PROTEINS: Structure, Function, and Genetics 11:281–296 (1991)

Protein Folding and Association: Insights From the Interfacial and Thermodynamic Properties of Hydrocarbons

Anthony Nicholls, Kim A. Sharp, and Barry Honig
Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032

ABSTRACT **We demonstrate in this work that the surface tension, water-organic solvent, transfer-free energies and the thermodynamics of melting of linear alkanes provide fundamental insights into the nonpolar driving forces for protein folding and protein binding reactions. We first develop a model for the curvature dependence of the hydrophobic effect and find that the macroscopic concept of interfacial free energy is applicable at the molecular level. Application of a well-known relationship involving surface tension and adhesion energies reveals that dispersion forces play little or no net role in hydrophobic interactions; rather, the standard model of disruption of water structure (entropically driven at 25°C) is correct. The hydrophobic interaction is found, in agreement with the classical picture, to provide a major driving force for protein folding. Analysis of the melting behavior of hydrocarbons reveals that close packing of the protein interior makes only a small free energy contribution to folding because the enthalpic gain resulting from increased dispersion interactions (relative to the liquid) is countered by the freezing of side chain motion. The identical effect should occur in association reactions, which may provide an enormous simplification in the evaluation of binding energies. Protein binding reactions, even between nearly planar or concave/convex interfaces, are found to have effective hydrophobicities considerably smaller than the prediction based on macroscopic surface tension. This is due to the formation of a concave collar region that usually accompanies complex formation. This effect may preclude the formation of complexes between convex surfaces.**

Key words: hydrophobicity, surface tension, protein folding, protein association

INTRODUCTION

For many years hydrophobicity has been assumed to be the major interaction stabilizing globular proteins as well as protein-protein and protein-ligand association and binding. The hydrophobic effect was thought to result from the entropically unfavorable ordering of water molecules around nonpolar solutes. In recent years both of these widely accepted hypotheses have been questioned, based in large part on studies of the temperature dependence of hydrocarbon solubility in water and on corresponding studies of protein unfolding.[1,2] For example, the attraction of hydrocarbons to one another in the aqueous phase has been attributed to dispersion attractions between hydrocarbons rather than to the standard explanation, which is based on the disruption of bulk water structure and dynamics.[1,3] Similarly, enhanced packing (dispersion) interactions in the solidlike protein interior have been proposed as a major driving force in protein folding. Thus, although by no means abandoned, the classical picture of protein folding as resulting from the entropically driven reduction in nonpolar surface area seems far less secure than it was a few years ago. Nevertheless, one of the central conclusions of this study is that many elements of the classical picture are correct.

In order to evaluate the relative free energy contributions of different forces to protein stability and association, we define a hypothetical unfolding pathway in which these interactions are modified as folding proceeds. The folded protein itself is described as a closely packed solidlike entity.[4,5] In general, the protein will contain cavities that may be empty or contain water. In the first step of unfolding, the solid core is assumed to melt into a liquidlike state. Packing interactions are weakened in this process and, concomitantly, the mobility of the amino acid side chains is increased. This may also be accompanied by relief of strain energy if the native conformations of side chains are distorted to optimize packing in the fully folded state. In the second step the liquidlike interior unfolds, thus creating an interface between nonpolar groups and water. The formation of the interface can be described in terms of two types of interactions involving: (1) packing of

Received May 17, 1991; revision accepted July 9, 1991.
Address reprint requests to Barry Honig, Dept. of Biochemistry and Molecular Biophysics, Columbia University, 630 W. 168th St., New York, NY 10032.

nonpolar groups against water, i.e., dispersion interactions; and (2) perturbation of the structure of water at the interface, i.e., the classical model for hydrophobicity.

The pathway defined above divides the unfolding process into a melting step and a step involving changes in interfacial interactions between amino acid side chains and water. In addition, if the interior of the protein is less than optimally packed, i.e., contains cavities, the energetics of cavity formation in organic solids becomes relevant. A considerable amount of data is available on the thermodynamics of hydrocarbon melting and on hydrocarbon/water interfacial free energies. In this study we show that these can be used to provide insights into the energetic basis of protein folding and association.

Interfacial free energies are generally viewed as macroscopic properties, and the validity of applying them to individual molecules has been a question of some concern. Nevertheless, surface area—hydrophobicity relationships, which are a measure of oil/water interfacial free energy, have been widely used in studies of biological macromolecules. In two recent publications,[6,7] we arrived at a new estimate of the interfacial free energy derived from solubility data of about 47 cal/mole/Å^2. It was found that much of the discrepancy between this value and that of 72 cal/mole/Å^2, which corresponds to macroscopic oil/water surface tension,[8] could be accounted for with a simple model for the curvature dependence of the hydrophobic effect.[6] The fact that the microscopic and macroscopic values can be reconciled suggests that the same basic interactions are dominant in both size domains and, consequently, that hydrophobic interfacial free energy is a valid concept at the molecular level. In this work we extend the model to provide a detailed description of the dependence of hydrophobic interfacial free energy on local curvature. Application to the association and binding reactions of real proteins reveals new insights as to the factors that lead to the formation of stable complexes.

THEORY AND METHODS

An Interfacial Tension Model for the Hydrophobic Effect

We first note that there is a precedent for treating hydrophobicity as a form of surface tension,[9,10] and that surface tension is expected to vary with curvature. Indeed, Tolman[11] used thermodynamic arguments to derive the following simple expression for the microscopic surface tension of a surface with curvature (1/R):

$$\gamma(R)/\gamma(\infty) = 1/(1 + 2\delta/R) \quad (1)$$

where $\gamma(R)$ is the surface tension of spherical particle of radius R and $\gamma(\infty)$ is the macroscopic surface tension; δ is a constant of molecular dimensions, which is related to the thickness of the boundary layer. Eyring and coworkers[12] found that the microscopic surface tension of a number of liquids as obtained from Equation (1) predicted solubilities of various gases that were in good agreement with exprimental data. This result demonstrates the feasibility of applying the Tolman expression on the scale of individual molecules. It should be pointed out in this regard that Sinanoglu[13] demonstrated some time ago that the concept of surface tension could be extended down to molecular dimensions.

Equation (1) in principle provides a means of obtaining hydrophobicities as a function of curvature, but two difficulties are encountered. First, the constant δ is not well defined, and, second, the Tolman correction is applicable to spherical interfaces, but not to complex interfacial geometries formed by proteins. In order to derive a well-defined model that can be applied to complex systems, we introduce the postulate that the free energy contribution associated with a single water molecule is related to the number of other water molecules that can be packed around it. An immediate consequence is that the energy of a water molecule near a nonpolar surface increases relative to that in bulk because other water molecules are excluded by the nonpolar surface. To complete the model we need a way to quantitate the number of water molecules that can be packed around a particular surface. For this we use the concept of accessible area, defined as the locus of the centre of water probe sphere in contact with the surface.[14] Accessible area has been widely used as an extensive measure of the hydrophobic effect and the values for the *microscopic* surface tension given above relate to accessible area. At this point no further information as to water structure, hydrogen bonding, etc., is introduced into the model. Its validity is considered further in the discussion.

The Microscopic Surface Tension of a Uniformly Curved Surface

We illustrate in Figure 1a,b why one might expect the surface tension at a uniformly curved surface to be different than that of a smooth plane. Clearly a water molecule in contact with the surface of a plane is able to interact with fewer neighboring water molecules than if it is in contact with a convex spherical surface. Stated another way, a probe water molecule has a smaller accessible area when in contact with a plane than when in contact with a sphere. If we make the simple assumption that hydrophobicity is related to the free energy penalty associated with reducing the contact area between water molecules, the expectation is that planes are more hydrophobic than spheres. More generally, the hydrophobicity at a particular point on a nonpolar surface (the "local" hydrophobicity) is expected to be proportional to the accessible area of a probe water molecule lost due to the surface.

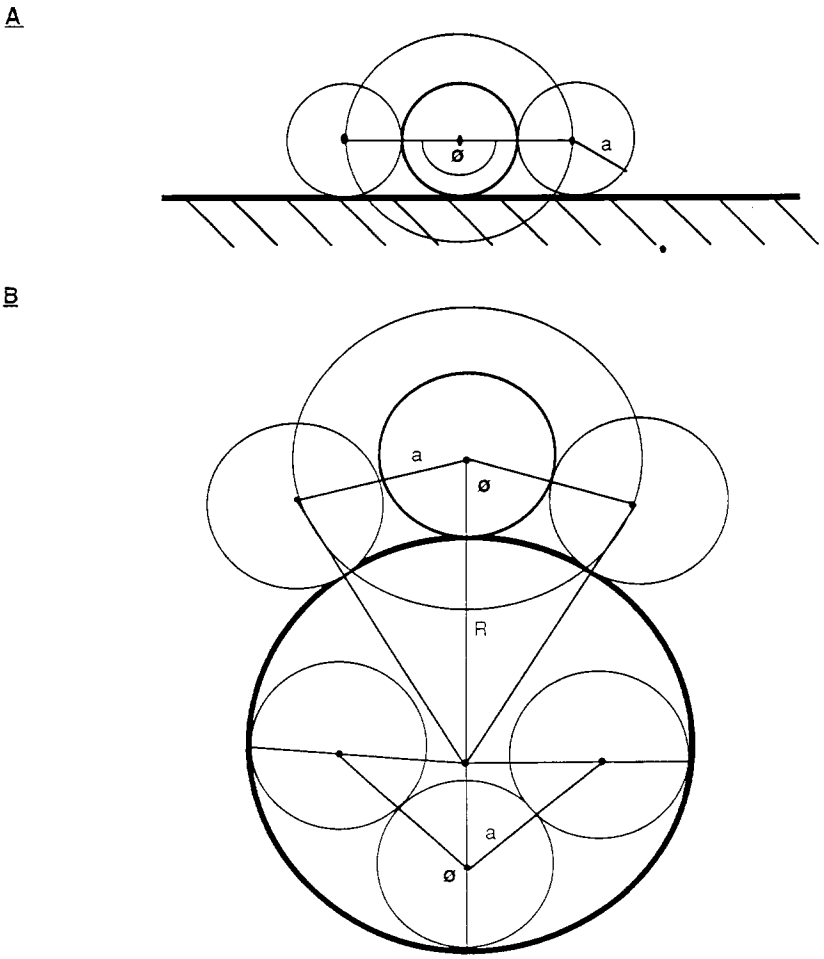

Fig. 1. Dependence of inaccessible area on curvature. (**A**) Illustrates that the accessible area of a water molecule pressed up against a plane is $2\pi a^2$. (**B**) Shows that the same water up against a spherical surface will have either more or less accessible area depending on whether the surface is convex or concave. The inaccessible solid angle ϕ is shown and by simple geometry turns out to be $2\pi(1+a/R)$ or $2\pi(1-a/R)$ for the convex or concave case, respectively.

In order to obtain a quantitative measure of the local hydrophobicity, we place the center of a probe water molecule at a point on the solvent accessible surface of the nonpolar solute. The local hydrophobicity at that point is proportional to the solid angle on the accessible surface of the *probe water molecule* removed by the nonpolar surface. We use the accessible surface of the probe water since this represents the locus of possible positions of other water molecules in contact with it. For a water molecule in bulk, 4π radians are accessible so that the solid angle removed is zero. For a plane the value is just 2π. Figure 1b shows that for a convex surface, with curvature (1/R), the solid angle excluded is $2\pi/(1+a/R)$, where a is the radius of the water molecule. If we assume a linear relationship between the angle excluded and the energy penalty for a water molecule up against the oil/water interface, then we arrive at

$$\gamma(R)/\gamma(\infty) = 1/(1 + a/R). \quad (2)$$

This expression has the same form as the Tolman equation above, except that 2δ has been replaced by a, the radius of water. It is striking that such a similar form can be reached via two very different derivations. The immediate advantage of our approach is that there are no unknown constants. Furthermore, one is not limited to convex surfaces: the same principle holds with concave surfaces, as is shown in Figure 1b. Here one obtains:

$$\gamma(R)/\gamma(\infty) = 1/(1 - a/R) \quad (3)$$

i.e., one merely inverts the sign of R, as one might expect for a negatively curved surface. We note that for concave surfaces we would predict an *increase* in surface tension relative to a plane. Also, when the radius of curvature of the concave surface is less than 2a, all 4π solid angle of the water's environment are excluded, i.e., the maximum surface tension for a concave surface would be predicted to be twice that of the planar surface.

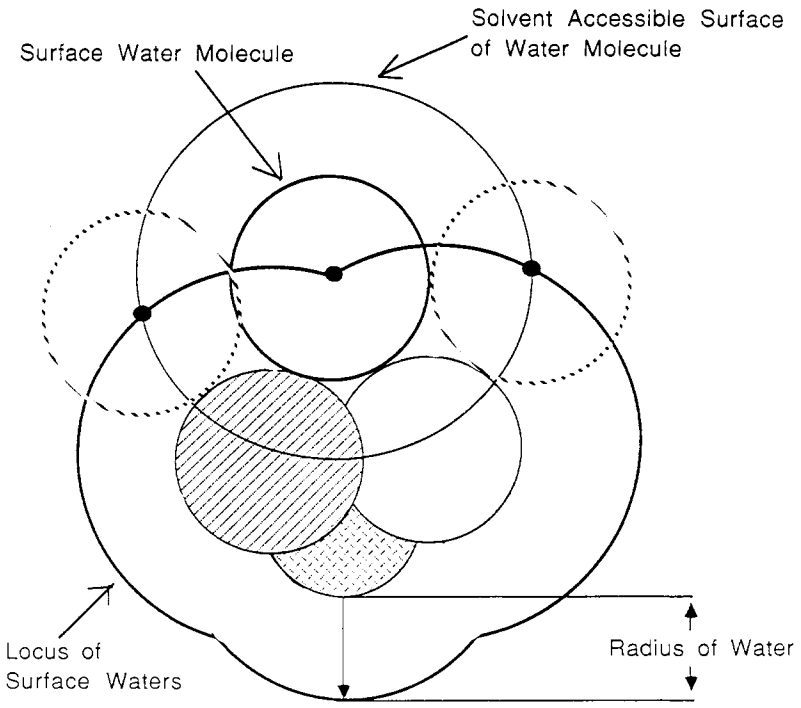

Fig. 2. Calculation of the inaccessible solid angle at any point on an arbitarily shaped accessible surface. A test sphere twice the radius of water is placed at each point on the accessible surface and points on that test sphere are then determined to be inside or outside the accessible volume. The ratio of points inside to the total number on the test sphere, multiplied by 4π, is the inaccessible solid angle.

The Microscopic Surface Tension of an Arbitarily Shaped Surface

Equation (2) was derived by assuming that the energy penalty for a water molecule in contact with a nonpolar sphere was proportional to the solid angle excluded by the sphere. With this assumption the generalization to a nonspherical surface is simply to calculate the excluded solid angle, ϕ_i, at each point, i, on the accessible surface. This can be used to define a curvature correction factor (relative to a planar surface), c_i, at each point that is given by $c_i = \phi_i/2\pi$. Figure 2 illustrates this concept in two dimensions. The local surface tension is then given by $\gamma_i = \gamma(\infty)c_i$. Note that c_i is a measure of purely local curvature for simple surfaces, but in general it depend on nonlocal interactions (up to distances of two water diameters, see below).

We can calculate a free energy of transfer of a hydrocarbon solute from pure hydrocarbon liquid to water, ΔG_{ow}, (ow denotes oil to water) starting with the macroscopic hydrocarbon oil-water surface tension, $\gamma_{ow}(\infty)$, as follows. The average curvature of the solute, C, is determined by calculating the curvature factor, c_i, at sample points distributed over the accessible surface of a hydrocarbon solute, giving:

$$C = \Sigma A_i c_i / A \tag{4}$$

where A is the total accessible area of the molecule and A_i is the area of the surface element associated with each sample point. The transfer energy is now:

$$\Delta G_{ow} = \gamma_{ow}(\infty)AC \tag{5}$$

Numerical Methods

The algorithm used to calculate the curvature correction over the surface of a molecule is a variant on that developed by Shrake and Rupley[15] to calculate accessible area. The latter places at each atom a set of points to represent the surface of a sphere of radius equal to that of water plus that of the atom. Then each of these points is tested to determine whether they lie within all similar spheres centered at other atoms. Those that do not represent the solvent accessible area for that atom. The procedure is repeated for each atom to give the total accessible area. Our variation, as illustrated in Figure 2, is to take each accessible surface point, place a set of points to represent a sphere of radius twice that of water about it, and determine what fraction of such points lie inside all atoms expanded by the radius of water. This fraction multiplied by 4π is the inaccessible solid angle for that point on the accessible surface, and the inaccessible angle divided by 2π is the c_i discussed above.

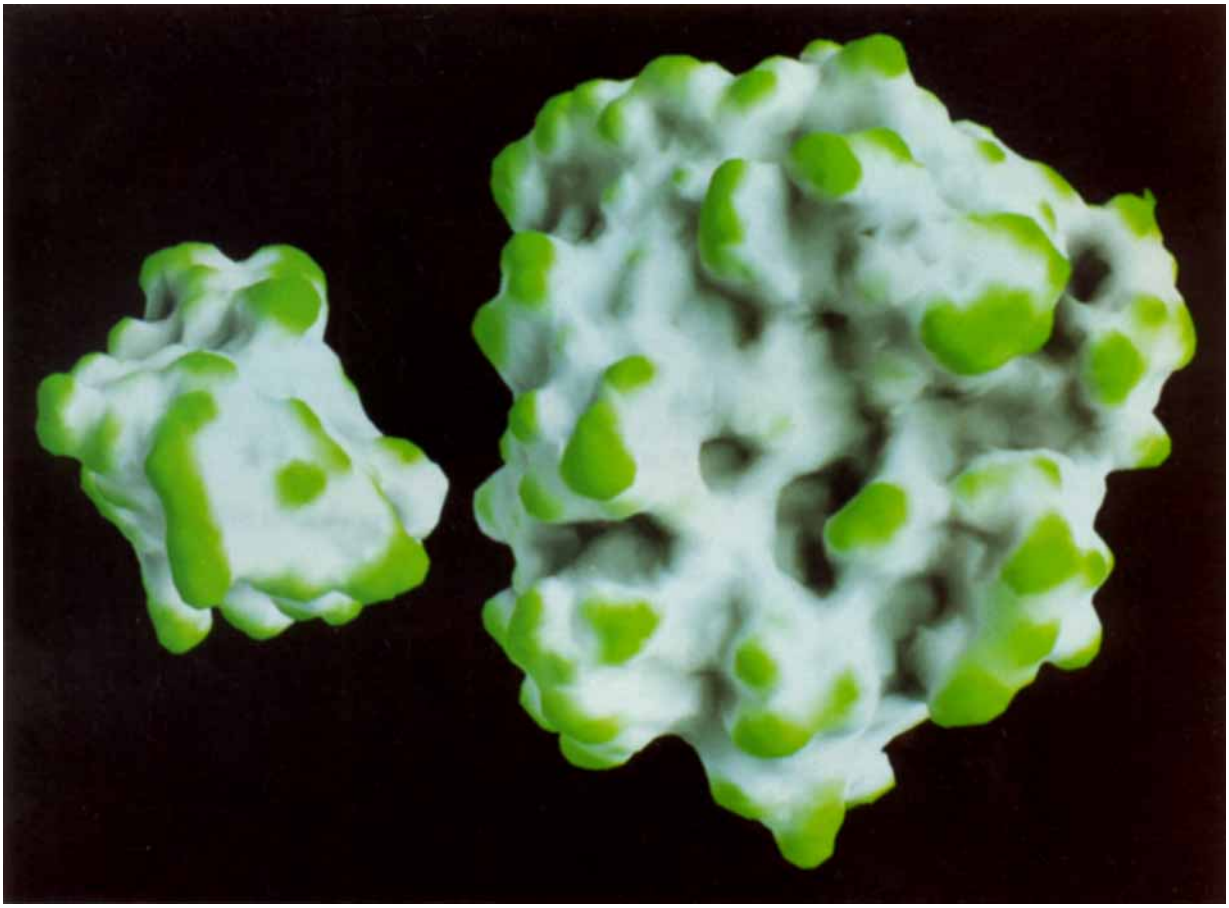

Fig. 3. The molecular surface of Trypsin and PTI. The surface is color coded such that the most convex part of the surfaces is green, the most concave is grey, and planar parts are white. Colors are linearly interpolated for other values of curvature. The most concave part of the Trypsin surface (right) corresponds upon binding to the most convex part of the PTI surface (left) shown.

We have spent some time modifying the basic approach outlined above to improve the efficiency of the algorithm, details to be published elsewhere. At present, running on a Silicon Graphics 220, the procedure calculates the curvature correction of trypsin in about six minutes, using a dot density of 642 per test sphere.

In addition to quantitative use, the curvature corrections so calculated can also be used qualitatively in graphical display. In this we note a similarity to the approach taken by Connolly.[16] Each point on the molecular surface[17] is uniquely associated with a point on the accessible surface for which a curvature correction can be calculated. Thus it is possible to map this quantity, which as we have seen is like a local hydrophobicity, on to the molecular surface. We have developed a program for the Silicon Graphics IRIS, (Nicholls and Honig, unpublished), which can rapidly generate and display a solid rendering of the molecular surface, such that this local hydrophobicity can be illustrated via color coding.

Figure 3 is an example of this approach for the binding surfaces of Trypsin and and Protein Trypsin Inhibitor (PTI). White represents the base hydrophobicity, i.e., of a plane with an excluded solid angle of 2π. Green and grey represent surfaces with less than, or greater than, 2π steradians excluded respectively. Green tends to correspond to convex parts of the surface and gray to concave, but the correspondence is not exact since the excluded solid angle at any point is a function not just of the infinitesimal curvature at that point, but, as noted above, it also can depend on the presence of surface up to two water diameters away.

Hydrocarbon Thermodynamic Data

Accessible surface areas, corrected liquid to water, and gas to water transfer energies for n-alkanes were taken from Sharp et al.[7] Liquid alkane-water surface tensions were taken from Aveyard and Haydon.[8] Liquid alkane-vapor surface tensions and alkane melting enthalpies were taken from the CRC

Handbook of Chemistry and Physics.[18] Alkane vaporization free energies were taken from Ben-Naim and Marcus.[19]

RESULTS AND DISCUSSION

Bridging Microscopic and Macroscopic Hydrophobicity

As discussed above and in a previous work,[6] the value for the microscopic hydrocarbon/water surface tension based on alkane solubility data is about 47 cal/mol$Å^2$. This is still significantly less than the macroscopic value of 72 cal/mol/$Å^2$. In order to determine whether curvature effects can account for the difference, we have used alkane solubility data together with our local curvature model to obtain a value for the macroscopic surface tension, $\gamma(\infty)$. That is, we are attempting to predict a macroscopic number based entirely on microscopic information.

As can be seen in Figure 4, the average curvature correction, C, of linear alkanes ranges from 0.56 for methane, 0.72 for decane, to a limiting value of about 0.8 for a CH_2 group in longer alkanes (the decane value includes the effects of the terminal methyl groups and is thus smaller than the value derived from the CH_2 groups alone). Equation (5) was used in conjunction with these curvatures and the corrected alkane water/transfer free energies and surface areas reported in Table 1 of Sharp et al.[7] to yield a prediction for $\gamma(\infty)$ based on each alkane.

The predicted value for the macroscopic surface tension for the longer alkanes is about 59 cal/mole/$Å^2$. To make a meaningful comparison with the macroscopic measurement of surface tension, two further corrections must be applied. First, the microscopic surface tension is defined here in terms of accessible surface area, whereas the macroscopic surface tension refers to a smooth, or projected planar surface area. If we approximate the hydrocarbon side of a planar macroscopic oil-water interface as a close-packed array of 2Å radius spheres (a reasonable radius for a methyl group), the accessible surface area is about 20% larger than the projected area (see Fig. 5). The effect is to increase the predicted surface tension by 20%. Second, since the interface between two liquids is not mechanically constrained, unlike the solute/water interface, the former will experience thermal fluctuations, or capillary waves. Whereas such fluctulations increase the total surface area at the interface, they also contribute entropically. The net effect, of course, is to reduce the interfacial free energy, otherwise such fluctuations would not persist. The theory of Buff et al.[20] gives an estimate of a 5% reduction in free energy due to capillary wave fluctuations. Applying both the roughness and fluctuation, we arrive at 67 cal/mole/$Å^2$ as the value of the macroscopic surface tension predicted from alkane solubility data. This number is in excellent agreement with the experimentally determined macroscopic value of 72cal/mole/$Å^2$, as shown in Figure 4.

It is, of course, possible that the agreement is fortuitous, but at this stage it does seem reasonable to suggest that the geometric model proposed here provides a good approximation to hydrophobic surface tension on both a microscopic and macroscopic scale. Support for the model rests upon its closeness in form to Tolman's thermodynamically derived relationship and in its success in reproducing the predictions of scaled particle theory.[6] The existence of a model that successfully bridges the macroscopic and microscopic regimes strongly indicates that interfacial hydrophobic free energies may be applied with some confidence on the dimensions of individual molecules, once curvature effects are properly described.

Transferring Amino-Acid Side Chains Into Water

In this section we discuss the second step in our hypothetical unfolding pathway, the transfer of amino acid side chains from the liquidlike protein interior into water. The interfacial model for hydrophobicity discussed so far appears applicable to this step, i.e., the process can be described in terms of changes in oil/water surface tension or, alternatively, in terms of transfer free energies of hydrophobic groups. Table I reports the transfer free energies of alkanes between different phases. The contribution for the liquid to water transfer is found to be about 1.38 kcal/mole per CH_2 group. This rather large value would appear to provide the dominant free energy contribution of nonpolar amino acids to protein folding. This contention is based in part on the agreement between changes in protein stability and changes in solvent transfer free energy for point mutations.[7] Additional support is provided by the small contributions available from the melting/freezing process as described in the next section.

We now consider the relationship between dispersion and packing interactions and the phenomenon of hydrophobicity. The surface tension between immiscible liquids such as a liquid alkane (1), and water (2), γ_{12}, can be used to obtain information about the forces acting between the molecules through the well-known relation:[21]

$$\gamma_{12} = \gamma_1 + \gamma_2 - W_{12} \quad (6)$$

where $\gamma 1$ and $\gamma 2$ are the surface tensions of the alkane and water at the vapor interface respectively, and W_{12} is the work of adhesion between alkane and liquid. Equation (6) states that the energy of forming an interface between two liquids is the sum of the energies needed to form an interface between each liquid and vapor, plus the work of bringing two completely separated such surfaces into contact. Typical values for a liquid water/alkane system are

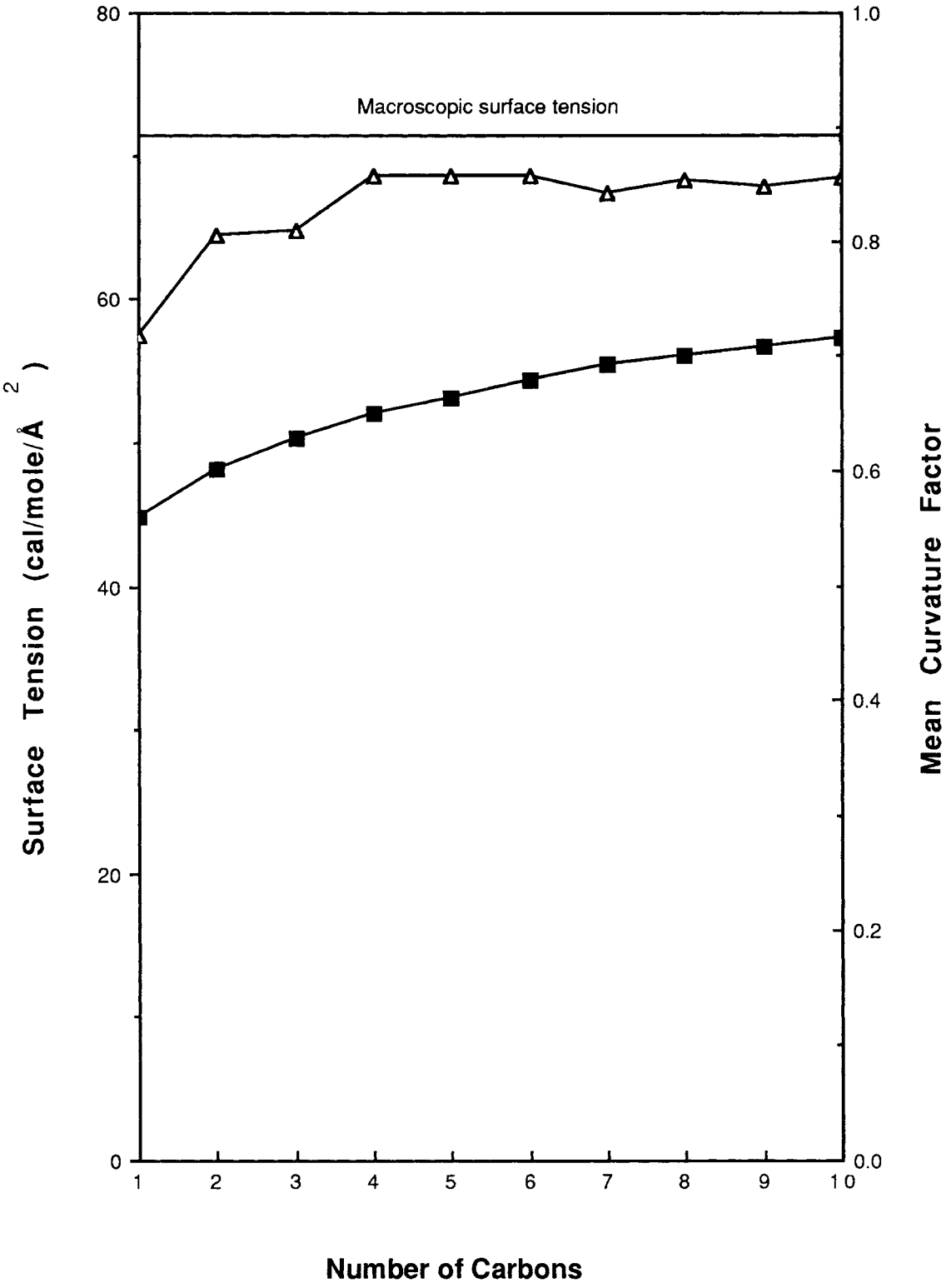

Fig. 4. Prediction of the macroscopic liquid alkane surface tension from alkane solubility data. (■); Mean curvature, calculated using Equation (4). (△): The curvature corrected interfacial energy. Experimental value of 72 cal/mole/Å^2 is indicated on the figure.

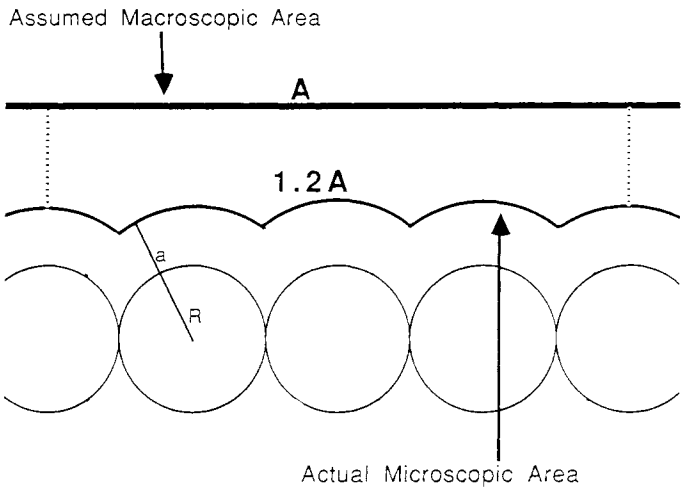

Fig. 5. The "roughness" of an oil-water interface. The upper area (**A**) corresponds to the macroscopic-surface area, while the lower microscopic accessible area is larger. Calculations done with spheres of 2 Å radius in close contact, and a water probe radius of 1.4 Å, suggest a factor of 1.2 between the two areas.

TABLE I. Phase Transfer Free Energies for n-Alkanes*

n	Area†(Å²)	L− > W†	V− > W†	L− > V‡	S− > V§
1	130	3.29	2.63	0.66	0.15
2	163	4.72	3.34	1.38	−0.21
3	194	6.01	4.07	1.94	0.03
4	224	7.53	4.89	2.64	1.29
5	252	8.97	5.57	3.40	1.23
6	282	10.35	6.29	4.05	1.95
7	311	11.59	6.90	4.69	2.55
8	340	13.02	7.70	5.33	3.93
9	369	14.36	8.38	5.98	4.21
10	399	15.73	9.10	6.63	5.08
	slope"	1.38	0.72	0.66	0.6
	R^2	0.999	0.999	0.999	0.943

*Energies are in kcal/mole, n: number of carbon atoms, L: liquid alkane, W: water, V: vapor, S: solid alkane, arrow indicates direction of transfer.
†Data taken from Sharp et al.[7]
‡Data taken from Ben Naim and Marcus.[29]
§Sum of the liquid to vapor transfer energy, and solid to liquid transfer data from Figure 10.
"Slope is best fit value per CH_2 group from linear regression, with correlation coefficient R^2.

γ_1 = 30 cal/moleÅ², γ_2 = 104 cal/mole/Å² [18] and γ_{12} = 72 cal/mole/Å² [8], which gives an alkane/water adhesion energy of 62 cal/mole/Å². The equivalent of Equation (6) for the alkane/alkane adhesion energy ($W_{11} = 2\gamma_1$ since, $\gamma_1 = \gamma_2$ and $\gamma_{11} = 0$), gives 60 cal/mole/Å², i.e., almost the same. In other words, the attractive forces between alkanes, per unit area, are about equal to those between alkane and water. Since the only interactions alkanes can effectively make, either with themselves or with water alkanes, are dispersion interactions, the important consequence is that *dispersion forces make no net contribution to the hydrophobic effect for liquid hydrocarbons.* It follows that they make little or no contribution to the liquid to water step in the unfolding pathway.

If dispersion effects make no net contribution to hydrophobicity, one is left with the classical model of an entropically driven process involving the ordering of water molecules. The major argument against this model has been that hydrophobicity increases with temperature and reaches a maximum when it is enthalpically rather than entropically driven. However, this difficulty can be easily resolved by invoking the phenomenon of entropy/enthalpy compensation.[6] Water molecules can adapt to the presence of a nonpolar interface either by sacrificing a hydrogen bond and thus increasing entropy, or by retaining hydrogen bonds at an entropic cost. The latter is the dominant response at 25°C, whereas the former becomes increasingly important as the temperature rises. In either case it is the dis-

ruption of water structure and dynamics that drives hydrophobicity rather than the attractive enthalpic interactions between hydrocarbon molecules.

This type of entropy/enthalpy compensation appears to account for the applicability of our simple geometric model for hydrophobic surface tension on both the microscopic and macroscopic scale. As we have pointed out previously,[6] two lines of evidence might suggest that the model should have failed. First, in contrast with alkane solubility measurements at 25°C, macroscopic surface tension is an enthalpically driven process. Second, consistent with the thermodynamic evidence, computer simulations have found that water molecules around small spherical cavities maintain their hydrogen bonds, whereas one hydrogen bond is sacrificed near planar surfaces. The existence of a smooth *free energy* transition between the microscopic and macroscopic regimes despite inconsistencies between the observed enthalpies and entropies (and underlying water structure and dynamics) provides direct evidence for the compensation phenomenon. We suggest that the same type of compensation occurs as the temperature is increased.

Melting the Solidlike Protein Interior

We have defined the first step in our hypothetical unfolding pathway as the melting of the closely packed protein interior. This event involves a weakening of dispersion interactions and a concomitant increase in side-chain mobility. The thermodynamics of alkane chain melting can provide information on the free energy changes associated with these processes.

In Figure 6 the melting temperature, T_m, density, and heat of fusion, ΔH_m, for alkanes ranging from methane to n-octacosane are plotted against number of carbon atoms. The free energy of melting at 25°C is obtained from relation $\Delta G(25) = \Delta H_m(1-298/(T_m+273))$, and is also plotted on the figure. The free energy varies quite linearly with number of carbons over the entire range, spanning melting points from below −100°C to above 30°C, and for densities ranging from 0.4 to 0.8 g/ml. A linear regression analysis gives values of $\Delta G(25°) = 0.14$ kcal/mole/CH_2, $\Delta H = 0.59$ kcal/mole/CH_2, and $\Delta S_m = 1.5$ cal/mole/°K/CH_2 for alkane melting.

The appreciable enthalpy of melting (≈ 0.59 kcal/mole/CH_2) is presumably due to better packing in the solid than in the liquid state. However, this is largely offset by the entropy increase upon melting ($-T\Delta S \approx -0.45$ kcal/mole/CH_2 at 25°C). The entropy change for alkane melting provides an estimate of the cost of losing side-chain mobility upon packing or association. It is interesting to note that the value of 1.5 eu per rotatable torsion angle is not too different from the figure of 2.2 eu that one estimates from the reduction of three conformers per bond (i.e., g^+, g^-, and t) to one.[22]

Overall, the free energy of alkane melting at room temperature is rather small. The figure of 0.14 kcal/mole/CH_2 should be compared to a value of 1.38 kcal/mole/CH_2 for the hydrophobic effect derived above from Table IV; i.e., packing contributes about 10% of hydrophobicity. It should also be noted that the free energy of melting per methylene group is effectively constant over nearly a doubling of density. *Entropy-enthalpy compensation effects make arguments based on density alone an unreliably guide to the strength of interactions within a material.* It should be emphasized that the enthalpic contribution of dispersion forces to protein folding will be substantial, but since they are offset to a great extent by the freezing of side chain motion, the consequences, for free energy, of close packing should be viewed as a second-order effect. It is, of course, possible that dispersion and chain entropy effects in protein hydrophobic core groups are radically different from those for alkanes, but this seems unlikely due to their nonspecific nature.

Cavity Formation in Proteins

An internal space is defined as a cavity if it is large enough to accommodate the van der Waals sphere of a small atom or water (e.g., O = 1.4Å), and it is not connected to the outside.[23] In spite of their high packing densities, many native proteins appear to have at least one cavity.[23] Site-directed mutagenesis, particularly to make amino acid side chains smaller, may also introduce cavities into proteins.

In order to estimate the energy involved in creating a cavity, we consider the thermodynamics of cavity formation in a nonpolar solid, such as a hydrocarbon. One approach would be to use the hydrocarbon/vacuum surface tension times the appropriate cavity area. The difficult problem of measuring the solid/vacuum surface free energy of a hydrocarbon arises, however. Another approach is to use the free energy of sublimation for a cavity-size hydrocarbon. This is possible because the number of intermolecular contacts lost when a cavity is formed is equal to the number of contacts lost when a molecule vaporizes. From Table I, the sublimation free energy is about .6 kcal/mole per CH_2 group, which provides an estimate for the cost of cavity formation. This number is nearly identical to a previous estimate based on the distribution of empty cavities in proteins.[23]

Changes in Protein Stability Resulting From Amino Acid Substitution

In a recent work we demonstrated that the volume-corrected transfer free energies of hydrophobic amino acids from octanol to water were closely correlated with observed changes in stability resulting from amino acid substitution. This result suggests that treating the protein interior as a liquid hydrocarbon is a reasonable approximation. In this sec-

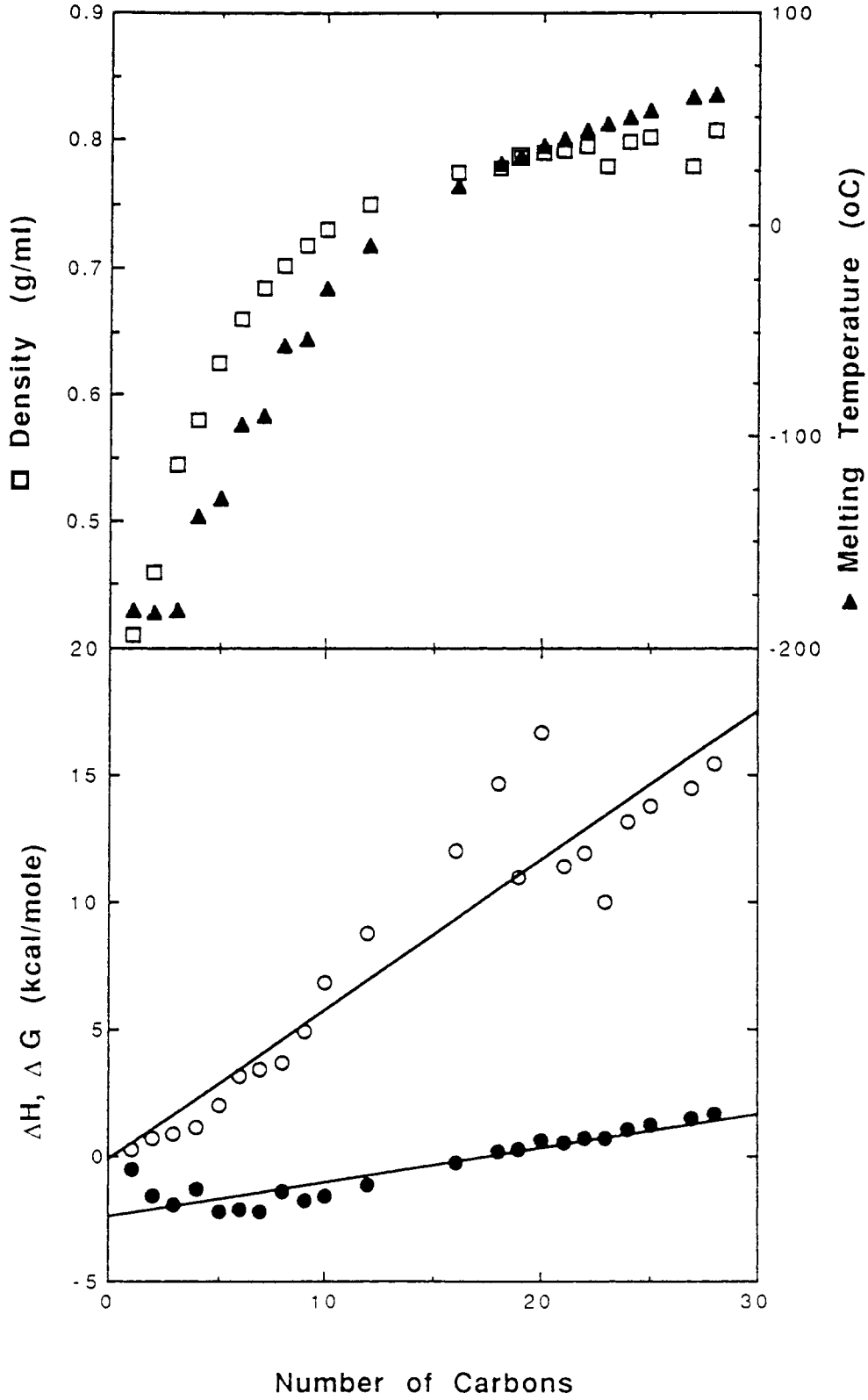

Fig. 6. Thermodynamics of n-alkane chain melting. Upper panel shows melting temperature (▲) and density (□) as a function of chain length. Lower panel shows enthalpy of melting (○) and free energy of melting at 25°C (●).

TABLE II. Comparison of Calculated and Measured Changes in Protein Stability Due to Mutation of Hydrophobic Amino Acids*

Mutant	Calculated change in stability due to: Hydrophobicity	+	Melting	+	Cavity	Number of examples†	Measured change in stability Mean	Min	Max
L− > A	3.87		4.26		6.06	13	3.18	1.6	5.8
I− > A	3.84		4.23		6.03	8	3.63	2.5	5.1
V− > A	2.48		2.74		3.94	13	2.32	0	4.7
L− > V	1.39		1.52		2.12	2	0.85	0.75	0.95
I− > V	1.36		1.49		2.09	9	1.16	0.5	1.8

*Energies are in kcal/mole, where positive values indicate destrabilization of the mutant with respect to the wild type.
†Data taken from Kellis et al.,[24] Shortle et al.,[25] Sandberg and Terwilliger.[28]

tion we consider the effects that were ignored in our previous treatment, the solid to liquid transition and the possible effects of cavities that were formed as a result of the substitution. We consider a wild-type protein in which a mutation is made from a large nonpolar amino acid to a smaller one, e.g., Ile− > Ala.

We first define an unfolding pathway in aqueous solution, keeping account of the *difference* in free energy between wild-type and variant at each step. We treat the general case where the mutation to a smaller residue leaves an empty cavity whose dimension ranges from zero (total repacking) to a size equal to the loss of area and volume that occurs in the unfolded state. It is convenient to define a reference state for the mutant protein in which no cavities have been introduced. If a cavity is left by the mutation, we add in the additional cavition formation energy.

Step 1: *Cavity formation*—The number derived above is .6 kcal/mole per CH_2 group. Step 2. *Melting of the solid protein core*—From Figure 4, the free energy of melting of long hydrocarbons at 25°C is about .15 kcal/mole per CH_2. Step 3. *Unfolding of the liquidlike core*—The octanol-water transfer free energies of amino acids that are used in the analysis are reported in Table 4 of Sharp et al.[7] It is worth noting that the change in free energy per CH_2 group varies from 1.24 to 1.36 kcal/mole, close to the value of 1.38 reported for the alkane/water transfer free energies reported in Table I of this work. The relevant transfer free energies are reported in the column two of Table II. Adding .15 kcal for the melting of each CH_2 group increases the predicted effects on stability for each mutation as shown in the third column of Table II. The value of .6 kcal/mole per CH_2 group derived above for the sublimation free energy provides a further increment to our estimate of the destabilizing effect of particular mutation. (Column 4 of Table II) The use of sublimation free energies assumes implicitly that the opening up of a cavity in a protein frees the motion of side chain rotors, as would occur in the sublimation of a crystal. If side chain motion is still restricted in the vicinity of a protein cavity, enthalpies rather than free energies of sublimation should be used. The enthalpies of sublimation of hydrocarbons are approximately 1.8 Kcal/mol per CH_2. Thus, the maximum predicted effect of a single methylene on protein stability is 3.2 kcal/mole (1.38 kcal/mole-hydrophobic effect, 1.8 kcal/mole-enthalpy of vaporization).

As shown previously and is evident from Table II, the octanol/water transfer free energies with or without the melting correction agree quite well with experimental values for barnase[24] and S. Nuclease.[25] The largest mutational effects may suggest the presence of cavities. Values significantly smaller than predicted might be accounted for by a relief in strain energy brought about by the smaller mutation or by mutations in which the amino acid is not completely buried. In any case, our analysis suggests that a liquid hydrocarbon model for the protein interior is a good first approximation for treating the effects of mutations of hydrophobic amino acids on protein stability. Packing corrections are found to be small, whereas cavity effects can be significant if a large unfilled void is formed by the mutation. However, the range of values obtained in the experiment suggests that partial or complete repacking is the general expectation.

Substrate Binding and Macromolecular Association

In this section we consider how curvature affects hydrophobic contributions to binding energies. In particular it is of interest to determine how curvature varies over the surface of a protein and what effects this variation has on the magnitude of hydrophobic contributions to binding energies. The focus of this work is on hydrocarbons where we have seen that interfacial free energy has a well-defined meaning. The question then arises as to whether alkane/water interfacial free energy has any relevance to regions of the protein surface that contain polar and/or charged groups.

Two strategies have been used in the past in applying hydrophobicity/surface area relationships to mixed polar and nonpolar interfaces. Some workers have used total surface area; others have calculated

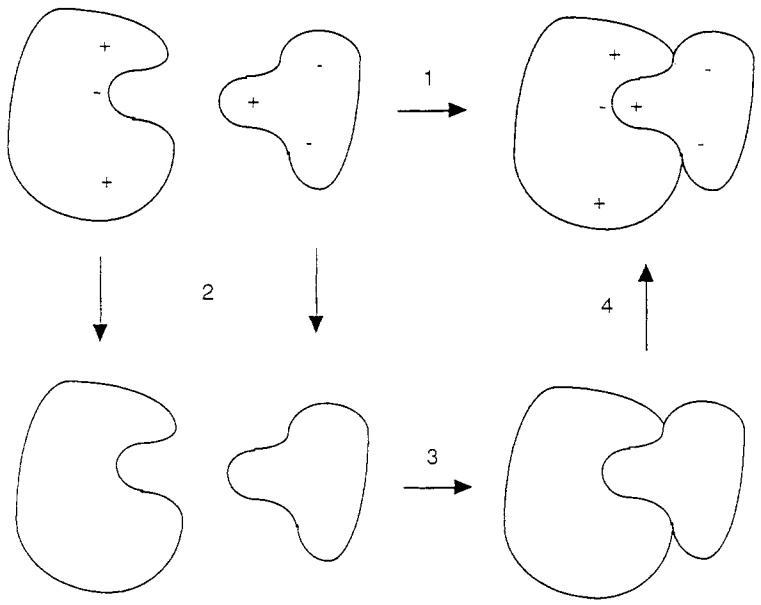

Fig. 7. Born cycle for binding separating polar from non-polar contributions. Step 1. Binding of the two fully charged molecules, including all electrostatic and nonpolar or hydrophobic interactions. Steps 2 and 4. Discharging of reactants, and charging of product complex respectively. Step 3. Binding of discharged molecules, involving only nonpolar interactions.

hydrophobic free energy only from nonpolar buried area. The latter procedure is certainly the most physically meaningful but suffers from some uncertainty regarding, for example, the extent to which the hydrophobicity of a nonpolar group is affected by a neighboring charged group. Nevertheless, as a first approximation in evaluating the energetic effects of hydrophobicity, only nonpolar surface area should be treated. In this regard aromatic groups should be treated differently than alkanes since they have a significant quadrupole moment and consequently are less hydrophobic than alkanes. Reflecting this is the fact that benzene/water surface tension is about 50 cal/mole/Å^2, i.e., significantly smaller than the alkane/water value.

Despite the above discussion, a simple Born cycle makes it possible to apply a simple interfacial free energy treatment to the entire (polar + nonpolar) area of a molecule. The basic idea is to divide any process into purely electrostatic and nonelectrostatic steps. If a solute is described in terms of charges embedded in a nonpolar cavity, electrostatic free energies can be obtained by calculating the charging (or discharging energies of the cavity[26] (see also below). The remaining interactions all involve nonpolar entities and can be obtained from the interfacial free energy relationships described in this study.

Consider, for example, the binding reaction shown in Figure 7 (which could also be viewed as representing a folding process). The binding reaction can be separated into three distinct steps where in steps 2 and 4 the reactants are discharged and the products recharged, respectively. The electrostatic free energies associated with these two processes can be calculated directly.[26,27] Step 2 now represents the hypothetical aggregation of two alkane-like bodies that have the shapes of the molecules involved but are completely nonpolar. The contribution of this step to binding can be obtained directly from the change in the interfacial free energy that accompanies the reaction where the entire interacting surface is considered.

Since hydrophobicity depends on local curvature, the expectation is that binding energies will depend on the curvature of the molecules that are associating. We used the algorithm described above to obtain local and average curvatures for a number of systems. We first determined the accuracy of our numerical procedure by calculating the curvature correction for a 2 Å radius sphere. We found that our algorithm was accurate to within about 1% even at a low resolution of about one dot/Å^2.

Table III gives the average curvature for several molecules calculated at densities of 5 dots/Å^2. We have found that at this level of resolution, increasing the density of points affects results by less than 1%. The curvature of benzamidine is close to that of alkanes, but clearly the curvature approaches unity as the size of the molecule increases. Not surprisingly, proteins appear to have average curvatures comparable to those of appropriately sized spheres (e.g., the curvature of a 25Å radius sphere is 0.96). However, as can be seen in Figure 3, proteins can

TABLE III. Curvature Correction for Different Molecules*

	Benzamidine	PTI	Trypsin	MnSOD Monomer
Surface area (Å^2)	285	3902	9061	9323
Curvature correction	0.696	0.924	0.997	0.981

*Trypsin crystal structure taken from Marquart et al.[29] Manganese Superoxide Dismutase structure courtesy of Martha Ludwig.

have highly convex and concave regions, which on average produce a curvature correction close to unity. We note that it is possible to exceed unity for the correction factor. For instance, if there are internal cavities in the molecule, i.e., volumes that can hold water molecules completely isolated from bulk water, these will be concave and hence have curvature corrections greater than unity.

In illustrating the convex and concave nature of the molecular surface, Figure 3 also clearly depicts the 'lock and key' nature of the interaction between PTI and the Trypsin binding site. The two molecules have been arranged as if the complex had been split open and the two rotated 90° in opposite directions. It is clear that the curvature colorings appear as mirror images of each other but reversed in sign, i.e., grey on trypsin where there is green on PTI. This raises the possibility of using local curvature in the general problem of predicting binding sites.

When considering the free energies of association of hydrophobic surfaces a simple argument predicts that a concave and convex surface of equal but opposite curvature will interact more strongly than two planar surfaces of the same area. This is because the hydrophobic energy of a concave surface rises faster as a function of curvature than the energy of an equivalent convex surface drops, i.e.,

$$1/(1+a/R) + 1/(1-1a/R) = 2/(1 - (a/R)^2) > 2. \quad (7)$$

Thus one might expect average hydrophobicities for the binding of a substrate into a concave binding site to exceed the planar value of 60 cal/mole/Å^2. Of course, since the effective surface tension of concave regions is high, some energy has already been invested in producing such a surface.

In order to determine what happens in real systems, we have studied three protein complexes in detail. These include benzamidine/trypsin, PTI/trypsin, and the dimeric protein, Mn superoxide dismutase (Mn-SOD). Trypsin/PTI and trypsin/benzamidine are examples of a convex molecule binding to a highly concave site, whereas the Mn-SOD dimer involves two largely planar surfaces. We have also constructed a fourth complex consisting of benzamidine in contact with *a* region of the surface of trypsin away from the binding pocket. This is to illustrate the effect of bringing two *noncomplementary* surfaces together. Calculations of accessible area and curvature corrections were carried out as described above. The effective hydrophobicity is obtained as follows: The hydrophobic energy of the individual subunits (1 and 2) is just $\gamma(\infty)A_1C1$ and $\gamma(\infty)A_2C_2$ whereas that for the complex (12) is $\gamma(\infty)A_{12}C_{12}$, where C is the curvature correction defined in Equation 4. The hydrophobic contribution to binding is given by the change in surface free energy:

$$\Delta G = \gamma(\infty)\{A_{12}C_{12} - A_1C_1 - A_2C_2\} = \gamma_{eff}A_{buried} \quad (8)$$

where $A_{buried} = A_1 + A_2 - A_{12}$ is just the net area buried.

Two distinct contributions to binding can be identified: that from atoms completely buried in the interface (inner) and that from atoms that are either partially buried or whose curvature correction is modified by the formation of the interface (peripheral). The latter arises because atoms near the boundary between two interacting subunits, though not completely buried, will find themselves in a concave "collar" region after the complex has been formed (see Fig. 8).

Table IV presents the results for the four complexes. The effective hydrophobicity. γ_{eff} and the net buried accessible area for both inner and peripheral regions is tabulated. For trypsin/PTI the effective hydrophobicity of the inner atoms (62.7 cal/mole/Å^2) slightly exceeds $\gamma(\infty)$ (59 cal/mole/Å^2). Thus the prediction made above concerning the interaction between concave and convex regions is borne out, although in this case the effect is quite small. In contrast benzamidine binds to a deep cleft in trypsin with an extremely high effective hydrophobicity (104 cal/mole/Å^2), but here the effective hydrophobicity for the interaction is less than the macroscopic value. The Mn-SOD dimer has two interacting surfaces that are essentially planar and the binding energy of the reaction is driven by an effective hydrophobicity that is equal to $\gamma(\infty)$.

The effective hydrophobicity for peripheral areas, by contrast with inner areas, is much less than $\gamma(\infty)$ for all the complexes considered. Their contribution to binding will thus be proportionately less. (Note that in this region A_{12} is not equal to zero because atoms in the collar region are partially exposed in the complex.)

Considering now the sum of inner and peripheral areas, we find that the effective hydrophobicity for

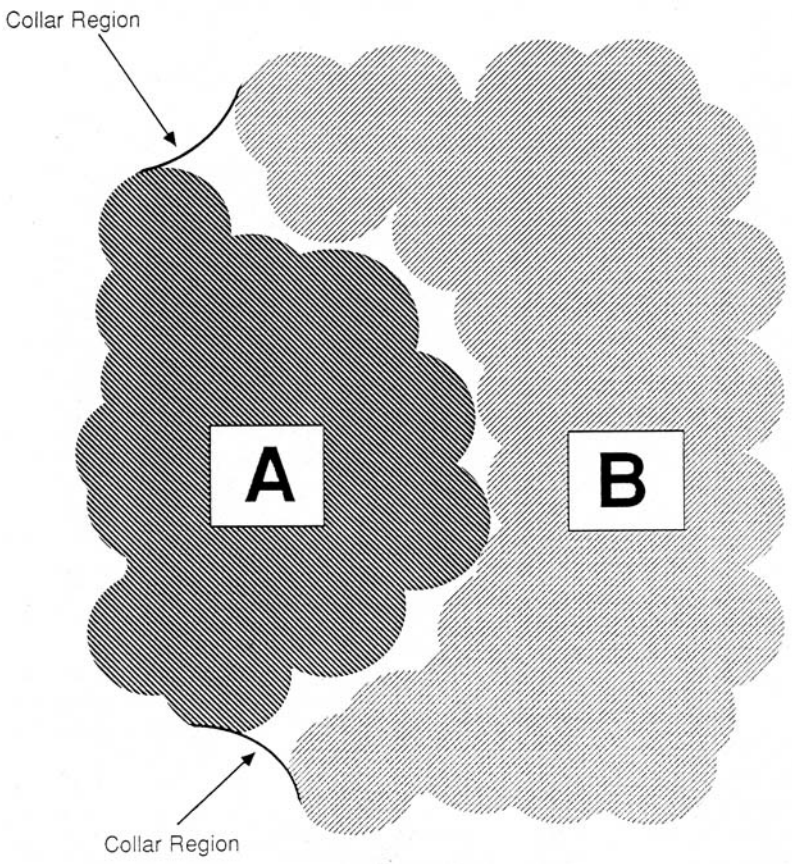

Fig. 8. The collar region upon complex formation. Illustration that when concave and convex macromolecules bind, there is a concave "collar" region formed on the periphery. This has the effect of reducing the binding affinity, using the model of hydrophobicity presented in the text.

TABLE IV. Buried Area and Effective Hydrophobicities in Complex Formation*

	Region					
	Inner		Peripheral		Total	
Complex	A_{buried}	γ_{eff}	A_{buried}	γ_{eff}	A_{buried}	γ_{eff}
PTI	388	53	406	35	794	43
Trypsin	295	76	365	45	659	59
PTI + Trypsin	682	63	771	40	1453	50
MnSOD Monomer A	296	60	612	46	909	50
MnSOD Monomer B	268	59	630	44	897	49
MnSOD dimer	564	60	1242	45	1806	50
Benzamidine (in pocket)	207	42	62	31	269	40
Trypsin	54	105	37	73	91	92
Benzamidine + Trypsin	261	55	99	47	359	53
Benzamidine (on surface)	0	0	165	28	165	28
Trypsin	28	76	80	40	108	50
Benzamidine + Trypsin	28	76	245	32	273	36

*All areas are in Angstroms squared. PTI: pancreatic trypsin inhibitor, MnSOD: manganese superoxide dismutase, A_{buried} and γ_{eff} are the net change upon complex formation and the effective hydrophobicity of the reaction, respectively. See text for details.

each of the stable complexes is close to 50 cal/mole/Å^2. However, for the hypothetical benzamidine/trypsin complex in which the inhibitor is on the surface rather than in the binding pocket, the effective hydrophobicity is significantly decreased, to 36 cal/mole/Å^2. In this case there is a strong "collar" effect

that is not balanced by a large attractive contribution from a concave/convex interface.

Our analysis of the PTI/trypsin, SOD, and benzamidine/trypsin complexes thus leads to a number of insights as to how subunit interfaces affect binding energies. It appears very difficult to achieve interfacial free energies corresponding to macroscopic surface tension due to the inevitable formation of a concave collar region whenever a complex is formed. This reduces the net hydrophobic interaction substantially. The effect can be offset if the binding site is concave (as in trypsin) or if a large essentially planar surface is available for binding (as in SOD). In either case, once curvature is accounted for, net hydrophobicities of about 50 cal/mole/Å^2 are obtained. This value is quite close to the hydrophobicity of linear hydrocarbon solutes, suggesting that it is a good approximation for many biological applications.

The hydrocarbon solute value, however, will be inappropriate for applications in which the binding site is not known, for example, in docking problems. As was the case in the hypothetical complex where benzamidine was placed on the surface rather than in the trypsin binding site, effective hydrophobicities much smaller than 50 cal/mole/Å^2 are possible if an incorrect configuration is generated. In general it appears that curvature effects may play a role in distinguishing stable from unstable complexes. The methods described in this work show how it is possible to account for these effects.

CONCLUSIONS

1. Interfacial free energy is a valid concept at both the macroscopic and microscopic levels.
2. Organic solvent to water transfer free energies satisfactorily account for the effects of nonpolar amino acids on protein stability.
3. Hydrophobicity at 25°C is due primarily to the disruption (entropic or enthalpic) of water structure and dynamics. Dispersion forces play a negligible role.
4. Hydrophobicity, as characterized by organic liquid to water transfer free energies, is a major driving force in protein folding.
5. The close packing of the hydrophobic core contributes little to protein stability because increased dispersion interactions are counterbalanced by the entropically unfavorable freezing of side chain motion.
6. Protein surfaces are on average quite planar.
7. Stable protein complexes are characterized by effective hydrophobicities of about 50 cal/mole/Å^2. This reflects a balance between a strong interactions involving buried atoms and weaker interactions in a concave collar region that surrounds the buried atoms.
8. It appears necessary to have a convex/concave or planar/planar subunit interaction for a stable complex to be formed.

ACKNOWLEDGMENTS

This work was supported in part by research grants from the NSF (PCM88-05439). ONR (N00014-90-J-1713), and NIH (GM-30518 and GM (41371). We thank Martha Ludwig for supplying us with Manganese Superoxide Dismutase coordinates.

REFERENCES

1. Privalov, P.L., Gill, S.J., Stability of protein structure and hydrophobic interaction. Adv. Prot. Chem. 39:191–234, 1988.
2. Murphy, K.P., Privalov, P.L., Gill, S.J. Common features of protein unfolding and dissolution of hydrophobic compounds. Science. 247:559–561, 1990.
3. Creighton, T. Stability of folded conformations. Curr. Opinion Struct. Biol. 1:5–16, 1991.
4. Richards, F.M. The interpretation of protein structures: total volumes, group volume distributions and packing density. J. Mol. Biol. 82:1–14, 1974.
5. Klapper, M.H. On the nature of the protein interior. Biochim. Biophys. Acta. 229:557–566, 1970.
6. Sharp, K.A., Nicholls, A., Fine, R.M., Honig, B. Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects. Science. 252:106–109, 1991.
7. Sharp, K.A., Nicholls, A., Friedman, R., Honig, B. Extracting hydrophobic free energies from experimental data: relationship to protein folding and theoretical models. Biochemistry (in press), 1991.
8. Aveyard, R., Haydon, D.A. Thermodynamic properties of aliphatic hydrocarbon/water interfaces. J. Coll. Int. Sci. 20:2255–2261, 1965.
9. Tanford, C.H. Interfacial free energy and the hydrophobic effect. Proc. Natl. Acad. Sci. 76:4175–4176, 1979.
10. Sinanoglu, O. Microscopic surface tension down to molecular dimensions and microthermodynamic surface areas of molecules of clusters. J. Chem. Phys. 75:463–468, 1981.
11. Tolman, R.C. The effect of droplet size on surface tension. J. Chem. Phys. 17:333, 1949.
12. Choi, D.S., Jhon, M.S., Eyring, H. Curvature dependence of the surface tension and the theory of solubility. J. Chem. Phys. 53:2608, 1970.
13. Sinanoglu, O. What size cluster is like a surface? Chem. Phys. Lett. 81:188–189, 1981.
14. Lee, B., Richards, F.M. The interpetation of protein structures: Estimation of static accessibility. J Mol. Biol. 55: 379–400, 1971.
15. Shrake, A., Rupley, J.A. -solvent accessible surface area calculation. J. Mol. Biol. 79:351–371, 1973.
16. Connolly, M.L. Measurement of protein surface shape by solid angle. J. Mol. Graphics. 4:3–6, 1986.
17. Richards, F.M. Areas, volumes, packing and protein structure. Ann. Rev. Biophys. Bioeng. 6:151–176, 1977.
18. CRC. "CRC Handbook of Chemistry and Physics" (Lide, D.R., ed.) Boca Raton: CRC Press, 1990.
19. Ben-Naim, A., Marcus, Y. Solvation thermodynamics of nonionic solutes. J. Chem. Phys. 81:2016–2027, 1984.
20. Buff, F.P., Lovett, R.A., Stillinger. F.H. Interfacial density profile for fluids in the critical region. Phys. Rev. Lett. 15:621–623, 1965.
21. Adamson, A.W. "Physical Chemistry of Surfaces." New York: John Wiley & Sons, 1976.
22. Novotny, J., Bruccoleri, R.E., Saul, F.A. On the attribution of binding energy in antigen-antibody complexes McPC 603, D1.3, and HyHel-5. Biochemistry. 28:4735, 1989.
23. Rashin, A., Iofin, M., Honig, B. Interval cavities and buried waters in globular proteins. 25(3619):1986.
24. Kellis, J.T., Nyberg, K., Fersht, A.R. Energetics of complimentary side-chain packing in a protein hydrophobic core. Biochem. 28:4914–4922, 1989.
25. Shortle, D., Stites, W.E., Meeker, A.K. Contributions of the large hydrophobic amino acids to the stability of staphylococci nuclease. Biochem. 29:8033–8041, 1990.
26. Gilson, M.K., Honig, B. Calculation of the total electro-

static energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. Proteins. 4:7–18, 1988.
27. Sharp, K., Honig, B. Electrostatic interactions in macromolecules: theory and applications. Ann. Rev. Biophys. Biophys. Chem. 19:301–332, 1990.
28. Sandberg, W.S., Terwilliger, T.C. Energetics of repacking a protein interior. Proc. Natl. Acad. Sci. USA. 88:1706–1710, 1991.
29. Marquart, Walter, Deisenhofer, J., Bode, Walter. The geometry of the reactive site and of the peptide groups in trypsin, typsinogen and complexes and inhibitors. Acta Crystallogr, Sect B. 39:480, 1983.