment proximal to Ile[6] and Ile[6′] as critical for the conformational specificity of the (α′-SS-α′)$_2$ protein. Curiously, the two side chains do not make extensive van der Waals contact but instead are bridged by Val[9] suggesting this side chain is also critical to tight side chain packing. Consistent with this prediction is the observation that a Leu[9] → Ala[9] substitution in α′-SS-α′ results in conversion from a highly specific structure to one of low to moderate structural specificity (data not shown). These results suggests that β-branched side chains are responsible for the structural specificity and a proximal side chain (Val[9]) is contributing in a synergistic manner to this structural specificity.

When comparable substitutions of α-SS-α are made for Leu[13], quite different structural specificity outcomes are observed.[15,56] In sharp contrast to Leu[6], substitution of Leu[13] with β-branched amino acids isoleucine and valine fails to improve the conformational specificity of the protein. However, the substitution of α-SS-α with a bulky aromatic phenylalanine residue, Leu[13] → Phe[13], results in a protein of high structural specificity. Examination of the α′-SS-α′ structure, which also has a Leu[6,6′] → Ile[6,6′] substitution, shows Phe[13] and Phe[13′] side chains packing with each other and with the Phe[17] and Phe[17′], and these aromatic groups are clearly positioned on the hydrophobic surface of the putative dimer interface. Large upfield and downfield shifts for resonances of three of the four benzyl rings is also consistent with close interaction of these side chains. The side chains of His[10,10′], Phe[13,13′], and Phe[17,17′] form an aromatic core, and this core clearly contributes to conformational specificity.

Finally, amino acid substitutions for d-position Leu[20,20′] in the prototype α-SS-α with β-branched isoleucine and valine and aromatic phenylalanine fail to improve structural specificity. Remarkably these same substitutions in the conformationally specific α′-SS-α′ results in complete loss of structural specificity, indicating that this is a key residue for maintaining conformational integrity. These data from the two proteins, α-SS-α and α′-SS-α′, suggest that, once conformational specificity is achieved, Leu[20] is important for maintenance of this structure. However, this site alone cannot drive a specific conformation from the prototype α-SS-α sequence.

Role of the Disulfide Bond. The α′-SS-α′ disulfide bond is not responsible for driving the magnetic asymmetry of (α′-SS-α′)$_2$ since converting the protein to (α′-SH)$_4$ by chemical reduction results in significant ^{15}N HSQC correlation changes for only the N-terminal residues extending from Cys[1]-Trp[7] and Cys[1′]-Trp[7′]. As expected the glycine ^{15}N HSQC correlations in the reduced (α′-SH)$_4$ form are overlapped, indicating that they are in similar magnetic environments unlike that in the oxidized form. Glu[5,5′], Ile[6,6′], and Trp[7,7′] also show significant shift changes upon chemical reduction, indicating either a local change in structure or motional changes for this part of the protein. ^{15}N HSQC correlations from Lys[8]-Leu[31], Lys[8′]-Leu[31′] show relatively small chemical shift perturbations, indicating that side chain packing is responsible for the magnetic asymmetry of the chemically symmetric α-helices and not disulfide bond formation. Thus, the disulfide covalent bond is certainly dispensable in future designs.

Magnetic asymmetry of the α′-SS′-α′ molecule in the chemically reduced form is somewhat surprising, given the

(56) Gibney, B. R.; Rabanal, F.; Skalicky, J. J.; Wand, A. J.; Dutton, P. L. *J. Am. Chem. Soc.* **1999**, *121*, 4952−60.

observation that several other α-helix bundle proteins show magnetic symmetry for chemically identical constituent peptides. Magnetic symmetry of the constituent peptides of a multihelix bundle protein could arise from symmetrically arranged monomers or exchange between asymmetric monomers that is fast on the proton chemical shift time scale. Several previously characterized multihelix systems are believed to have magnetic symmetry due to the former. For example, two different length peptides that correspond to the DNA binding domain of GCN4 self-associate in solution and result in magnetically equivalent α-helices.[17,57] The de novo designed four α-helix bundle proteins α_1 (12 residues)[58] and α_1B (16 residues)[59] self-assemble in solution and show only a set of resonances for a single peptide, indicating all helices are magnetically equivalent.

The Histidine Micro-Environments are Similar. The histidine apparent pK_a's of 5.3−5.6 (Supporting Information Table S2) are significantly more acidic than that of free histidine (pK_a = 6.1) or the Gly-His-Gly peptide (pK_a = 6.9).[60] These results clearly show that the neutral imidazole is stabilized relative to the imidazolium cation and this is entirely consistent with the burial of the imidazole ring in a low dielectric interface between the two helices composing the covalent α′-SS-α‘ unit or at the interface of the (α′-SS-α‘)$_2$ dimer. The NMR resonances broaden considerably for α′-SS-α‘ at pH < 5.0 which is consistent with the formation of an unfavorable imidazolium cation in the protein interior and ensuing chemical exchange between structures containing buried and neutral imidazole side chains versus those having solvated imidazolium side chains. Protonation of glutamate side chains at pH < 5.0 may also lead to nonspecific aggregation since these negative charges are important for balancing the lysine positive charges and furthermore contribute to water solubility of the protein. The secondary structure of the protein as monitored by CD spectroscopy is not changing over the pH range of 4.8−8.0, suggesting that the overall secondary structure is not changed over this pH range (data not shown). Thus, the histidine titration data support the structural model that places these side chains at solvent inaccessible intra- and interhelical interfaces.

The His^{24} residue behaves differently from the other three histidines. His^{24} exists in several different conformations that interconvert on the millisecond time scale at pH 6.50 as exemplified by the presence of multiple and broad $N^{\epsilon 2}-H^{\epsilon 1}$, $N^{\epsilon 2}-H^{\delta 2}$, and $N^{\delta 1}-H^{\epsilon 1}$ correlations in the ^{15}N HMBC spectrum (see Figure 2). The ^{13}C HSQC correlations for His^{24} are also consistent with this view. The correlation for His^{24} $C^{\epsilon 1}-H^{\epsilon 1}$ is strong and sharp yet that of His^{24} $C^{\delta 2}-H^{\delta 2}$ is in the noise of the constant time ^{13}C HSQC spectrum. It is intriguing to speculate that this side chain, which is located at the putatively disordered intermolecular interface, is reporting on the different intermolecular conformations that are energetically accessible.

Apo-(α′-SS-α′)$_2$ Does Not Contain a Classic Heme Binding Cavity. The tight hydrophobic packing in the vicinity of the heme ligands His$^{10,10'}$ and His$^{24,24'}$ results in the absence of a classic heme binding pocket such as those observed in globular heme proteins, e.g., apomyoglobin, apocytochrome b_{562}.[61] Despite being devoid of cavities in which to incorporate the heme macrocycles, (α′-SS-α′)$_2$ binds four hemes with high affinity, two with K_d values of 10−20 nM and two with K_d values of 5−15 μM. This would suggest that the affinity is primarily due to the thermodynamically favorable histidine ligands themselves rather than the presence of a well-formed cavity for the burial of the prosthetic group. Nevertheless, the burial of the heme groups in a low dielectric hydrophobic core raises their redox potentials, from −130 to −280 mV relative to bis-imidazole ligated iron(protoporphyrin IX) with further elevation due to both oxidized heme−oxidized heme at both heme sites and to oxidized heme-arginine charge−charge electrostatic interactions local to the His$^{24,24'}$ site.[62] While incorporation of the four macrocycles causes reorganization of the hydrophobic core leading to a multistructured or disordered heme protein, the scaffold retains its helical content and aggregation state illustrating the robustness of this architecture for the construction of maquettes.

Structural Dichotomy of (α′-SS-α′)$_2$. The current design produced a nativelike interface between the covalent pair of α-helices but has failed to produce a unique interface between the two covalent units comprising the four α-helix bundle. The absence of intermonomer NOEs is presumably due to the absence of significant intensity arising from one dominant dimer−dimer interface. Dynamic nulling of the NOE is ruled out by the similarity of ROESY and NOESY spectra. These observations lead to the conclusion that the interface between the disulfide bridged di-α-helical monomers in the four-helix bundle is disordered and that the activation barrier between the various conformers is quite small. There is no evidence for the formation of a single dominant interface over the temperature range of 5−65 °C and pH range of 5.5−7.5. Thus, the structure displays a very unusual dichotomy of structure with one interface having a highly specific, nativelike structure and the adjacent interface that is averaged over several different packing interfaces. A similar situation has apparently been observed previously in a β-sheet protein of de novo design.[63] Since the α′-SS-α′ protein shows a single set of narrow resonances there is no motion slower than ∼1 kHz assuming the proton chemical shift difference between the different states is on the order of 500 Hz. Narrow line widths are consistent with an absence of intermediate exchange. This view is consistent with a narrow and deep free-energy well for α′-SS-α′ without low-energy barriers separating states close in free energy. The energetics associated with the dimer interface appear to be quite different and can best be viewed as a free-energy surface that is relatively flat with multiple and roughly equivalent energy minima with low-energy barriers between them. This observation has implications for the use of line shape and chemical shift dispersion analysis for judging the attainment of nativelike protein structures in solution.

Amide Hydrogen Exchange. The presence of amide hydrogens that exchange with solvent at rates significantly slower than their intrinsic chemical exchange rate is a hallmark of folded proteins[64] and has been proposed as a criterion for nativelike structure in designed proteins.[65] (α′-SS-α′)$_2$ shows a very simple pattern of amide hydrogen-exchange protection that is largest near the midpoint of each helix (see Table 2). The slowest exchanging amide proton in α′-SS-α′ has a protection

(57) Oas, T. G.; McIntosh, L. P.; EK, O. S.; Dahlquist, F. W.; Kim, P. S. *Biochemistry* **1990**, *29*, 2891−4.
(58) Ciesla, D. J.; Gilbert, D. E.; Feigon, J. *J. Am. Chem. Soc.* **1991**, *113*, 3957−3961.
(59) Osterhout, J., J. J.; Handel, T.; Na, G.; Toumadje, A.; Long, R. C.; Connolly, P. J.; Hoch, J. C.; Johnson, J., W. C.; Live, D.; DeGrado, W. F. *J. Am. Chem. Soc.* **1992**, *114*, 331−37.
(60) Markley, J. L. *Acc. Chem. Res.* **1975**, *8*, 70−80.
(61) Feng, Y.; Sligar, S. G.; Wand, A. J. *Nat. Struct. Biol.* **1994**, *1*, 30−5.
(62) Gibney, B. R.; Rabanal, F.; Reddy, K. S.; Dutton, P. L. *Biochemistry* **1998**, *13*, 4635−4643.
(63) Ilyina, E.; Roongta, V.; Mayo, K. H. *Biochemistry* **1997**, *36*, 5245−50,
(64) Englander, S. W.; Kallenbach, N. R. *Q. Rev. Biophys.* **1983**, *16*, 521−655.
(65) Betz, S. F.; Raleigh, D. P.; DeGrado, W. F. *Curr. Opin. Struct. Biol.* **1993**, *3*, 601−610.

factor of $\sim 5.0 \times 10^4$, corresponding to an apparent ΔG_{unf} of −6.56 kcal/mol. This value is considerably smaller than −22.7 kcal/mol (ΔG_{unf}) derived from isothermal chemical unfolding studies of α′-SS-α′,[15] indicating that a global unfolding transition[66,67] is not rate-controlling for amide hydrogen exchange. Contributions to amide hydrogen exchange from end-fraying of α-helices and a monomer−dimer exchange that is fast relative to amide hydrogen exchange will increase the velocity of the amide hydrogen exchange. The latter assumes that amide exchange will be faster from the monomer in comparison to that from the dimer. Summing of individual amino acid free energies associated with α-helix formation[4,5] commencing with Leu$^{31,31'}$ and ending at the slowest Phe$^{17,17'}$ results in ∼−6.5 kcal/mol free energy. Summing from the N-terminus gives a smaller value of ∼−5.0 kcal/mol; however, the energetic contribution from the disulfide bridge is difficult to estimate. The summed free energy values for α-helix formation are quite similar to the experimentally derived free energies (4.94−6.56; Table 2) for the central three turns, Leu$^{14,14'}$-Leu$^{23,23'}$, of the α-helix suggesting the rate-limiting step for amide hydrogen exchange is end-fraying of the peptide and not from a cooperative global unfolding event.

Comparison of Non-Nativelike α-SS-α and Nativelike α′-SS-α′ Structure. Attainment of *nativelike* protein structure is clearly a requisite for using designed proteins as meaningful models for natural proteins. Nativelike structure is assessed using energetics and cooperativity of the folding−unfolding transition, NMR spectral characteristics, amide hydrogen-exchange rates, hydrophobic dye binding, and resistance to proteolytic hydrolysis.[65]

The unfolding free energy (ΔG_{unf}) does not correlate with nativelike structure in the (α-SS-α)$_2$ family of proteins; however, the molar co-solvation (*m*) value is well correlated. Isothermal chemical denaturation of (α-SS-α)$_2$ reveals a ΔG_{unf} of 14.4 kcal/mol and *m*-value of 1.4 kcal/mol M at 50 °C and a ΔG_{unf} of 16.9 kcal/mol and an *m*-value of 1.94 at 25 °C. Similarly, a ΔG_{unf} of 23.0 kcal/mol and an *m*-value 3.2 kcal/mol M are found for (α′-SS-α′)$_2$ at 50 °C and ΔG_{unf} of 22.7 kcal/mol and an *m*-value of 3.25 kcal/mol M at 25 °C. The larger *m*-values for (α′-SS-α′)$_2$ suggest that this protein exposes more hydrophobic surface area upon unfolding than (α-SS-α)$_2$.[68] This is consistent with hydrophobic groups being effectively sequestered from solvent in the folded state, like that observed in native protein structure. In contrast, the ΔG_{unf} does not correlate with nativelike structure in this family of proteins.[56] Thus the relatively large *m*-value for (α′-SS-α′)$_2$ suggests this property is a good measure of nativelike structure.

Binding of the hydrophobic dye, 8-anilino-1-naphthalene sulfonic acid (ANS) is often used as a qualitative criterion for assessing the attainment of nativelike protein structure. Natural proteins do not typically bind the dye, and non-nativelike proteins avidly bind the dye. The nativelike (α′-SS-α′)$_2$ protein does not bind ANS, whereas the multistructured prototype (α-SS-α)$_2$ also does not bind the dye, suggesting this assay is inadequate for discriminating between nativelike and non-nativelike structure. The lack of significant ANS binding to the non nativelike (α-SS-α)$_2$ protein may reflect a mixture of different and nativelike structures in solution that results in the observed broad NMR resonances for the distribution of structures; however, each of the constituent protein structures may display well packed hydrophobes that limit the amount of hydrophobic dye binding.

Another often-used criterion for attainment of nativelike structure is the presence of a single set of NMR resonances with large chemical shift dispersion. Natural proteins will usually populate a single conformation in solution that is separated sufficiently in free energy from alternate conformers, resulting in a single set of disperse NMR resonances. Thus, designed proteins showing a single set of resonances with a large (compared to that of random coil) range of chemical shifts is a good indicator of nativelike protein structure. This is illustrated quite dramatically with comparison of the ^{13}C HSQC spectra for nativelike (α′-SS-α′)$_2$ and non-nativelike (α-SS-α)$_2$ proteins.[15,16] The attainment of a single, nativelike conformation for (α′-SS-α′)$_2$ is quite evident.

There is clearly not a best, single approach for assessing nativelike structure in designed proteins. These results suggest that a nativelike protein structure in solution, as assessed by NMR spectroscopy, displays anomalies, such as the lack of intermonomer NOEs that are atypical of a natural protein. Single polypeptide maquette designs are currently being pursued to avoid this nonspecific intermolecular interface. The promise of using simple model proteins to simplify the study of larger and more complex protein systems can only be fully realized in the context of a well-packed protein structure that behaves like a natural protein.

Acknowledgment. This work was supported by NIH grants GM35940 and U.S. Army Research Office equipment grant DAAH04-96-1-0312 to A.J.W., NIH grant GM41048 to P.L.D., and NRSA grants GM18121 to J.J.S. and GM17816 to B.R.G. The Washington University Mass Spectrometry Center provided mass spectrometry services. The authors are grateful for Bill DeGrado and Jim Lear for aid in the analytical ultracentrifugation experiments. We thank Peter Flynn, Ernie Fuentes, and Jeffrey Urbauer for many useful discussions.

Supporting Information Available: One table of measured amide hydrogen-alpha hydrogen scalar coupling constants and residual NH dipolar couplings, three tables of aromatic resonance assignments, one table summarizing observed interstrand NOEs, and four figures describing the use of coproporphyrin and pyrene to test the topology of the bundle (PDF). This material is available free of charge via the Internet at http://www.pubs.acs.org

JA983309F

(66) Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. *Proteins* **1994**, *20*, 4−14.
(67) Bai, Y.; Sosnick, T. R.; Mayne, L.; Englander, S. W. *Science* **1995**, *269*, 192−7.
(68) Myers, J. K.; Pace, C. N.; Scholtz, J. M. *Protein Sci.* **1995**, *4*, 2138−48.
(69) Nicholls, A.; Sharp, K. A.; Honig, B. *Proteins* **1991**, *11*, 281−96.