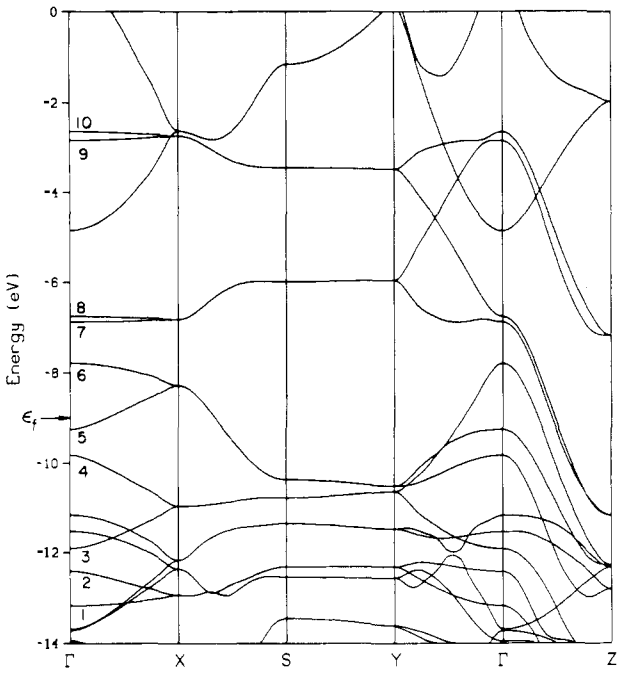


Figure 14. Calculated band structure for **11(33)**.

ordering is as follows (in eV): (**1**) 0.61, (**3**) 0.0, (**9**) 1.19, (**10**) 1.27, (**11(31)**) 0.95, (**11(32)**) 0.97, (**11(33)**) 0.86. These energies are subject to the errors associated with the EH method. As we have stated in our previous papers on this subject[1] these instabilities (relative to graphite) do not preclude the existence of these nets because the energy values (obtained by a very approximate MO method) only represent thermodynamic stability and do not take into account kinetic stability. Organic chemistry is full of examples of compounds that are thermodynamically less stable but exist anyway because of large kinetic barriers to their rearrangement. All of the nets discussed here should have large kinetic barriers for rearrangement to diamond or graphite. Many bonds have to be broken to get from one structure to another.

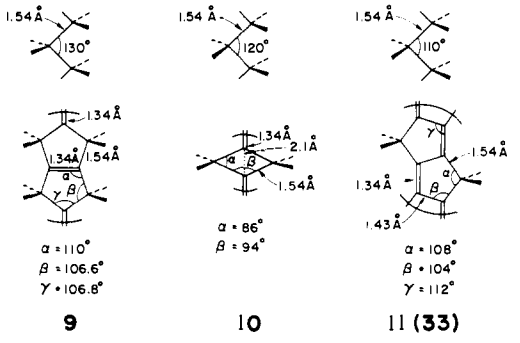
It is tempting to speculate that conducting 3,4-connected nets such as those advocated earlier[1] or here might be formed as local regions during the catalyzed high-pressure conversion of graphite to diamond or during the low-pressure reverse (and less rewarding) process.

Acknowledgment. Kenneth Merz thanks all of the group members for helping in familiarizing him with solid-state theory and with the programs used during the course of this research. We also acknowledge Prof. A. F. Wells for providing us with some useful comments on the derivation of nets. We also thank the Clardy group for supplying us with their version of the PLUTO program. Our research at Cornell was generously supported by the National Science Foundation through research grant DMR821722702 to the Materials Science Center. We are grateful to Jane Jorgensen and Elisabeth Fields for their expert drawings.

Appendix

The extended Hückel method[20] in the tight binding approximation[21] was used in all calculations. The parameters used for carbon and hydrogen were the following: $H_{ss} = -21.4$ eV, $H_{pp} = -11.4$ eV, $\zeta_{s,p} = 1.625$; $H_{ss} = -13.6$ eV, $\zeta_{s,p} = 1.3$, respectively. To calculate average properties for the 1-D cases a 100 point k-point set was used and for the 3-D examples a 27 point k-point set was chosen.[22]

The geometries used for the calculations are given below.



(20) Hoffmann, R. *J. Chem. Phys.* **1963**, *39*, 1397. Hoffmann, R.; Lipscomb, W. N. *J. Chem. Phys.* **1963**, *36*, 3179; **1962**, *37*, 2872.
(21) Whangbo, M.-H.; Hoffmann, R. *J. Am. Chem. Soc.* **1978**, *100*, 6093.
(22) Pack, J. D.; Monkhorst, J. H. *Phys. Rev. B* **1977**, *16*, 1748.

Design of a 4-Helix Bundle Protein: Synthesis of Peptides Which Self-Associate into a Helical Protein

Siew Peng Ho and William F. DeGrado*

Contribution from E. I. du Pont de Nemours & Company, Central Research & Development Department, Experimental Station, Building 328, Wilmington, Delaware 19898.
Received February 18, 1987

Abstract: An incremental synthetic approach is described for the design of a 4-helix bundle protein. On the basis of secondary structure prediction rules and model building, two amphiphilic 16-residue peptides, α_1A and α_1B, were designed to form α-helices that would cooperatively tetramerize (give stable 4-helix structures) in solution. The peptides were synthesized by chemical methods, and their ability to form stable helical tetramers was confirmed by molecular weight determinations and circular dichroism studies in the presence and absence of denaturant. The free energy of tetramerization of both peptides was determined to be on the order of −20 kcal/mol. In the second stage of the work, short peptidic links were inserted between the sequence of two α_1B peptides in an attempt to design a covalent cross-link between two of the helical pairs in the 4-helix bundle structure. Two peptides, α_1B-Pro-α_1B and α_1B-Pro-Arg-Arg-α_1B, were synthesized, and their tendency to form dimeric aggregates (4-helix structures) was probed. The peptide α_1B-Pro-α_1B was found to give trimeric aggregates rather than the expected dimeric structures. Incorporation of charged arginine residues in the loop achieved the desired result: the ensuing peptide, α_1B-Pro-Arg-Arg-α_1B, forms stable helical dimers in solution.

The design of large molecules with defined conformational properties is a necessary first step in the de novo design of macromolecular receptors and catalysts. To date, most of the molecules that have been designed to mimic the properties of

0002-7863/87/1509-6751$01.50/0

enzymes and receptors have utilized cyclodextrin,[1] spherand,[2] cyclophane,[3] or crown ether[4] frameworks. In contrast, there have been few attempts to design from first principles proteins with predetermined structures and properties. The main difference between these two approaches is that in organic systems, one seeks to enforce through cyclization and the use of rigid aromatic subunits the one conformation that gives maximum recognition and catalysis. Proteins, however, are composed of flexible chains, and for a 100 residue protein molecule without any disulfide cross-links, there are in theory 10^{100} possible backbone conformations[5] (this number is larger than the estimated number of atoms in the universe[5]). The principles underlying the processes by which proteins adopt a relatively well defined set of conformations out of such a large number of possibilities are only beginning to be understood.

Empirical secondary structure prediction schemes[6] and conformational analysis algorithms[7] are some of the methods that have been developed for structure prediction in proteins. In addition systematic synthetic approaches have been devised to elucidate the mechanisms by which proteins fold into their three-dimensional structures. For instance, designed, synthetic peptides have been used to show how hydrophobic periodicity[8] in a protein sequence influences the formation of simple secondary structures such as amphiphilic α-helices and β-sheets.[9] Together with hydrophobic periodicities, empirical prediction methods have been very effective tools for the design of peptides comprising a single secondary structural feature.[10] This approach has been taken one step further by the design of α-helical coiled-coils.[11] A logical extension to this synthetic approach is the design and synthesis of multiple copies of secondary structures that will pack in a predictable manner to give a protein with a predetermined three-dimensional shape.[12] This paper describes the first and second stages (Figure 1, a and b) in an incremental approach aimed at the design and synthesis of a 4-helix bundle protein.

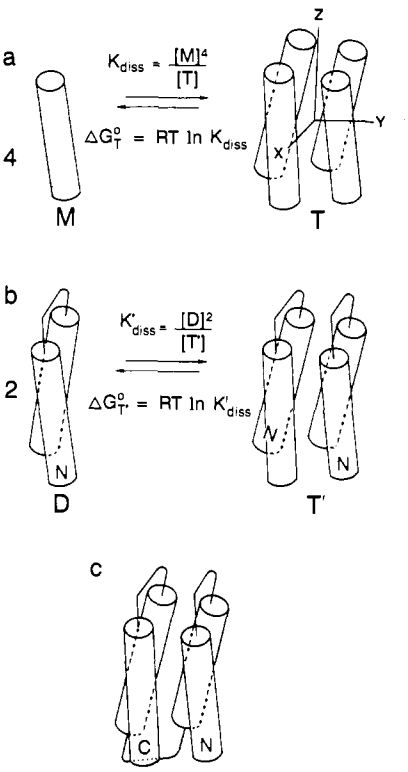


Figure 1. Incremental approach to the design of a 4-helix bundle protein. (a) The sequence of an amphiphilic helix is designed with the aim of obtaining stable tetrameric aggregates in solution. (b) Once the sequence of the helix has been optimized, sequences for the loops can be evaluated. With use of the best helix sequence obtained in step (a) various designed loop sequences are evaluated on the basis of the stability of the resulting dimer of helical hairpins. (Another possible arrangement would place the loops at opposite ends of the bundle.) (c) The ultimate 4-helix bundle is constructed from 4 optimized helices and 3 optimized loops. Dissociation constants and free energies are experimentally determined as a measure of the stability of the tetramers and dimers. For conceptual clarity the monomeric species in a and b (left side of the equilibrium) are depicted as cylinders. This is not meant to imply that they are predominantly helical—CD measurements indicate that the monomeric forms of these peptides are predominantly in random coil configurations.

In collaboration with Eisenberg and his co-workers,[12a] we have designed a 4-helix bundle protein (shown schematically in Figure 1c) that is an idealized version of a naturally occurring class of proteins which includes myohemerythrin, apoferrittin, tobacco mosaic virus coat protein, and cytochrome C'.[13] These proteins are comprised of four helices connected together by three loops. The helices are packed nearly antiparallel to one another, crossing at an angle of approximately 20°. With this packing arrangement the helices diverge from a point of closest approach, giving rise to cavities that can accommodate metals or heme moieties. The proteins in this structural class serve different functional roles and have little sequence homology, and yet their folding patterns are strikingly similar. This suggests that hidden beneath the complexities of the individual amino acid sequences of these proteins is a common, highly degenerate code that allows them to adopt similar conformations.

Inspection of interhelical packing in naturally occurring 4-helix bundles shows that the structural cores (i.e., the portions of the proteins which appear to be involved in structural stabilization and not in ligand binding) of these protein motifs are composed almost entirely of close-packed apolar residues. This is often accomplished by interdigitation between residues on neighboring

(1) (a) Tabushi, I.; Kuroda, Y.; Yamada, M.; Higashimura, H.; Breslow, R. *J. Am. Chem. Soc.* **1985**, *107*, 5545–5546. (b) D'Souza, V. T.; Hanabusa, K.; O'Leary, T.; Gadwood, R. C.; Bender, M. L. *Biochem. Biophys. Res. Commun.* **1985**, *129*, 727–732. (c) Tabushi, I.; Kuroda, Y. *Adv. Catal.* **1983**, *32*, 417–466. (d) Breslow, R.; Bovy, P.; Lipsey Hersch, C. *J. Am. Chem. Soc.* **1980**, *102*, 2115–2117.

(2) (a) Cram, D. J.; Lam, P. Y. S.; Ho, S. P. *J. Am. Chem. Soc.* **1986**, *108*, 839–841. (b) Cram, D. J.; Katz, H. E. *J. Am. Chem. Soc.* **1983**, *105*, 135–137.

(3) (a) Lutter, H.-D.; Diederich, F. *Angew. Chem.* **1986**, *98*, 1125–1127. (b) Schurmann, G.; Diederich, F. *Tetrahedron Lett.* **1986**, *27*, 4249–4252.

(4) (a) Sasaki, S.; Shionoya, M.; Koga, K. *J. Am. Chem. Soc.* **1985**, *107*, 3371–3372. (b) Sasaki, S.; Kawasaki, M.; Koga, K. *Chem. Pharm. Bull.* **1985**, *33*, 4247–4266. (c) Chao, Y.; Weisman, G. R.; Sogah, G. D. Y.; Cram, D. J. *J. Am. Chem. Soc.* **1979**, *101*, 4984–4958.

(5) Creighton, T. E. *Proteins*; W. H. Freeman and Co.: New York, 1983; pp 161.

(6) (a) Chou, P. Y.; Fasman, G. D. *Adv. Enzymol.* **1978**, *47*, 45–148. (b) Sueki, M.; Lee, S.; Powers, S. P.; Denton, J. B.; Konishi, Y.; Scheraga, H. A. *Macromolecules* **1984**, *17*, 148–155. (c) Scheraga, H. A. *Pure Appl. Chem.* **1978**, *50*, 315–324. (d) Kabsch, W.; Sander, C. *FEBS Lett.* **1983**, *155*, 179–182. (e) Lim, V. I. *J. Mol. Biol.* **1974**, *88*, 873–894.

(7) (a) Karplus, M. *Adv. Biophys.* **1984**, *18*, 165–190. (b) Levitt, M. *Annu. Rev. Biophys. Bioeng.* **1982**, *11*, 251–271. (c) Weiner, P. K.; Langridge, R.; Blaney, J. M.; Schaefer, R.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 3754–3758.

(8) (a) Eisenberg, D.; Weiss, R. M.; Terwilliger, T. C. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 140–144. (b) Eisenberg, D.; Weiss, R. M.; Terwilliger, T. C.; Wilcox, W. *Faraday Symp. Chem. Soc.* **1982**, *18*, 109–120.

(9) (a) DeGrado, W. F.; Lear, J. D. *J. Am. Chem. Soc.* **1985**, *107*, 7684–7689. (b) Brack, A.; Spach, G. *J. Am. Chem. Soc.* **1981**, *103*, 6319–6323.

(10) (a) Mutter, M.; Altmann, K. H.; Muller, K.; Vuilleumier, S.; Vorherr, T. *Helv. Chim. Acta* **1986**, *69*, 985–995. (b) Kaiser, E. T.; Kézdy, F. J. *Science* **1984**, *223*, 249–255. (c) Kaiser, E. T.; Kézdy, F. J. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 1137–1143.

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

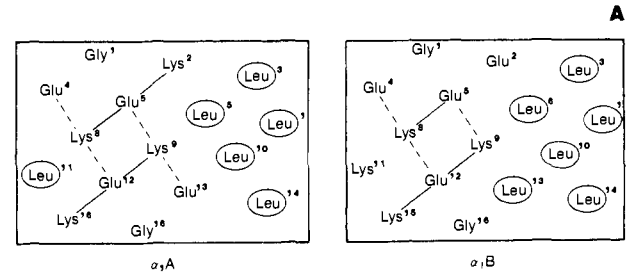
(12) (a) Eisenberg, D.; Wilcox, W.; Eshita, S. M.; Pryciak, P. M.; Ho, S. P.; DeGrado, W. F. *Proteins* **1986**, *1*, 16–22. (b) Erickson, B. W.; Daniels, S. B.; Reddy, P. A.; Unson, C. G.; Richardson, J. S.; Richardson, D. C. *Current Communications in Molecular Biology*; Cold Spring Harbor Laboratory: New York, 1986; pp 53–57. (c) Utter, M. *Angew. Chem., Int. Ed. Engl.* **1985**, *24*, 639–653. (d) Kullman, W. *J. Med Chem.* **1984**, *27*, 106–115. (e) Moser, R.; Thomas, R. M.; Gutte, B. *FEBS Lett.* **1983**, *157*, 247–251. (f) Gutte, B.; Daumigen, M.; Wittschieber, E. *Nature (London)* **1979**, *281*, 650–655.

(13) (a) Weber, P. C.; Salemme, F. R. *Nature (London)* **1980**, *287*, 82–84. (b) Argos, P.; Rossman, M. G.; Johnson, J. E. *Biochem. Biophys. Res. Commun.* **1977**, *75*, 83–86.

helices.[14] Residues protruding from positions $i-4$, i, and $i+4$ form a ridge (a four-type ridge) which packs against residues $k-3$, k, and $k+3$ (a three-type ridge) of a neighboring helix.[14] This and similar packing patterns have been predicted on the basis of model building and energy minimization studies.[15] Thus, a major common force influencing the formation and stability of the 4-helix bundle appears to be the packing of apolar side chains into semiregular arrays. Favorable interactions between the helical macrodipole[16] are also thought to stabilize the 4-helix bundle geometry. Finally, the loops between the helices probably play a role in facilitating the formation of 4-helix bundles by diminishing the loss in configurational entropy which occurs when four helices are brought together.

Design of a 4-Helix Bundle. Our design of a synthetic 4-helix bundle seeks to maximize the pseudo-222 symmetry found in the natural bundles and includes four helices of identical sequence.[12a] The use of symmetry simplifies the design: therefore a sequence is sought which, upon application of a 222 symmetry operator, gives rise to a tightly packed 4-helix bundle with a cyclic interaction pattern. The choice of identical helical sequences also facilitates the use of an incremental, experimental strategy in the process of designing and optimizing the protein as illustrated in Figure 1. In the first step of this process the helix is designed and synthesized. If helix–helix packing is indeed the dominant force in determining the structure of the 4-helix bundle, and the sequence of the peptide is properly designed, four such peptides should assemble into a tetramer as illustrated in Figure 1a. The free energy for the folding process can be evaluated from the experimentally determined dissociation constant for tetramerization. After optimization of the helical sequence, loop sequences can then be examined for their ability to induce a hairpin bend (Figure 1b). The effects of the various loop designs are studied while keeping the helical sequence constant. In this way, the individual requirements for the design of a 4-helix bundle (helix formation, helix–helix packing, loop formation) can be separately evaluated and optimized.

The strategy used in the design of the helices in the 4-helix bundle has been described previously.[12a] To summarize briefly, the helices should be composed of amino acid residues which strongly favor helix formation. The helices should be amphiphilic: i.e., they should have an apolar face to interact with neighboring helices and a polar face to maintain water solubility of the ensuing aggregate. Driven by hydrophobic interactions, proper packing of the apolar side chains should play the dominant role in stabilizing the structure. In addition, the acetyl group was added at the N-terminus to stabilize helix formation.[17] Glycine, a strong helix breaker,[6] was included as the first and last residue to set the stage for adding a hairpin loop between the helices. Following these criteria, the sequence of α_1A was designed, primarily by using CPK and Kendrew models. The apolar side chains were arranged such that they can interdigitate most effectively when in a 4-helix bundle, thus stabilizing this structure over other possible alternative aggregates. As many favorable electrostatic side chain/side chain interactions as possible were also included. The design was subsequently examined by using computer graphics and energy refinement techniques to confirm that a reasonable structure could indeed be obtained. The sequence of α_1A is shown in Figure 2B. Its helical net diagram projection (Figure 2A) illustrates the positions from which the various side chains project from the surface of a helix. In this illustration, the side chain positions are plotted on a cylinder with the same dimensions as the helix, and then the cylinder is cut along its length, opened, and laid flat with the helical axis along the Z direction. Potential electrostatic



B

α_1	Ac-Glu Leu Leu Lys Lys Leu Leu Glu Glu Leu Lys Gly-COOH
α_1A	Ac-Gly Lys Leu Glu Glu Leu Leu Lys Lys Leu Leu Glu Glu Leu Lys Gly-COOH
α_1B	Ac-Gly Glu Leu Glu Glu Leu Leu Lys Lys Leu Lys Glu Leu Leu Lys Gly-$CONH_2$
α_2B (P)	α_1B — Pro — α_1B
α_2B(PRR)	α_1B — Pro — Arg — Arg — α_1B

Figure 2. (A) Helical net diagrams of α_1A and α_1B. Hydrophobic leucine residues are circled while possible salt bridges are represented by the solid and broken lines. (B) The amino acid sequences of α_1,[12a] α_1A, α_1B, α_2B(P), and α_2B(PRR).

interactions along three-type ridges are indicated by solid lines, and those along four-type ridges are indicated by dotted lines. The hydrophobic leucyl residues are circled and segregate on one face of the helix. On the basis of an analysis of α_1A with computer graphics, it appears that leucine at position 11 of the sequence is excessively exposed to solvent while the carboxylate of Glu 13 is partially buried in the structure of the tetramer. Also it has recently been demonstrated for the N-terminal helix of ribonuclease A that it is destabilizing to place negative charges near the N-terminus or postively charged residues near the C-terminus of the structure.[17,18] Therefore, an improved sequence (α_1B) was obtained by replacing Leu 11 with Lys and Glu 13 with Leu. In addition, Lys 2 of α_1A was changed to a Glu, and the C-terminal carboxylate was changed to a carboxamide. The synthesis and characterization of these two peptides, α_1A and α_1B, will be the subject of the first portion of this paper. The properties of these 16-residue peptides will be compared with those of a shortened 12-residue version of α_1A described earlier.[12a]

The second stage of the design process involves the design of a loop to connect two helices of optimized sequence (α_1B). On the basis of an examination of CPK models, a single proline residue appeared capable of serving as a suitable link if the C- and N-terminal glycine residues are slightly unwound (Figure 2B). This amino acid was chosen because of its high tendency to break helices and its frequent occurrence at turns.[6] In addition, a second loop was evaluated which contained two additional arginyl residues after the proline, α_2B(PRR). The extra charge is expected to favor the formation of a hairpin loop due to its high degree of hydrophilicity.[19]

Experimental Section

Materials and Methods. Peptide syntheses were carried out with a Beckman Model 990B peptide synthesizer programmed as described previously.[20] Amino acid derivatives were from Bachem and were checked by TLC[20b] prior to use. The peptides were cleaved from their resins with use of anhydrous HF/anisole (9:1) at 0 °C for 60 min. High-performance liquid chromatography was accomplished with a Du Pont Instruments Series 8800 HPLC equipped with a Du Pont Instruments variable-wavelength UV spectrophotometer. Amino acid analyses were obtained with a Beckman 119 Cl ion exchange system with nin-

(14) (a) Chothia, C. *Annu. Rev. Biochem.* **1984**, *53*, 537–572. (b) Chothia, C.; Levitt, M.; Richardson, D. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 4130–4134.

(15) (a) Chou, K. C.; Nemethy, G.; Scheraga, H. A. *J. Am. Chem. Soc.* **1984**, *106*, 3161–3170. (b) Chou, K. C.; Nemethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1983**, *87*, 2869–2881.

(16) (a) Hol, W. G. J.; Halie, L. M.; Sander, C. *Nature (London)* **1981**, *294*, 532–536. (b) Sheridan, R. P.; Levy, R. M.; Salemme, F. R. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 4545–4549. (c) Wada, A. *Adv. Biophys.* **1976**, *9*, 1–63.

(17) Shoemaker, K. R.; Kim, P. S.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Nature (London)* **1987**, *326*, 563–567.

(18) (a) Shoemaker, K. R.; Kim, P. S.; Brems, D. N.; Marqusee, S.; York, E. J.; Chaiken, I. M.; Stewart, J. M.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 2349–2353. (b) Kim, P. S.; Bierzynski, A. Baldwin, R. L. *J. Mol. Biol.* **1982**, *162*, 187–199.

(19) (a) Rose, G. R. *Nature (London)* **1978**, *272*, 586–590. (b) Rose, G. R.; Gierasch, L. M.; Smith, J. A. *Adv. Protein Chem.* **1985**, *37*, 1–109.

(20) (a) Barany, G.; Merrifield, R. B. *Peptides, Analysis, Synthesis and Biology*; Academic: New York, 1979; Vol. 2. (b) DeGrado, W. F.; Kaiser, E. T. *J. Org. Chem.* **1982**, *47*, 3258–3261.

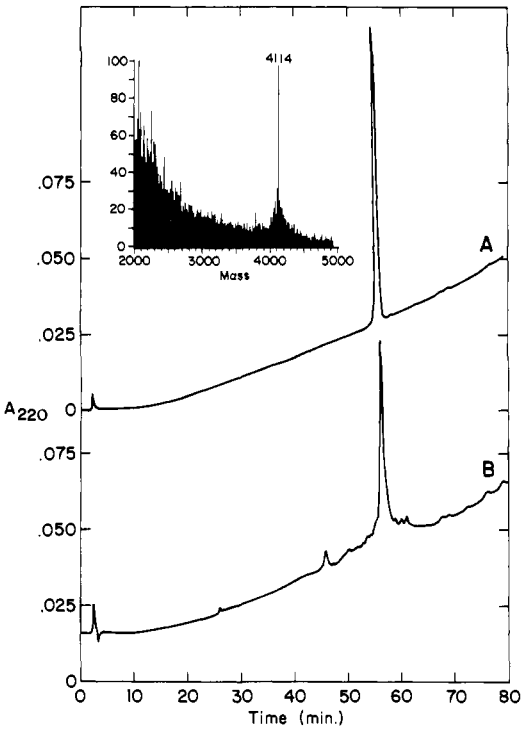


Figure 3. HPLC chromatograms of crude α_2B(PRR) (panel B) and the peptide after reverse phase purification (panel A). The chromatography was run by using an analytical Hamilton PRP1 column and a linear, 80 min gradient of 0–100% aqueous acetonitrile containing 0.1% TFA. The FABMS of the pure peptide at a thousand resolution is shown in the inset.

hydrin detection. Positive fast atom bombardment mass spectra were obtained at unit resolution on a VG ZAB-E instrument at an accelerating voltage of 8 kV. An Ion Tech gun produced the 8 kV, 2 mA Xe atom beam. Circular dichroism spectra were recorded on a Jasco 500 spectropolarimeter interfaced with an IBM AT computer. Spectra were taken with the use of a 0.1 mm path length cell and were signal averaged four times to maximize the signal-to-noise ratio. Single wavelength measurements were recorded with the use of 0.1 mm to 10 cm path length cells.

Peptide Syntheses. α_1A was synthesized by using the same methods as previously described for α_1.[12a] It was purified in a single step by reverse-phase HPLC with the use of a 2.5 by 25 cm Hamilton PRP1 column and a linear 36 min gradient of 32 to 42% aqueous acetonitrile containing 0.1% TFA (32 to 42% B) at a flow rate of 10 mL/min: yield 23% based on the loading of the first amino acid on the resin; amino acid analysis $Glu_{4.08}$ (4), $Gly_{2.01}$ (2), $Leu_{5.95}$ (6), $Lys_{3.95}$ (4); fast atom bombardment mass spectroscopy (FABMS, average chemical mass) expected 1882.1, found 1882.2.

α_1B, α_2B(P), and α_2B(PRR) were synthesized by similar methods, beginning with a 1% cross-linked polystyrene divinylbenzene benzhydrylamine polymer (Bachem, substitution level 0.37 mmol/g). It was found that as the chain was elongated the coupling reactions required increasingly longer times to come to completion as assessed by quantitative ninhydrin and picric acid titration methods. Typically, each residue was coupled three times in threefold excess as its symmetric anhydride in a mixture of DMF and DMSO (1:1). Only by repeated couplings in solvents which are highly disruptive of hydrogen bonding was it possible to obtain high yields in the coupling reactions. α_1B was purified in a single step with the use of a 60 minute gradient of 35 to 41% B: yield 50%; amino acid analysis $Glu_{4.03}$ (4), $Gly_{2.02}$ (2), $Leu_{5.98}$ (6), $Lys_{3.95}$ (4); FABMS expected 1881.2, found 1881.1.

α_2B(P) was purified by using a 40 min gradient of 40 to 48% B: yield 10%; amino acid analysis $Glu_{8.45}$ (8) $Pro_{0.83}$ (1), $Gly_{4.11}$ (4), $Leu_{12.46}$ (12), $Lys_{8.42}$ (8); FABMS expected 3801.6, found 3801.3.

α_2B(PRR) was purified by using a 36 min gradient of 40 to 46% B: yield 30%; amino acid analysis $Glu_{8.08}$ (8), $Pro_{1.05}$ (1), $Gly_{4.00}$ (4), $Leu_{11.81}$ (12), $Lys_{7.85}$ (8), $Arg_{2.22}$ (2); FABMS expected 4113.0, found 4112.7. Figure 3 illustrates the purity of crude, unfractionated α_2B(PRR), the peptide after HPLC purification, and a FABMS of the purified material.

Size Exclusion Chromatography. Size exclusion chromatography of the peptides was carried out by using a 1.6 cm by 90 cm column of Sephadex G50F. The eluent was monitored by measuring the absorbance

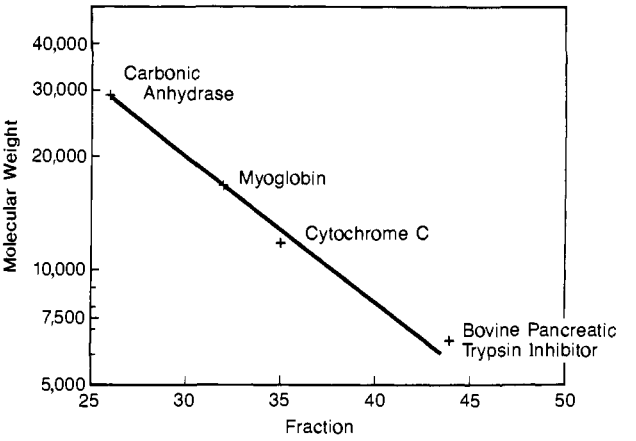


Figure 4. Calibration curve for the 1.6 cm × 90 cm G50F Sephadex column used in the size exclusion chromatography of the peptide aggregates.

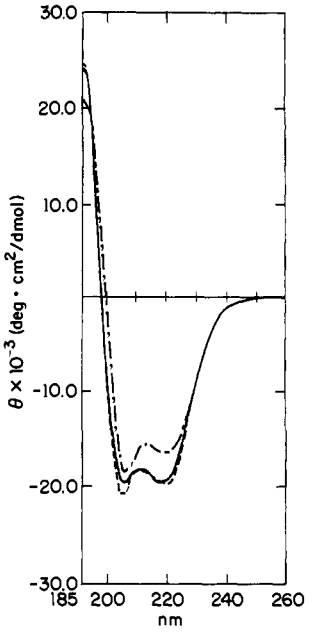


Figure 5. CD spectra of α_1A (dotted, dashed line), α_1B (dashed line), and α_2B(PRR) (solid line) at a concentration of 1 mg/mL in 10 mM MOPS at pH 7.0. Spectra were recorded in a 0.1 mm path length cell.

Table I. Size Exclusion Chromatography Data[a]

peptide	theor mol wt (daltons)	exptl aggregate mol wt (daltons)	degree of aggregation
α_1A	1880	7700	4.1
α_1B	1880	8300	4.4
α_2B(P)	3800	12600	3.3
α_2B(PRR)	4110	9200	2.2

[a] We estimate the method to be accurate to within 10%.

Scheme I

$$P_n \rightleftharpoons nP_{mon} \qquad K_{diss} = [P_{mon}]^n/[P_n]$$

at 220 nm. The peptides or protein standards (0.1 to 0.5 mg) were applied to the column in 0.5 mL of 0.5 M NaCl, 0.05 M 3-[*N*-morpholino]propanesulfonic acid (MOPS) buffer, pH 7.0, and eluted with the same buffer at a flow rate of 0.3 mL/min. Figure 4 illustrates the calibration curve for the column. The peptides were applied to the column at a concentration of approximately 0.05 mM. Their apparent molecular weights determined by interpolation from the standard curve were divided by their monomolecular weights to give the aggregation numbers reported in Table I.

CD Spectra of Peptides. The CD spectra of the peptides α_1A, α_1B, and α_2B(PRR) in 0.01 M MOPS solution (1 mg/mL of peptide concentration) at neutral pH are illustrated in Figure 5.

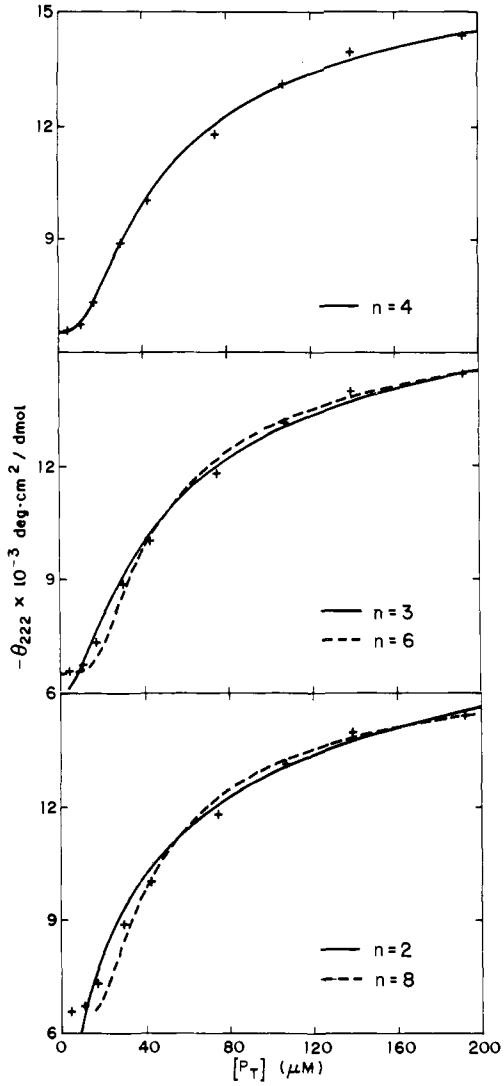


Figure 6. Concentration dependence of the ellipticity of α_1A at 222 nm. The lines are computer-generated, theoretical curves describing various monomer–*n*mer equilibria. The top panel shows the monomer–tetramer equilibrium, the middle panel monomer–trimer and monomer–hexamer equilibra, and the bottom panel monomer–dimer and monomer–octamer equilibria.

Concentration Dependence of CD Spectra. The concentration dependence of the CD spectra of the peptides in 0.15 M NaCl and 0.01 M MOPS at pH 7.0 was analyzed according to Scheme I by using eq 1,[9a] where P_n, P_{mon}, and P_T are the self-associated, monomeric, and total

$$[P_T] = \left\{ \frac{(\theta_{obsd} - \theta_{mon})K_{diss}}{n(\theta_n - \theta_{mon})(1 - ((\theta_{obsd} - \theta_{mon})/(\theta_n - \theta_{mon})))^n} \right\}^{1/(n-1)} \quad (1)$$

concentrations of the peptides, respectively, n is the degree of association, θ_{obsd} is the ellipticity observed at 222 nm, θ_{mon} is the ellipticity of the monomer, and θ_n is the ellipticity of self-associated form of the peptide. The values of n, θ_{mon}, θ_n, and K_{diss} were determined from the experimental data by using the nonlinear regression computer program MLAB (Gary Knott, National Institutes of Health, Bethesda, Maryland) and eq 1. Alternatively, the experimental data were analyzed according to various monomer–*n*mer equilibria by holding n constant at a given integral value and allowing θ_{mon}, θ_n, and K_{diss} to vary (Figure 6). The goodness of the fit was evaluated visually, by examining the sum of the squares, and by determining whether the fitted values of θ_{mon}, θ_n, and K_{diss} are physically reasonable.

CD Experiments in the Presence of Guanidine Hydrochloride. Stock solutions containing 0.15 M NaCl and 0.01 M MOPs with no guanidine, 4 M guanidine, and 8 M guanidine were individually titrated to pH 7.0. Solutions of intermediate guanidine concentrations were obtained from these stock solutions. The peptide concentration dependence of θ_{222} at various guanidine concentrations (an example of which is shown in Figure 7 for α_1A) was similarly analyzed with eq 1.

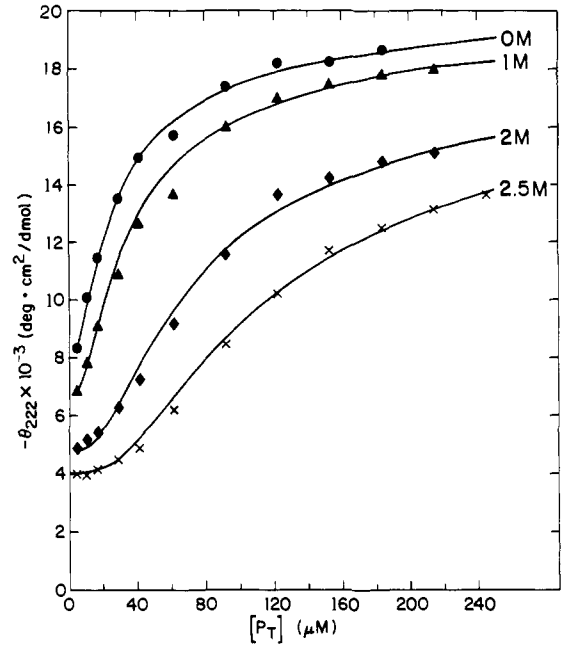


Figure 7. Peptide concentration dependence of θ_{222} for α_1A at various guanidine concentrations. The data (represented by symbols) are optimally described by theoretical monomer–tetramer equilibria (solid lines).

Results

Characterization of α_1A and α_1B. Figure 5 illustrates the spectra for α_1A and α_1B at fairly high peptide concentrations where they are predominantly in the aggregated state (see within). The spectra are typical of helical proteins[21] showing double minima at 208 and 220 nm and a maximum at 191 nm.

The self-association of α_1A and α_1B was investigated by using size exclusion chromatography. Their hydrodynamic behavior is consistent with the formation of helical tetramers (Table I). The concentration dependence of their CD spectra is also consistent with tetramer formation. Figure 6 illustrates the concentration dependence of the ellipticity at 222 nm for α_1A in 0.15 M NaCl and 0.01 M MOPS at pH 7.00. At high peptide concentrations this peptide's CD spectrum is typical of an α-helical protein (Figure 5), while at lower concentrations the spectrum shows considerably lower helical content (data not shown). Such aggregation-induced secondary structure formation is commonly observed with amphiphilic peptides and can be conveniently used to determine the degree of association and the free energy for the folding process.[9a,12a] Analysis of the data in Figure 6 as described in the Experimental Section gives the degree of aggregation as 4.25 for α_1A, consistent with it forming a tetramer. To confirm this the data were fit to various monomer–*n*mer schemes (Figure 6). The data are optimally fit to a monomer–tetramer equilibrium with a value of −21 200 deg·cm^2/dmol for the ellipticity of the tetramer and −8700 deg·cm^2/dmol for that of the monomer and a dissociation constant of 2.38×10^{-14} M^{-3}. The dissociation constant corresponds to a free energy of tetramerization of −18.6 kcal/mol ($RT \ln K_{diss}$), with a 1M standard state.

α_1B aggregates failed to dissociate at concentrations that could be conveniently monitored by CD spectroscopy. Therefore, to determine the free energy of tetramerization of this peptide, a chaotropic denaturant had to be included in the buffer. This decreased the stability of the folded, aggregated form of the peptide, allowing accurate determinations of K_{diss} at experimentally accessible peptide concentrations. From the values of K_{diss} thus obtained at various guanidine concentrations, an extrapolation of the data could be made to zero denaturant concentration. Several different models for extrapolating to zero denaturant concentration have been described in the literature,[22] the most

(21) Greenfield, N.; Fasman, G. D. *Biochemistry* **1969**, *8*, 4108–4116.

Table II. Experimentally Determined Ellipticities and Free Energies of Tetramerization for α_1A and α_1B in the Presence of Various Concentrations of Guanidine

peptide	[guan] (M)	$-\theta_{mon}$ (deg·cm/dmol)	$-\theta_{tet}$ (deg·cm^2/dmol)	$-\Delta G°$ (kcal/mol)
α_1A	0	8.2	20.7	18.6
	1	6.7	20.6	18.1
	2	4.8	20.0	16.7
	2.5	4.0	20.6	15.7
α_1B	1.5	6.8	23.5	17.8
	2.0	6.0	23.6	16.7
	2.5	5.5	22.1	16.3

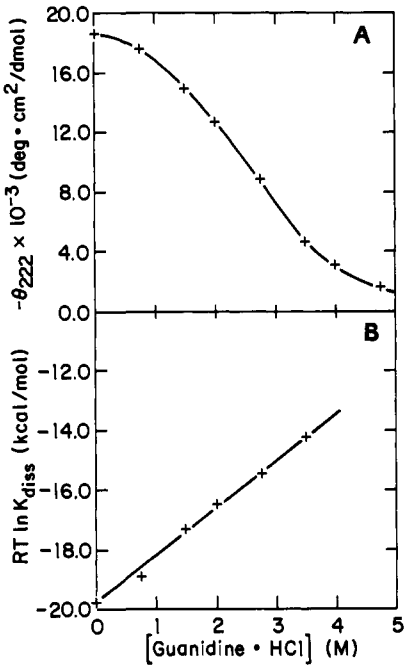


Figure 8. Guanidine denaturation curve (A) and a plot of free energy vs. the guanidine concentration (B) for α_1A.

widely used of which is the linear method.[22a] To determine whether this method might be applicable to α_1B the behavior of α_1A was first examined. Figure 7 illustrates the peptide concentration dependence of the ellipticity of α_1A at various concentrations of guanidine ranging from 0 to 2.5 M. As the concentration of guanidine is increased, the curves are displaced toward higher peptide concentrations, indicating a decrease in the stability of the tetramer with respect to the monomer. Each of the curves could be optimally described as a cooperative monomer–tetramer equilibrium. The calculated value for the ellipticity of the tetramer (−21 000 ± 1 000 deg·cm^2/dmol) was found to be independent of the guanidine concentration, while the ellipticity of the monomer and the dissociation constants were found to vary markedly with guanidine (Table II). The ellipticity of the monomer decreased linearly with increasing guanidine and followed the equation θ_{mon} = 1700[Guan] − 8300 (correlation coefficient 0.998). The tetramer dissociation constants increased logarithmically with the guanidine concentration. We interpret these results to indicate that the structure of the tetramer is invariant with respect to the guanidine concentration but that the stability of this structure relative to that of the monomer is decreased as the guanidine concentration is increased.

From the variation of θ_{mon} and θ_{tet} as a function of guanidine concentration, it is possible to determine dissociation constants for tetramerization at a single peptide concentration, eliminating the need to measure complete peptide concentration dependences at each guanidine concentration. Figure 8 illustrates the guanidine denaturation curve for α_1A measured at a peptide concentration of 1.15 × 10^{-4} M and the free energy of tetramerization (RT ln K_{diss}) as a function of the guanidine concentration. The data were well described by a linear equation of the form RT ln K_{diss} = 1.5[Guan] − 19.7. The intercept of −19.7 kcal/mol is the free energy of tetramerization extrapolated to zero guanidine concentration, and it is in good agreement with the value of −18.6 kcal/mol determined above. In each case the values of RT ln K_{diss} measured from a complete examination of the peptide concentration dependence (Table II) agreed within a single kcal/mol with the value calculated from the guanidine denaturation curve at a single peptide concentration.

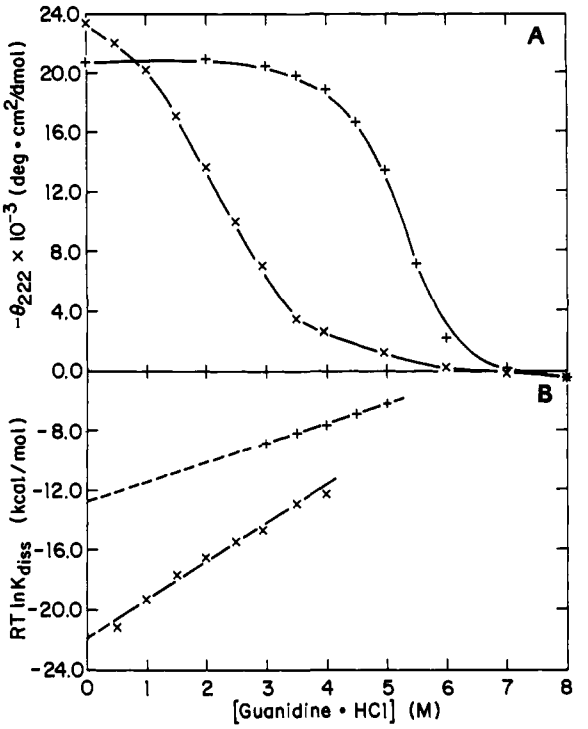


Figure 9. (A) Guanidine denaturation curves for α_1B (X) and α_2B(PRR) (+). (B) Free energy vs. guanidine concentration plots for α_1B and α_2B(PRR). The lines are least-squares fits to the data. Extrapolated portions of the data are represented by the broken lines.

The free energy of dissociation of α_1B was determined by a similar process. At guanidine concentrations of 1 M or higher, peptide concentration/ellipticity isotherms could readily be measured and were again optimally described by a cooperative monomer-to-tetramer equilibrium. The ellipticity of the tetramer was invariant between 1.0 and 2.5 M guanidine (−23 000 ± 1 000 deg·cm^2/dmol, Table II), and the ellipticity of the monomer varied linearly with the guanidine concentration in a manner similar to that of α_1A (θ_{mon} = 1700[Guan] − 9400, correlation coefficient 0.998). The guanidine denaturation curve for α_1B (Figure 9A) was measured at a constant peptide concentration of 1.07 × 10^{-4} M and analyzed as for α_1A (Figure 9B), giving a slope of 2.57 kcal/mol/mol guanidine) and an intercept of −21.9 kcal/mol. Thus, at room temperature and in the absence of denaturant, tetramerization of α_1B is 2 to 3 kcal/mol more favorable than tetramerization of α_1A.

Characterization of α_2B(P) and α_2B(PRR). Size exclusion chromatography of α_2B(P) (Table I) demonstrated that this peptide was forming trimers (depicted pictorially in Figure 10) rather than the desired dimer of helical hairpins and was hence not studied further. In a subsequent design attempt, two extra arginyl residues were inserted after the proline with the expectation that the charge would result in destabilization of the trimeric structure while favoring the formation of a hairpin loop. α_2B-(PRR) does indeed appear to have a dimeric structure as determined by its elution with an apparent molecular weight of 9200. It also appears to form helices based on the shape and magnitude of its CD spectrum (Figure 5). It was not possible to obtain a

(22) (a) Schellman, J. A. *Biopolymers* **1978**, *17*, 1305–1322. (b) Pace, C. N. *CRC Crit. Rev. Biochem.* **1975**, 1–43. (c) Pace, C. N.; Vanderburg, K. E. *Biochemistry* **1979**, *18*, 288–292.

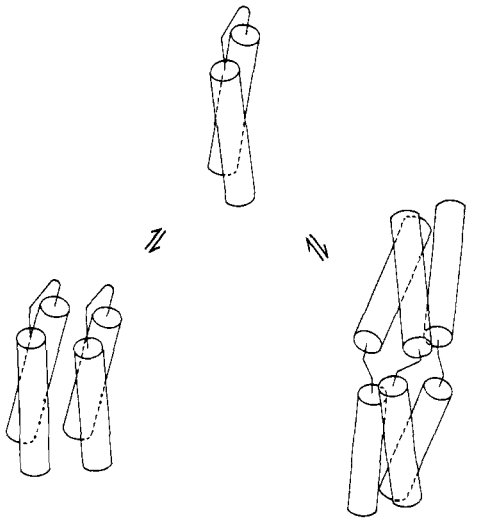


Figure 10. Diagrammatic representation of trimer formation by α_2B(P) instead of the desired dimer of helical hairpins.

dissociation constant for the dimerization of this peptide in the absence of denaturants, because it dissociates at too low a peptide concentration to be measured experimentally. However, in the presence of guanidine at a concentration of 4 M or greater, the ellipticity showed a marked dependence on the peptide concentration (Figure 11). Analysis of the data in Figure 11 gives the degree of aggregation as 2.2, indicating that this peptide indeed forms dimers. As illustrated in Figure 11 the data are optimally fit to a monomer–dimer equilibrium, alternate equilibria giving less satisfactory fits to the data. Determination of the dissociation constant at 4.5 and 5.0 M guanidine indicates that the ellipticity of the dimer (−21 000 ± 1 000 deg·cm^2/dmol) is invariant with respect to the guanidine concentration (the same value is obtained for the peptide in the absence of guanidine) and that the ellipticity of the monomer is well predicted by the equation for α_1B. Analysis of the guanidine denaturation curve (Figure 9) shows that the free energy for dimerization is linear between 3 and 5 M guanidine concentration, with an intercept of −12.8 kcal/mol, and a slope of 1.3 kcal/(mol·(mol of guanidine)$^{-1}$).

Discussion

Protein design requires consideration of various aspects involved in protein folding and stability. For a 4-helix bundle these include helix formation, helix termination, helix–helix packing, and loop formation: a successfully designed sequence must satisfy each of these often overlapping functional roles. Currently these processes are poorly understood[23] so it appeared unlikely that a sequence could be designed in a single step which would properly fulfill each of these requirements and fold into a stable protein of predetermined structure. We therefore chose the incremental experimental approach described here. This approach has the advantage that each step in the folding process can be dissected and analyzed separately.

The formation of helices is a necessary and perhaps initial step in tetramer formation by α_1A or α_1B. As is the case for other natural peptides,[24] monomeric helix formation by α_1A and α_1B in dilute aqueous solution is somewhat unfavorable. The CD spectra for these peptides extrapolated to infinite dilution indicate that they have a low but significant helical content, and the helical content decreases as a function of the guanidine concentration. In the absence of guanidine, both peptides α_1A and α_1B are approximately one-third as helical in the monomeric state as compared to the tetrameric state (−8000 deg·cm^2/dmol for the monomer vs. −22 000 deg·cm^2/dmol for the tetramer). This

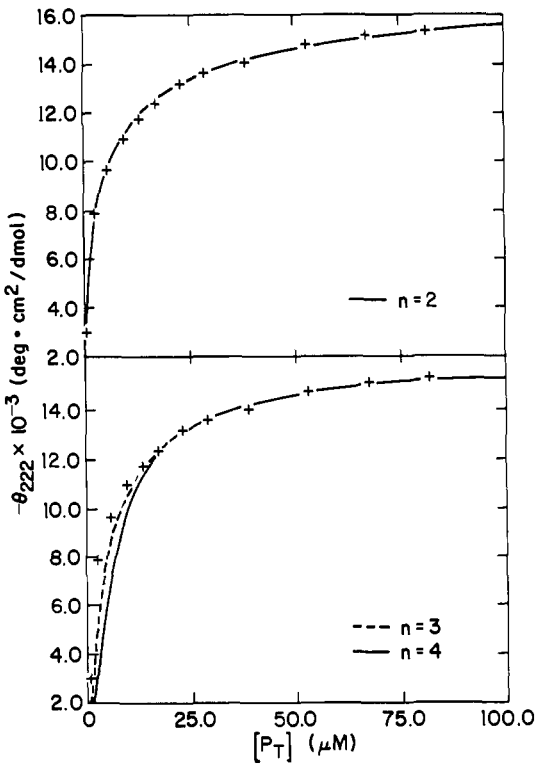


Figure 11. Peptide concentration dependence of θ_{222} for α_2B(PRR) at 4.5 M guanidine. Computer-generated monomer–dimer (top panel) and monomer–trimer and monomer–tetramer curves (bottom panel) are shown superposed on the data.

corresponds to helical contents of approximately 25% and 75% for the monomers and the tetramers, respectively (compared to the ellipticity of poly-L-lysine·HCl[21]). This would suggest that the helices in the tetrameric state are on the order of 12 residues in length. It is not unreasonable to assume that helix formation by peptides of this length is an all-or-nothing process. Then the equilibrium constant for helix formation in the monomeric state would be 0.5, i.e., helix formation is only slightly unfavorable and on the order of a single kcal/mol. The helical contents for the monomeric forms of α_1A and α_1B are identical within experimental error. This would suggest that the difference in stability of the tetramers formed by these two peptides does not derive from a difference in helical stability but rather from a difference in interhelix stabilizing interactions.

The free energies of tetramerization for α_1A and α_1B appear to be consistent with apolar interactions between the isobutyl side chains of the leucyl residues being the major driving force for helix formation and self-association. The value of −20 kcal/mol for α_1A corresponds to −0.83 kcal/mol per leucine: if a correction is made for the unfavorability of helix formation this value increases to −0.9 to −1.0 kcal/mol, close to the value estimated for the transfer of a leucine side chain from water to the interior of a protein (−1.2 kcal/mol[25]). Also, in models it appears that a leucyl side chain was excessively exposed in α_1A and that it might be more efficiently buried in α_1B. Indeed the tetramer formed by α_1B is 2 to 3 kcal/mol more stable than α_1A. The magnitude of the difference is again consistent with apolar interactions being the cause of the increased stability.

The effect of peptide length of the helix-forming segment is illustrated by comparing the properties of α_1 with α_1A. α_1 is a shortened, 12-residue version of α_1A which folds into tetramers with a net stability of −11.4 kcal/mol.[12a] Examination of models indicates that extending the helix formed by this peptide by just one turn greatly increases the available surface for interaction with neighboring helices. Also, the monomeric helical content of α_1 is less than that for α_1A, indicating that helix formation is

(23) (a) Wetlaufer, D. B. *The Protein Folding Problem*; Westview: Boulder, CO, 1984. (b) Ghelis, C.; Yon, J. *Protein Folding*; Academic: New York, 1982.

(24) Bierzynski, A.; Kim, P. S.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 2470–2474.

(25) Guy, H. R. *Biophys. J.* **1985**, *47*, 61–70.

less favorable for α_1 than for α_1A.

The difficulties encountered in the design of a loop illustrate the usefulness of an incremental experimental approach. In light of the studies with α_1B it was clear that trimer formation by α_2B(P) most probably represents two flaws in the design: (1) the loop sequence failed to direct helix termination and loop formation sufficiently well and (2) there exists an unforeseen trimeric aggregate with lower lying energy than the desired dimer. Had the properties of α_1B not been first established it would have seemed equally probable that the helical sequence was at fault and not the loop sequence. Subsequently, two arginyl residues were included after the prolyl residue in the linking sequence. Formation of a trimer such as that illustrated in Figure 10 would be inhibited by electrostatic repulsions between the guanidinyl moieties on adjacent chains. Also, turns in natural proteins are known to be located at predominantly hydrophilic loci in protein sequences.[19] Thus, the linking sequence of α_2B(PRR) has been designed not only to stabilize the desired structure but also to destabilize other possible aggregates. This peptide indeed forms a dimer with hydrodynamic properties consistent with it forming a 4-helix bundle.

The effect of adding the linking sequence in α_2B(PRR) can be seen by comparing the guanidine denaturation curves for α_1B and α_2B(PRR) (Figure 9A). At similar peptide concentrations, an approximately 3 M higher guanidine concentration is required to unfold α_2B(PRR) as compared to α_1B. Thus, the linking sequence stabilizes the formation of the four-helix structure when the concentration of the peptides are reasonably dilute ($\leq$1 mg/mL). However, a careful thermodynamic analysis demonstrates that the loop sequence could be substantially improved. Analysis of the guanidine dependence of the dissociation constants for aggregation of α_2B(PRR) and α_1B indicates that α_2B(PRR) forms a less stable structure than that formed by the unlinked helices; extrapolation to zero guanidine concentration gives stabilities of -22 and -13 kcal/mol for α_1B and α_2B(PRR), respectively (1 M standard state). The value for α_2B(PRR) is rather uncertain and is probably a lower limit[22c] because it was obtained by extrapolation over a wider range of guanidine concentration (0 to 3 M) than the range for which experimental data were available (3 to 5 M). (In the case of horse myoglobin, for example, extrapolation over a 1M guanidine concentration range underestimated the stability of the protein by almost 30%[22c]). It is therefore more valid to compare the linked and unlinked peptides at a concentration of guanidine for which there exist good data for both peptides. At 4 M guanidine the free energies of association for the linked and unlinked helices are -7.6 and -11.6 kcal/mol, respectively. This corresponds to -2.9 kcal/mol ($-11.6/4$) of energy per monomer for the unlinked helices. A peptide containing two monomers connected by an optimal loop should thus have a free energy of association of -5.8 kcal/mol, discounting any additional structural stabilization, or entropic stabilization due to the cratic contributions to the free energy. Instead we observe a value of -3.8 kcal/mol ($-7.6/2$) for α_2B-(PRR), indicating that the structure is slightly destabilized rather than stabilized by the loop. There are several possible causes for this slight (2.0 kcal/mol) destabilization. The charged groups of neighboring loops might interact unfavorably, the conformation of the loop might be somewhat strained, or the loop might constrain the helices to lie in a somewhat distorted geometry as compared to the unlinked helices. A very small change in the rotation about even a single side chain could easily account for an effect of this magnitude. We are currently designing other loops that will help to distinguish between these possibilities.

The hydrodynamic and thermodynamic studies presented in this paper demonstrate that it has been possible to design a peptide which forms a tetramer of α-helical segments. The fact that these helices can be connected by relatively short loops without altering their hydrodynamic and spectroscopic properties would tend to support the idea that the helices are indeed arranged antiparallel to one another as indicated in Figure 1. However, a convincing demonstration that the peptides form helical bundles must await a crystallographic investigation of the structure.

The 4-helix bundle geometry provides an attractive structure for designing synthetic binding sites and catalysts. The packing of the helices at 20° angles causes the helices to flare at either end of the structure, creating a cavity. By extending the length of the helices described herein it should be possible to increase the size of the cavity, providing a semirigid matrix for positioning functional groups appropriate for binding and catalysis.

Acknowledgment. We thank Drs. David Eisenberg, Zelda Wasserman, Ray Salemme, Pat Weber, Carl Pabo, and James Lear for many helpful discussions. We extend our gratitude to Dr. Barbara Larsen for FAB mass spectral analysis of the higher molecular weight peptides. We also acknowledge the assistance of Rosemarie Raffaele and Keith Smithyman in this work.

Chemical and Mutagenic Analysis of Aminomethylphosphonate Biodegradation

L. Z. Avila, S. H. Loo, and J. W. Frost*

Contribution from the Department of Chemistry, Stanford University, Stanford, California 94305. *Received August 19, 1986*

Abstract: Utilization of aminomethyl-, *N*-methylaminomethyl-, *N,N*-dimethylaminomethyl-, and *N*-acetylaminomethylphosphonate by *Escherichia coli* as a sole source of phosphorus during growth resulted in the extracellular generation of *N*-methylacetamide, *N,N*-dimethylacetamide, trimethylamine, and *N*-methylacetamide, respectively. Product identification relied on synthesis of ^{13}C-enriched aminomethylphosphonates followed by ^{1}H NMR analysis of products isolated from the biodegradation of the labeled and unlabeled phosphorus sources. To circumvent the requirement of an intact cell for carbon to phosphorus bond degradation, transposon mutagenesis was exploited as a complement to the chemical analysis. *E. coli* K-12 were infected with λ::Tn5. Colonies resistant to kanamycin were selected and then screened for loss of the ability to use ethylphosphonate as a sole source of phosphorus. The mutant identified, *E. coli* SL724, was also unable to degrade aminomethylphosphonates. This combination of chemical and mutagenic analysis points toward a shared mechanism between alkyl- and aminomethylphosphonate biodegradation.

Widespread use of agricultural chemicals such as *N*-phosphonomethylglycine (**1**) draws attention to the need for understanding degradation and detoxification of these chemicals in the environment. Of particular relevance is the ability of gram-

0002-7863/87/1509-6758$01.50/0