

A desolvation barrier to hydrophobic cluster formation may contribute to the rate-limiting step in protein folding

JEFFREY A. RANK¹ AND DAVID BAKER²

¹Department of Physics, University of Washington, Seattle, Washington 98195

²Department of Biochemistry, University of Washington, Seattle, Washington 98195

(RECEIVED July 16, 1996; ACCEPTED November 5, 1996)

Abstract

To gain insight into the free energy changes accompanying protein hydrophobic core formation, we have used computer simulations to study the formation of small clusters of nonpolar solutes in water. A barrier to association is observed at the largest solute separation that does not allow substantial solvent penetration. The barrier reflects an effective increase in the size of the cavity occupied by the expanded but water-excluding cluster relative to both the close-packed cluster and the fully solvated separated solutes; a similar effect may contribute to the barrier to protein folding/unfolding. Importantly for the simulation of protein folding without explicit solvent, we find that the interactions between nonpolar solutes of varying size and number can be approximated by a linear function of the molecular surface, but not the solvent-accessible surface of the solutes. Comparison of the free energy of cluster formation to that of dimer formation suggests that the assumption of pair additivity implicit in current protein database derived potentials may be in error.

Keywords: hydrophobic interaction; potential of mean force; protein folding

The hydrophobic interaction is extremely important in the folding and stabilization of proteins (Kauzmann, 1959; Dill, 1990), but is relatively poorly understood. In spite of this poor understanding, computer modeling of protein folding has developed substantially over the past 10 years. One approach is to simulate protein folding in the presence of explicitly modeled water molecules, which, depending upon the accuracy of the water model, should at least in part reproduce the hydrophobic interaction. Molecular dynamics simulations with an explicit solvent model have provided valuable insights into possible folding/unfolding trajectories (Daggett & Levitt, 1994; Karplus & Sali, 1995). However, because of the complexity of the processes being simulated and the absence of an obvious reaction coordinate, most studies have focused on qualitative features of the trajectories and not sought to calculate free energy changes (the study of Brooks and Boczko [1995] is an exception). A second approach avoids the complexity and high computational demands of simulations with hundreds of explicit water molecules by deriving effective potentials from the distribution of amino acid residues in known protein structures (reviewed in Jones & Thornton, 1996). Although not modeled explicitly, the hydrophobic interaction dominates such "knowledge-based" potentials. These relationships are not true interaction potentials in a rigorous sense, but they have been extremely useful in a wide variety of applications, including protein fold recognition (Jones &

Thornton, 1996). Pair additivity is usually assumed in calculating the total energy or score: the energy is the sum of the solvation and interaction energies of individual and pairs of residues (Sippl, 1995).

In this paper, we separate hydrophobic core formation from the multitude of other complex interactions involved in protein folding by investigating a considerably simplified problem: the association of methane molecules in water. We employ the widely used TIP4P model of water (Jorgenson et al., 1983), and standard free energy perturbation methods (Beveridge & DiCapua, 1989) to calculate the free energy of interaction of pairs, triplets, and larger numbers of methane molecules as a function of separation distance. These studies help bridge the rather large gap between computer studies of nonpolar solvation (Pratt & Pohorille, 1992) and pair interactions of nonpolar solutes (Jorgenson et al., 1988) and studies of protein folding (Daggett & Levitt, 1994; Brooks & Boczko, 1995; Karplus & Sali, 1995). The simplicity of the problem relative to protein folding allows the calculation of free energies, and perhaps more importantly, facilitates the understanding of the physical principles underlying the interactions. The results have implications for both the modeling of hydrophobic interactions and the understanding of the rate-limiting step in protein folding.

Results and discussion

Standard free energy perturbation methods were used to calculate the free energy of association or potential of mean force (PMF) for spherical nonpolar solutes of different sizes and numbers in TIP4P water. The four reactions studied are described schematically in Figure 1. These include the dimerization of two methane mol-

Reprint requests to: David Baker, Department of Biochemistry University of Washington, Seattle, Washington 98195; e-mail: dabaker@u.washington.edu.

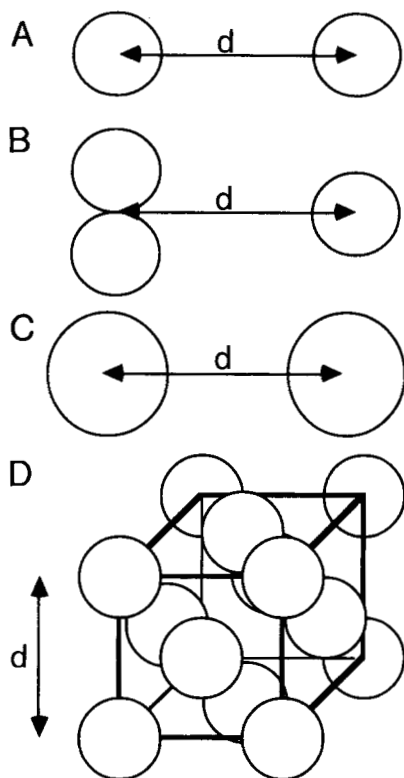


Fig. 1. Schematic of simulations. Solute configurations and reaction coordinates for each of the four simulations are shown. **A:** Methane dimer. **B:** Methane trimer. **C:** Big solute dimer. **D:** Methane cluster. The vdW radii of the solutes are 1.85 Å in A, B, and D, and 2.6 Å in C.

ecules (Fig. 1A), the association of a single methane with a preformed methane dimer along the symmetry axis of the dimer to form a trimer (Fig. 1B), the dimerization of two larger nonpolar solutes (Fig. 1C), and the association of 14 methane molecules arrayed on the vertices of a face-centered cubic (fcc) lattice to form a close-packed cluster (Fig. 1D).

Figure 2A shows the free energy of association or PMF for two methane molecules in TIP4P water at 25 °C. As noted previously in similar calculations (Jorgenson et al., 1988), there is a pronounced minimum at contact, and a second minimum for a solvent-separated pair. A virtually identical curve was obtained previously (Smith & Haymet, 1993) using SPC water and molecular dynamics rather than Monte Carlo sampling: the minima and maxima are in precisely the same positions, and the differences in free energy between the solvent-separated and contact minima (0.59 ± 0.06 kcal/mol) are very similar. Thus, the methane methane PMF is not sensitive to the differences in the water models. The results also do not depend on the size of the water box: a significantly larger box was used in the current study (396 water molecules instead of 106 in the molecular dynamics study).

Calculating the interactions between methanes in water can be separated into two parts: determining the free energy cost of forming the relevant cavities, and determining the solute–solvent interaction energy (Ben-Naim, 1987). An elegant theory based on geometric considerations—the scaled particle theory—describes many properties of liquids, including the free energy of formation of spherical cavities (Reiss, 1965). Not surprisingly, for a water-sized solvent at atmospheric pressure, scaled particle theory predicts that the free energy of cavity formation will be dominated by a term proportional to the surface area of the cavity; the work due to the volume change only becomes significant for cavity sizes on the order of microns or larger. An overly simplified but nevertheless useful interpretation is that the cost of cavity formation is proportional to the number of lost water–water interactions at the surface of the cavity; formation of cavities in water is unfavorable because of the strong interactions between water molecules. Although the generalization of the theory to nonspherical cavities is quite problematic, the scaled particle theory result for spherical cavities provides a rough estimate of the scale factor relating the free energy of cavity formation to the surface area of the cavity (see Materials and methods).

It has been noted previously (Jackson & Sternberg, 1994) that the molecular surface, as defined originally by Richards (1977) is a much better measure of the number of lost water–water interactions than the more frequently used solvent-accessible surface.

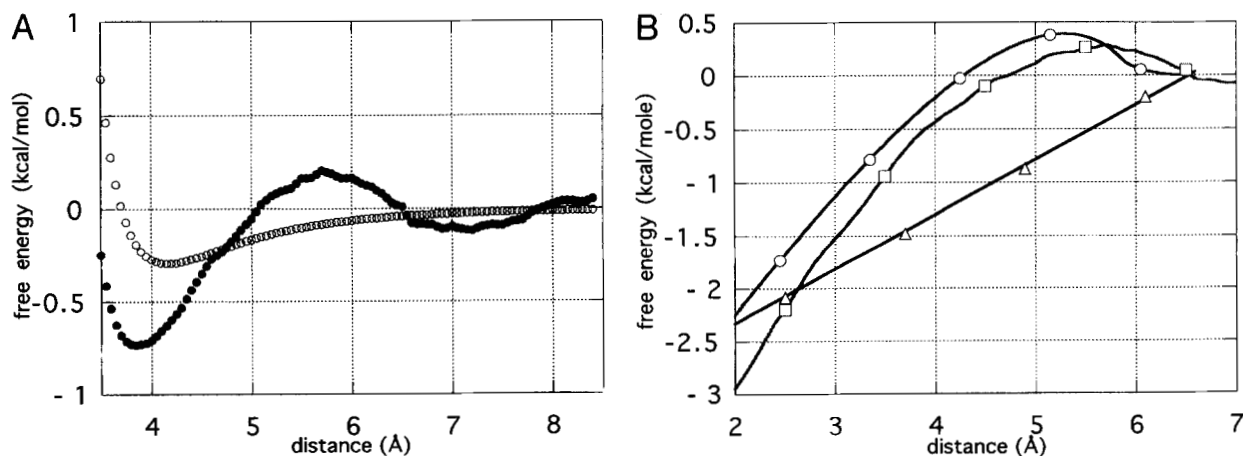


Fig. 2. Methane dimer PMF. **A:** Comparison of the PMF (closed symbols) to the vdW interaction between the solutes (open symbols). **B:** Comparison of the solvent contribution to the PMF (squares) to the molecular surface area (circles) and the solvent-accessible surface area (triangles).

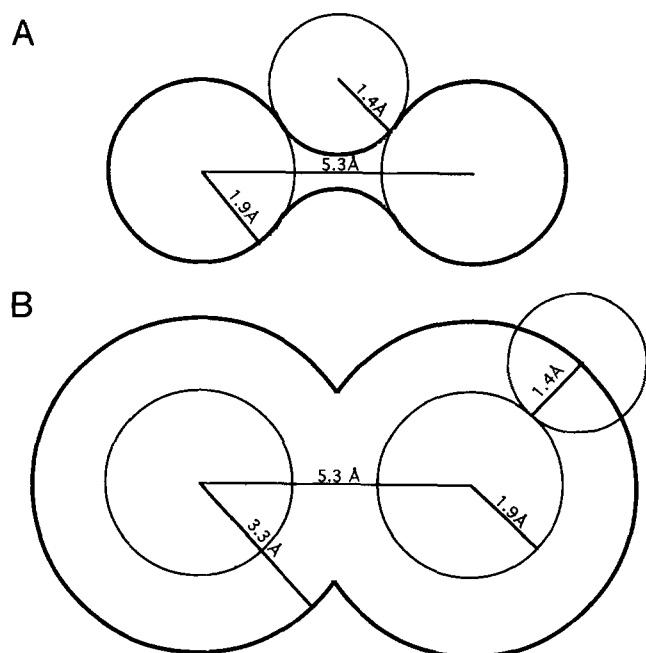


Fig. 3. Comparison of solvent-accessible and molecular surfaces. The outlines of two methane molecules (radius 1.9 Å) and two water molecules (radius 1.4 Å) are indicated in thin lines. The molecular surface (A) and solvent-accessible surface (B) are indicated by thick lines.

The molecular surface is essentially the sum of the portions of the surfaces of water molecules not in contact with other water molecules due to the presence of the solutes (Fig. 3A). It was found that the changes in the molecular surface follow the methane-methane PMF much more closely than the changes in the solvent-accessible surface (Jackson & Sternberg, 1994).

The changes in molecular surface are best compared not with the complete PMF, but with the solvent contribution to the PMF [the difference between the PMF and the van der Waals (vdW) interaction between the methanes]. As shown in Figure 2B, the

similarity between the calculated free energy changes and the molecular surface is quite strong. In contrast to the molecular surface, the change in the solvent-accessible surface (Fig. 2B, triangles) does not resemble the solvent contribution to the PMF; the barrier at intermediate solute separations is completely missed. The molecular surface and the solvent-accessible surface at the barrier are indicated schematically in Figure 3; the molecular surface is at a maximum because of the extended solvent-excluded region between the methanes, whereas the solvent-accessible surface is considerably less than that of the fully separated solutes (Wood & Thompson, 1990; Jackson & Sternberg, 1994).

Simple potential functions used in protein folding studies often assume pair additivity of interactions between nonpolar groups (Sippl, 1995). To investigate potential many-body interactions, we simulated the formation of a trimer of methane molecules. The reaction coordinate is the distance between a methane molecule and two methane molecules held fixed at the origin (Fig. 1B). The PMF was calculated in exactly the same manner as for the dimer. As shown in Figure 4A, the depth of the free energy minimum at contact is only slightly greater than that of a single dimer (Fig. 2A), and is considerably less than the sum of the interactions between the two pairs of dimers (Fig. 4A) expected if pair additivity were to hold. It should be noted, however, that the absolute values of the free energies depend on the setting of the interactions to zero at very long separations; the errors associated with this zeroing may be as much as 0.2–0.3 kcal/mol.

There is a substantial barrier to association of the methanes at a separation of ~ 5.6 Å (Fig. 4A). The similarity between the solvent contribution to the trimer PMF and the molecular surface (Fig. 4B) suggests that the barrier results from an effective increase in cavity size due to exclusion of water molecules from the volume between the methanes. Again, the barrier is not reflected in the solvent-accessible surface area (Fig. 4B, triangles).

To investigate further the degree to which the interactions between nonpolar solutes can be approximated by the molecular surface area, we simulated the association of two larger solutes (Fig. 1C). Both the vdW interaction parameters of the solutes were increased to correspond with the increase in both the radius and well depth of the vdW interaction in noble gases of increasing

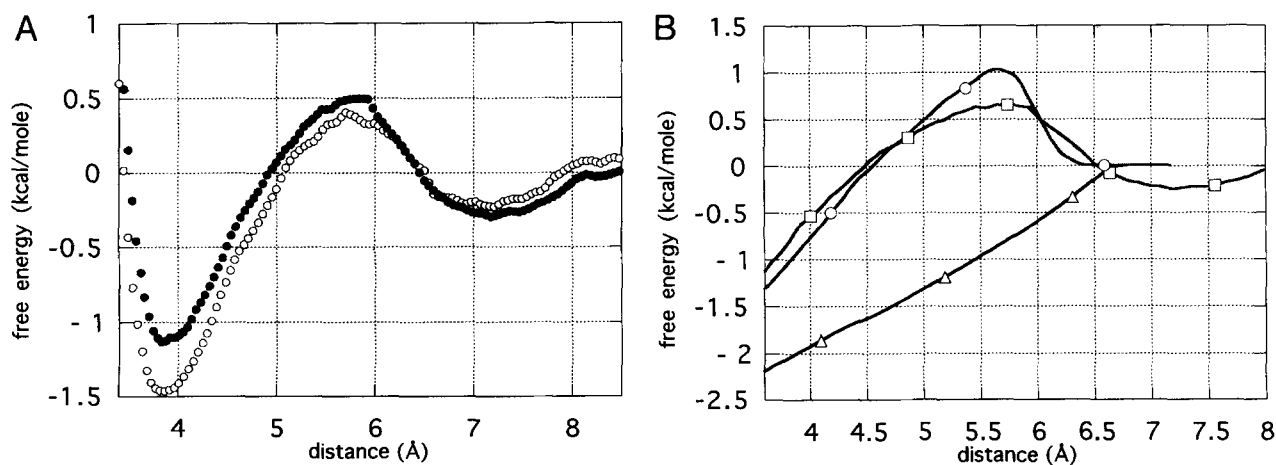


Fig. 4. Methane trimer PMF. **A:** Comparison of the trimer PMF (closed symbols) to the PMF expected were the interactions between methanes pair additive (open symbols). **B:** Comparison of the solvent contribution to the PMF (squares) to the molecular surface area (circles) and the solvent-accessible surface area (triangles). The x axis is the distance between methane centers.

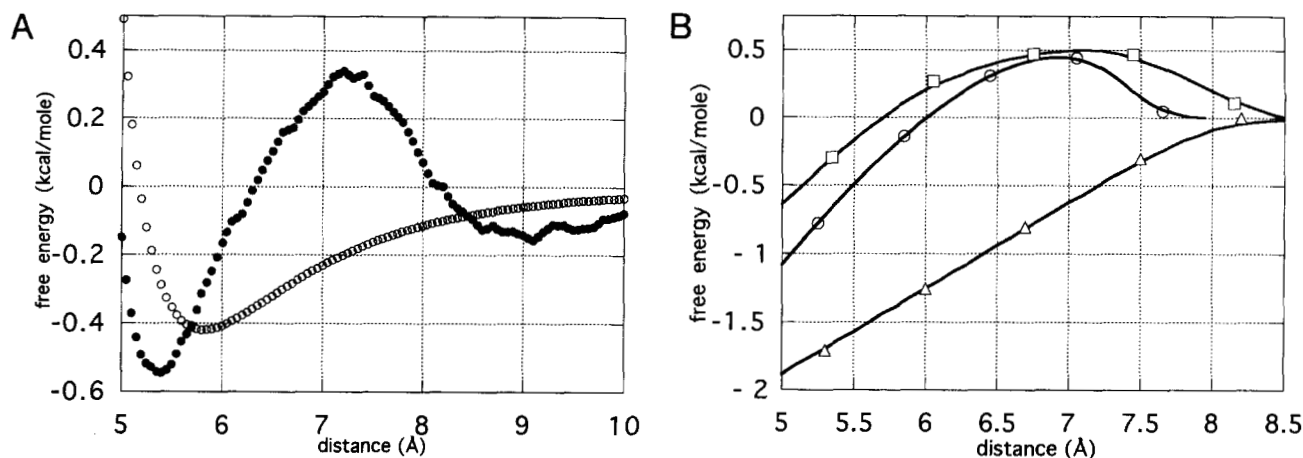


Fig. 5. Big solute dimer PMF. **A:** Comparison of the big (2.6-Å radius) solute PMF (closed symbols) to the vdW interaction between the solutes (open symbols). **B:** Comparison of the solvent contribution to the PMF (squares) to the molecular surface area (circles) and the solvent-accessible surface area (triangles).

molecular weight. Consistent with the increase in the solubility of the noble gases with increasing size, the free energy minimum is more shallow for the big solutes than for the methane pair despite the increase in the strength of the solute–solute vdW interaction (compare Fig. 5A to Fig. 2A).

The overall features of the solvent contribution to the PMF are again captured fairly well by the solute molecular surface (Fig. 5B), but the latter predicts a somewhat more attractive interaction. The discrepancy may reflect the increase in the solvent–solute interaction strength for the big solutes (see Materials and methods); the scale factor relating molecular surface area to free energy from scaled particle theory incorporates only the cost of cavity formation, not solvent–solute interactions.

To further explore multibody interactions and possible interactions in protein hydrophobic core formation, we simulated the formation of a cluster of 14 methane molecules. The importance of adequate sampling for accurate estimation of free energy changes

greatly limits the trajectories that may be considered for calculation of the free energy of cluster formation. To simplify calculations, we chose to simulate the formation of a cluster of 14 methane molecules arranged on the vertices of a single unit cell of an fcc lattice (Fig. 1D). The fcc lattice provides a one-dimensional reaction coordinate and allows a high packing density for the spherical solutes.

The methane cluster PMF shown in Figure 6A (closed circles), the average of two PMFs calculated using different initial conditions (see Materials and methods), is our best estimate of the free energy changes accompanying the uniform expansion of the methane molecules from densely packed to relatively uniformly distributed throughout the solvent. There is a pronounced minimum at close to vdW contact between the methanes (lattice spacing 5.6 Å), a large barrier at a lattice spacing of ~ 8.1 Å, and a second minimum at ~ 10 Å. The PMF expected if the interactions between each pair of methanes were identical to those of an isolated pair of

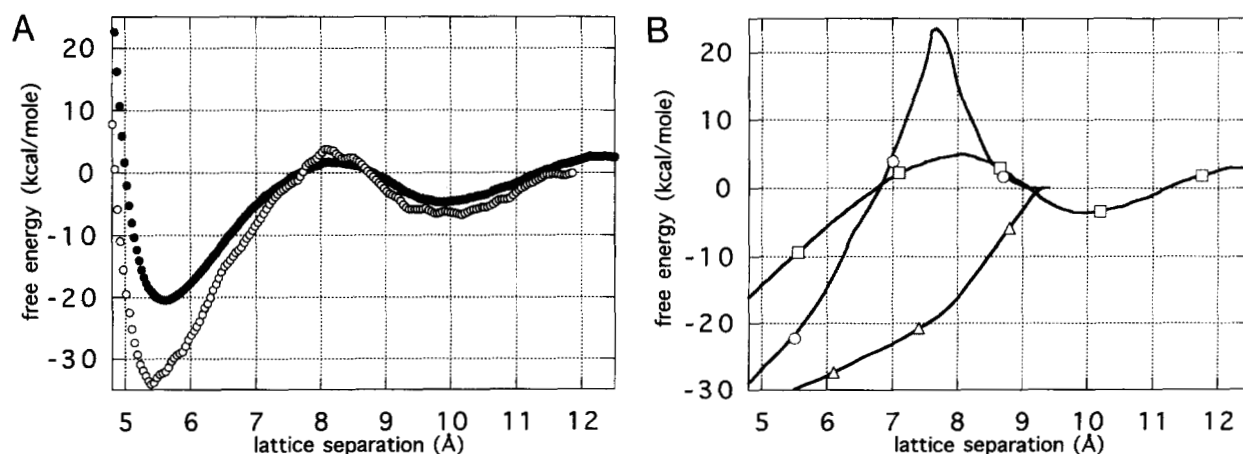


Fig. 6. Methane cluster PMF. **A:** Comparison of free energy of cluster formation (closed symbols; the average methane cluster PMF from Fig. 8 is shown) to the free energy expected if the interactions between methanes were pair additive (open symbols; the calculation assumes that contributions from interactions between methanes separated by more than 8.4 Å are negligible). **B:** Comparison of the solvent contribution to the PMF (squares) to the molecular surface area (circles) and the solvent-accessible surface area (triangles).

methanes (Fig. 2A) is shown in Figure 6A (open circles). Despite the uncertainties resulting from the incomplete convergence of the methane cluster PMF calculations (see Materials and methods), there appears to be significant departure from pair additivity (compare Fig. 6A, open circles to the curves in Fig. 8). Interestingly, methane aggregation was found to occur spontaneously in a molecular dynamics study of a system composed of 18 methanes and 107 water molecules (Wallqvist, 1991), but because free energy changes were not calculated in the earlier study, it is difficult to make detailed comparisons.

What contributes to the barrier at ~ 8 Å? The solvent contribution to the interaction is shown in Figure 6B along with the molecular surface of the methanes. The decrease in the molecular surface after ~ 8 Å suggests that solvent molecules can only penetrate the cluster after this point. Inspection of snapshots of the water distributions around the methanes also suggests that water penetration is the critical step (Fig. 7). Water molecules are excluded from the methane cluster before (Fig. 7A) and at (Fig. 7B) the barrier, but not later (Fig. 7C). The increase in size of the cavity from which solvent is excluded is substantial in going from the close-packed state to the barrier (compare Fig. 7A to 7B).

The barrier in Figure 6 is undoubtedly an overestimate of the actual barrier to cluster formation because of the high symmetry maintained throughout the calculation. The decrease in the free energy which accompanies the penetration of water molecules as the lattice spacing is increased beyond 8 Å is in part a consequence of the symmetry; were the spatial relationships between methane molecules less uniform, the penetration of water would occur more gradually. However, any trajectory that brings 14 solutes together simultaneously will necessarily have a water-excluding point with an expanded effective cavity size before contact is reached. It is worth emphasizing that the presence of the barrier is not an artifact of the symmetry: significant barriers are observed in both the dimer and trimer calculations. The symmetry may actually increase the resemblance of the problem to protein folding: in computer simulations of protein unfolding, considerable expansion of the protein occurs prior to solvent penetration (Daggett & Levitt, 1993; Li & Daggett, 1994) and hydrogen-deuterium exchange experiments on small proteins show that the exchange of amide protons of core residues following the penetration of water into the core can be a highly cooperative event (Kiefhaber & Baldwin, 1995; Yi & Baker, 1996).

The simple exponential kinetics observed for the folding and unfolding of small proteins suggests that there is a sizable free energy barrier blocking transitions between folded and unfolded states. Might water-exclusion effects similar to those we observe in the formation of methane clusters contribute to such a barrier? A somewhat oversimplified statement of the analogy is that the high "microscopic" surface tension of water may oppose not only the initial expansion of the cluster (the unfolding of a protein), but also the final stage of cluster formation (folding of a protein) because of the transient increase in the molecular surface at the point at which water molecules are excluded from the interior of the cluster (protein core). Measurements of the pressure dependence of the rates of folding and unfolding of staphylococcal nuclease support this view: there is a considerable increase in volume of the protein solvent system at the transition state (Vidugiris et al., 1995). A barrier consisting of an expanded water-excluding state (which might resemble Fig. 6B) is also consistent with the finding that truncation of side chains in protein hydrophobic cores slows the rate of folding more than mutations elsewhere (Serrano et al.,

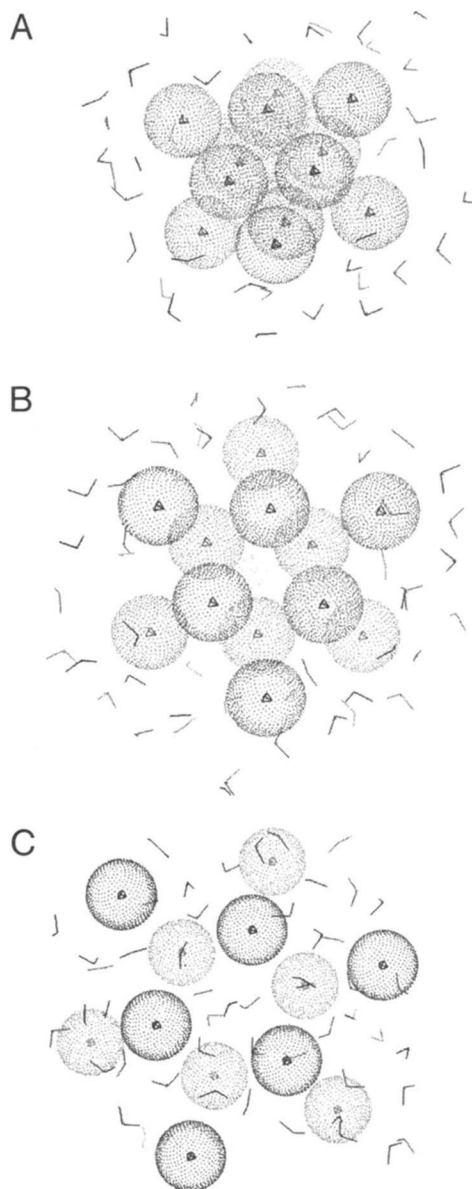


Fig. 7. Snapshots of the water configuration around the methane cluster at different stages of association. **A:** At the free energy minimum at contact. **B:** At the free energy barrier. **C:** At the solvent separated minimum. A thin section containing only a subset of the solvent and solute molecules in the cluster is shown for clarity (the three views have the same thickness).

1992; Itzhaki et al., 1995): such mutations in the protein interior would almost certainly reduce the effective cavity size and hence the free energy of the expanded state less than that of the unfolded state.

Current theoretical views of the folding process envision an energetically downhill search for the native state in a "folding funnel" (Dill, 1987; Wolynes et al., 1995) in which the free energy barrier to folding is primarily due to the loss of chain configurational entropy. An increase in the free energy of solvent-protein or solvent-solvent interactions during folding (the increase in effective cavity size accompanying solvent exclusion as suggested here and/or the desolvation of polar groups [Waldburger et al., 1996]) would create a bump in such a funnel that could be missed in

simple models of folding that neglect the solvent. It seems possible that desolvation and loss of chain configurational entropy both could contribute to the free energy barrier to folding; for example, both effects would disfavor the approach of core hydrophobic side chains prior to the formation of substantial favorable interactions.

The agreement between the free energy changes calculated using the explicit water simulation and the much simpler molecular surface calculation is quite encouraging for the computational modeling of protein folding because it appears that, at least for hydrophobic interactions, an explicit solvent model is not required for approximating free energy changes. However, the comparisons shown in Figures 4A and 6A suggest that the pair additivity assumed in the modeling of hydrophobic interactions using current database-derived potentials could be in error. Induced polarization effects neglected in the TIP4P model could lead to further deviations from pair additivity. Our results further dramatize the conclusion (Tunon et al., 1992; Jackson & Sternberg, 1994) that the molecular surface area, not the solvent-accessible surface area, should be used in approximating hydrophobic interactions.

Materials and methods

Monte Carlo simulations

The TIP4P model (Jorgenson et al., 1983) was used for water, and a united atom approximation for the solute molecules. The solute-solute and solute-solvent interactions were modeled by a Leonard Jones potential:

$$U_{LJ}(r) = 4\epsilon([\sigma/r]^{12} - [\sigma/r]^6) \quad (1)$$

with parameters $\sigma_{m-m} = 3.71 \text{ \AA}$, $\epsilon_{m-m} = 0.294 \text{ kcal/mol}$ (Jorgenson et al., 1988) for the methane molecules (Figs. 2, 4, and 6), and $\sigma_{b-b} = 5.18 \text{ \AA}$ and $\epsilon_{b-b} = 0.42 \text{ kcal/mol}$ for the "big" solutes (Fig. 5). The parameters used for solute-solvent interactions were determined using the Lorentz-Berthelot mixing rules (Allen & Tildesly, 1986). All solvent interactions were quadratically damped to zero between 8.0 and 8.5 \AA . Monte Carlo calculations were performed using a $20 \times 20 \times 30 \text{ \AA}$ box containing 396 water molecules for the calculations described in Figures 2, 3, and 4. A $25 \times 25 \times 25 \text{ \AA}$ box containing 515 water molecules was used for the methane cluster calculation (Fig. 6). All calculations used an NPT ensemble at 25 °C and 1 atm with periodic boundary conditions.

PMFs between solute molecules were computed using the perturbation method (Beveridge & DiCapua, 1989):

$$\Delta G(d) = -kT \ln \langle \exp\{-\Delta E_i(d)/kT\} \rangle_b \quad (2)$$

where ΔG is the change in free energy and ΔE_i the change in energy of state i , for a change in the solute separation d . $\langle \rangle_b$ denotes a Boltzmann-weighted average over all states of the unperturbed system. The reaction coordinates for the different simulations are indicated in Figure 1; a perturbation step size of 0.05 \AA was used in the first three calculations and a step size of 0.038 \AA in the fcc calculation. Double-wide sampling and preferential sampling of waters around the solutes were used to increase sampling efficiency (Allen & Tildesly, 1986). Solutes were placed at a random position in one of 25 pre-equilibrated water boxes, high energy waters were removed and the system was equilibrated for 1–1.5 million MC moves except in the methane cluster calculation described below. Sampling was then conducted over an additional 2–3 million moves for the methane dimer, 2 million moves for the methane trimer, and 6–9 million moves for the big solute dimer.

Solvent moves consisted of both a rotation between -15 and 15° around one of the three axes parallel to the box edges, and a translation of between -0.15 and 0.15 \AA in each of the x , y , and z directions. A volume move was performed every 2,375 moves: the center of mass coordinates of all molecules were multiplied by a scale factor between 0.994 and 1.006. Solutes were subjected to a random translation of between -0.08 and 0.08 \AA in each of the x , y , and z directions every 90 moves. Because of the larger box size in the fcc simulation (Figs. 1D, 6), solute and volume changes instead were attempted every 110 and 2,875 moves, respectively. Acceptance rates for solvent moves were about 40%.

A standard procedure for calculating errors in free energy perturbation calculations is to calculate averages of the quantities of interest over blocks of 100,000–300,000 cycles, and then to calculate the standard deviations of these "block averages" (Allen & Tildesly, 1986). As will be described in more detail elsewhere, we found that such a procedure underestimated errors significantly; therefore, in all calculations reported here, free energies and standard deviations were calculated from a minimum of 10 independently equilibrated runs except for the methane cluster calculation. The average error per angstrom for each PMF was estimated by adding the standard deviation of the mean free energy change for each point in quadrature and then dividing by the total length of the PMF; the errors were 0.044, 0.058, 0.064 kcal/(mol \AA) for the dimer, trimer, and big solute calculations, respectively, with the errors 1.5–2 times larger for points at the longest separations than for those at the shortest.

Slow solvent penetration into the methane cluster and trapping of high-energy waters within the cluster complicated the methane cluster PMF calculation. The existence of energy barriers too large to be crossed during the course of a simulation is clearly a problem for free energy calculations, as there is a possibility that important regions of phase space are sampled inadequately. To test the equilibration of the system, free energy changes were calculated using four different protocols. In one set of runs, the 36 highest-energy water molecules were removed after introduction of the methane cluster. Two PMFs were then calculated after equilibrating the system for 1 and 5 million moves, respectively (only five independent runs were performed for the calculations with 5 million equilibration moves). The PMF changed significantly after the increased equilibration (Fig. 8, triangles), suggesting the system had not fully equilibrated in 1 million moves. In a second set of runs, 46 waters were removed initially and two PMFs were again calculated after 1 and 5 million equilibration moves. The results (Fig. 8, squares) suggest clearly that equilibration is again incomplete. The statistical errors within the sets of independent runs conducted for each of the four simulation protocols were at least an order of magnitude smaller than the systematic differences between the sets due to differences in initial conditions and equilibration times. Analysis of water configurations suggested that removal of 36 waters results in initial overpenetration of solvent in the methane cluster, whereas removal of 46 waters gives initial underpenetration.

Figure 8 indicates clearly that equilibration for as many as 5 million moves is insufficient for complete relaxation of the system. However, the PMF values calculated after 5 million equilibration moves (Fig. 8, open triangles and squares) are more similar than those calculated after only 1 million moves (Fig. 8, closed triangles and squares). Given this convergence and the lack of sufficient computing power to lengthen equilibration times substantially, in Figure 6 we use the average (Fig. 8, diamonds) of the two longer equilibration runs as a best estimate to the methane cluster PMF.

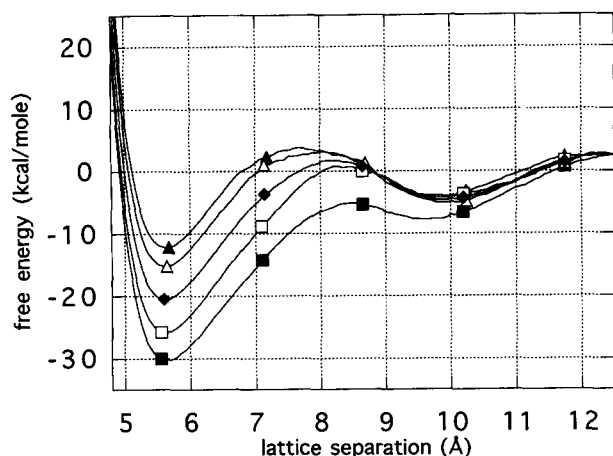


Fig. 8. Dependence of cluster PMF on initial conditions and equilibration time. Triangles, 36 high-energy waters removed before equilibration; squares, 46 waters removed; closed symbols, 1 million attempted equilibration moves; open symbols, 5 million attempted moves; diamonds, the average of the two 5 million equilibration calculations.

Surface area calculations

The changes in the molecular surface and the solvent-accessible surface (Richards, 1977) were calculated using simple analytic expressions for the dimerization reactions (see below) and using the MSCON (Connolly, 1983) and NACCESS (Hubbard & Thornton, 1993) programs for the multibody reactions. For comparison to the PMF calculations, surface area changes were converted to free energy changes using scale factors of $0.025 \text{ kcal}/\text{Å}^2$ and $0.110 \text{ kcal}/\text{Å}^2$ for the solvent-accessible surface and the molecular surface, respectively (Jackson & Sternberg, 1994). Only qualitative agreement between the PMF calculations and the surface area calculations should be expected; our conclusions are independent of the precise values of the scale factors. In all surface calculations, a probe radius of 1.4 Å was used to approximate the size of a water molecule.

Analytical expression for the molecular surface area of two spherical solutes

The molecular surface area (13, 14) of two solutes of radius R_s , separated by a distance x can be separated into two parts. For

$x > 2R$, the two solutes are separated completely and the molecular surface is the surface area of the two solute spheres (see Equation 6). For $x \leq 2R$, there are two separate terms. The first is the portion of the solute surfaces that can contact the solvent probe (first term in Equation 6). The second is the area of the surface formed by the probe as it is revolved around the line connecting the solute centers while in contact with both solutes (Fig. 9). This double contact point area can be expressed as:

$$A_2 = (2)2\pi \int_{z_1}^{z_1+z_2} f(y) dL, \quad (3)$$

where

$$f(y) = \sqrt{R^2 - (x/2)^2} - \sqrt{R_p^2 - y^2} \quad (4)$$

is the distance from the line through the solute centers to the surface of the solvent probe (Fig. 9),

$$dL = \sqrt{1 + f'(y)^2} dy \quad (5)$$

is the differential arc length along the curve $f(y)$, and R_p is the solvent probe radius. Integrating and simplifying yields the second term in Equation 6 for the molecular surface area:

$$M_A(x) = \begin{cases} 4\pi R_s^2 \left\{ 1 + \frac{x/2}{R} \right\} \\ + 4\pi R_p \{ \alpha(x) \sqrt{R^2 - (x/2)^2} - z_2(x) \}; & x \leq 2R \\ 8\pi R_s^2; & x > 2R \end{cases}, \quad (6)$$

where the angle α and the distances z_1 and z_2 indicated in Figure 9 are functions of the solute separation distance x :

$$\alpha(x) = \begin{cases} \cos^{-1} \left(\sqrt{1 - \frac{(x/2)^2}{R^2}} \right); & x \leq 2\sqrt{R^2 - R_p^2} \\ \cos^{-1} \left(\frac{R^2 + R_p^2 - ((x/2) - z_1(x))^2}{2RR_p} \right); & x > 2\sqrt{R^2 - R_p^2} \end{cases} \quad (7)$$

$$z_1(x) = \begin{cases} 0; & x \leq 2\sqrt{R^2 - R_p^2} \\ \sqrt{R_p^2 - R^2 + (x/2)^2}; & x > 2\sqrt{R^2 - R_p^2} \end{cases}$$

$$z_2(x) = \frac{R_p}{R} (x/2) - z_1(x). \quad (8)$$

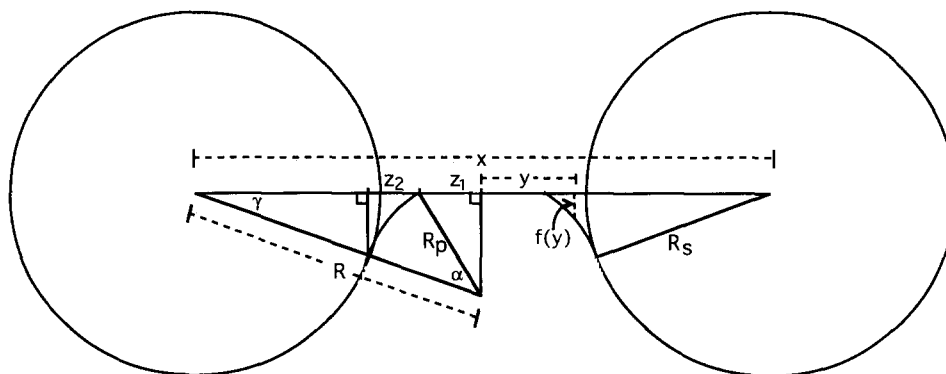


Fig. 9. Diagram for analytical molecular surface calculation. The probe is in contact with both solutes and penetrates beyond the line connecting the solute centers. R_s , solute radius; R_p , probe radius; x , distance between the solutes.

Acknowledgments

We thank W. Reinhardt and T. Lybrand for helpful advice at the beginning of this work; J. Solomon for assistance with parallel computing implementation and resources through DOE grant #DE-FG03-92ER25134; D. Alonso for help in surface area calculations; and K. Dill, D. Eisenberg, D. Teller, W. Parson, A. Li, C. Bystroff, W. Reinhardt, and V. Daggett for comments on the manuscript. J.R. was supported by an NIH molecular biophysics predoctoral fellowship. This work was supported by young investigator awards to D.B. from the NSF and the Packard foundation.

References

- Allen MP, Tildesly DJ. 1986. *Computer simulations in liquids*. Oxford: Clarendon Press.
- Ben-Naim A. 1987. *Solvation thermodynamics*. New York: Plenum Press.
- Beveridge DL, DiCapua FM. 1989. Free energy via molecular simulation. *Annu Rev Biophys Chem* 18:431-492.
- Brooks CL, Boczko EM. 1995. First principles calculation of the folding free energy of a three-helix bundle protein. *Science* 269:393-396.
- Connolly ML. 1983. Analytical molecular surface calculation. *J Appl Crystallogr* 16:548-558.
- Daggett V, Levitt M. 1993. Protein unfolding pathways explored through molecular dynamics simulations. *J Mol Biol* 232:600-619.
- Daggett V, Levitt M. 1994. Protein folding \leftrightarrow unfolding dynamics. *Curr Opin Struct Biol* 4:291-295.
- Dill K. 1987. The stabilities of globular proteins. In: Oxender DL, Fox CF, eds. *Protein engineering*. pp 187-192.
- Dill K. 1990. Dominant forces in protein folding. *Biochemistry* 29:7133-7155.
- Hubbard SJ, Thornton JM. 1993. *NACCESS* computer program. London: University College, Department of Biochemistry and Molecular Biology.
- Itzhaki LS, Otzen DE, Fersht AR. 1995. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: Evidence for a nucleation-condensation mechanism for protein folding. *J Mol Biol* 254:260-288.
- Jackson RM, Sternberg MJ. 1994. Application of scaled particle theory to model the hydrophobic effect: Implications for molecular association and protein stability. *Protein Eng* 7:371-384.
- Jones DT, Thornton JM. 1996. Potential energy functions for threading. *Curr Opin Struct Biol* 6:210-216.
- Jorgenson WL, Buckner JK, Boudon S, Tirado-Rives J. 1988. Efficient computation of absolute free energies of binding by computer simulations. Application to the methane dimer in water. *J Chem Phys* 89:3742.
- Jorgenson WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. 1983. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926-935.
- Karplus M, Sali A. 1995. Theoretical studies of protein folding and unfolding. *Curr Opin Struct Biol* 5:58-73.
- Kauzmann W. 1959. Some factors in the interpretation of protein denaturation. *Adv Protein Chem* 14:1-63.
- Kiefhaber T, Baldwin RL. 1995. Kinetics of hydrogen bond breakage in the process of unfolding of ribonuclease A measured by pulsed hydrogen exchange. *Proc Natl Acad Sci USA* 92:2657-2661.
- Li A, Daggett V. 1994. Characterization of the transition state of protein unfolding by use of molecular dynamics: Chymotrypsin inhibitor 2. *Proc Natl Acad Sci USA* 91:10430-10434.
- Pratt LR, Pohorille A. 1992. Theory of hydrophobicity: Transient cavities in molecular liquids. *Proc Natl Acad Sci USA* 89:2995-2999.
- Reiss H. 1965. Scaled particle methods in the statistical thermodynamics of fluids. *Adv Chem Phys* 9:1-84.
- Richards FM. 1977. Areas, volumes, packing and protein structure. *Annu Rev Biophys Chem* 6:151-176.
- Serrano L, Matouschek A, Fersht AR. 1992. The folding of an enzyme. III. Structure of the transition state for unfolding of barnase analysed by a protein engineering procedure. *J Mol Biol* 224:805-818.
- Sippl MJ. 1995. Knowledge-based potentials for proteins. *Curr Opin Struct Biol* 5:229-235.
- Smith DE, Haymet ADJ. 1993. Free energy, entropy, and internal energy of hydrophobic interactions: Computer simulations. *J Chem Phys* 98:6445-54.
- Tunon I, Silla E, Pascual-Ahuir JL. 1992. Molecular surface area and the hydrophobic effect. *Protein Eng* 5:715-716.
- Vidugiris GJA, Markley JL, Royer CA. 1995. Evidence for a molten globule-like state in protein folding from determination of activation volumes. *Biochemistry* 34:4909-4912.
- Waldburger CD, Jonsson T, Sauer RT. 1996. Barriers to protein folding: Formation of buried polar interactions is a slow step in acquisition of structure. *Proc Natl Acad Sci USA* 93:2629-2634.
- Wallqvist A. 1991. Molecular dynamics study of a hydrophobic aggregate in an aqueous solution of methane. *J Phys Chem* 95:8921-8927.
- Wolyne PG, Onuchic JN, Thirumalai D. 1995. Navigating the folding routes. *Science* 267:1619-1620.
- Wood RH, Thompson PT. 1990. Differences between pair and bulk hydrophobic interactions. *Proc Natl Acad Sci USA* 87:946-949.
- Yi Q, Baker D. 1996. Direct evidence for a two-state protein unfolding transition from hydrogen-deuterium exchange, mass spectrometry, and NMR. *Protein Sci* 5:1060-1066.