# SOLVENT ACCESSIBILITY, PROTEIN SURFACES, AND PROTEIN FOLDING

Arthur M. Lesk, *Fairleigh Dickinson University, Teaneck, New Jersey, 07666 U. S.A.*
Cyrus Chothia, *William Ramsay, Ralph Foster, and Christopher Ingold Laboratories, University College, London, WC1E 6BT, England*

ABSTRACT Studies of the native structures of proteins, together with measurements of the thermodynamic properties of the transition between unfolded and native states, have defined the major components of the forces that stabilize native protein structures. However, the nature of the intermediates in the folding process remains largely hypothetical. It is a fairly widespread and not implausible assumption that the intermediates in the folding of a monomeric protein contain the same kinds of secondary and tertiary structures that appear in the native conformation, and that, although unstable, their lifetimes are prolonged by forces similar to those that stabilize the native structure. We wished to examine what happens if, during the folding of a monomeric protein, regions of secondary structure come together to form an intermediate of reduced instability. We applied calculations of accessible surface area (a measure of hydrophobic stabilization) and parameterized nonbonded energy calculations (measuring the strengths of van der Waals forces) to identify the kinds of stabilizing interactions that might be available to such an intermediate. First, we analyzed the total buried surface area of two types of proteins into contributions from formation of secondary structure alone, interaction of pairs of secondary-structural elements, the formation of the complete secondary structure without the turns, and the complete native structure. The formation of secondary structure alone, without tertiary-structural interactions, buries roughly half the surface that the complete structure does. We then analyzed in more detail the approach of two $\alpha$-helices to form a complex, as an illustrative example of the nature of the interaction between compact structural units which remain fairly rigid during their interaction. Many features of the results are not limited to the interaction of $\alpha$-helices. (The results therefore neither confirm nor refute the hypothesis that $\alpha$-helices are intermediates in the folding of proteins.) We find that the first forces to be felt upon approach arise from solvent exclusion from hydrophobic side chains. These forces do not impose the stringent geometric conditions on the relative position and orientation of the two helices as does the close packing which optimizes the van der Waals interactions at shorter distances apart. Therefore there appears to be a range of distances in which hydrophobic interactions could create a nonspecific complex between two helices in which the side chains might have sufficient time to seek the proper interdigitation observed in the native structure, where the two helices are in intimate contact. Indeed, we find that only in the final stages of approach is the native geometry the most stable; in the region in which solvent-exclusion forces predominate, the conformation with helix axes parallel is more stable than the native conformation, in the cases we examined. The computational experiments we have performed were an attempt to describe how certain components of the interaction energy vary with the relative position and orientation of the interacting units. Although we used the best numerical estimates of the energy components that were available to us, the qualitative features of our results—which we consider to be the most important features— do not depend on the precision of the parameters.

## INTRODUCTION

The combination of the study of high-resolution crystal structures of proteins and the measurement of the thermodynamic properties of the transition between unfolded and native

states has clarified the nature of the major components of the forces that stabilize native protein structures. Although the nature of the intermediates in folding remains largely hypothetical, it is a plausible assumption that a monomeric protein, as it folds, proceeds through transient intermediates that possess the same types of secondary and tertiary structures as a native protein, and that would be stabilized (in a relative sense) by the same types of forces that are active in producing the true stability of the native structure (Karplus and Weaver, 1976; Baldwin, 1978, 1979).

The relationship between the structure of the native protein and the structure of intermediates in its folding pathway cannot be described with confidence. On the one hand, the forces that stabilize secondary structures in native proteins should certainly produce transiently stable regions of secondary structures from the unfolded state. Moreover, examinations of native structures have often suggested plausible folding pathways, in the form of a sequence of successively larger portions of the native structure that could form consecutively. But, in the one case for which hard structural information is available—pancreatic trypsin inhibitor—the folding does not proceed in this manner (Creighton, 1978).

We therefore wished to consider the general question of the extent to which forces that stabilize secondary and tertiary structures in native proteins are in fact available to plausible intermediates.

We regard the major thermodynamic components of the folding process as follows: (*a*) The internal rotational degrees of freedom of the backbone and many of the side chains are frozen into restricted conformations, with a concomitant large decrease in entropy. (*b*) The exclusion of water from contact with the residues in the interior of the protein contributes a large increase in entropy. (*c*) The close packing of protein interiors—the achievement of densities similar to those of crystals of amino acids—produces a state of relatively low enthalpy. Factors *b* and *c* "pay for" factor *a*. The net result is an intrinsic stability (~7–17 kcal/mol) that applies to the whole protein in its native state but not to intermediates (Privalov and Khechinashivili, 1974). In analyzing the stabilities of hypothetical intermediates, we rely as much as possible on the correlations between structural and energetic features. These include correlation between buried surface area and hydrophobic stabilization, and between packing density and van der Waals interactions.

Although for typical monomeric proteins no portion of the native structure is itself more than transiently stable at room temperature, it is possible to estimate the contributions of different aspects of the native structure to stability by calculations of buried surface area. For two $\alpha$-helical proteins (sperm whale myoglobin and *Chironomus* erythrocruorin), and a $\beta$-sheet protein (one domain of the Bence-Jones protein), we have analyzed the contributions to the burying of surface area of different parts of the structure, excised from the native molecule but retaining the same structure.

In addition, we have studied the forces active during an encounter between two $\alpha$-helices in aqueous solution. Contacts between pairs of helices are a recurrent theme in protein structures (Chothia et al., 1977). The globins, in particular, have a tertiary structure that consists almost entirely of helix-helix contacts; and the pattern of residue-residue interactions at homologous positions is preserved by evolution (Lesk and Chothia, 1980). We attribute this conservation to the necessity to maintain close-packed contacts in the face of the substitution of side chains by mutations; indeed, although in the globins the nature of the contact regions tends to be maintained, the molecules can tolerate rather large shifts in the relative positions and orientations of pairs of helices in contact. The extent to which this specificity extends to intermediates in the folding is unclear.

Upon investigating the forces between two helices as a function of the distance apart, we find that there is an interesting relationship between the separation of the helices, the ranges of the different types of forces between them, and the geometrical specificity of the interaction. As two helices approach each other from a long distance, the first forces to become active are solvent exclusion forces which are geometrically rather unspecific. In this range the native conformation is less stable than an alternative in which the helix axes are parallel, or, at least, the parallel-axis configuration buries more surface area. Only when the helices are almost in contact does the native structure assert its greater stability, on the basis of van der Waals interactions. This suggests that the helical complex between the two helices may be sterically unspecific, but that its stabilization by hydrophobic forces may give the side chains the opportunity to explore each other to find a good fit.

## MATERIALS AND METHODS

The coordinates for sperm whale metmyoglobin were given to us by Dr. T. Takano. Those for *Chironomus* erythrocruorin were from Dr. W. Steigemann. We used coordinates for the Bence-Jones protein from the Brookhaven National Laboratories Protein Data Bank.

Accessible surface area (Lee and Richards, 1971) were computed in a manner described previously (Chothia, 1976), using a program written by Dr. M. Levitt. Nonbonded interaction energies were calculated by a program also written by Dr. M. Levitt (Levitt, 1974).

For calculations of nonbonded energies, the helices were brought to the desired relative position and orientation with side chain conformations unchanged from the native structure, and the subjected to 30 cycles of energy refinement.

## RESULTS

### *Analysis of Buried Surface Area in Different Portions of the Native Structure.*

Globins—$\alpha$-Helical Proteins The globins are characterized by eight helices comprising approximately three-quarters of the residues in the ~150 residue polypeptide chains. The structure of the globins are stabilized by homologous sets of helix-helix contacts, the major ones of which are contacts between helices A and H, B and E, B and G, F and H or F'F and H, and G and H. (The F' helix is in the EF portion of mammalian globins, homologous to the beginning of the F helix of nonmammalian globins. Although these residues in mammalian globins were not originally described as helical, it is correct to regard the F helix of nonmammalian globins as broken into two contiguous helices, the axes of which make an angle of ~135° [Lesk and Chothia, 1980]).

We calculated the accessible surface area of the following states of the molecules: (*a*) the fully unfolded state; (*b*) the states in which only one helix is formed, for each helix; (*c*) individual pairs of helices in contact in the native structure; (*d*) the entire set of helices, assembled in their native structure, but without the turns; and (*e*) the complete native structure. Fig. 1 reports the accessible surface areas, in a chart that facilitates gauging the progress of the molecule as it buries sufficient surface to fold stably, assuming that at least some intermediates have some structural features in common with the native structure. As found by Richards and Richmond (1978) for myoglobin, and by Chothia (1976) for other proteins, the secondary structure alone, without any tertiary-structural interactions, buries about half the total surface area ultimately sequestered in the native molecule.

Fig. 1, containing the results for sperm whale myoglobin, and Fig. 2, containing the results for *Chironomus* erythrocruorin, suggest that the pattern is common to all globins.

A Domain of the Bence-Jones Protein, a $\beta$-sheet Protein In this case, we calculated

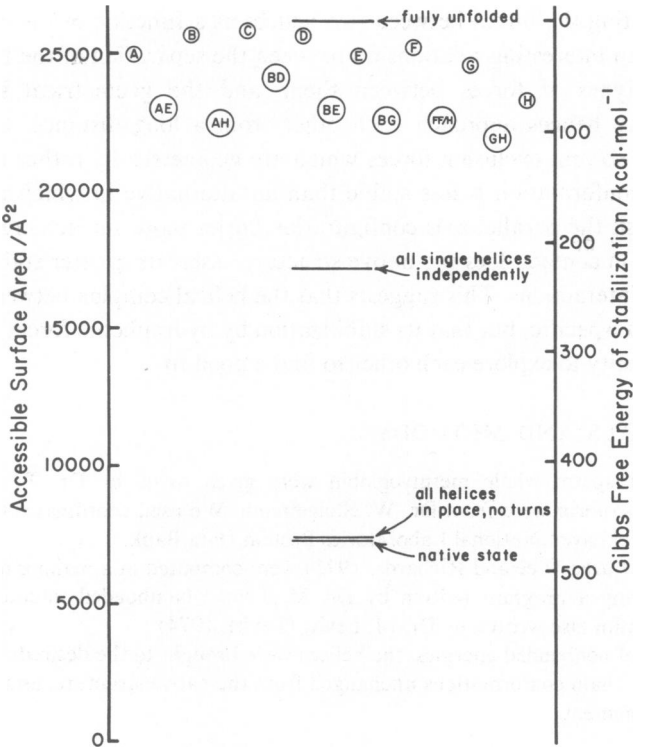


Figure 1 Analysis of contributions to burying of surface from different secondary and tertiary structural elements of sperm whale myoglobin.

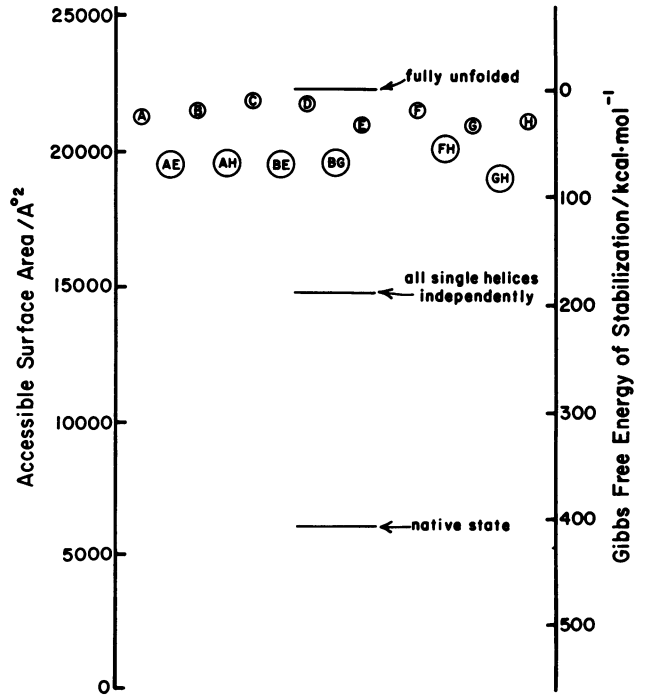


Figure 2 Analysis of contributions to burying of surface from different secondary and tertiary structural elements of *Chironomus* erythrocruorin.

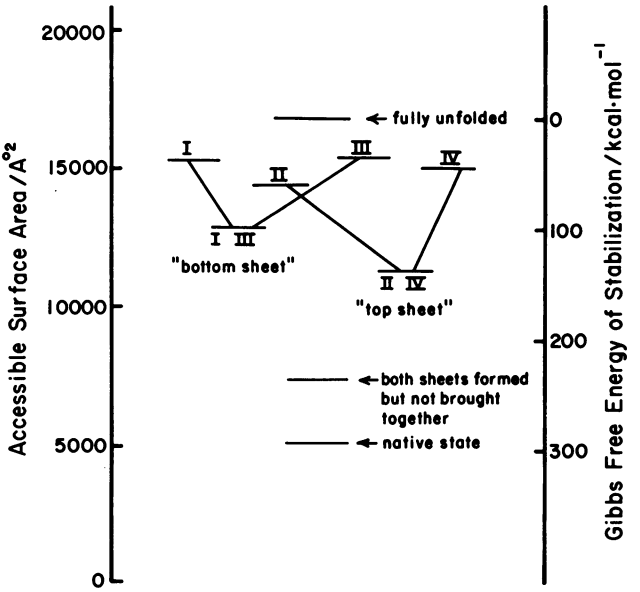


Figure 3 Analysis of contributions to burying of surface from secondary and tertiary structural elements of the monomer of the Bence-Jones protein. Region definitions: I: residues 2–27; II: residues 28–56; III: residues 57–80; IV: residues 81–106.

the accessible surface area of (*a*) the fully unfolded state; (*b*) the formation of individual "hairpins" each containing two strands of sheet in proximity and held together by hydrogen bonds; (*c*) the formation of each of the two sheets; (*d*) bringing together the two sheets to form the native structure. Again, roughly half the surface is buried by the secondary structural elements ("hairpins") themselves. Fig. 3 charts these results.

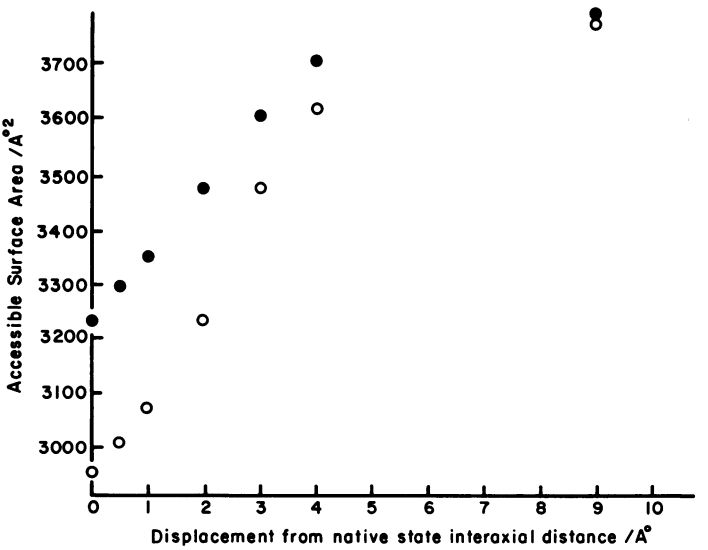


Figure 4 The variation of accessible surface area with inter-helix axis separation for the B and G helices of sperm whale myoglobin Abscissa: displacement, in Å, from the observed separation. ● = native structure, O = helices parallel.

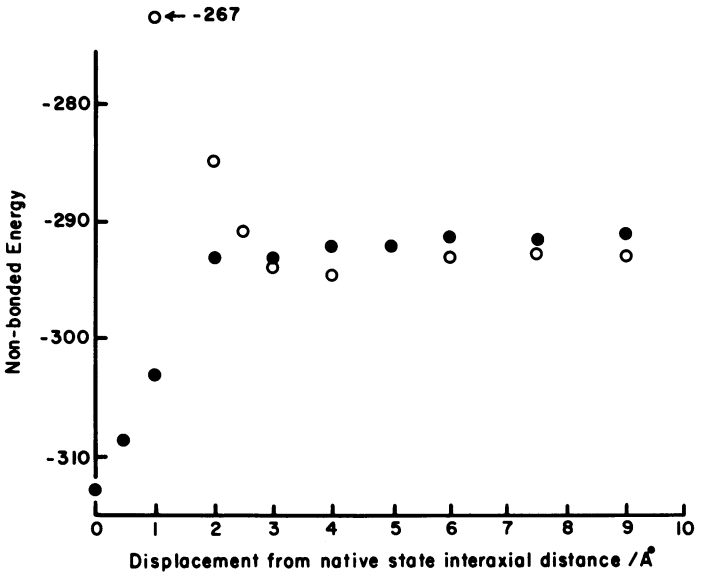


Figure 5 Variation of nonbonded energy with inter-helix axis separation for the B and G helices of sperm whale myoglobin.

*Experiments in Bringing Helices Together*

For the B/E and B/G pairs of helices in sperm whale myoglobin, we computed the accessible surface area and nonbonded energy as a function of distance apart, as the helices were moved in and out along the line perpendicular to both helix axes, We compared two orientations of the helices: the native conformation, and a nonnative conformation attained by a rotation around the line perpendicular to both axes until the helix axes became parallel.

Figs. 4 and 5 report the results for the B/G helices, and Figs. 6 and 7 report the results for the B/E helices. (In the latter case there was a Leu-Leu collision in the parallel axis orientation that could be relieved by a rotation of one side chain into another allowed conformation. Aside from this we made no deliberate changes in the side chain conformations from those of the native state.)

In the region between ~7 Å displacement from the native separation down to ~2 Å displacement, hydrophobic forces dominate the interaction, and the parallel-axis state would seem to be more stable than the native. It is clear why this is so: the angles between the helix axes in the native state are in the range $\Omega \simeq -60°$ for both B/E and B/G pairs; thus the helices cross obliquely. In the parallel-axis state, many more residues are involved in the occluding surfaces of the pair of helices.

At ~2 Å displacement, the nonbonded energy begins to change very rapidly, as the side chains begin to encounter each other and to interact strongly. Now the native state shows a much lower energy than the parallel-axis conformation.

As noted by Richards and Richmond (1978), the first burying of surface from water begins when the helices are 6–7 Å displaced from their native interaxial distance (depending somewhat on the relative orientation.) In this range, the side chains are not in contact: not only is there no steric interference between side chains on different helices, but even the attractive component of the nonbonded energy has not changed from its value for larger separations. Physically, there may well be a difference between monolayers of water trapped between protein surfaces, about to be excluded from the interhelix region of space, and bulk

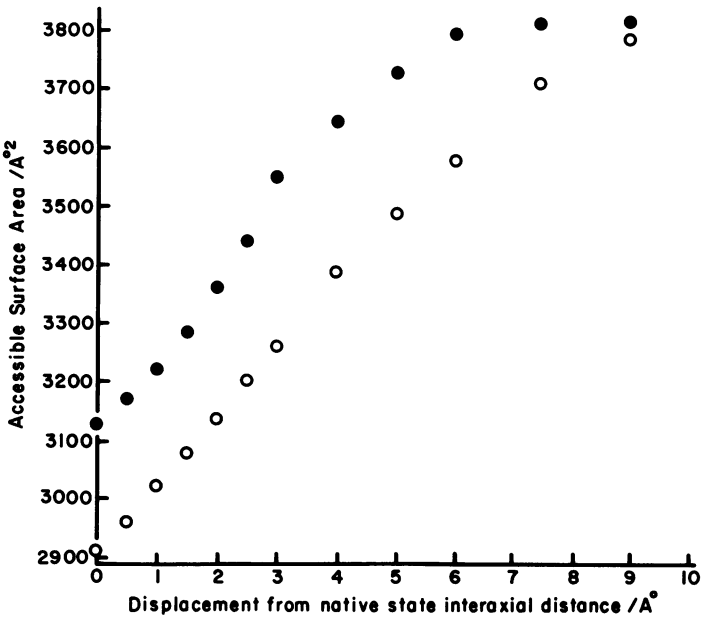


Figure 6 The variation of accessible surface area with inter-helix axis separation for the B and E helices of sperm whale myoglobin.

water. However, these differences are not reflected in our calculations, which report only the accessibility to individual water molecules and ignore water-water interactions.

Not only are hydrophobic forces not specific for the native orientation of the helices, they do not resolve the choice of which portion of the hydrophobic patch on a helix surface will form the interhelix contact in the native state. Fig. 8 shows the distribution of hydrophobic

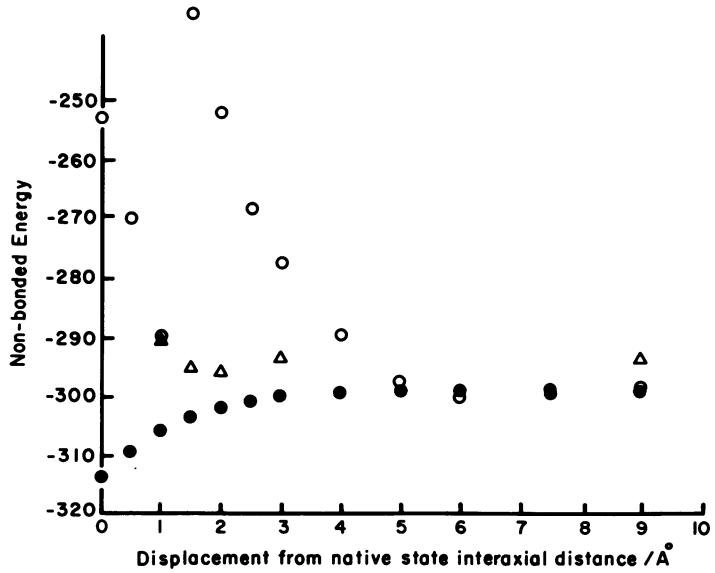


Figure 7 Variation of nonbonded energy with inter-helix axis separation for the B and E helices of sperm whale myoglobin. ● = native structure, O = helices parallel, Δ = helices parallel, collision relieved by change of side chain conformation.

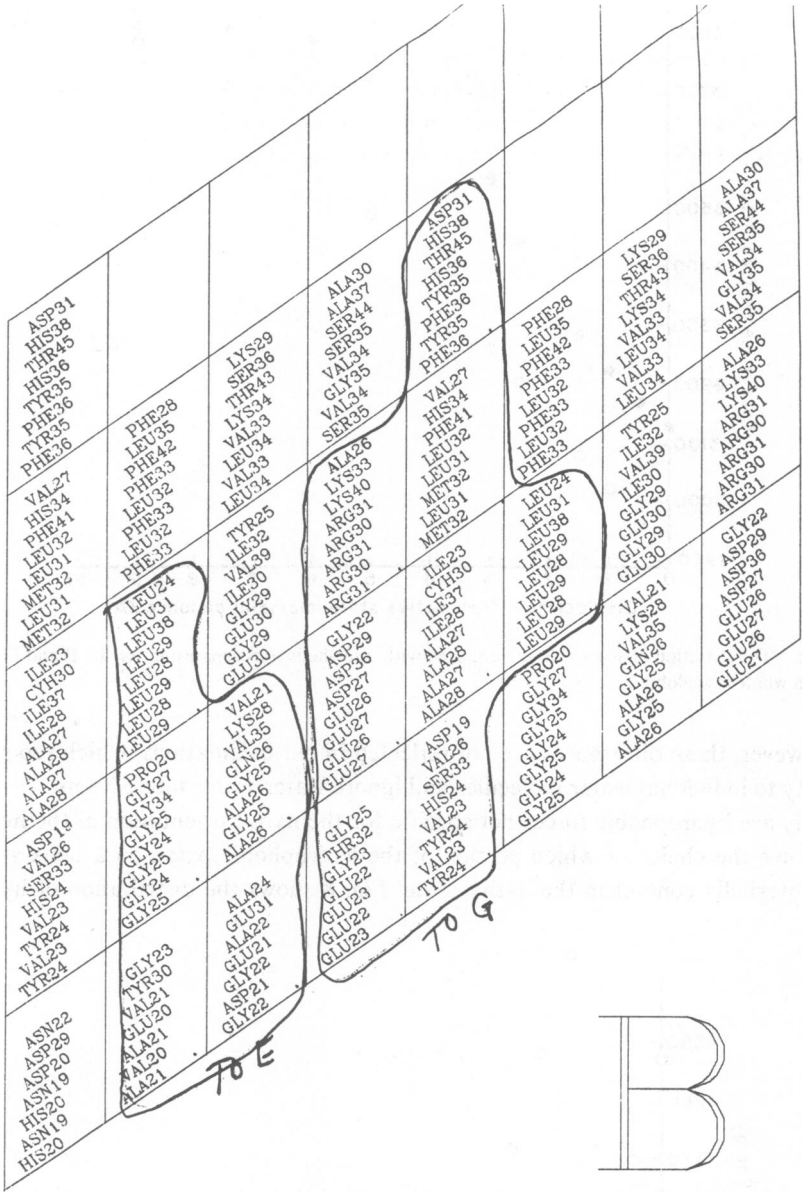


Figure 8 The distribution of amino acids on the surface of the B helix of globins. The surface has been unrolled through 720°, giving a view of all portions of the surface. The hydrophobic patches that make contact with the E and G helices are outlined.

residues on the surface of the B helix in sperm whale myoglobin. In the native structure, this helix forms important contacts with the E and the G helix. Different portions of the hydrophobic patch participate in the B/E contacts in the native structure, and these contacts depend on the precise surface topography of the areas within the hydrophobic patch. But there does not appear to be any mechanism whereby the correct portion of the patch could be chosen by hydrophobic forces alone in a preliminary collision between two helices at the relatively distant separations in which hydrophobic forces dominate the energy of interaction.

The B/E and B/G contacts are typical of helix contacts in globins and in other proteins, except that they are somewhat larger than average. (The mean size of the region of a helix surface that interacts with another secondary-structural element is 4.7 residues.) On the average, one-quarter of the amino acids that form a helix contact are charged or polar. The charged distal end of the side chain of an asp or glu residue at the periphery of a contact may protrude into the solvent, while the proximal, hydrocarbon, portion packs against other side chains.

## CONCLUSIONS

We have examined some of the implications of a model for protein folding that postulates the formation and encounter of units of secondary structure to form productive intermediates. We make the assumption that the forces that stabilize the intermediates are similar in nature to the forces that stabilize the final native structure. We find that:

(*a*) Hydrophobicity, although a driving force for the creation of productive intermediates, is nonspecific with respect to both the mutual orientation of a pair of helices and the selection of the proper portion of the total hydrophobic surface patch that will form the contact in the native structure.

Analysis of the burying of surface area suggests that hydrophobic forces could account for the considerable—although marginal and transient—stability of such intermediates. This result, in connection with the different dependence on helix separation of hydrophobic and van der Waals forces, suggests how two helices might have the opportunity to explore each others' surfaces to find the good fit required to form the native state. Hydrophobic forces might keep the two helices in the same vicinity (to within their range of ~8 Å displacement from native separations) while the van der Waals, forces, which become active at separations of 2 Å or less from the final state, determine the final geometry.

We express the caution, however, that we need a clearer understanding of what is physically happening in the region in which bulk water but not all individual water molecules is excluded from the interface between two approaching secondary structures.

(*b*) The greater stability of the native geometry depends critically on the precise shapes of the surfaces of secondary structures in contact.

We have found that in the globins, the pathway of evolution has been constrained by the necessity to maintain close-packed interfaces between helices (Lesk and Chothia, 1980). This suggests another caution, that the study of the details of the surface-surface recognition process in the region in which the secondary structural elements are in intimate contact may require an explicit treatment of the dynamics of the side chains with all atoms, or at least nonhydrogen atoms, present. Simplifications that are attractive may not reproduce certain essential features of this, the final stage of the unfolding process.

This work was supported by grants to Dr. Peter Pauling from the U. S. National Institute of General Medical Sciences, 1-R01-GM25435-01 and the Science Research Council, and to A. Lesk from the National Science Foundation, DCR 74-24390.

*Received for publication 7 December 1979.*

## REFERENCES

Baldwin, R. L. 1978. The pathway of protein folding. *Trends Biochem. Sci.* 3:66–68.

Baldwin, R. L. 1979. The mechanism of folding of ribonucleases A and S. *Proceedings of the Regensburg Conference.* In press.

Chothia, C. 1976. The nature of the accessible and buried surfaces in proteins. *J. Mol. Biol.* **105**:1–14.
Chothia, C., M. Levitt, and D. Richardson. 1977. Structure of proteins: packing of $\alpha$-helices and pleated sheets. *Proc. Natl. Acad. Sci. U.S.A.* **74**:4130–4134.
Creighton, T. E. 1978. Experimental studies of protein folding and unfolding. *Prog. Biophys. Mol. Biol.* **33**:231–297.
Karplus, M., and D. L. Weaver. 1976. Protein-folding dynamics. *Nature (Lond.).* **260**:404–406.
Lee, B., and F. M. Richards. 1971. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**: 379–400.
Lesk, A. M., and C. Chothia. 1980. How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *J. Mol. Biol.* **136**:225–270.
Levitt, M. 1974. Energy refinement of hen egg-white lysozyme. *J. Mol. Biol.* **82**:393–420.
Privalov, P. L., and N. N. Khechinashvili. 1974. A thermodynamic approach to the problem of stabilization of globular protein structure, a calorimetric study. *J. Mol. Biol.* **86**:665–684.
Richards, F. M., and T. Richmond. 1978. Solvents, interfaces and protein structures. *Ciba Found. Symp.* **60**:23–37.

# DISCUSSION

*Session Chairman*: V. Adrian Parsegian *Scribe*: James B. Matthew

PARSEGIAN: We will start with a question from an anonymous referee. "Your paper attempts to give theoretical support to the hypothesis of formation of secondary structures as intermediates in the folding process. You examine this by calculating the hydrophobic interaction between helices using the surface energy approach as a function of distance in order to obtain an energy that decreases with distance. I believe that the same results would have been obtained for the distance dependence of any two compact substructures not necessarily helices".

LESK: The referee is correct. We are not trying to demonstrate that folding intermediates are helices on the basis of energetic calculations. As any neutral species come together form distant separations there will be forces of attraction which increase in strength. What is really bothering us is why the tertiary structural interactions seem to require the surface topography of the helix.

ACKERS: I have a comment on your evaluation of the energetics of hydrophobic solvent accessibility effects. The value that is used for assessing the free energy of solvent effects in your paper and earlier related work is 22–25 cal/Å$^2$, which is derived from the correlations of free energy of transfer of amino acid side chains from $H_2O$ to ethanol (Nozaki and Tanford). The available exposed area has been very difficult to evaluate in any large protein system. In the past it has been impossible to get very accurate experimental data which permit one to tell whether these values are meaningful and whether the inside of the protein is like ethanol or not.

Recently it has been possible to measure accurately values for self assembly of proteins. In the cases of assembly of hemoglobin, dimer into tetramers, one would estimate from solvent accessibility data that energy of stabilization of deoxy Hb, which has the larger buried surface area, should be ~20 kcal/mol greater than the oxy form; but the experimental values are only 5–6 kcal/mol. So, there is a discrepancy of about a factor of 4 in those energies. Two possible ways of explaining this are: (1) the dominant factors in that stability are not the hydrophobic stabilization effects which Chothia et al. have proposed (in fact this turns out to be the case if you look more deeply into the pattern of thermodynamic effects); (2) the scaling factor which you use and which has been used in the past may be quite wrong.

My question is, if the scaling factor is incorrect by a factor of four, would this alter any conclusions regarding the general roles of the hydrophobic effect in relation to the other effects? I also have a question for Dr. B. K. Lee with respect to this. Is his value for the constant relating the free energy of cavity surface change similar to the value one gets from looking at ethanol-water transfer energies or is it quite different?

B. K. LEE: The value I discussed earlier is ~75 cal/Å$^2$ or so. The difference comes from neglecting the enthalpy term in the interaction between the apolar group and solvent.

LESK: The most important aspects of our results are qualitative and do not depend on the numerical details of the relationship between buried surface area and changes in thermodynamic properties. To respond to Dr. Ackers: (*a*) We wished to consider how plausible it is to assume that intermediates in folding have secondary and tertiary structures similar to portions of the final, native structure. We showed that as two helices approach each other from large separations, cases with the relative orientation with axes parallel show greater burial of surface than those in which the relative helical axis orientation is skewed as in the native state. We infer that at such separations the