Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 4545–4549, August 1982
Biochemistry

α-Helix dipole model and electrostatic stabilization of 4-α-helical proteins

(electrostatic interactions/protein structure)

ROBERT P. SHERIDAN*, RONALD M. LEVY*†, AND F. R. SALEMME‡§

*Department of Chemistry, Rutgers University, New Brunswick, New Jersey 08903; and ‡Department of Biochemistry, University of Tucson, Tucson, Arizona 85717

Communicated by Frederic M. Richards, April 12, 1982

ABSTRACT **A simple dipole model is developed for estimation of the electrostatic interaction energy between α-helices in proteins. This model is used to estimate the electrostatic stabilization in a recurrent protein tertiary structural motif, an array of four closely packed α-helices. It is found that, for the proteins examined (cytochrome c′, hemerythrin, myohemerythrin, cytochrome b_{562}, and a T4 phage lysozyme domain), their common antiparallel arrangement of adjacent helices confers a stabilization of 5–7 kcal/mol (1 cal = 4.18 J). In contrast, a similarly packed array of parallel helices is relatively destabilized by 20 kcal/mol. These results show that helix–dipole interactions are important in the stabilization of this structural motif. These effects are discussed both in the context of folding pathways for 4-α-helical proteins and the stabilization of the higher aggregates.**

X-ray crystallography has shown that many proteins incorporate α-helices as the predominant form of secondary structure. Although α-helical proteins show considerable structural diversity (1), the recurrence of a small number of observed interhelical geometries suggests that interactions between helices play a determinative role in protein structural organization. Consequently, investigations of the physical origin of these preferred arrangements are relevant to a fundamental understanding of protein structure.

Thus far, most studies have been concerned with evaluation of the hydrophobic stabilization resulting from the transfer of the α-helix surface from an aqueous to a hydrophobic environment upon association of helix pairs (2, 3, 4). The concomitant requirements for intimate side-chain packing appear consistent with only few interhelical geometries (2), thus accounting reasonably well for the interactions actually seen in proteins. Nevertheless, the α-helix backbone possesses a large net dipole moment (5, 6, 7), a property suggesting that helix electrostatic interactions may also be important in the stabilization of α-helical proteins. Hol *et al.* (8) have recently calculated the electrostatic interactions between the peptide groups of both α-helices and β-sheets in various proteins and demonstrated that secondary structures pack in a manner to provide significantly favorable electrostatic energy.

In the present communication we focus on the role of helix–dipole interactions in the stabilization of a recurrent protein tertiary structural motif. This is a nearly parallel arrangement of 4-α-helices (9) which is seen in cytochrome c′, cytochrome b_{562}, hemerythrin, apoferritin, and a domain of T4 phage lysozyme (Fig. 1). It is shown that a simple dipole model for the α-helix provides a reliable estimate of the electrostatic interaction energies between pairs of α-helices, and we apply this treatment to several 4-α-helical proteins. It is found that, owing to the adjacent antiparallel sense of the helices, dipole interactions confer significant electrostatic stabilization to the tertiary structures of these proteins. These results suggest that helix–dipole interactions may be important both in the folding pathway for these proteins and in their subsequent association in multimeric aggregates.

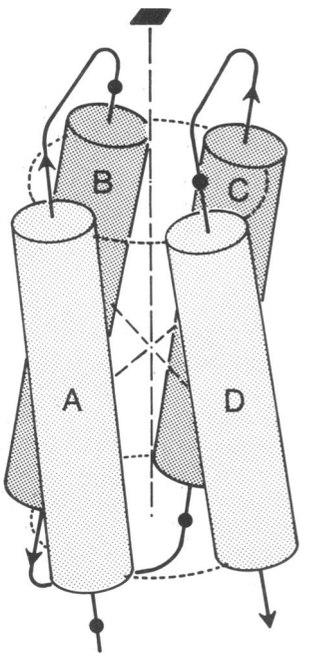

FIG. 1. Schematic illustration of the recurrent 4-α-helical structural motif. Four sequentially connected α-helices are packed in an approximately 4-fold symmetric array. Adjacent helices interact at angles of ≈18° and have closest interhelix axis approach distances of ≈10 Å; diagonally related helices are ≈14 Å apart. Various observed structures include slight distortions in interhelical distance and angle or truncations of the helices (9).

METHODS

Coordinates for 4-α-helical proteins were obtained from the Brookhaven Protein Data Bank (10). The choice of residues for the helix termini were taken from the Data Bank entry. Coordinates of idealized symmetric arrays were generated by positioning four regular 25-residue α-helices (1.53-Å rise; 98.2° per residue; 3.66 residues per helical turn) on a framework having the mean geometry of the observed structures (9).

All-atom computations of helix electrostatic potential assumed polyalanine helices that were terminated by acetyl (Ac)

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: TMV, tobacco mosaic virus.

† To whom reprint requests should be addressed.
§ Visiting Associate Professor, Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511.

and methylamino (MeNH) groups. Each atom was represented by a charge placed at the atomic center. The charges (in units of electron charge) were derived from minimal basis set calculations on amino acid models (11): Ac, C_α −0.208, H_α 0.067, C 0.310, O −0.303; Ala, N −0.378, H 0.201, C_α 0.062, H_α 0.086, C_β −0.179, H_β 0.067, C 0.310, O −0.303; Gly, N −0.378, H 0.201, C_α −0.002, H_α 0.086, C 0.310, O −0.303; MeNH N −0.378, H 0.201, C_α −0.024, H_α 0.067.

Dipole computations of electrostatic potential and interaction energies approximated the helix as two charges (±0.25 electron) lying on the helix axis, the positive charge displaced 0.3 Å from the NH_2-terminal C_α in the N → C direction and the negative charge displaced 1.3 Å from the helix COOH-terminal C_α in the N → C direction. The helix axes were determined from C_α positions by the method of Chothia *et al.* (3).

RESULTS

Line Dipole Model of an α-Helix. Electrostatic potentials of the α-helix have been calculated from line dipoles (6), from models which deal only with the atoms of the peptide link (8), and from all-atom models (7). In all cases it was suggested that the electrostatic potential of the helix could be approximated by a positive charge at the NH_2 terminus and equal negative charge at the COOH-terminus. For the present application, the positions of these equivalent charges were located more exactly by matching the electrostatic potential of such simple dipole models to the electrostatic potential of all-atom models of regular polyalanine helices of various lengths. One such match is shown in Fig. 2. Except for very short helices (<7 residues), charges of + and − 0.25 electron lying on the helix axis as described above provide a reasonable description of the electrostatic potential of the helix.

In order to determine the validity of the dipole model in computations of the electrostatic potential for observed nonideal α-helix geometries, a comparison was made of the electrostatic interaction energy between pairs of helices in several proteins by using both the all-atom and the simple dipole models. Fig. 3 shows the correlation of the electrostatic energy between pairs of α-helices calculated using the simple dipole model versus the all-atom results. Considering the simplicity of the model, it provides a good estimate of the interaction energy between the pairs of helices, even though many of the interactions are rather short range. Most of the scatter is due to the fact that the electrostatic potential of the all-atom model is not cylindrically symmetrical. The fact that the placement of equivalent charges in the simple dipole model is based on an idealized regular helix, whereas the real helices are somewhat irregular, appears to be a much less important source of error.

Electrostatic Stabilization of 4-α-Helical Proteins. In proteins that incorporate a 4-α-helical motif, the helices are packed in an array of approximately square cross section, with adjacent helix axes inclined at an angle of 18 ± 5° (9). Fig. 1 illustrates an idealized arrangement in which the helices are of equal length and most closely approach each other midway along their axes. The observed structures can be regarded as various distortions or truncations of the idealized arrangement (9). Although current treatments of interhelical packing interactions suggest nearly equivalent extents of hydrophobic stabilization irrespective of the relative N-to-C senses of the adjacently packed helices (2, 3), the observed structures invariably pack with adjacent helices antiparallel. Table 1 gives estimates of the electrostatic stabilization afforded by this arrangement for both real and idealized structures, based on the helix–dipole model. The magnitude of the energies require further comment.

The dipole moment of the peptide group in the all-atom model is about 62% of the experimental value (5). Also, polarization by hydrogen bonding increases the dipole moment of the peptide group (12, 13), by a factor of ≈1.25. Thus, the equivalent charges at the termini should be about 0.25 × (1.6 × 1.25) = ±0.5 electron. [This is close to the estimate of Hol *et al.* (6) based on a line dipole model for the helix.] Therefore, the energies of interaction should be increased by a factor of $(1.6 \times 1.25)^2$. Finally, the interaction energy calculated from the simple dipole model is ≈80% of that calculated from the all-atom model (see Fig. 3 legend). Thus, for a dielectric constant of 1, the energies from the simple dipole model must be in-

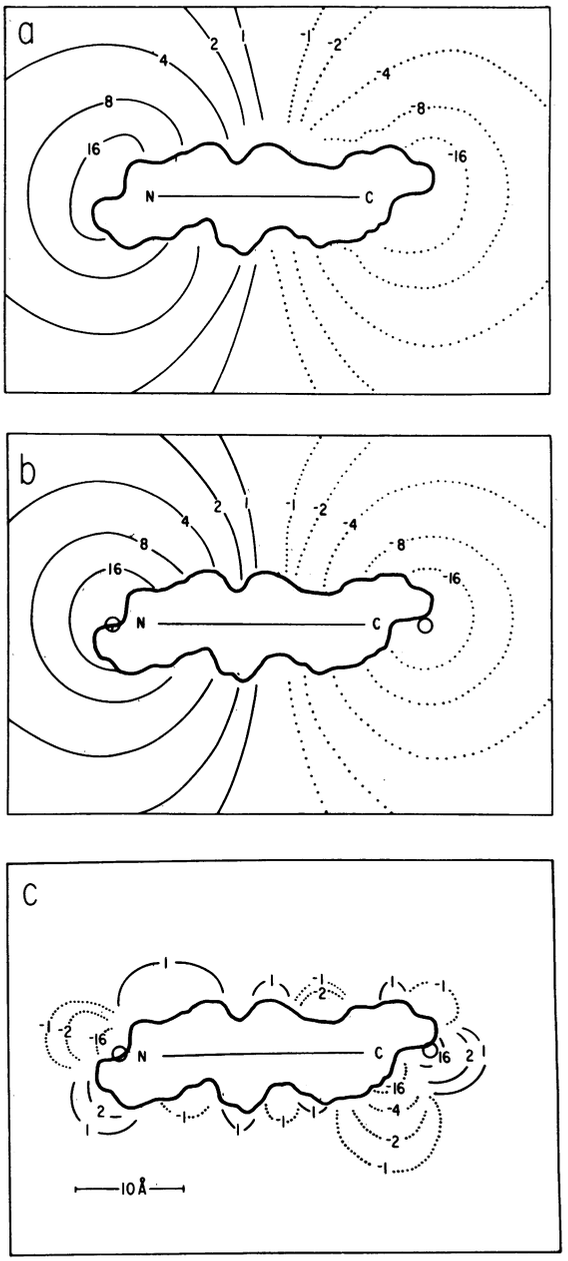

FIG. 2. (*a*) Electrostatic potential of the all-atom model for the regular Ac-Ala_{17}-MeNH helix. The plane of cross section is through the helix axis. Contour levels show kcal/mol. (*b*) Electrostatic potential of the simple dipole model. The magnitude and position of the charges were chosen such that the electrostatic potential of the simple dipole model closely matched the potential from the all-atom model. (*c*) Difference in electrostatic potential: all-atom model minus simple dipole model. Most of the small differences between the two models is a reflection of the fact that the more exact all-atom model does not have cylindrical symmetry.

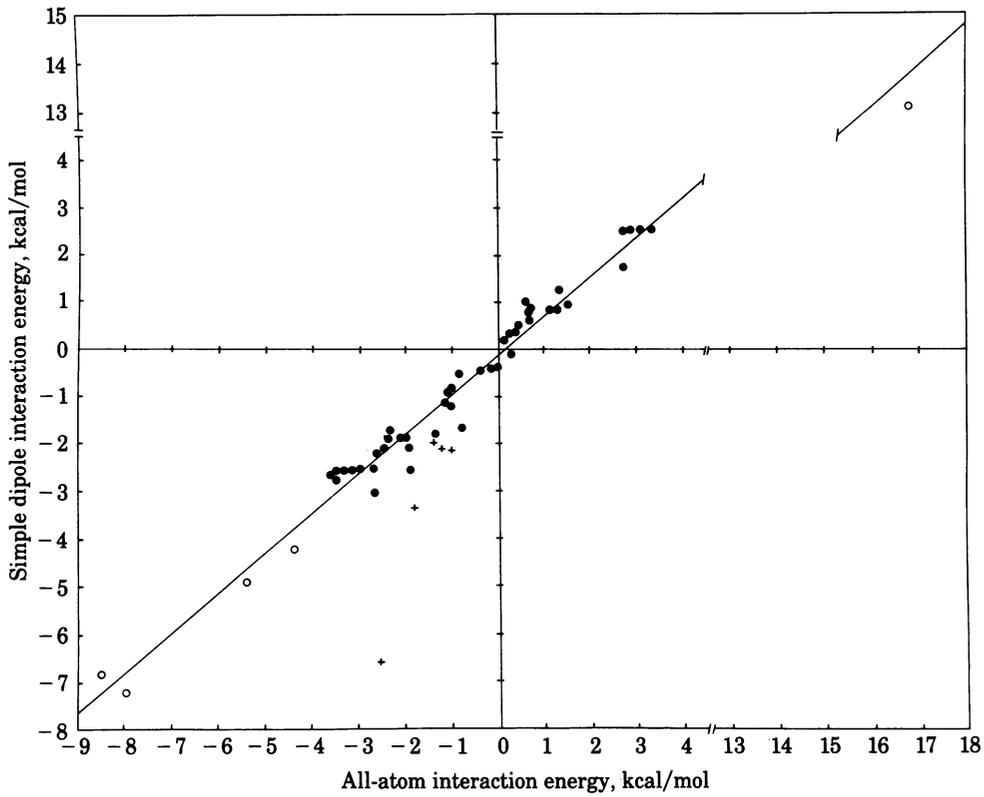

FIG. 3. Correlation in helix–helix electrostatic interaction energy between the all-atom model and the simple dipole model. The proteins containing the helices used for this determination were myoglobin, triose phosphate isomerase, cytochrome *c'*, cytochrome b_{562}, and T4 lysozyme. The last three contain 4-α-helical arrays. Parallel and antiparallel 4-α-helical arrays consisting of regular Ac-Ala_{23}-MeNH helices were also included. ●, Pairwise interactions; ○, total energies for 4-α-helical bundles; +, pairwise interactions between helices separated by two or fewer residues (of which there are none in the 4-α-helical bundles). The linear regression for 54 points (not including +) gives a slope of 0.81 (correlation = 0.99).

creased by approximately $(1.6 \times 1.25)^2 \times 1.2 = 4.8$. However, the dielectric constant in the interior of proteins is usually taken as 4–5 (14), so that the consideration of the dielectric effect approximately compensates for the absence of charge scaling in our results.

For the six proteins listed in Table 1, the electrostatic energy stabilizing the 4-α-helical bundles varies between −4.8 and −7.0 kcal/mol. From comparison of this result with the electrostatic energy of the idealized antiparallel array [−6.8 kcal/mol (Table 1)], it is apparent that, although there is a spread

Table 1. Simple dipole helix–helix interaction energy

	kcal/mol for 4-α-helical bundle proteins*							
	Cytochrome *c'*	Hemerythrin† (*Themiste*)	Hemerythrin† (*Phascolopsis*)	Cytochrome–b562	Myohemerythrin†	T4 lysozyme	Parallel bundle‡	Anti-parallel bundle‡
Residues in helix§								
A	5–23	21–37	18–38	2–19	18–38	93–106	1–25	1–25
B	42–54	41–64	40–62	24–45	40–62	115–123	26–50	26–50
C	79–100	69–86	69–87	62–86	69–87	126–134	51–75	51–75
D	106–117	90–103	88–105	88–108	93–110	143–155	76–100	76–100
Interaction:								
AB	−2.0	−2.0	−2.9	−2.5	−2.9	−0.9	2.5	−2.5
AC	1.1	1.0	1.5	1.3	1.3	0.9	1.5	1.5
AD	−0.4	−2.1	−1.9	−1.8	−1.8	−2.2	2.5	−2.5
BC	−2.2	−2.2	−2.3	−2.7	−2.4	−1.5	2.5	−2.5
BD	0.9	1.1	1.3	1.7	1.3	1.0	1.5	1.5
CD	−2.2	−1.5	−2.7	−2.9	−2.1	−2.1	2.5	−2.5
Total	−4.8	−5.7	−7.0	−6.9	−6.6	−4.8	13.0	−7.0

* Coordinates (except for cytochrome *c'* and regular bundles) from Brookhaven Protein Data Bank (10).
† Only coordinates of C_α available.
‡ Consists of four regular Ac-Ala_{23}-MeNH helices arranged such that the axes of adjacent helices are 9.8 Å away at the point of closest approach and tilted relative to each other by 18°. The pairwise energies of adjacent all-atom helices may vary up to ±12% with rotation around the helix axis.
§ Choice of termini and residue numbering as in Data Bank entry. A ±2-residue change in termini choice gives ≤20% change in interaction energy.

in the electrostatic energies for the real α-helical proteins owing to distortions of the idealized structure and differences in helix length, the antiparallel helix packing makes a significant contribution to the stabilization of the native structures. In contrast, a similar parallel packing of 4-α-helices (never observed) would be less stable by almost 20 kcal/mol. It should be noted that Hol *et al.* (8) reported a value of −14.3 kcal/mol for the electrostatic stabilization of the 4-α-helical array in tobacco mosaic virus (TMV) based on dielectric constant $\varepsilon = 1$ for proteins. Use of a more realistic dielectric constant, $\varepsilon = 4$–5, brings the estimate of the electrostatic energy of the 4-α-helical bundle in TMV into closer agreement with the results presented in Table 1.

DISCUSSION

The free energy of protein tertiary structural stabilization generally amounts to 10–20 kcal/mol (15). In addition to the helix–dipole electrostatic interactions considered here, both interhelical van der Waals forces and electrostatic interactions between charged surface side chains must play a role in stabilization of the protein structure. However, current treatments of interhelical packing (2, 3) suggest little difference in the magnitude of van der Waals or solvent exclusion effects which depend upon the relative senses of helices in a given geometrical arrangement. Similarly, although it is known that charged side-chain interactions generally stabilize folded proteins (16), they cannot account for the recurrence of the antiparallel arrangement. The proteins sharing this motif have diverse origins, and the sequential arrangement of charged amino acids bears no systematic relationship to the sense of the macroscopic helix dipoles (9). Moreover, surface exposure of charged side chains in proteins situates them in a solvent medium of high dielectric constant, so that the magnitude of the electrostatic effects are significantly smaller than helix–dipole interactions in the low-dielectric protein interior (14). Whereas both interhelical packing and side-chain electrostatic effects would be equivalent, on the average, for helices in parallel and antiparallel array, the corresponding helix–dipole interactions would destabilize the molecule by 15–20 kcal/mol (Table 1), an amount comparable to the molecules' total stabilization energy. Consequently, it appears that, in addition to the previously described packing and chiral effects (9), helix–dipole interactions play a major role in stabilization of the 4-α-helical structural motif.

In addition to the stabilization afforded the native protein structure, helix–dipole interactions also may be important in the process of folding 4-α-helical proteins and in their subsequent association as multimeric aggregates. In the former case, it is frequently assumed that the formation of hydrogen-bonded secondary structures constitutes an early step in the process of protein folding. As shown in Fig. 4, the frequent occurrence of short interhelical connections in 4-α-helical proteins would be expected to be entropically favorable to the electrostatic association of sequentially adjacent helices. The interesting point is that, although many alternative pairwise arrangements are possible, the most direct routes of association (i.e., a, c, g, i, and a, d, h, i in Fig. 4) correspond to intermediates of successively increasing electrostatic stabilization and lead ultimately to the same final result. This suggests that the incorporation of dipole interactions may prove useful in the further development of protein folding algorithms (17), particularly as a means for estimating relative stabilities of both intermediates and final folded structures. Moreover, such approaches are not restricted to 4-α-helical proteins but appear to be equally applicable to other α-helical proteins such as myoglobin, whose tertiary structure corresponds essentially to a superhelically folded hairpin array of α-helices (2).

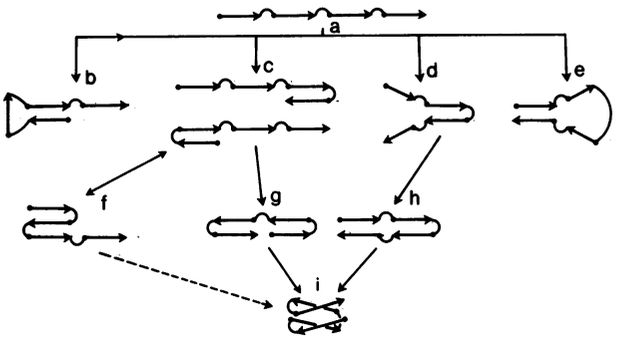

FIG. 4. Helix–dipole interactions during folding of 4-α-helical proteins. Part a shows an initial folding state in which α-helices have formed with the associated production of helix dipoles (arrows). Interhelical connections in observed 4-α-helical proteins are typically quite short, thus entropically favoring the formation of adjacently paired helices whose interactions are stabilized by electrostatic interactions (c and d) and disfavoring improperly paired (b) or remote interactions (e). Subsequent adjacent interactions (f, g, h) produce arrangements of enhanced electrostatic stabilization. However, g and h lead directly to the final structure (i) with a correspondingly large increase in electrostatic stabilization, whereas f or its variants must undergo rearrangements to additional intermediates having less stabilization.

An interesting property of the 4-α-helical motif is its frequent occurrence in multimeric aggregates having quite different subunit interaction patterns. For example, cytochrome *c'* is a symmetric dimer (18), hemerythrin is an octamer with 422 symmetry (19), apoferritin is a 24-subunit array with cubic (432) symmetry (20), and TMV is a radially symmetric disc (21) or helical arrangement (22). Although these structural variations reflect detailed differences in both subunit geometries and the nature of specific interactions formed between molecules within a given complex, all of these structures incorporate principal subunit interactions which result in the same pairwise antiparallel helix–dipole arrangement seen within an individual 4-α-helical array. The recurrence of this common feature among structures whose subunits are related by different symmetry operations is a result of the symmetry of the individual 4-α-helical subunits. As shown in Fig. 5*a*, individual subunits correspond to dipole arrays with 222 symmetry. Owing to this high internal symmetry, intersubunit antiparallel arrays are produced in multimeric structures whose symmetry-related subunits are generated by either 2-fold rotations parallel to any of the array diad axes or by pure translations. For example, Fig. 5*b* illustrates the dipole arrangement in the cytochrome *c'* dimer in which subunits are related by a diad axis parallel to the long axis of the monomer array. In hemerythrin (Fig. 5*c*) and apoferritin (Fig. 5*d*), in contrast, principal diad-related subunit interactions involve 2-fold rotations parallel to face diad axes of the monomer. Finally, in TMV (Fig. 5*e*), adjacent subunits related by radial or helical symmetry can be viewed essentially as translational operations on the individual monomers (e.g., by unrolling the helix to form a sheet of monomers). Consequently, it appears that the same pairwise dipole interactions that are important in the stabilization and organization of the individual monomer arrangements make similarly important contributions to their intimately packed aggregate structures.

CONCLUSION

It has been shown here that the electrostatic energy associated with helix–dipole interactions can be accurately estimated using a simple dipole model. This may be of particular use both in estimating the energy of intermediates generated by computational protein folding alogrithms and in estimating the mag-

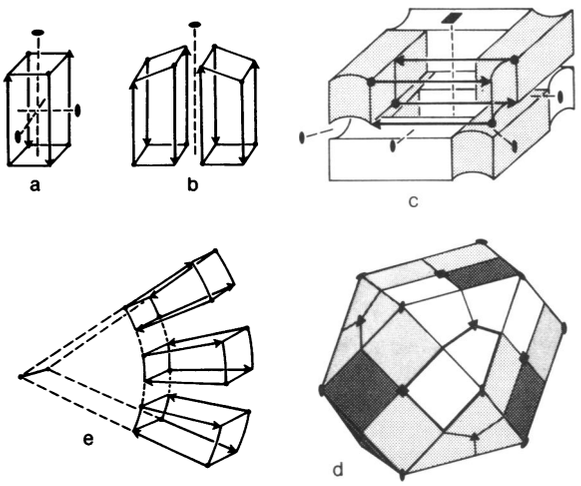

FIG. 5. Dipole interactions in aggregate 4-α-helical structures. (*a*) Dipole symmetry in a monomeric 4-α-helical subunit (with twist removed for clarity). The monomer dipole array has 2-fold rotational symmetry axes (dashed) centrally located on each face and along the long axis of array. (*b*) Arrangement found for cytochrome *c'*. Subunits are related by a 2-fold axis parallel to the long-axis diad of the monomer. (*c*) Arrangement of subunits in hemerythrin. Rotation about a face-diad of the monomer produces the most extensive subunit interactions between the two 4-fold symmetric layers of the octamer. (*d*) The 24 subunit arrangement in apoferritin. This structure is represented as a cubeoctahedron, a polyhedron with six square and eight equilateral triangular faces. As shown by the dark outline, each subunit occupies one-fourth of each square face and one-third of each triangular face. The structure is shaded as trimers whose subunits make pairwise antiparallel interactions with diad-related subunits from adjacent trimers at the 12 vertices of the polyhedron. (*e*) Dipole interactions in a section of one layer of the TMV disc or helix structure. Adjacent molecules in a single layer (shown) or in multiple layers all incorporate pairwise antiparallel helix interactions at the molecular interfaces.

nitude of electrostatic stabilization effects for proteins whose structures are not known at atomic resolution.

Application of the dipole model to the recurrent 4-α-helical structural motif shows that helix–dipole interactions are a major factor in the stabilization of this arrangement. Furthermore, owing to the symmetry properties of this array, it has been shown that helix–dipole interactions similarly may contribute to the stabilization of the diverse aggregates of these molecules.

This work was supported by grants from The Petroleum Research Fund as administered by the American Chemical Society and by National Institutes of Health Grant GM-30580. Work in the laboratory of F.R.S. was supported by National Institutes of Health Grant GM-21534. Computer time was supplied by the Rutgers University Center for Computer and Information Services and by the University of Arizona Computer Center. R.M.L. is an Alfred P. Sloan Foundation Fellow.

1. Richardson, J. S. (1981) *Adv. Protein Chem.* **34,** 167–339.
2. Richmond, T. J. & Richards, F. M. (1978) *J. Mol. Biol.* **119,** 537–555.
3. Chothia, C., Levitt, M. & Richardson, D. (1980) *J. Mol. Biol.* **145,** 215–250.
4. Lesk, A. M. & Chothia, C. (1980) in *Proteins and Nucleoproteins Structure, Dynamics and Assembly*, ed. Parsegian, V. A. (Rockefeller University Press), pp. 35–44.
5. Wada, A. (1976) *Adv. Biophys.* **9,** 1–63.
6. Hol, W. G. J., van Duijnen, P. Th. & Berendsen, H. J. C. (1978) *Nature (London)* **273,** 443–446.
7. Sheridan, R. P. & Allen, L. C. (1980) *Biophys. Chem.* **11,** 133–136.
8. Hol, W. G. J., Halie, L. M. & Sander, C. (1981) *Nature (London)* **294,** 532–536.
9. Weber, P. C. & Salemme, F. R. (1980) *Nature (London)* **287,** 82–84.
10. 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.
11. Hayes, D. M. & Kollman, P. A. (1976) *J. Am. Chem. Soc.* **98,** 3335–3345.
12. Sheridan, R. P., Peters, N., Lee, R. H. & Allen, L. C. (1979) *Biopolymers* **18,** 2451–2458.
13. Mehler, E. (1980) *J. Am. Chem. Soc.* **102,** 4051–4056.
14. Matthew, J. B., Hanania, G. I. H. & Gurd, F. R. N. (1979) *Biochemistry* **18,** 1919–1928.
15. Privalov, P. L. (1979) *Adv. Protein Chem.* **33,** 167–241.
16. Wada, A. & Nakamura, H. (1981) *Nature (London)* **293,** 757–758.
17. Cohen, F. E., Richmond, T. J. & Richards, F. M. (1979) *J. Mol. Biol.* **132,** 275–288.
18. Weber, P. C., Howard, A., Xuong, N. H. & Salemme, F. R. (1981) *J. Mol. Biol.* **153,** 399–424.
19. Stenkamp, R. E., Sieker, L. C., Jensen, L. H. & McQueen, J. E., Jr. (1978) *Biochemistry* **17,** 2499–2504.
20. Banyard, S. H., Stammers, D. K. & Harrison, P. M. (1978) *Nature (London)* **271,** 282–284.
21. Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R. & Klug, A. (1978) *Nature (London)* **276,** 362–368.
22. Stubbs, G., Warren, S. & Holmes, K. (1977) *Nature (London)* **267,** 216–221.