Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 7315–7319, August 1992
Biophysics

Role of loop–helix interactions in stabilizing four-helix bundle proteins

(helix–helix interaction/loop–loop interaction/helix–loop interaction/energy minimization)

KUO-CHEN CHOU*, GERALD M. MAGGIORA*, AND HAROLD A. SCHERAGA†

*Computational Chemistry, Upjohn Laboratories, Kalamazoo, MI 49001; and †Baker Laboratory of Chemistry, Cornell University, Ithaca, NY 14853-1301

Contributed by Harold A. Scheraga, May 4, 1992

ABSTRACT **One of the critical issues regarding proteins with a four-helix bundle motif is which interactions play the major role in stabilizing this type of folded structure: the interaction among the four α-helices or the interaction between the loop and helix segments. To answer this question, an energetic analysis has been carried out for three proteins with a four-helix bundle—namely, methemerythrin, cytochrome *b*-562, and cytochrome *c*′. The structures on which the analysis has been made were derived from their respective crystallographic coordinates. All three proteins have long helices (16–26 residues) and most of their loops are short (3–5 residues). However, it was found in all three proteins that loop–helix interactions were stronger than helix–helix interactions. Moreover, not only the nonbonded component but also the electrostatic component of the interaction energy were dominated by loop–helix interactions rather than by interhelix interactions, although the latter involve favorable helix–dipole interactions due to the antiparallel arrangement of neighboring helices. The results of the energetic analysis indicate that the loop segments, whether they are in a theoretical model or in real proteins, play a significant role in stabilizing proteins with four-helix bundles.**

A previous theoretical model study (1) indicated that the left-handed twisted feature that occurs in many antiparallel four-helix bundle proteins (2–5) could be explained in terms of nonbonded interactions between the constituent helices and that the antiparallel arrangement of the helices is due in part to electrostatic interactions. Inclusion of the effect of the loops connecting the adjacent helices made the model more realistic, and the calculated result improved somewhat as well (6), leading to a much better four-fold symmetrical array, as observed (3, 4). In addition, a detailed energetic analysis (7), in which the various nonbonded and electrostatic interaction energy terms in a theoretical four-helix bundle model were calculated explicitly, showed that the interactions of loops with helices play a significant role in stabilizing the four-helix bundle structure. In the theoretical model, each of the four helices was formed by poly(Ala)$_{12}$, and each of the three loops was formed by poly(Ala)$_{10}$. It was found there that the loop–helix interaction (not only the nonbonded but also the electrostatic energy) was stronger than the interhelix interaction. Consequently, it seems that the stabilization energy of the four-helix bundle motif results from interactions between the α-helices and the loops.

However, it might be argued that such a conclusion cannot be generalized for *proteins* with a four-helix bundle for the following two reasons: (*i*) When the bundle is modeled with loops 10 residues long, with helices of a length of 12, as adopted in ref. 7, then it is hardly surprising that the loops play a significant role. Real proteins, which fold as regular four-helix bundles, usually consist of long helices [e.g., 16–24 residues in methemerythrin (8)] and short loops (3–5 residues). (*ii*) The use of a poly(Ala) sequence for the helices might lead to different interaction energies among the four helices than would be observed in a protein. Therefore, to find an appropriate answer for such a question, some further calculations based on real four-helix proteins were carried out.

As is well known, the determination of the structures of loops is a difficult problem that must be addressed in predicting the three-dimensional structures of proteins (9, 10) and in the study of protein folding (11–13). Loops are also thought to assume important roles in molecular function and biological recognition (14). Therefore, an investigation of the interactions between loops and the other regular structural elements of a protein should provide new insights for studying these problems. The present work was carried out in an attempt to reveal the role of loops in stabilizing the four-helix bundle motif in proteins. For this purpose, a detailed energetic analysis was carried out on three four-helix bundle proteins—namely, methemerythrin, cytochrome *b*-562, and cytochrome *c*′—based on their crystal structures. All three proteins have long helices (16–26 residues), which are typical for proteins with four-helix bundles, and most of their loops are short (3–5 residues).

METHODS

The computational procedure consists of the following three steps: (*i*) generation of the initial structures for each of the proteins investigated, (*ii*) energy refinement of each protein molecule, and (*iii*) energy-component analysis. The computations were carried out by using the residue geometry and the energy parameters found in ECEPP/2, which is an updated version, modified by Némethy *et al.* (15), of the original ECEPP (empirical conformational energy program for peptides) algorithm developed by Momany *et al.* (16). The energy is calculated as the sum of electrostatic, nonbonded, hydrogen-bond, and torsional energies. Energy minimizations are based on a general unconstrained optimization algorithm (17). The computations were carried out with one processor on an IBM 3090/400J computer at Upjohn Laboratories. The standard conventions for nomenclature of peptide conformations (18) have been followed.

Generation of the Initial Structures. The non-hydrogen atom coordinates of methemerythrin have been determined by Stenkamp *et al.* (8) at 2.0 Å resolution, those of cytochrome *b*-562 by Lederer *et al.* (19) at 2.5 Å resolution, and those of cytochrome *c*′ by Finzel *et al.* (20) at 1.7 Å resolution. All three proteins assume a typical four-helix bundle motif. Based on the non-hydrogen atom coordinates taken from the Brookhaven Protein Data Bank (21), a set of dihedral angles for the backbone and side chains of each of these proteins was generated. This step included an optimal

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: ECEPP, empirical conformational energy program for peptides.

fitting operation in which the dihedral angles were varied and the ECEPP residue geometries were used to generate a molecule whose coordinates matched the x-ray coordinates as closely as possible by minimizing the corresponding root-mean-square deviations. The backbone and side-chain non-hydrogen atom root-mean-square deviations between the structures thus generated and those determined crystallographically were 0.19 Å for methemerythrin, 0.17 Å for cytochrome *b*-562, and 0.21 Å for cytochrome *c'*. All hydrogen atoms were generated according to the ECEPP geometry of L-amino acids (16). Dihedral angles related to positions of hydrogen atoms were assigned according to data on energy-minimized amino acid residues (22). Although such an assignment is somewhat arbitrary, it was corrected by subsequent energy minimization (see the following section).

Energy Minimization. The structures obtained above were used as starting points for energy minimizations of the whole protein. The ionizable side chains were kept in their neutral forms. For each of the three structures, energy minimization was carried out in two successive steps. In the first step, all backbone dihedral angles were held fixed, and the energy was minimized with respect to the side-chain dihedral angles only in order to eliminate side-chain atomic overlaps. In the second step, the energy was minimized with respect to *all* dihedral angles. In both steps, no constraint to the x-ray structure was imposed. The total central processing unit time varied for the different proteins, ranging from 108 to 356 minutes.

Energy-Component Analysis. In order to compare and analyze the various interactions that contribute to the overall stability of the four-helix bundle proteins investigated, it is useful to separate the total conformational energy, E_{tot}, into components as defined below (12):

$$\left\{\begin{array}{ll} E^{\alpha}_{intra} = & \text{Sum of the energies of the four individual constituent } \alpha\text{-helices} \\ E^{\alpha}_{inter} = & \text{Total intersegment energy among the four } \alpha\text{-helices} \\ E^{\alpha}_{tot} = & E^{\alpha}_{intra} + E^{\alpha}_{inter} \\ E^{loop}_{intra} = & \text{Sum of the energies of the individual loop segments} \\ E^{loop}_{inter} = & \text{Total intersegment energy among all the loop segments} \\ E^{loop}_{tot} = & E^{loop}_{intra} + E^{loop}_{inter} \\ \varepsilon = & \text{The interaction energy between the loop segments and the } \alpha\text{-helices of the molecule} \\ E_{tot} = & E^{\alpha}_{tot} + E^{loop}_{tot} + \varepsilon. \end{array}\right. \quad [1]$$

These energies include both the backbone and side chains. In the above expressions, the N- and C-terminal portions of the chain are treated as loops and, hence, their contributions to the interaction energy are included.

RESULTS AND DISCUSSION

After energy minimization, the root-mean-square deviation based on all backbone and side-chain non-hydrogen atoms between the x-ray crystallographic and the energy-minimized structures for methemerythrin is 0.86 Å, for cytochrome *b*-562 is 0.52 Å, and for cytochrome *c'* is 1.10 Å. This indicates that the crystal structures (with ECEPP geometry) are quite close to the energy-minimized structures for all three four-helix bundle proteins, although it was observed that, during the course of energy minimization, the energy dropped significantly (by 3–5 orders of magnitude).

If the ϕ, ψ dihedral angles of the helix-forming residues are taken to be $-68 \pm 30°$ and $-38 \pm 30°$, respectively (23), it was found after energy minimization that the sequence positions of the four main α-helices are those listed in Table 1, together with those of the three internal loops, and the N- and C-terminal peptide segments in each of the three four-helix bundle proteins studied here. The corresponding stereo cartoon ribbon drawings for methemerythrin, cytochrome *b*-562, and cytochrome *c'* are shown in Figs. 1–3, respectively.

Table 2 lists the geometric parameters characterizing the three four-helix bundle proteins. As shown in Table 2, the orientation angles (1) for all adjacent pairs of helices lie within the range $-155 \pm 15°$, indicating that the four-helix bundle of the three proteins assumes the usual left-handed twist (1, 3). The left-handed twist feature of these three bundle proteins is also clearly shown by the stereo ribbon drawings in Figs. 1–3. Furthermore, as given in Table 2, the bundle twisted angle (6) $\Theta = -9°$, $-12°$, and $-15°$, for methemerythrin, cytochrome *b*-562, and cytochrome *c'*, respectively, indicating that the amount of twist of the four-helix bundles is in the order: cytochrome *c'* > cytochrome *b*-562 > methemerythrin.

The various energetic terms defined in Eq. 1 are given in Table 3. Those terms whose definitions are not given explicitly in Eq. 1 are defined in the footnotes of Table 3. It can be seen that the intrahelix energy is much stronger than the intraloop energy in all of the three four-helix bundle proteins investigated. However, the magnitude of the loop–helix interaction energy is not only much larger than that of the interloop interaction energy but also larger than that of the interhelix interaction energy. For methemerythrin and cytochrome *c'*, this is also true for both the nonbonded and electrostatic components.

It has been pointed out that an antiparallel arrangement of neighboring α-helices is favored by electrostatic interactions between the helices (1, 24) due to the large dipole moment of the α-helices (25–28). However, as shown in Table 3 for methemerythrin, cytochrome *b*-562, and cytochrome *c'*, $\varepsilon_{ES} = -30.7$, -45.1, and -52.6 kcal/mol, respectively, which are much greater than the corresponding $\varepsilon^{\alpha}_{ES} = -3.5$, -11.0, and -6.5 kcal/mol. This indicates that the electrostatic interaction energy in these four-helix bundle proteins is due primarily to interactions between loops and helices rather than interactions among the four helices themselves. Thus, it is seen that loop–helix interactions play a significant role in stabilizing the four-helix bundle structure not only for an earlier theoretical model (6), in which all helices and loops are formed from poly(Ala) sequences, but also for real four-helix bundle proteins.

Table 1. Computed sequence position of the four-α-helix bundle proteins investigated

Protein	Helix 1	Helix 2	Helix 3	Helix 4	Loop 1	Loop 2	Loop 3	N attachment	C attachment
Methemerythrin	19–37	41–64	70–85	91–108	38–40	65–69	86–90	1–18	109–113
	(19–37)	(41–64)	(70–85)	(91–109)	(38–40)	(65–69)	(86–90)	(1–18)	(110–113)
Cytochrome *b*-562	3–19	23–42	56–80	84–104	20–22	43–55	81–83	1–2	105–106
	(3–19)	(23–40)	(56–80)	(84–104)	(20–22)	(41–55)	(81–83)	(1–2)	(106)
Cytochrome *c'*	5–30	40–57	79–102	104–124	31–39	58–78	103	1–4	125–128
	(5–30)	(40–58)	(79–102)	(104–125)	(31–39)	(59–78)	(103)	(1–4)	(126–128)

Values in parentheses pertain to the original x-ray structures.

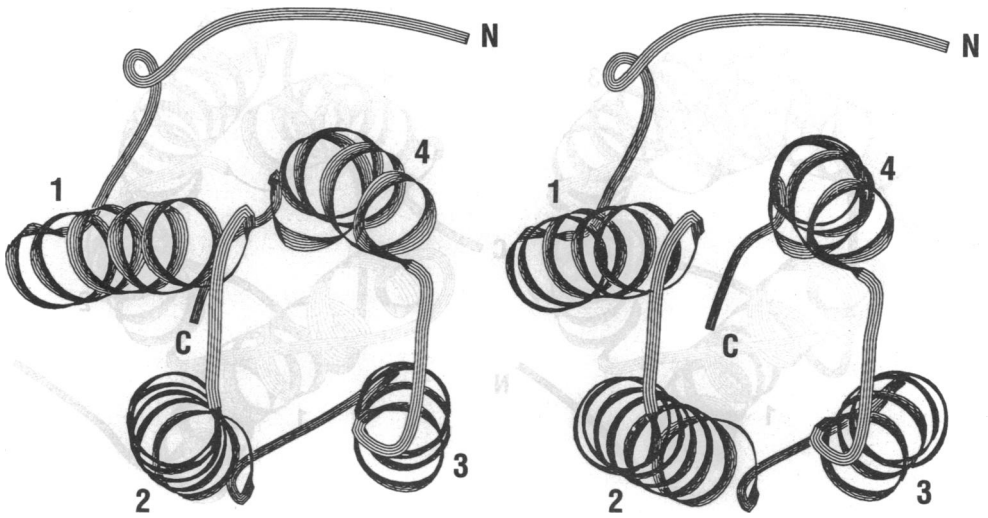

FIG. 1. Stereo cartoon ribbon drawing of the energy-minimized methemerythrin. This is an antiparallel left-handed twisted bundle protein. The four α-helices, numbered 1, 2, 3, and 4, are connected sequentially through the loops. N and C represent the N and C termini of the entire polypeptide chain, respectively.

It should be noted that, according to the procedure described in ref. 7 and used here, the polypeptide chain is divided into segments (i.e., helices and loops) by making cuts at the appropriate C′—N peptide bonds. Next, the intersegment energy for any pair of segments (helix–helix, loop–loop, or loop–helix) is calculated as the sum over all pairs of atoms in the two respective segments. Although all interactions must be included in the calculations, they do not all contribute to the stabilization: the 1–4 and some 1–5 interactions (16) at the adjacent site of two segments are either independent of conformation or vary very little with conformation. In view of this, an additional calculation has been carried out on methemerythrin to illustrate this point. In this calculation, the entire polypeptide chain was placed in an extended conformation—i.e., all its backbone dihedral angles were fixed at 180°—and the energy was minimized by varying only its side-chain dihedral angles. For such an extended chain, it was found that the interloop and interhelix energies were almost zero as expected, and the total loop–helix interaction energy was only −22 kcal/mol, which is much smaller in magnitude than −253.0 − (−152.4) = −100.6 kcal/mol, the difference between the loop–helix interaction energy and the interhelix interaction energy in the corresponding four-helix bundle folded form. This indicates that, even if an adjustment is made for "boundary effects," it will not change the general conclusion that the loop–helix interactions contribute more favorable interaction energy than the helix–helix interactions in stabilizing the four-helix bundle structure for the three cases examined here.

Gilson and Honig (29) have examined the effect of water on the interactions in a four-helix bundle. They have shown that this solvent has two destabilizing effects on the electrostatic components of interhelix interactions. It reduces the magnitude of interhelix dipole interactions and destabilizes packed structures owing to the desolvation of the helical dipoles upon helix association. Thus, the solvent would make the interhelical interactions even weaker than shown here. It remains to be seen how the solvent would affect the loop–helix interactions.

The reasons that loop–helix interactions are stronger than interhelix interactions in the absence of explicit solvation are not yet clear. It appears, however, that the effective interactive distances between atoms in helices and their adjacent loops are on average smaller than those between adjacent helices. Further studies, especially with solvation included explicitly, are needed to identify the origin of this behavior.

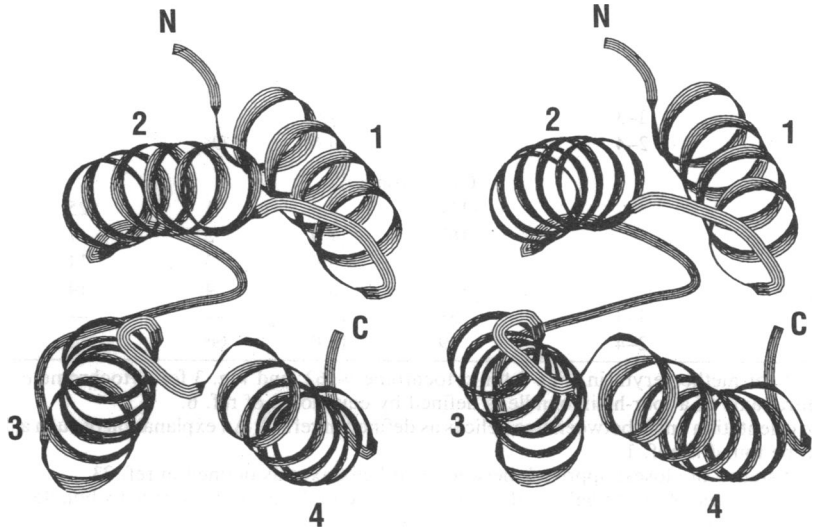

FIG. 2. Stereo cartoon ribbon drawing of the energy-minimized cytochrome *b*-562. See the legend to Fig. 1 for further explanation.

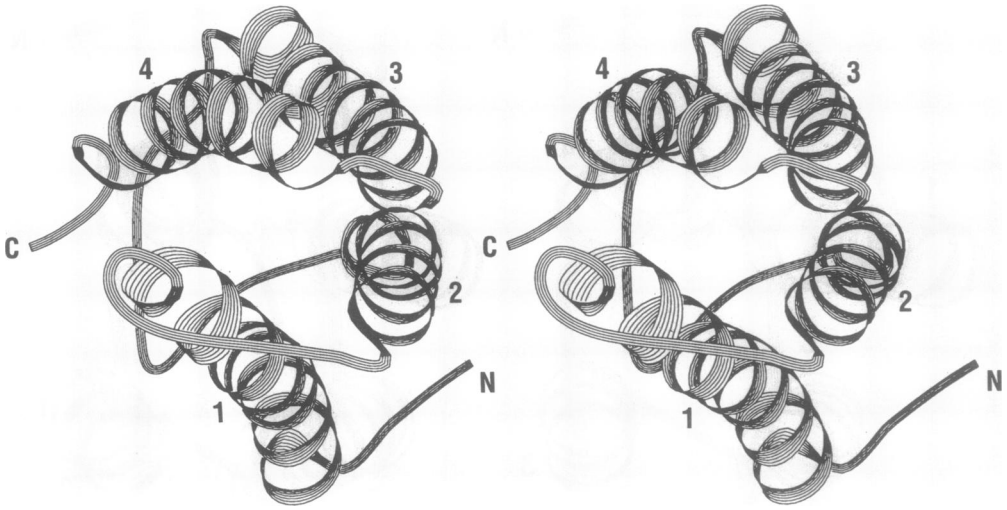

FIG. 3. Stereo cartoon ribbon drawing of the energy-minimized cytochrome *c'*. See the legend to Fig. 1 for further explanation.

For the present, we note that an effective dielectric constant is used in ECEPP to take account of the solvent effect in an approximate way. In addition, the solvent effect is simulated to some extent by taking the ionizable groups as neutral.

CONCLUSION

In a four-helix bundle protein, as far as interactions within a polypeptide segment are concerned, the intrahelix interaction energy is much stronger than the intraloop interaction energy; however, as far as interactions between two polypeptide segments are concerned, loop–helix interactions play a significant role in stabilizing protein structure. As is true for interhelix interactions, nonbonded interactions play a dominant role in loop–helix interactions. In both cases, the electrostatic part of the interaction energy is small but is larger between loops and helices than between the helices themselves, although the latter involves the interaction of the large dipole moments due to the antiparallel arrangement of neighboring α-helices (25–28). The conclusion holds true not

Table 2. Calculated geometric parameters characterizing the four-helix bundle proteins investigated

Relationship between helices[a]				Relationship between helix and the central axis[b]		
Adjacent pair	Diagonal pair	Ω_0,[c] deg	D,[d] Å	Helix	Ω_i,[e] deg	R_i,[f] Å
		Methemerythrin				
1–2		−158	8.9	1	13	6.1
2–3		−169	10.6	2	10	6.6
3–4		−168	11.7	3	10	8.2
4–1		−169	9.4	4	4	7.6
	1–3	23	15.0	—	—	
	2–4	14	14.2	Θ[g]	−9	
		Cytochrome *b*-562				
1–2		−162	8.7	1	18	7.2
2–3		−167	9.1	2	8	5.9
3–4		−170	7.7	3	9	6.8
4–1		−151	9.3	4	13	4.7
	1–3	26	14.0	—	—	
	2–4	20	10.6	Θ[g]	−12	
		Cytochrome *c'*				
1–2		−154	9.3	1	25	7.3
2–3		−167	7.0	2	7	5.8
3–4		−167	8.0	3	14	5.5
4–1		−144	8.8	4	14	5.5
	1–3	38	13.0	—	—	
	2–4	19	12.0	Θ[g]	−15	

[a]See Fig. 1 for methemerythrin, Fig. 2 for cytochrome *b*-562, and Fig. 3 for cytochrome *c'*.
[b]The central axis of a four-helix bundle is defined by equation 3 of ref. 6.
[c]Ω_0 is the orientation angle between two helices as defined in ref. 23. An explanation of such a definition can also be found in ref. 1.
[d]D is the distance of closest approach between two helix axes as defined in ref. 23.
[e]Ω_i is the tilted angle of the ith helix with respect to the central axis of the four-helix bundle, as defined by equation 6 of ref. 6.
[f]R_i is the closest approach (23) of the ith helix axis to the central axis of the four-helix bundle.
[g]Θ is the twisted angle of the four-helix bundle, as defined by equation 5 of ref. 6.

Table 3. Energetic terms of the four-α-helix bundle proteins investigated

	Energy, kcal/mol												
	α-Helix set					Loop set							
	Intrahelix	Interhelix			Total	Intraloop	Interloop			Total	Loop–helix interaction			Total
Protein	E^{α}_{intra}	$\varepsilon^{\alpha}_{ES}$	$\varepsilon^{\alpha}_{NB}$	E^{α}_{inter}	E^{α}_{tot}	E^{loop}_{intra}	ε^{loop}_{ES}	ε^{loop}_{NB}	E^{loop}_{inter}	E^{loop}_{tot}	ε_{ES}	ε_{NB}	ε	E_{tot}
Meth	−447.8	−3.5	−148.9	−152.4	−600.2	−98.2	−0.9	−24.7	−25.6	−123.8	−30.7	−222.3	−253.0	−977.0
Cyt *b*-562	−553.5	−11.0	−169.5	−180.5	−734.0	−54.3	−1.3	−8.8	−10.1	−64.4	−45.1	−146.8	−191.9	−990.3
Cyt *c*′	−445.4	−6.5	−156.5	−163.0	−608.4	−96.4	−2.9	−31.3	−34.2	−130.6	−52.6	−181.8	−234.4	−973.4

Meth, methemerythrin; Cyt *b*-562, cytochrome *b*-562; Cyt *c*′, cytochrome *c*′. One calorie = 4.18 J. The energetic terms are defined as follows (see also Eq. 1): $\varepsilon^{\alpha}_{ES}$, electrostatic interhelix energy; $\varepsilon^{\alpha}_{NB}$, nonbonded interhelix energy; E^{α}_{inter}, $E^{\alpha}_{inter} = \varepsilon^{\alpha}_{ES} + \varepsilon^{\alpha}_{NB}$; ε^{loop}_{ES}, electrostatic interloop energy; ε^{loop}_{NB}, nonbonded interloop energy; E^{loop}_{inter}, $E^{loop}_{inter} = \varepsilon^{loop}_{ES} + \varepsilon^{loop}_{NB}$; ε_{ES}, electrostatic loop–helix interaction energy; ε_{NB}, nonbonded loop–helix interaction energy; ε, $\varepsilon = \varepsilon_{ES} + \varepsilon_{NB}$; $E_{tot} = E^{\alpha}_{tot} + E^{loop}_{tot} + \varepsilon$.

only for a theoretical four-helix bundle model in which all helices and loops are formed from poly(Ala) sequences but also for real four-helix bundle proteins. Such a finding will have some impact for an understanding of intersegment interactions in a protein during its folding process (13).

We thank Dr. K. Kirschner for sending us a preprint of ref. 13 and Dr. D. C. Rohrer for the cartoon ribbon drawings. This work was supported by Computer Chemistry, Upjohn Laboratories, and by the National Institute of General Medical Sciences (GM-14312), the National Science Foundation (DMB90-15815), and the National Foundation for Cancer Research.

1. Chou, K. C., Maggiora, G. M., Némethy, G. & Scheraga, H. A. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 4295–4299.
2. Argos, P., Rossmann, M. G. & Johnson, J. E. (1977) *Biochem. Biophys. Res. Commun.* **75,** 83–86.
3. Weber, P. C. & Salemme, F. R. (1980) *Nature (London)* **287,** 82–84.
4. Richardson, J. S. (1981) *Adv. Protein Chem.* **34,** 167–339.
5. Banner, D. W., Kokkinidis, M. & Tsernoglou, D. (1987) *J. Mol. Biol.* **196,** 657–675.
6. Carlacci, L. & Chou, K. C. (1990) *Protein Eng.* **3,** 509–514.
7. Carlacci, L. & Chou, K. C. (1990) *Protein Eng.* **4,** 225–227.
8. Stenkamp, R. E., Sieker, L. C. & Jensen, L. H. (1983) *Acta Crystallogr. Sect. B* **39,** 697–703.
9. Carlacci, L., Chou, K. C. & Maggiora, G. M. (1991) *Biochemistry* **30,** 4389–4398.
10. Palmer, K. A. & Scheraga, H. A. (1991) *J. Comput. Chem.* **12,** 505–526.
11. Thornton, J. M., Sibanda, B. L., Edwards, M. S. & Barlow, D. J. (1988) *Bio Essays* **8,** 63–69.
12. Chou, K. C. & Carlacci, L. (1991) *Proteins Struct. Funct. Genet.* **9,** 280–295.
13. Urfer, R. & Kirschner, K. (1992) *Protein Sci.* **1,** 31–45.
14. Leszczynski, J. F. & Rose, G. D. (1986) *Science* **234,** 849–855.
15. Némethy, G., Pottle, M. S. & Scheraga, H. A. (1983) *J. Phys. Chem.* **87,** 1883–1887.
16. Momany, F. A., McGuire, R. F., Burgess, A. W. & Scheraga, H. A. (1975) *J. Phys. Chem.* **79,** 2361–2381.
17. Gay, D. M. (1983) *Assoc. Comput. Mach. Trans. Math. Software* **9,** 503–524.
18. IUPAC-IUB Commission on Biochemical Nomenclature (1970) *Biochemistry* **9,** 3471–3479.
19. Lederer, F., Glatigny, A., Bethge, P. H., Bellamy, H. D. & Mathews, F. S. (1981) *J. Mol. Biol.* **148,** 427–448.
20. Finzel, B. C., Weber, P. C., Hardman, K. D. & Salemme, F. R. (1985) *J. Mol. Biol.* **186,** 627–643.
21. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *J. Mol. Biol.* **112,** 535–542.
22. Vásquez, M., Némethy, G. & Scheraga, H. A. (1983) *Macromolecules* **16,** 1043–1049.
23. Chou, K. C., Némethy, G. & Scheraga, H. A. (1984) *J. Am. Chem. Soc.* **106,** 3161–3170.
24. Sheridan, R. P., Levy, R. M. & Salemme, F. R. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 4545–4549.
25. Wada, A. (1976) *Adv. Biophys.* **9,** 1–63.
26. Hol, W. G. J., van Duijnen, P. T. & Berendsen, H. J. C. (1978) *Nature (London)* **273,** 443–446.
27. Sheridan, R. P. & Allen, L. C. (1980) *Biophys. Chem.* **11,** 133–136.
28. Hol, W. G. J., Halie, L. M. & Sander, C. (1981) *Nature (London)* **294,** 532–536.
29. Gilson, M. K. & Honig, B. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 1524–1528.