Iterative Protein Redesign

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

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

Received September 16, 1998

Abstract: An iterative redesign protocol for the transformation of a non-native peptide into a series of nativelike proteins derived from elementary considerations of biological evolution coupled with ^{1}H NMR as an artificial selection criterion is presented. Each of three heptad *d* position leucines in the helix−helix interfaces of the prototype heme protein maquette, $[H10H24]_2$ or $(\alpha\text{-SS-}\alpha)_2$, were replaced in a unit modification per helix by more conformationally restricted β-branched and aromatic amino acids. The secondary structure content (evaluated by circular dichroism and infrared spectroscopies), solvent accessibility of the tryptophan residues (measured by fluorescence spectroscopy), global stability (quantitated by isothermal chemical denaturation), and degree of conformational specificity (determined by ^{1}H NMR spectroscopy) of the resultant peptides were determined. Improvement in the degree of conformational specificity was utilized as a selection criterion to choose three of the nine singly modified peptides for a second unit modification per helix. Five of the resultant seven doubly modified peptides were nativelike, as determined by NMR spectroscopy. One of the doubly modified peptides was chosen for a third unit modification per helix, which resulted in three triple variants with low conformational specificity. The 20 proteins synthesized fold into discrete, stable four-α-helix bundles but with differing stabilities ($-\Delta G^{H_2O}$ from 10.50 to 22.73 kcal/mol) and varying degrees of conformational specificity (multistructured to singular solution structure). The singly, doubly, and triply modified (per helix) peptides can be mapped onto a contiguous segment of sequence space, providing the first experimental map of this vast molecular terrain. The energetic contours of sequence space are revealed in terms of both global folding energies ($-\Delta G^{H_2O}$) and degree of conformational specificity within the hydrophobic core. Remarkably, six of the peptides studied (30%) contain uniquely structured hydrophobic cores amenable for NMR structural determination. The map of sequence space readily identifies a plastic site within the protein, a position which can be occupied by various amino acids with retention of a uniquely structured global fold, thereby providing a possible route for iterative redesign toward chemical enzymatic function.

Biological evolution of proteins in a population of individuals proceeds by discrete mutational steps, with each subsequently evaluated for fitness through selective pressures. Examination of successful protein mutations though time maps functional sequence space with the evolutionary history of the protein, providing clues to the final design.[1] At each step, the protein primary sequence must successfully encode not only the structural or catalytic function for which the protein is required but also the overall protein fold, global stability, solubility, supramolecular interactions, cell localization, and final degradation.[2] This array of ancillary functions which are held in the protein primary amino acid sequences are also subject to evolutionary pressures and tend to obscure the engineering specifications which promote the primary function of a protein, e.g., chemical catalysis.

The minimalist hierarchical design[3] and chemical synthesis[4] of proteins offers an approach which avoids many of the ancillary biochemical functions, thus yielding more direct insight by providing both successful and unsuccessful (biologically incompetent) sequences. This approach is proving insightful in identifying the determinants that yield distinct secondary structural motifs and some simple tertiary and quaternary topologies.[5−8] Additionally, the expansion of protein design to include the incorporation of metals and biological cofactors yields a more complete understanding of cofactor assembly and incorporation and promises future chemical reactivity of the types observed by natural enzymes, e.g., Lewis acid hydrolysis and organic substrate oxidation.[9−25] Furthermore, structure-based

* Address correspondence to P. Leslie Dutton, B501 Richards Bldg., Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA 19104. Phone: (215) 898 5668. Fax: (215) 573 2235. E-mail: dutton@mail.med.upenn.edu.

† Present address: Department de Quimica Organica, Universitat de Barcelona, Marti I franques, 1-11, 080028 Barcelona, Spain.

‡ Present address: Departments of Chemistry, Biological Sciences, and Biophysical Sciences and the Center for Structural Biology, State University of New York at Buffalo, Buffalo, NY 14260.

(1) Maynard Smith, J. *Nature* **1970**, *225*, 563−564.

(2) Anfinsen, C. B. *Science* **1973**, *181*, 223−230.

(3) (a) DeGrado, W. F.; Wasserman, Z. R.; Lear, J. D. *Science* **1989**, *243*, 622−628. (b) 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. (c) O'Neil, K. T.; DeGrado, W. L. *Science* **1990**, *250*, 646−651.

(4) (a) Merrifeld, R. B.; Stewart, J. M. *Nature* **1965**, *207*, 522−523. (b) Gutte, B.; Merrifeld, R. B. *J. Am. Chem. Soc.* **1969**, *91*, 501−502.

(5) (a) Harbury, P. B.; Zhang, T.; Kim, P. S.; Alber, T. *Science* **1993**, *262*, 1401−1407. (b) Harbury, P. B.; Kim, P. S.; Alber, T. *Nature* **1994**, *371*, 80−83.

(6) (a) Schenck, H. L.; Gellman, S. H. *J. Am. Chem. Soc.* **1998**, *120*, 4869−4870. (b) Kortemme, T.; Ramirez-Alvarado, M.; Serrano, L. *Science* **1998**, *281*, 253−256.

(7) Smith, C. K.; Regan, L. *Science* **1995**, *270*, 980−982.

(8) Kamtekar, S.; Schiffer, J. M.; Xiong, H.; Babik, J. M.; Hecht, M. H. *Science* **1993**, *262*, 1680−1685.

(9) Choma, C. T.; Lear, J. T.; Nelson, M. J.; Dutton, P. L.; Robertson, D. E.; DeGrado, W. F. *J. Am. Chem. Soc.* **1994**, *116*, 856−865.

10.1021/ja9833117 CCC: $18.00
Published on Web 05/12/1999

redesign methodologies[26−30] have had notable success in redesigning nativelike proteins, even to the point of producing hyperthermophiles, given an input backbone fold. De novo design has shown success in designing proteins from scratch with a nativelike structure, particularly with four-α-helix bundle protein architecture;[31−37] however, no protocols for transforming non-native peptides into nativelike proteins without the aid of a backbone structure have been developed.

Even once discovered, these designed nativelike sequences often remain as inadequately understood isolated points within the vast uncharted terrain of sequence space, with little evidence for the derivation of their conformational specificity. As a result, protein designers have a limited understanding of the consequences of stepwise evolutionary-like mutational changes on such structures.[38] Additionally, despite the widespread application of site-directed mutagenesis studies to natural protein structure and function, no map of sequence space exists to aid the protein designer. Thus, with the current insufficient understanding of sequence space, the probability of designing synthetic proteins with preselected nativelike characteristics appears to be small, but this really has not been investigated experimentally.[39] Furthermore, the probability of evolving or rationally designing an entatic domain[40] capable of promoting preselected enzyme catalysis remains unknown.

The minimalist protein design approach provides a mechanism that potentially avoids a common limitation of retrospective views of protein evolution—incompleteness—by providing insight into both successful and, equally important, unsuccessful mutations which lead to biological incompetence. Herein, we use the prototype four-α-helix bundle heme protein maquette scaffold to sketch a limited stepwise mutational sortie around the sequence of a peptide which encodes a non-native four-α-helix bundle protein on a contiguous map of sequence space. A primitive iterative redesign protocol grounded in elementary evolutionary considerations, coupled with ^{1}H NMR as a selection criterion, is developed to transform a non-native peptide into a nativelike four-α-helix bundle protein whose solution structure has been solved by NMR methods and is presented in the accompanying paper.[41] A consequence of the iterative redesign protocol is an exploration of sequence space delineating the dependence of structural stability and specificity on each of the hydrophobic core modifications studied. A high density of uniquely folded proteins and a plastic site that tolerates a wide variety of hydrophobic side chains while maintaining conformational specificity are identified. Finally, multiple stepwise paths to uniquely folded proteins as well as mutational pathways around otherwise non-native proteins are modeled in sequence space.

Materials and Methods

Peptide Synthesis. All peptides were synthesized on a continuous-flow PerSeptive Biosystems Pioneeer solid-phase synthesizer using the Fmoc/^{t}Bu protection strategy[42] with NovaSyn PR-500 resin at 0.2 mmol scale. Single extended coupling cycles (60 min) were employed for OPfp/HOBt chemistry, while single standard coupling cycles (30 min) were used for OH/HATU-based reactions. The side-chain protecting groups used are as follows: His (^{t}Boc); Lys (^{t}Boc); Glu (O^{t}Bu); Cys

(10) Robertson, D. E.; Farid, R. S.; Moser, C. C.; Urbauer, J. L.; Mulholland, S. E.; Pidikiti, R.; Lear, J. D.; Wand, A. J.; DeGrado, W. F.; Dutton, P. L. *Nature* **1994**, *368*, 425−431.

(11) (a) Rabanal, F.; DeGrado, W. F.; Dutton, P. L. *J. Am. Chem. Soc.* **1996**, *118*, 473−474. (b) Rabanal, F.; DeGrado, W. F.; Dutton, P. L. *Tetrahedron Lett.* **1996**, *37*, 1347.

(12) (a) Gibney B. R.; Mulholland S. E.; Rabanal F.; Dutton P. L. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 15041−15046. (b) Mulholland, S. E.; Gibney, B. R.; Rabanal, F.; Dutton, P. L. *J. Am. Chem. Soc.* **1998**, *120*, 10296−10302.

(13) Gibney, B. R.; Johansson, J. S.; Rabanal, F.; Skalicky, J. J.; Wand, A. J.; Dutton, P. L. *Biochemistry* **1997**, *36*, 2798−2806.

(14) Sharp, R. E.; Rabanal, F.; Moser, C. C.; Dutton, P. L. *Proc. Natl. Acad. Sci. U.S.A.* 10465−10470.

(15) Sharp, R. E.; Diers, J. R.; Bocian, D. F.; Dutton, P. L. *J. Am. Chem. Soc.* **1998**, *120*, 7103−7104.

(16) (a) Rau H. K.; DeJonge N.; Haenel W. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11526−11531. (b) Rau H. K.; Haehnel W. *J. Am. Chem. Soc.* **1998**, *120*, 468−476.

(17) (a) Mutz, M. W.; McLendon, G. L.; Wishart, J. F.; Gaillard, E. R.; Corin, A. F. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 9521−9526. (b) Mutz, M. W.; Case, M. A.; Wishart, J. F.; Ghadiri, M. R.; McLendon, G. L. *J. Am. Chem. Soc.* **1999**, *121*, 638−639.

(18) (a) Rojas, N. R.; Kamtekar, S.; Simons, C. T.; McLean, J. E.; Vogel K. M.; Spiro, T. G.; Farid, R. S.; Hecht, M. H. *Protein Sci.* **1997**, *6*, 2512−2524. (b) Scott, M. P.; Biggins, J. *Protein Sci.* **1997**, *6*, 340−346.

(19) (a) Pinto, A. L.; Hellinga, H. W.; Caradonna, J. C. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 5562−5567. (b) Coldren, C. D.; Hellinga, H. W.; Caradonna, J. P. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 6635.

(20) (a) Benson, D. E.; Wisz, M. E.; Liu, W.; Hellinga, H. W. *Biochemistry* **1998**, *37*, 7070−7076. (b) Wisz, M. E.; Garrett, C. Z.; Hellinga, H. W. *Biochemistry* **1998**, *37*, 8269−8277.

(21) D'Auria, G.; Maglio, O.; Nastri, F.; Lombardi, A.; Morelli, G.; Paolillo, L.; Pedone, C.; Pavone, V. *Chem. Eur. J.* **1997**, *3*, 350−362.

(22) (a) Arnold, P. A.; Shelton, W. R.; Benson, D. R. *J. Am. Chem. Soc.* **1997**, *119*, 3181−3182. (b) Arnold, P. A.; Benson, D. R.; Brink, D. J.; Hendrich, M. P.; Jas, G. S.; Kennedy, M. L.; Petasis, D. T.; Wang, M. X. *Inorg. Chem.* **1997**, *36*, 5306−5315.

(23) Ghadiri, M. R.; Choi, C. *J. Am. Chem. Soc.* **1990**, *112*, 1630−1632.

(24) (a) Klemba, M.; Regan, L. *Biochemistry* **1995**, *34*, 10094−10100. (b) Farinas, E.; Regan, L. *Protein Sci.* **1998**, *7*, 1939−1946.

(25) (a) Dieckmann, G. R.; McRorie, D. K.; Tierney, D. L.; Utschig, L. M.; Singer, C. P.; O'Halloran, T. V.; Penner-Hahn, J. E.; DeGrado, W. F.; Pecoraro, V. L. *J. Am. Chem. Soc.* **1997**, *119*, 6195−6196. (b) Dieckmann, G. R.; McRorie, D. K.; Lear, J. D.; Sharp, K. A.; DeGrado, W. F.; Pecoraro, V. L. *J. Mol. Biol.* **1998**, *280*, 897−912.

(26) Lazar, G. A.; Desjarlais, J. R.; Handel, T. M. *Protein Sci.* **1997**, *6*, 1167−1178.

(27) (a) Dahiyat, B. I.; Mayo, S. L. *Science* **1997**, *278*, 82−87. (b) Dahiyat, B. I.; Mayo, S. L. *Protein Sci.* **1996**, *5*, 895−903. (c) Malakauskas, S. M.; Mayo, S. L. *Nat. Struct. Biol.* **1998**, *5*, 470−475.

(28) Jiang, X.; Bishop, E. J.; Farid, R. S. *J. Am. Chem. Soc.* **1997**, *119*, 838−839.

(29) (a) Harbury, P. B.; Tidor, B.; Kim, P. S. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 8408−12. (b) Harbury, P. B.; Plecs, J. J.; Tidor, B.; Alber, T.; Kim, P. S. *Science* **1998**, *282*, 1462−1467

(30) Raleigh, D. P.; Betz, S. F.; DeGrado, W. F. *J. Am. Chem. Soc.* **1995**, *117*, 7558−7559.

(31) Hill, R. B.; DeGrado, W. F. *J. Am. Chem. Soc.* **1998**, *120*, 1138−1145.

(32) Dolphin, G. T.; Brive, L.; Johansson, G.; Baltzer, L. *J. Am. Chem. Soc.* **1996**, *118*, 11297−11298.

(33) (a) Munson, M.; Balasubramanian, S.; Fleming, K. G.; Nagi, A. D.; O'Brien, R. O.; Sturtevant, J. M.; Regan, L. *Protein Sci.* **1996**, *5*, 1584−1593. (b) Munson, M.; O'Brien, R.; Sturtevant, J. M.; Regan, L. *Protein Sci.* **1994**, *3*, 2015−2022.

(34) (a) Betz, S. F.; DeGrado, W. F. *Biochemistry* **1996**, *35*, 6955−6962. (b) Betz, S. F.; Liebman, P. A.; DeGrado, W. F. *Biochemistry* **1997**, *36*, 2450−2458.

(35) Gibney, B. R.; Rabanal, F.; Skalicky, J. J.; Wand, A. J.; Dutton, P. L. *J. Am. Chem. Soc.* **1997**, *119*, 2323−2324.

(36) (a) Schafmeister, C. E.; Miercke, L. J. W.; Stroud, R. M. *Science* **1993**, *262*, 734−738. (b) Schafmeister, C. E.; LaPorte, S. L.; Miercke, L. J. W.; Stroud. R. M. *Nat. Struct. Biol.* **1997**, *4*, 1039−1046.

(37) Roy, S.; Ratnaswamy, G.; Boice, J. A.; Fariman, R.; McLendon, G.; Hecht, M. H. *J. Am. Chem. Soc.* **1997**, *119*, 5302−5306.

(38) Cordes, M. H. J.; Davidson, A. R.; Sauer, R. T. *Curr. Opin. Struct. Biol.* **1996**, *6*, 3−10.

(39) Mathematical experiments exploring sequence space have been performed, see: (a) Lau, K. T.; Dill, K. A. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 638−642. (b) Dill, K. T. *Biochemistry* **1990**, *29*, 7133−7155. (c) Shakhnovich, E. I.; Gutin, A. M. *Nature* **1990**, *346*, 773−775. (d) Lipman, D. J.; Wilbur, W. J. *Proc. R. Soc. London B* **1991**, 7−11. (e) Macken, C. A.; Perelson, A. S. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 6191−6195.

(40) Vallee, B. L.; Williams, R. J. P. *Proc. Natl. Acad. Sci. U.S.A.* **1968**, *59*, 498.

(41) Skalicky, J. J.; Gibney, B. R.; Rabanal, F.; Bieber-Urbauer, R. J.; Dutton, P. L.; Wand, A. J. *J. Am. Chem. Soc.* **1999**, *121*, 4941−4951.

(42) Bodanszky, M. *Peptide Chemistry: A Practical Approach*, 2nd ed.; Springer-Verlag: New York, 1993.

(Trt); Arg (Pmc). After peptide assembly, the N-termini were manually acetylated and thoroughly washed with DMF followed by CH_2Cl_2. The peptides were simultaneously cleaved/deprotected using 90:8:2 (v/v/v) trifluoroacetic acid/ethanedithiol/water for 2 h. Crude peptides were 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 aqueous−acetonitrile gradients containing 0.1% (v/v) TFA. The N-terminal cysteine residues of purified peptides were air oxidized to disulfides in 100 mM ammonium carbonate buffer, pH 9.5 (5 h), which was followed by analytical HPLC. The identities of the resulting di-α-helical disulfide-bridged peptides were confirmed with laser desorption mass spectrometry, performed at the Protein Chemistry Laboratory of the University of Pennsylvania.

Solution Molecular Weight Determination. Size exclusion chromatography was performed on a Beckman System Gold HPLC system equipped with a diode array detector using a Pharmacia Superdex 75 column eluted with aqueous buffer (10 mM KP_i, 100 mM KCl, pH 8.0) at a flow rate of 0.75 mL/min. The column was standardized with the following globular proteins (MW): aprotinin (6.5 kDa), horse heart cytochrome *c* (12.1 kDa), myoglobin (16.7 kDa), chymotrypsinogen (25.0 kDa), ovalbumin (43.0 kDa), and bovine serum albumin (67.0 kDa). The eluents were monitored at 220 nm and solution molecular weights determined by interpolation.

Circular Dichroism Spectropolarimetry. CD spectra were recorded on an AVIV 62DS spectropolarimeter using rectangular quartz cells of 0.2 cm path length with a 10-s averaging time. Thermal control was maintained by a thermoelectric module with a Neslab CFT-33 refrigerated recirculating water bath as a heat sink. Peptide concentrations were between 2.5 and 3.0 μM (four-helix bundle) as determined spectrophotometrically using $\epsilon_{280} = 5600\ M^{-1}\ cm^{-1}$ helix^{-1} for Trp.

Infrared Spectroscopy. Fourier transform infrared spectra were recorded on a Bruker IFS 66 FT-IR spectrometer equipped with a Globar source, a KBr beam splitter, a mercury−cadmium-telluride detector, and an attenuated total reflectance cell.

Fluorescence Spectroscopy. Steady-state measurements of intrinsic tryptophan and bound ANS fluorescence were recorded with an ISS K2 multifrequency cross-correlation phase and modulation spectrofluorometer in quartz cells of 1.0 cm path length within a thermostatically controlled cell holder (25 ± 0.1 °C). Excitation and emission slit widths were 2 nm.

Denaturation Studies. Peptide denaturation curves at 25 °C were fit to a dimer folded to two-monomer unfolded equilibrium[43] using a nonlinear least-squares routine to the following equation:

$$\text{fraction folded} = 1 - (\exp(-\Delta G_{unf}/RT)/4P) \times [(1 + 8P/\exp(-\Delta G_{unf}/RT))^{1/2} - 1]$$

where P is the molar concentration of total monomeric protein and $\Delta G_{unf} = \Delta G^{H_2O} + m$[Gdn·HCl], m being the cosolvation term, which is a measure of the cooperativity of the transition, and [denaturant] the concentration of denaturant (M).

NMR Spectroscopy. NMR experiments were performed using identical conditions on a Varian INOVA-600 spectrometer. One-dimensional spectra were acquired with 8192 complex points using a spectral width of 7200 Hz per FID. The peptide samples were prepared at 500 μM di-α-helical monomer concentration (250 μM four-helix bundle) in 20 mM phosphate (pH 7.25), 50 mM KCl, and 8% D_2O. Proton chemical shifts were referenced to an external sample of DSS at 0.00 ppm. The NMR data were processed on a SGI Crimson computer using the FELIX95 software (Biosym Technologies, San Diego, CA).

Results

Evolution-like Iterative Protein Redesign. The *apo* form of our prototype heme protein maquette, $[H10H24]_2$, folds into a compact state containing a high degree of secondary structure but displays characteristics of a poorly ordered and non-native hydrophobic core.[10] $[H10H24]_2$ is comprised of four identical 31-amino-acid peptides, (α-SH), composed of ≈3.5 heptad repeats (heptad *c* to heptad *a*, Glu5 to Leu31) assembled as a four-α-helix bundle from a pair of disulfide-linked, di-α-helical monomers, (α-SS-α). Figure 1 demonstrates the global architecture of this maquette architecture using the NMR-derived structure of $(\alpha'\text{-SS-}\alpha')_2$, IFL in the present nomenclature, modeled in an anti topology.[41,44]

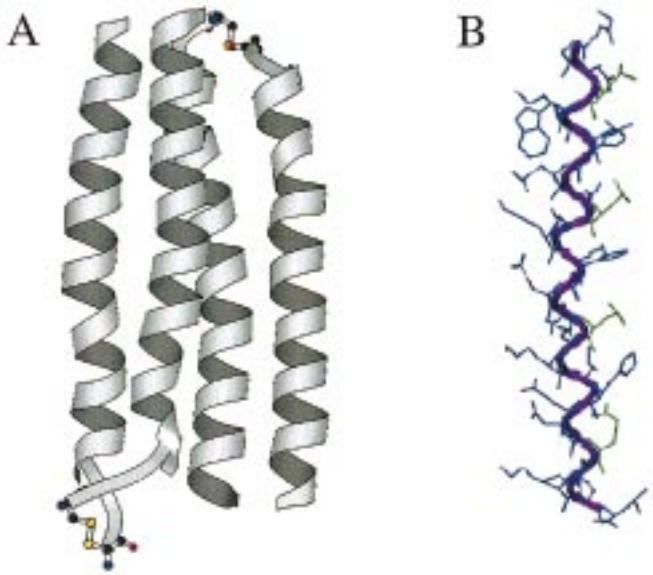

Figure 1. Solution NMR structure of (A) $[H10H24\text{-}L6I,L13F]_2$ (IFL), $[(Ac\text{-}CGGGE\mathbf{I}WKL{\cdot}HEE\mathbf{F}LKK{\cdot}FEE\mathbf{L}LKL{\cdot}HEERLKK{\cdot}L\text{-}CONH_2)_2]_2$, modeled in the anti global topology, and (B) a single modeled α-helix of the prototype $[H10H24]_2$ with the heptad *d* positions in green for clarity. The heptad *d* positions utilized for the iterative redesign protocol are shown in boldface in the primary sequence for clarity.

Our iterative redesign strategy is based on simple, conservative amino acid alterations within the hydrophobic core and does not require detailed knowledge of the backbone structure or the relative orientation of the two monomers in the dimeric four-α-helix architecture (disulfide loops on same side, syn, or opposite sides, anti),[34] which may be variable between peptide sequences. Based on biological evolution, it is composed of single mutational steps followed by evaluation providing an artificial selective pressure. We have restricted our attention to three hydrophobic heptad *d* positions per helix, occupied in the prototype by leucines whose side-chain entropy[45] may cause the non-native behavior of the prototype heme protein maquette, $[H10H24]_2$. We have reserved both the histidines in the hydrophobic *a* positions (heme ligation) and the arginine at *d* position 27 (heme redox potential modulation), derived by analogy to the cytochrome bc_1 complex.[46,47] Our alterations are architecturally rather unsophisticated because any single change

(43) Mok, Y.-K.; De Prat Gay, G.; Butler, P. J.; Bycroft, M. *Protein Sci.* **1996**, *8*, 310.

(44) Skalicky, J. J.; Bieber, R. J.; Gibney, B. R.; Rabanal, F.; Dutton, P. L.; Wand, A. J. *J. Biomol. NMR* **1998**, *11*, 227−228.

(45) Creamer, T. P.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *92*, 5937−5941.

(46) (a) Xia, D.; Yu, C.-A.; Kim, H.; Xia, J.-Z.; Kachurin, A. M.; Zhang, L.; Yu, L.; Deisenhofer, J. *Science* **1997** *277*, 60−66. (b) Kim, H.; Xia, D.; Yu, C.-A.; Xia, J.-Z.; Kachurin, A. M.; Zhang, L.; Yu, L.; Deisenhofer, J. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, 95, 8026−8033. (c) Iwata, S.; Lee, J. W.; Okada, K.; Lee, J. K.; Iwata, M.; Rasmussen, B.; Link, T. A.; Ramaswamy, S.; Jap, B. K. *Science* **1998**, *281*, 64−71.

(47) Zhang, Z.; Huang, L.; Shulmeister, V. M.; Chi, Y.-I.; Kim, K. K.; Hung, L.-W.; Crofts, A. R.; Berry, E. A.; Kim, S.-H. *Nature* **1998**, *392*, 677−684.

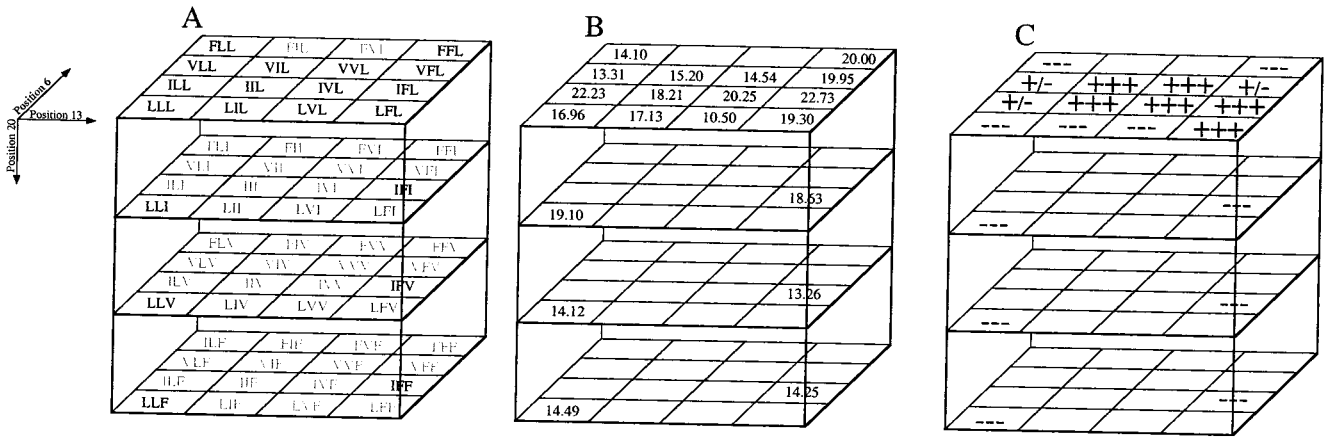

Figure 2. Cubic segment of sequence space explored in terms of (A) sequence, (B) $-\Delta G^{H_2O}$, and (C) conformational specificity. Each peptide is represented by a three-letter abbreviation derived from the amino acids at positions 6, 13, and 20 in panel A (the legend for panels B and C), with the peptides synthesized and studied in black. The prototype is LLL (top level, front row, left). Single sequence modifications can be traced as moves of a rook in chess. Thus, any member of the front row, top level is a single modification from LLL at position 13. Diagonal moves require multiple modifications. The stability of each peptide ($-\Delta G^{H_2O}$ in kcal/mol at 25 °C) is given in panel B. The degree of conformational specificity, multistructured (− − −), approaching singularity (+/−), or uniquely structured (+++), of each sequence derived from NMR is mapped in panel C.

Table 1. Peptide Characterization

	molecular weight			circular dichroism				fluorescence	
				molar ellipticity (deg·cm²/dmol)					
peptide	monomer mass (amu)	gel permeation aqueous (amu)	infrared amide I′ (cm^{-1})	Θ_{222}	Θ_{208}	ratio	% α-helix	tryptophan λ_{em}^{max} (nm)	ANS intensity (counts/s)
				Prototype					
LLL ([H10H24]$_2$)	7575	20100	1649	25800	26800	0.95	80.5	349	1138
				Single Variants					
ILL	7575	19900	1647	26100	21100	1.23	81.6	348	445
VLL	7547	20000	1649	26600	23700	1.11	82.9	345	630
FLL	7643	19800	1648	24400	24500	1.00	76.4	346	753
LIL	7575	19400	1650	24900	26700	0.93	77.7	347	443
LVL	7547	18200	1648	24200	22900	1.06	75.7	348	462
LFL	7643	19900	1649	25100	25600	0.98	78.4	345	377
LLI	7575	19800	1647	25920	28700	0.93	81.0	344	748
LLV	7547	20000	1649	25300	25600	0.99	79.2	349	642
LLF	7643	19600	1649	25700	28600	0.90	80.3	346	932
				Double Variants					
IIL	7575	19800	1649	26400	24400	1.08	82.5	348	583
IVL	7547	19500	1649	25600	24900	1.03	80.1	347	517
IFL	7643	19000	1651	27500	24300	1.13	85.9	344	369
VIL	7547	18100	1649	23500	24400	0.96	73.3	347	973
VVL	7518	17800	1650	24700	23800	1.04	77.3	347	473
VFL	7614	18200	1649	23700	24700	0.96	74.0	346	552
FFL	7712	18800	1648	27500	26900	1.02	86.0	346	1055
				TripleVariants					
IFI	7643	20000	1649	25300	24100	1.05	79.0	345	742
IFV	7547	19800	1648	25300	24100	1.05	79.0	348	832
IFF	7712	19600	1649	22700	22000	1.03	71.0	347	921

made in the primary sequence of one helix of [H10H24]$_2$ results in four identical alterations within the four-helix structure, and rather nonspecific hydrophobic packing forces at the helix interfaces were used to generate proteins of singular solution structure. We chose to restrict the substitution of the leucines to other hydrophobic but β-branched or aromatic amino acids, namely isoleucine, valine, and phenylalanine; hence, the binary pattern[8] of hydrophobic and hydrophilic amino acids (H/P) in each sequence is identical. Moreover, these four hydrophobic amino acids share degenerate codons (NUU and NUC), and so our modified peptides might represent viable neutral point mutations in natural proteins,[48] and the iterative redesign protocol may be amenable for future incorporation into combinatorial libraries. Limiting our inquiry to these four hydrophobic amino acids in three of the 27 helix positions reduced the size of sequence space from 1.3×10^{35} (20^{27}) peptides to the 64 peptides shown schematically in Figure 2A. Starting with [H10H24]$_2$ (here called LLL for the *d* position leucines at 6, 13, and 20), we probed this local protein sequence space by making nine peptides containing single conservative modifications per helix, evaluated using improvement in the ^{1}H NMR spectra, consistent with singular structure as a selection criterion since native-state natural proteins typically display high-quality ^{1}H NMR spectra. While ^{1}H NMR spectral quality was chosen as the selection criterion to provide a NMR solution structure

(48) Stryer, L. *Biochemistry*; W. H. Freeman and Co.: New York, 1988.

of one of these proteins, other criteria more relevant to biological evolution may be selected, i.e., resistance to proteolysis or catalytic rate. Modified sequences which did not map to the desired selection criterion were not pursued further since they were considered lethal mutations. Peptides with improved conformational specificity after a single modification were further evolved using a second iteration of redesign by incorporation of another unit modification per helix. Last, a third iteration of redesign was performed on one of the doubly modified variants. Thus, we traced an iterative redesign pathway in discrete steps from our conformationally nonspecific[49] prototype $[H10H24]_2$ to a series of modified peptides with uniquely structured hydrophobic cores, amenable for complete NMR structural characterization, and in the process mapped a contiguous segment of sequence space to aid in future protein design.

Characterization of the Di-α-helical Dimers. Secondary Structure. Table 1 summarizes the far-UV circular dichroism (CD) and FT-IR spectral results of each of the $[H10H24]_2$ variants (singly, doubly, and triply modified per helix), which are typical of highly α-helical peptides (74−86% helical content) with CD minima at 208 and 222 nm and FT-IR amide I′ bands between 1647 and 1651 cm^{-1}. Additionally, Table 1 lists the fluoresence emission maxima of the tryptophan residues of each peptide (λ_{max}^{em} of 344−349 nm), which are consistent with solvent-exposed heptad *e* positions and indicate that the hydrophobic core alterations have no large effect on the tryptophan microenvironment. Figure 3 shows the results from the VLL peptide, representative of the series. These spectroscopic data clearly demonstrate that these heptad *d* position modifications result in no significant alteration in the secondary structure content of the homotetrameric α-helical bundles.

Peptide Aggregation State. The association state of each of the $[H10H24]_2$ variants was evaluated using gel permeation chromatography (initial loading concentrations ranging from 5 to 200 μM). Table 1 shows that, at all concentrations studied by gel permeation chromatography, the peptides elute with apparent molecular weights between 17.8 and 20.1 kDa (14.9−15.2 kDa calculated for the four-helix bundles) based on a column standardized with globular proteins. These peptides coelute with $[H10H24\text{-}L6I,L13F]_2$, a structurally characterized four-α-helix bundle, whose four-α-helix bundle aggregation state has been determined by gel premeation chromatography, analytical ultracentrifugation, water-suppressed longitudinal encode−decode (SLED) experiments, and preliminary ^{15}N NMR relaxation measurements, indicating that the related proteins have similar hydrodynamic radii, as anticipated. As illustrated in Table 1, the gel permeation data demonstrate that these modifications in a heptad *d* position result in no significant alteration in the aggregation state preference of the bundle, demonstrating the robust nature of this maquette scaffold.

Global Thermodynamic Stability. Table 2 presents the global stability of each of the $[H10H24]_2$ variants assayed by isothermal chemical denaturation with guanidine hydrochloride followed by CD spectropolarimetry,[50] with representative data shown for VLL in Figure 3. Each peptide displays a fully reversible, cooperative unfolding transition consistent with a two-state unfolding process (dimer folded ↔ two unfolded monomers); the midpoint of the denaturation curves ($[Gdn]_{1/2}$), molar cosolvation terms (*m*), and $-\Delta G^{H_2O}$ values were highly dependent on the location and identity of the modification made

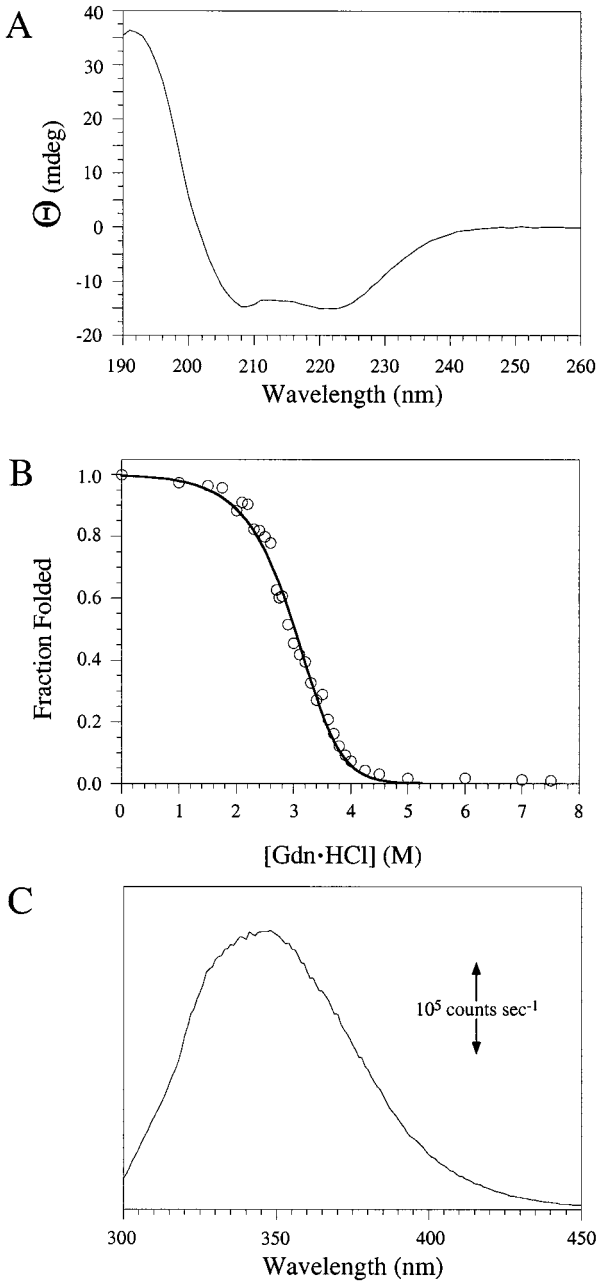

Figure 3. Representative characterization of the VLL peptide by far-UV/circular dichroism spectopolarimetry (secondary structure), isothermal chemical denaturation (global stability), and fluorescence spectroscopy (tryptophan microenvironment). The ^1H NMR spectrum of VLL, indicative of global conformational specificity approaching singularity, is given in Figure 4. The conditions of each experiment are as given in the text.

to the four helices. Since the *m*-value reflects the difference in solvent-accessible surface areas between the folded and unfolded states,[51] and hence the efficiency of hydrophobic core packing, larger *m*-values suggest peptides with more nativelike behavior.

Hydrophobic Dye Binding. The binding of the hydrophobic dye, 8-anilino-1-naphthalenesulfonic acid (ANS), to designed proteins is used as a qualitative probe for the presence of accessible hydrophobic regions due to poor hydrophobic core packing.[52] Substantial ANS fluorescence is observed within the

(49) Lattman, E. E.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 439−441.

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

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

(52) Betz, S. F.; Raliegh, D. P.; DeGrado, W. F. *Curr. Opin. Struct. Biol.* **1993**, *3*, 601−610.

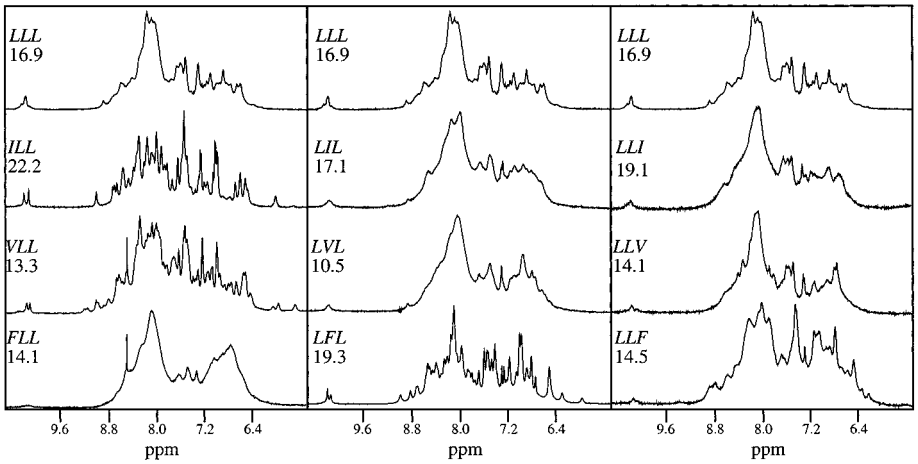

Figure 4. Hydrophobic core specificity, as determined by chemical shift dispersion in the ^{1}H NMR, and global stability ($-\Delta G^{H_2O}$) for each of the singly modified peptides studied. One-dimensional NMR spectra in the aromatic amide proton region of LLL, ILL, VLL, and FLL (left panel, from top); LLL, LIL, LVL, and LFL (middle panel, from top); and LLL, LLI, LLV, and LLF (right panel, from top). ^{1}H NMR experimental conditions are as given in Materials and Methods.

Table 2. Thermodynamic Evaluation of $[H10H24]_2$ Variants

peptide	m	$[Gdn]_{1/2}$	$-\Delta G^{H_2O}$ (25 °C)
	Prototype		
LLL ($[H10H24]_2$)	1.94	5.04	16.96
	Single Variants		
ILL	3.19	4.72	22.23
VLL	2.05	2.98	13.31
FLL	1.43	4.86	14.10
LIL	2.06	4.84	17.13
LVL	1.75	1.90	10.50
LFL	2.51	4.83	19.30
LLI	2.58	4.63	19.10
LLV	1.71	4.07	14.12
LLF	1.55	4.73	14.49
	Double Variants		
IIL	2.36	4.68	18.21
IVL	3.13	4.17	20.25
IFL	3.25	4.80	22.73
VIL	1.95	4.16	15.20
VVL	2.19	3.37	14.54
VFL	2.57	4.97	19.95
FFL	2.98	4.29	20.00
	Triple Variants		
IFI	2.96	3.80	18.63
IFV	1.75	3.36	13.26
IFF	1.79	3.83	14.25

low dielectric environment of protein cores which possess hydrophobic binding sites, e.g., apomyoglobin and bovine serum albumin, or disordered hydrophobic cores, e.g., molten globule folding intermediates, whereas ANS does not fluoresce to an appreciable extent in aqueous solution.[53] Table 1 compares the fluorescence intensity of each of the peptides, 300−1100 counts s^{-1}, to those of apomyoglobin and bovine serum albumin, 65 000 and 40 000 counts s^{-1}, under standard assay conditions (2 μM ANS with 40 μM peptide at pH 8.0, 10 mM KP_i, 100 mM KCl) illustrating that these peptides do not bind ANS with high affinity. The ANS fluorescence intensity of all of these designed peptides was close to the baseline and some 40−120-fold lower than that observed in apomyoglobin, indicating nativelike behavior for all the peptides studied and illustrating that ANS is not a sensitive probe of conformational specificity.

(53) (a) Semisotnov, G. V.; Rodionova, N. A.; Razgulyaev, O. I.; Uversky, V. N.; Gripas, A. F.; Gilmanshin, R. I. *Biopolymers* **1991**, *31*,119−128. (b) Stryer, L. *J. Mol. Biol.* **1965**, *13*, 482−495.

Global Conformational Specificity. The singularity or multiplicity of the tertiary fold(s) of each variant was qualitatively assessed with the amide and methyl proton line widths (see Figures 4−6) and the number of methyl correlations present in the natural abundance ^{13}C HSQC spectrum. Large chemical shift dispersion and narrow resonances (∼15 Hz for the amide protons) for proteins of these molecular weights and single sets of well-dispersed amide and methyl correlations is strong evidence for single solution conformations. Broad proton resonances and/or unexpected numbers of amide or methyl correlations strongly suggest multiple, slowly interconverting solution conformations. At this level of investigation, NMR spectroscopy will reveal the presence of multiple structural forms that are interconverting on time scales slower than ∼100 Hz. Faster interconversions will lead to averaged spectral parameters which may appear quite nativelike.

The prototype LLL is characterized by large proton line widths and very small proton and methyl carbon chemical shift dispersion. Nonspecific aggregation was ruled out as a source of the observed line broadening in LLL since the proton line widths are invariant over the wide ranges of concentration (25−500 μM), temperature (10−50 °C), and ionic strength (50−250 mM KCl). Furthermore, the gel permeation studies of Table 1 indicate that none of the poorly resolved spectra shown are the result of aggregation. Many of the LLL variants demonstrate remarkably narrow line widths and large chemical shift dispersion (selection criterion), characteristic of a single solution conformation, and clearly demonstrate unique magnetic environments for many of the methyl groups designed to pack into the hydrophobic core. The NOESY spectrum for the uniquely structured IFL peptide shows nearly all of the characteristic H^N-H^N NOEs expected for a helical protein, and IFL, ILL, and LFL proteins also display a large number of H^N-H^N NOEs, confirming their predominantly helical nature.

(i) First Iteration. We follow the sequence space mapped out in Figure 2A, starting with the prototype $[H10H24]_2$ (LLL at the front left of the top level). Moves for a single site alteration per helix (four per four-α-helix bundle) are those of the rook in chess. Compared with the stability of LLL ($-\Delta G^{H_2O} = 16.9$ kcal/mol), we find a substantial, 11.8 kcal/mol, range of stabilities within the nine single-site variants, as shown in Figure 2B. The greatest stability increase (5.3 kcal/mol) was elicited by an Ile at position 6 (ILL at second row, left on the top level), yielding $-\Delta G^{H_2O} = 22.2$ kcal/mol, which coincides with

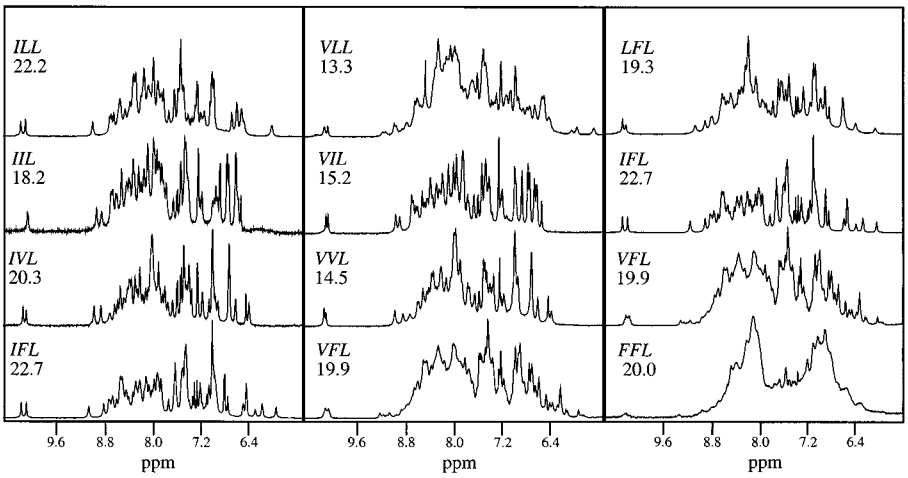

Figure 5. Determination, by ^{1}H NMR chemical shift dispersion, of the hydrophobic core specificity for each doubly modified peptide. One-dimensional NMR spectra in the aromatic amide proton region of ILL, IIL, IVL, and IFL (left panel, from top); VLL, VIL, VVL, and VFL (middle panel, from top); and LFL, IFL, VFL, and FFL (right panel, from top). The numbers in each panel are the $-\Delta G^{H_2O}$ values in kcal/mol. ^{1}H NMR experimental conditions are as given in Materials and Methods.

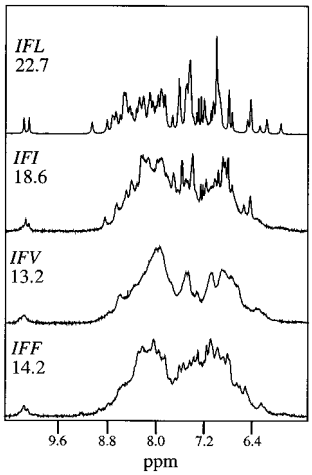

Figure 6. Triply modified peptide hydrophobic core specificity as evaluated by NMR spectroscopy. ^{1}H NMR spectra in the aromatic amine proton region of IFL, IFI, IFV, and IFF. The global stabilities of each four-α-helix bundle are given as the $-\Delta G^{H_2O}$ values in kcal/mol in each panel. ^{1}H NMR experimental conditions are as given in Materials and Methods.

improved NMR resolution (Figure 4) and hence better, but not yet unique, conformational specificity (indicated by +/− in Figure 2C). The greatest stability decrease (6.4 kcal/mol) resulted from placing a Val at position 13 (LVL), yielding $-\Delta G^{H_2O} = 10.5$ kcal/mol, with no observable improvement in conformational specificity. On the other hand, Val placed at position 6 (VLL) yields an improvement in conformational specificity despite significant (3.6 kcal/mol) destabilization. However, further NMR study (^{13}C HSQC) indicates that both ILL and VLL exist as a mixture of two conformations in solution (denoted as +/− in Figure 2C) in the ratios of 80:20 and 60:40, respectively. The only other single-position variant with an improved conformational specificity resulted from substitution with phenylalanine at position 13 (LFL); this was stabilized by 2.3 kcal/mol compared to the prototype and displayed a singular conformation by 2-D NMR (denoted +++ in Figure 2C). Similar examination of the position 20 variants failed to lead to improvement in conformational specificity, although LLI is stabilized by 2.1 kcal/mol over LLL. These data illustrate that single conservative modifications per helix that draw only on hydrophobic packing forces[54,55] are sufficient to transform the non-native LLL prototype peptide into uniquely structured proteins without resorting to non-hydrophobic interactions such as buried ion pairs,[56−58] helix caps,[59] structural disulfides,[60] or metal binding sites.[61]

(ii) Second Iteration. The three singly modified peptides with improved conformational specificity (ILL, VLL, and LFL) were further modified on each helix to yield seven double-variant peptides. Table 2 shows five peptides (IIL, IVL, IFL, VFL, and FFL) that are more stable than the LLL prototype, and Figure 5 demonstrates a separate set of five (IIL, IVL, IFL, VIL, and VVL) that fold to single solution conformations. VFL folds to a mixture of a major and a minor solution conformation, and only FFL is clearly multistructured in solution. We note that peptides containing a β-branched amino acid at position 6 either firmly possess conformational singularity (VVL, VIL, IVL, IIL, and IFL) or approach it (ILL, VLL, VFL), consistent with flexible leucine residues being responsible for the multistructured nature of the prototype. Since the Ile6 residues are not in direct van der Waal's contact in the NMR structure of ILF,[41] their sandwiching of the Leu9 residue may provide a hydrophobic contact important for the conformational specificity of IFL.

(iii) Third Iteration. One of the doubly modified peptides with high conformational specificity, IFL, whose solution structure has been determined by NMR methods,[41] was selected for a third cycle of modifications at position 20 since it already included alterations at positions 6 and 13 resultant from the first two iterations. The replacement of the leucine at position 20 in IFL with isoleucine, valine, and phenylalanine results in three triply modified peptides which show a complete loss of conformational specificity (IFI, IFV, and IFF, right panel, second column from front, top to bottom). Table 2 and Figure 6 illustrate that these peptides are both less stable than IFL and have low conformational specificity, as evaluated by ^{1}H NMR

(54) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151−176.
(55) Lim, W. A.; Sauer, R. T. *J. Mol. Biol.* **1991**, *219*, 359−376.
(56) Graddis, T. J.; Myszka, D. G.; Chaiken, I. M. *Biochemistry* **1993**, *32*, 12664−12671.
(57) O'Shea, E. K.; Lumb, K. J.; Kim, P. S. *Curr. Biol.* **1993**, *3*, 658−667.
(58) Schneider, J. P.; Lear, J. D.; DeGrado, W. F. *J. Am. Chem. Soc.* **1997**, *119*, 5742−5743.
(59) Dois, A. L.; Chakrabartty, A.; Klingler, T. M. Baldwin, R. L. *Biochemistry* **1994**, *33*, 3396−3403.
(60) Hecht, M. H.; Richardson, J. S.; Richardson, D. C.; Ogden, R. C. *Science* **1990**, *249*, 884−891.
(61) Regan, L.; Clarke, N. D. *Biochemistry* **1990**, *29*, 10878−10883.

spectroscopy. The lack of conformational specificity in these variants of IFL illustrates that the Leu20 residues in IFL are critical to the maintenance of the hydrophobic core packing, contrary to expectations based on the flexible nature of the leucine side chain on the α-helix backbone.

Discussion

Using a maquette scaffold, we have developed an iterative redesign protocol for the transformation of non-native peptides into proteins with nativelike properties. Employing simple conservative changes in an iterative manner within the hydrophobic core of a designed four-α-helix bundle, we have produced a large collection of nativelike designed proteins, a central challenge to molecular design. By viewing these proteins on a map of sequence space, we have provided protein designers with a guide to the design of nativelike proteins.

The peptide-derived data support the idea that the acquisition of conformational specificity is likely to rest on a threshold thermodynamic stability (ΔG^{thr}) which is sufficient for the peptide to fold, beyond which increases in ΔG^{H_2O} do not necessarily correlate with improvements in conformational specificity.[62] Several conformationally specific proteins (e.g., VVL, $-\Delta G^{H_2O} = 14.5$ kcal/mol, +++) are considerably less stable than other multistructured peptides (cf. FFL, $-\Delta G^{H_2O} = 20.5$ kcal/mol, − − −), illustrating the lack of correlation between thermodynamic stability and uniqueness of the solution conformation. Additionally, the molar cosolvation terms (*m*-values) and guanidine denaturation midpoints ($[Gdn]_{1/2}$ values) from which ΔG^{H_2O} is derived show no global correlation with the observed conformational specificity. Our data to date suggest an upper limit of −14.5 kcal/mol (VVL) for the threshold thermodynamic stability (ΔG^{thr}), sufficient to fold this peptide architecture into a four-α-helix bundle of singular solution conformation. Since the design of uniquely folded proteins does not require, but may lead to, substantial thermodynamic stability, computational and experimental methodologies for protein design should not solely optimize ΔG while striving to obtain nativelike structure.

The three sets of singly modified peptides (ILL,LIL, LLI; VLL, LVL, LLV; and FLL, LFL, LLF) demonstrate an interesting local correlation between the measured thermodynamic data and the observed conformational specificity. In all three sets of single variants with identical amino acid compositions, which ameliorates any effects due to peptide hydrophobicity or helical propensity,[3c,63−65] the sequence arrangement with the greatest *m*-value (ILL, VLL, and LFL) possesses the best NMR spectral characteristics, consistent with the attainment of conformational specificity driven by improvements in hydrophobic packing interactions. While larger *m*-values tend to lead to increases in ΔG^{H_2O}, the sequence with the highest ΔG^{H_2O} is not always the most conformationally specific, as it may possess a smaller *m*-value with a larger midpoint ($[Gdn]_{1/2}$). While *m*-values appear to correlate between sets of peptides with identical sequences (locally), the lack of a global correlation between nonidentical sequences suggests that sequence-derived changes in peptide chemical and physical properties obscure a direct correlation at this time.

The effects of the single modifications per helix are dependent on both the identity of amino acid substitution and its local sequence context, indicating that the design of uniquely structured proteins requires more chemical information than contained in the simple hydrophobic/hydrophilic (H/P) binary pattern of amino acids. We find that Ile is the most stabilizing amino acid at positions 6 (ILL) and 20 (LLI) but with position 13 most stable containing a Phe (LFL) with leucines in the remaining *d* positions; this indicates the importance of specifying explicitly both the sequence location and the identity of the amino acid in the design process. Furthermore, as indicated above, peptides with identical amino acid compositions and binary patterns can have a wide range of thermodynamic stabilities and conformational specificity outcomes. Additionally, the combination of two single modifications, VLL (destabilized relative to LLL with two solution conformations) and LVL (destabilized and multistructured) into a single peptide, VVL, results in both higher stability and conformational specificity due to changes in local packing context, which is recognized to influence secondary structure.[66] Thus, in contrast to the concept of additivity derived from surface alanine mutants of T4 lysozyme[67] and adventitiously observed in previous core variants of maquettes,[15] these double hydrophobic core-modified peptides are not simply the sum of the two single modifications in terms of either global unfolding energy or conformational specificity, illustrating the differences between surface and core mutations. Clearly, both the identity and the position of the amino acid substitutions made are important for generating a designed protein that folds to a single solution conformation. Thus, experimental[8] and computational approaches[39] to protein folding and design which ignore the chemical and physical identity of the particular amino acid at each position may be of limited utility in generating uniquely structured hydrophobic cores, the hallmark of native protein structure.

Mapping our peptides onto a segment of sequence space which reflects a single point in shape space[68] provides the first experimental view of both the global as well as the local details of this vast terrain. Our synthesis and detailed characterization of the peptides allows for a more facile recognition of discrete areas and patterns within the landscape than the stochastic points generated and studied by combinatorial chemistry[8,69] or the unsynthesized peptides in computational[26−30] methodologies. The energetic contour of the terrain and possible redesign pathways (direct and circuitous), as well as variations in the density of the chosen selection criterion (areas containing differing fractions of proteins with singular solution structure), become evident upon examination of even this minuscule segment of sequence space, illustrating its utility in guiding future protein design.

Consistent with our design, 100% of the proteins studied, shown in black in Figure 2A, fold into compact, stable four-α-helix bundle proteins compared to the 60% and 5% values derived from the binary pattern[8] and random[69,70] experimental combinatorial chemistry libraries, respectively. These values validate their respective designs, since all are higher than the 1−5% predicted from computational (the mathematical H/P model)[39a,71] studies. The area of sequence space which we investigated contains six peptides (30% of the 20 studied, 9.3%

(62) Gibney, B. R.; Rabanal, F.; Dutton, P. L. *Curr. Opin. Chem. Biol.* **1997**, *1*, 537−542.
(63) Creamer, T. P.; Rose, G. D. *Proteins* **1994**, *19*, 85−97.
(64) Zhou, N. E.; Kay, C. M.; Sykes, B. D.; Hodges, R. S. *Biochemistry* **1993**, *32*, 6190−6197.
(65) Chakrabartty, A.; Schellman, J. A.; Baldwin, R. L. *Nature* **1991**, *351*, 586−588.
(66) Minor, D. L. J.; Kim, P. S. *Nature* **1996**, *380*, 730−734.
(67) Zhang, X.-J.; Baase, W. A.; Matthews, B. W. *Protein Sci.* **1992**, *1*, 761−766.
(68) Schuster, P.; Fontana, W.; Stadler, P. F.; Hofacker, I. L. *Proc. R. Soc. London B* **1994**, *255*, 279−284.
(69) Davidson, A. R.; Sauer, R. T. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 2146−2150.
(70) Davidson, A. R.; Lumb, K. J.; Sauer, R. T. *Nat. Struct. Biol.* **1995**, *2*, 856−864.
(71) Bornberg-Bauer, E. *Biophys. J.* **1997**, *73*, 2393−2403.

of the 64 sequences in Figure 2) with singular solution structure (our desired selection criterion), suggesting that rational design of a synthetic protein with nativelike structural characteristics is not as difficult as traditionally believed whether using the de novo, combinatorial library, or computational design approaches.

Globally, the terrain is rather rugged in terms of both global peptide stabilities and degree of conformational specificity. Energetically, Figure 2B illustrates that the terrain of sequence space is quite irregular. Sequences encoding highly stable peptides (IVL, 20.25 kcal/mol) reside adjacent to those encoding much less stable peptides (LVL, 10.50 kcal/mol). Furthermore, it is clear from Figure 2C that adjacent sequences can have very different conformational specificities. The rugged terrain underscores the necessity for mapping sequence space, as it provides a more complete understanding of the successful protein designs.

Mutationally, the uniquely structured four-α-helix bundle proteins studied can be reached from the prototype LLL by several noncrossing redesign paths consisting of unit modifications per helix using only hydrophobic packing interactions. As shown in Figure 2A, IIL was directly reached in two rook-like moves (LLL → ILL → IIL) but can also be reached by alternate, more circuitous routes in sequence space, for instance LLL → VLL → VIL → IIL. However, the pathway of LLL → LIL → IIL is not viable, since it contains a step which does not meet the chosen selection criterion (LLL → LIL). Thus, our map of sequence space illustrates how iterative redesign (evolution) might utilize various pathways between two points circumventing intermediates which do not meet the selection criterion (lethal mutations), allowing for a more complete sampling of all possible sequences. This result is consistent with the observation of second site revertants in site-directed mutagenesis studies, which illustrate that lethal local mutations are of limited global consequence, as they can be circumvented, and furthermore, suggests that iterative redesign of other proteins toward nativelike structures and functions can be achieved through alternate networks with different starting points in sequence space.

Locally, fluctuations in the fraction of peptides with singular solution structures are readily observed when viewing the modifications in sequence space. The singly modified position 20 peptides (left, front, all levels) reside in an area devoid of conformational specificity, while the site composed of all the second site modifications of ILL is astonishing high, at 75% (top level, second row, left to right). Most notably, the broad range of tolerated modifications at position 13 in ILL with retention of singular solution structure implies considerable plasticity here when Ile is at position 6, a property of natural proteins derived from evolution by natural selection previously unobserved in a designed protein. Clearly, the highly stable global structure is able to compensate for the alterations at position 13 by side-chain and backbone structural rearrangements, which affect the magnitude of the global unfolding energy while retaining a discrete global fold and conformational specificity (the selection criterion). Thus, the conformational specificity of these peptides (IIL, IVL, and IFL) is determined (perhaps overdetermined) by information in other sites throughout the sequence, and these sequences represent a plateau or cluster of fitness in sequence space which encodes a series of uniquely folded four-α-helix bundle proteins. It is within this plateau that the plasticity of position 13 allows the opportunity for further iterations, leading to creation of addition functional capabilities such as prosthetic group binding and catalysis in this maquette scaffold. Consistent with this hypothesis is the observation that IXL proteins containing non-native amino acids (X = α-aminoisobutyric acid, L-*tert*-butylglycine, and L-norvaline) at position 13 fold into single solution conformations (unpublished results). Given the codon degeneracy of the four amino acids chosen, the plasticity observed in ILL, the observed circumvention of lethal points in sequence space, and the recently demonstrated self-replication[72] of coiled-coil peptides, it is reasonable to consider that iterative redesign ("chemical evolution") toward an entatic domain in synthetic peptides capable of promoting preselected enzyme catalysis may be feasible.

Acknowledgment. This work was supported by National Institues of Health grants to P.L.D. (41048 and 55876) and A.J.W. (35940), with NSRA postdoctoral grants to B.R.G. (17816) and J.J.S. (18121). F.R. was supported by a postdoctoral fellowship from the European Molecular Biology Organization. This work was also supported in part by the National Science Foundation (MRSEC IRG DMR96-32598). The authors wish to thank Bill DeGrado and the members of his laboratory for many stimulating discussions. Mass spectrometry was performed at the Protein Chemistry Laboratory of the University of Pennsylvania. The authors dedicate this work in loving memory to Lawrence John Gibney, Sr.

JA9833117

(72) Lee, D. H.; Granja, J. R.; Martinez, J. A.; Severin, K.; Ghadiri, M. R. *Nature* **1996**, *382*, 525−528.