

Temperature dependence of hydrophobic interactions: A mean force perspective, effects of water density, and nonadditivity of thermodynamic signatures

Seishi Shimizu and Hue Sun Chan^{a)}

Departments of Biochemistry and Medical Genetics and Microbiology, Faculty of Medicine, University of Toronto, Toronto, Ontario M5S 1A8, Canada

(Received 28 April 2000; accepted 21 June 2000)

Temperature-dependent properties of hydrophobic interactions are investigated by simulating the potential of mean force (PMF) between two methane-like solutes in TIP4P model water. Independent results from test particle insertion and free energy perturbation are compared to ensure that zero-PMF baselines are accurate. PMFs are computed under atmospheric pressure at five temperatures from 5 to 95 °C using constant-pressure simulations. The temperature dependence we observe does not agree with previous results from constant-volume simulations, highlighting the important effects of temperature-dependent water density on PMFs. Heat capacity changes upon association of two solutes are estimated at the PMF contact minimum, desolvation barrier, and the solvent (water)-separated minimum. The magnitude of the heat capacity change upon contact formation is much smaller than that predicted by the solvent accessible surface area (SASA). More surprisingly, the heat capacity change upon bringing two methanes from infinity to the desolvation barrier is large and positive. This implies that the thermodynamic signatures of the free energy barrier to desolvation have signs opposite to desolvation itself. This feature is not predicted by either SASA or a volume-based solvent exclusion model. The implications of these and other observations on implicit-solvent model potentials are discussed. Formulations based on thermodynamic perturbation and Widom's potential distribution theory are developed to relate PMF and hydration mean forces to the underlying structural properties of aqueous solutions. In particular, we provide a theoretical perspective to understand PMF in terms of local water density and the occurrences of configurations with highly unfavorable solute-solvent repulsive interactions. © 2000 American Institute of Physics. [S0021-9606(00)51435-4]

I. INTRODUCTION

Hydrophobic interaction has long been hypothesized to be one of the dominant driving forces in biomolecular processes such as protein folding.¹⁻³ A defining thermodynamic signature of hydrophobic interaction is a large increase in heat capacity when a nonpolar group is exposed to water; or equivalently, a large negative heat capacity change upon desolvation, as when a nonpolar solute is transferred from water to an oil solvent or to a gas phase.⁴⁻⁷ Inasmuch as hydrophobic interaction is a major contributor to protein stability, and the unfolding process may be viewed as a summation of solvation or transfer processes involving small constituent groups, heat capacity changes observed from small-molecule experiments provide a plausible rationalization of the large positive heat capacity change that accompanies protein unfolding^{2,8-12} and the related phenomenon of heat and cold denaturation.¹³⁻¹⁷

The small-solute transfer approach has been conceptually useful, but many issues remain to be clarified. For instance, analyses based on this approach have often depended on the use of random-coil-like chains to calculate denatured state properties,¹⁸ but experimental evidence indicates that

heat-denatured states are actually quite compact.¹⁹⁻²¹ Moreover, conventional approaches¹⁸ assume that the heat capacity gain upon protein denaturation arises almost exclusively from solvation effects. However, recent nuclear magnetic resonance experiments²² and theoretical considerations²³⁻²⁶ indicate that a significant fraction of this change may arise from conformational transitions in the denatured ensemble.

In the quest for tractable models to study complex biomolecular processes, implicit-solvent effective potentials^{23,27} have been constructed to circumvent the need to perform computationally intensive simulations with a large number of explicit water molecules. These approaches seek to approximate solvent (water)-mediated interactions by simple functional forms that depend on each chemical group's exposure to water. A number of measures have been used to quantify exposure. These include conventional surface area methods⁴⁻⁷ and volume-based solvent-exclusion approaches,^{23,24} among others. The interaction parameters in these models are often obtained from small-solute transfer experiments.^{18,23}

However, experimental measurements have demonstrated that the relationship between "bulk hydrophobic interaction" [exposure of hydrophobic residues from its pure phase to water] and "pair hydrophobic interaction" [potential of mean force (PMF) in water] is complex.²⁸ Computa-

^{a)} Author to whom correspondence should be addressed; electronic mail: chan@arrhenius.med.toronto.edu

tionally, it has been shown that the simple solvent accessible surface area approach fails to account accurately for the spatial dependence of PMF,^{29–31} and that water-mediated interactions not only depend on the exposed surface area, but are sensitive also to solute size and the curvature of the exposed surface.^{32–36} In view of these, while recognizing the desirability of implicit-solvent effective potentials, it is important to delineate their strengths and limitations.

To address some of these issues, we undertake to investigate the temperature dependence of hydrophobic PMFs between two small methane-like solutes. Recently, Lum *et al.*³⁷ proposed that hydrophobic mechanisms can depend drastically on the collective sizes of nonpolar groups involved; and they envisioned a novel “drying” mechanism for the attractive interactions between hydrophobic surfaces with area greater than $\sim 100 \text{ \AA}^2$. We do not address this issue in this paper. As a necessary first step, we focus here on two small solutes.

Insight into hydrophobic interaction has been gained by extensive experimental measurements of virial coefficients, Kirkwood–Buff integrals, and related spatially integrated quantities.^{3,28,38–40} However, the spatial dependence of PMF itself is currently inaccessible to direct experimental determination. Computer simulation is therefore an indispensable tool. Over the past 2 decades, a large number of publications have been devoted to PMF calculations.^{31,41–55} Among these, however, relatively few^{46,48,50,51,53} have tackled the temperature dependence of PMFs; and some of the key results from different groups appear to be contradictory (see below). Temperature dependence of hydrophobic effects has also been investigated by Skipper,⁵⁶ who simulated a methane solution (four methanes in 256 water molecules) at a fixed water density, and concluded that the tendency of methane molecules to aggregate increases with temperature from 275 to 317 K. This was followed by more extensive studies^{57–59} using similar systems of semiconcentrated methane solutions. However, these results are not straightforward to interpret in terms of two-methane interactions, because effects observed in these systems might have arisen from three- or four-methane interactions as well as solute configurational entropy.

Dang⁴⁶ calculated two-methane PMFs at two different temperatures (300 and 330 K). Different water densities were used for the two temperatures, presumably to maintain the two systems at essentially the same (atmospheric) pressure.⁵⁰ In contrast to Skipper,⁵⁶ Dang reported very little temperature dependence, and suggested that the discrepancy might be caused by Skipper’s use of a temperature-independent water density. Subsequently, Lüdemann *et al.*^{50,51} calculated two-methane PMFs from 250 to 500 K in an *NVT* ensemble (at a fixed water density), and reported a significant temperature dependence at the contact minimum. (The total number of molecules N , volume V , and temperature T are kept constant in an *NVT* ensemble.) They also compared two PMFs at the same temperature but with different water densities. In contrast to Dang,⁴⁶ they found no substantial pressure effects. In view of these disagreements, it is necessary to reexamine the issue of density dependence more carefully.

Other studies in the literature do not appear to support

the assertion of Lüdemann *et al.*⁵⁰ that the density dependence of PMF is negligible. Hydrophobic hydration has been shown to be density dependent,⁶⁰ and PMFs have been shown to be pressure dependent.^{54,55} Moreover, a theoretical consideration⁶¹ and an analysis of experimental data⁶² show that the difference between enthalpy measurements made at constant pressure versus that at constant volume can be of the same order of magnitude as the enthalpies themselves. Using a polarizable water model, Rick and Berne⁵³ have conducted *NPT* PMF simulations at three temperatures, and observed a significant difference between constant-volume and constant-pressure enthalpies of hydrophobic interaction. (The total number of molecules N , pressure P , and temperature T are kept constant in an *NPT* ensemble.) Therefore, based on these results, we expect significant differences in temperature dependence between PMFs obtained from constant-volume (constant-density) versus that from constant-pressure simulations. To test these ideas, we calculated PMFs at five temperatures in *NPT* ensembles.

The work presented in this paper is not limited to the calculation of model PMFs. We also seek to understand PMFs in terms of structural properties of water around the solutes. Detailed aspects of this type of calculation are likely to depend somewhat on the water model used.⁵² Therefore, it is imperative that efforts should be made to use model simulation as a vehicle to explore and infer general physical principles.

In their seminal work, Pratt and Chandler⁶³ employed the Ornstein–Zernike equation and a perturbation approximation to predict solute–solute PMFs from experimental water–water distribution functions. Their theory relates PMFs to the structure of pure water. In the present work, on the other hand, we are interested in developing formulations that relate PMFs to structural properties of water in the presence of the solutes. Previous approaches in this general direction include Geiger *et al.*,⁶⁴ who related structural properties of water clathrates to the heat capacity of hydrophobic interactions. Recently, systematic model simulations by Head-Gordon⁶⁵ confirmed the existence of various forms of water clathrates around hydrophobic solutes, and elucidated how their properties depend on the configuration of a solute pair. Studies by Pangali *et al.*⁴¹ and Zichi and Rosky⁶⁶ presented useful structural functions. These studies have been instrumental in relating the excess heat capacity and excess enthalpy of hydrophobic interaction to the structure of water. However, they did not attempt to quantitatively rationalize the PMF’s spatial dependence. More recently, a new information theory of hydrophobic interaction,⁴⁹ which may be viewed as an extension⁶⁷ of scaled particle theory,^{68–70} was successful in reproducing the PMF curves between two cavities or hard-sphere solutes, and solute–solvent attraction has been incorporated by perturbative approximation. But a direct relationship between PMF and the structure of water around the solute pair has not been investigated in this framework.

The organization of the rest of this paper is as follows. We first introduce the necessary theory and computational methods in Secs. II and III. Results of our systematic calculation of temperature-dependent PMFs are given in Sec. IV.

Their ramifications for protein folding are discussed in Sec. V. We present a physical perspective to rationalize some of our results in terms of hydration mean forces and local water density in Sec. VI. This is followed by a discussion of the relation between PMFs and potential distributions in Sec. VII. Section VIII explores the correlation between local water density and hydrogen bonding among water molecules. Finally, we conclude by summarizing our key findings in Sec. IX.

II. THEORY

We now derive two statistical mechanical formulations to connect local density of water to hydrophobic PMFs. They are based on (A) a mean force and (B) a potential distribution approach. They account for the “indirect” or hydration part of the PMF, which we denote as ΔG_h . The full PMF is the sum of ΔG_h and the direct solute–solute interaction. (The latter is modeled here by a Lennard-Jones potential). A connection between hydrogen bonding and local water density is outlined at the end of this section.

A. The hydration mean force

1. Relationship between water distribution and the average force exerted on a solute

Consider a system of N water molecules at positions $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$, and two methanes (labeled a and b) separated by a distance ξ . The change in hydration Gibbs free energy incurred by changes in solute–water interactions upon changing ξ from ξ_0 to ξ_1 equals

$$\Delta G_h = G_h(\xi_1) - G_h(\xi_0) = -kT \int_{\xi_0}^{\xi_1} d\xi \frac{\partial}{\partial \xi} \ln Z_{ab}(\xi), \quad (1)$$

where kT is Boltzmann constant times absolute temperature. Here the partition function

$$Z_{ab}(\xi) = \int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N e^{-\beta U_N} e^{-\beta U_a(\mathbf{r}_a)} e^{-\beta U_b}, \quad (2)$$

where $\beta = 1/kT$, U_N , $U_a(\mathbf{r}_a)$, and U_b are the interaction energy among all solvent (water) molecules, between solute a and all water molecules, and between solute b and all water molecules, respectively. We place solute b at the origin, and solute a at position \mathbf{r}_a . U_N , U_a , and U_b are functions of $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$, where U_a depends on \mathbf{r}_a as well. As the two solutes are in a homogeneous solvent, Z_{ab} is a function of $\xi = |\mathbf{r}_a|$. It follows that the component of Boltzmann-averaged force or the *mean force*⁷¹ exerted by the water molecules on solute a along the direction of \mathbf{r}_a is given by

$$kT \frac{\partial \ln Z_{ab}(\xi)}{\partial \xi} = \frac{1}{Z_{ab}(\xi)} \int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N \times \left[-\frac{\partial U_a}{\partial \xi} \right] e^{-\beta U_N} e^{-\beta U_a(\mathbf{r}_a)} e^{-\beta U_b}. \quad (3)$$

If the interactions between the solute a and the water molecules are additive, as they are in the model used here, $U_a = \sum_{i=1}^N u_{ai}(\mathbf{r}_a, \mathbf{r}_i)$, where u_{ai} is the potential energy between

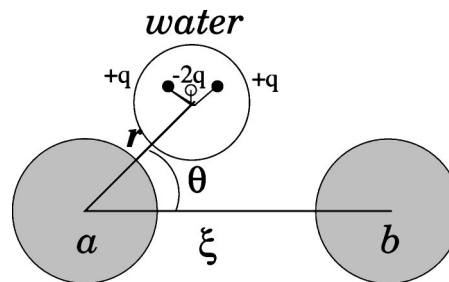


FIG. 1. Geometry of the model system. Each methane (shaded sphere) is represented by a united atom. The parameters (σ, ϵ) in the Lennard-Jones potential $4\epsilon[(\sigma/r)^{12} - (\sigma/r)^6]$, in units of (\AA , kcal/mol), for water–water, methane–methane, and water–methane interactions are (3.15365, 0.1550), (3.730, 0.294), and (3.44183, 0.2135), respectively. The charge q on an hydrogen atom is equal to $0.52e$, where e is the electronic charge (Refs. 84 and 113). This figure is drawn to scale. The radii of the spheres correspond to the molecules’ van der Waals radii. The positions of the charges (small circles) in the TIP4P water are depicted according to the same scale.

a and the i th water molecule. This means that the above integral can be written as a sum of N terms. Since the water molecules are indistinguishable,

$$\begin{aligned} kT \frac{\partial \ln Z_{ab}(\xi)}{\partial \xi} &= - \int d\mathbf{r}_1 \frac{\partial u_{a1}}{\partial \xi} \\ &\times \left\{ \frac{N}{Z_{ab}(\xi)} \int d\mathbf{r}_2 d\mathbf{r}_3 \cdots d\mathbf{r}_N e^{-\beta U_N} e^{-\beta U_a(\mathbf{r}_a)} e^{-\beta U_b} \right\} \\ &= - \int d\mathbf{r}_1 \frac{\partial u_{a1}}{\partial \xi} \rho(\mathbf{r}_a, \mathbf{r}_1), \end{aligned} \quad (4)$$

where $\rho(\mathbf{r}_a, \mathbf{r}_1)$ is the distribution function (number density) of water molecules at position \mathbf{r}_1 in the presence of solute a at \mathbf{r}_a and solute b at the origin. Using the fact that $\xi = |\mathbf{r}_a|$,

$$\frac{\partial u_{a1}}{\partial \xi} = \frac{\partial u_{a1}}{\partial r} \frac{\partial r}{\partial \xi} = \frac{\partial u_{a1}}{\partial r} \cos \theta, \quad (5)$$

where r is the distance between solute a and water 1, and θ is the angle between the intersolute vector $\mathbf{r}_b - \mathbf{r}_a$ and the direction $\mathbf{r}_1 - \mathbf{r}_a$ from solute a to water 1; see Fig. 1. This leads to the following connection between local water density and the hydration mean force induced by one solute on the other:

$$-\frac{\partial G_h}{\partial \xi} = - \int d\mathbf{r}_1 \rho(\mathbf{r}_a, \mathbf{r}_1) \frac{\partial u_{a1}}{\partial r} \cos \theta. \quad (6)$$

According to Eq. (6), the hydration mean force is generated by “collisions” between a solute and the water molecules around it, each water exerting a force of $-(\partial u_{a1}/\partial r)$ on the solute along the direction $\mathbf{r}_a - \mathbf{r}_1$. As noted above, the distribution function $\rho(\mathbf{r}_a, \mathbf{r}_1)$ in the integrand is the *local water density*—which is a property of the *water structure* around the solute.⁷² Here it tells us on average how many water molecules at position \mathbf{r}_1 are available to collide with the solute. The net hydration mean force is obtained by summing the force components in the \mathbf{r}_a direction (hence the $\cos \theta$ factor) from all *local* contributions. Equation (6) has been

used previously^{73,74} to calculate PMFs from simulated local water density. Naturally, dominant contributions are expected to come from water molecules very close to solute a , owing to the large magnitude of $-(\partial u_{a1}/\partial r)$ at small r . Therefore, in this view, penetration of water molecules into the core of solute a is the most significant process that contributes to the hydration mean force.

2. Angular dependence of the hydration mean force

In a spherical coordinate system that centers at the position of solute a ,

$$-\frac{\partial G_h}{\partial \xi} = -2\pi \int_0^\infty dr \int_0^\pi d\theta r^2 \sin \theta \rho(\mathbf{r}_a, \mathbf{r}_1) \frac{\partial u_{a1}}{\partial r} \cos \theta$$

$$= -\pi \int_0^\infty dr \int_0^\pi d\theta r^2 \sin 2\theta \rho(\xi, r, \theta) \frac{\partial u_{a1}}{\partial r}; \quad (7)$$

see Fig. 1. In this equation, the factor $\pi r^2 \sin 2\theta d\theta$ is the product of the projection $\cos \theta$ on the intersolute axis, and the circular area element $2\pi r \sin \theta (rd\theta)$ at θ . Since $\sin 2\theta$ is maximum (minimum) when $2\theta = \pi/2$ ($3\pi/2$), for a given local density ρ at a small r , the strongest contributions to the hydration mean force are from the directions $\theta = \pi/4$ (which favors dissociation of the solutes) and $3\pi/4$ (which favors association of the solutes).

B. A potential distribution approach

1. Potential distribution theory for PMFs

The mean force formulation above accounts for the derivative of PMF. We now use potential distribution theory^{75,76} to relate PMF directly to structural properties of water. To this end, the full PMF is expressed as a difference

$$\text{PMF} = \Delta G(\xi) = \mu_{ab}^*(\xi) - \mu_a^* \quad (8)$$

between two (pseudo⁷⁷) chemical potentials of solute a . Here $\Delta G(\xi) = \Delta G_h(\xi) + U_{ab}(\xi)$, where U_{ab} is the direct interaction between the two solutes; $\mu_{ab}^*(\xi)$ corresponds to inserting a at a fixed position ξ apart from a solute b that is already in the aqueous solution, whereas μ_a^* describes a fixed-position insertion into pure water. In the canonical ensemble, μ_a^* may be calculated by the standard relation

$$e^{-\beta\mu_a^*} = \int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N$$

$$\times \left(\frac{e^{-\beta U_N}}{\int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N e^{-\beta U_N}} \right) e^{-\beta U_a}$$

$$\equiv \langle e^{-\beta U_a} \rangle_N, \quad (9)$$

which is the average of $\exp(-\beta U_a)$ in pure water. This is the basis of the test-particle insertion approach to computing chemical potentials.⁷⁵⁻⁸⁰ On the other hand, if one is interested in a relationship between μ_a^* and the structure of water around the solutes rather than that of pure water, one may rewrite Eq. (9) as

$$e^{\beta\mu_a^*} = \int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N$$

$$\times \left(\frac{e^{-\beta(U_N+U_a)}}{\int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N e^{-\beta(U_N+U_a)}} \right) e^{\beta U_a}$$

$$\equiv \langle e^{\beta U_a} \rangle_{N,a}, \quad (10)$$

which requires averaging $\exp(\beta U_a)$ over water configurations in the presence of the solute. Widom^{75,76} showed that this averaging can be performed by first defining a distribution function

$$g_a(\epsilon) = \frac{\int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N \delta(\epsilon - U_a) e^{-\beta U_N} e^{-\beta U_a}}{\int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N e^{-\beta U_N} e^{-\beta U_a}} \quad (11)$$

of the potential energy ϵ between solute a and all water molecules. (δ is the Dirac delta function.) Then

$$e^{\beta\mu_a^*} = \int d\epsilon g_a(\epsilon) e^{\beta\epsilon}. \quad (12)$$

A similar potential distribution function

$$g_{ab}(\epsilon; \xi) = \frac{\int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N \delta(\epsilon - U_a) e^{-\beta U_N} e^{-\beta U_a} e^{-\beta U_b}}{\int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N e^{-\beta U_N} e^{-\beta U_a} e^{-\beta U_b}} \quad (13)$$

is defined for the hydration part of the two-solute $\mu_{ab}^*(\xi)$ to yield

$$e^{\beta[\mu_{ab}^*(\xi) - U_{ab}(\xi)]} = \int d\epsilon g_{ab}(\epsilon; \xi) e^{\beta\epsilon}. \quad (14)$$

It follows from Eqs. (8), (12), and (14) that

$$\Delta G_h(\xi) = kT \ln \int d\epsilon g_{ab}(\epsilon; \xi) e^{\beta\epsilon}$$

$$- kT \ln \int d\epsilon g_a(\epsilon) e^{\beta\epsilon}, \quad (15)$$

or, equivalently,

$$e^{\beta\Delta G_h(\xi)} = \frac{\int d\epsilon g_{ab}(\epsilon; \xi) e^{\beta\epsilon}}{\int d\epsilon g_a(\epsilon) e^{\beta\epsilon}}. \quad (16)$$

The hydration part of PMF is now seen to be determined by the change in the Boltzmann average of $\exp(\beta\epsilon)$ upon introducing the second solute b . Because this factor depends exponentially on $+\epsilon$, dominant contributions to these averages arise from configurations with high ϵ values, when there are strong repulsive interactions between solute a and the surrounding water molecules. In other words, as far as its hydration part is concerned, PMF is determined by the shift in the degree of water penetration into solute a upon the introduction of b . The relation

$$e^{\beta\Delta G_h(\xi)} - 1 = e^{-\beta\mu_a^*} \int d\epsilon [g_{ab}(\epsilon; \xi) - g_a(\epsilon)] e^{\beta\epsilon}, \quad (17)$$

which is readily obtained from Eqs. (8), (12), and (14), shows that $\Delta G_h(\xi)$ has the same sign as the integral of $(g_{ab} - g_a)\exp(\beta\epsilon)$.

2. Generalization to constant-pressure *NPT* ensembles

The potential distribution approach introduced above for canonical *NVT* ensembles can be readily extended to constant-pressure ensembles by using the *NPT* expressions of Shing and Chung^{78,81}

$$e^{\beta\mu_a} = \frac{\int dV \int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N e^{-\beta U_N} e^{-\beta PV}}{\int dV \int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N V e^{-\beta U_N} e^{-\beta U_a} e^{-\beta PV}}, \quad (18)$$

$$e^{\beta(\mu_{ab} - U_{ab})} = \frac{\int dV \int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N e^{-\beta U_N} e^{-\beta U_a} e^{-\beta U_b} e^{-\beta PV}}{\int dV \int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N V e^{-\beta U_N} e^{-\beta U_a} e^{-\beta U_b} e^{-\beta PV}},$$

where P is the (constant) pressure and V is the (variable) volume of the system, and μ_a and μ_{ab} are full chemical potentials. In general, a full chemical potential μ is related to its corresponding pseudochemical potential μ^* by $\mu = \mu^* + kT \ln \rho$, where ρ is the number density of the solute, and $kT \ln \rho$ is the ‘‘liberation’’ term that accounts for the center-of-mass translational freedom of the solute.⁷⁷ This implies that the pseudochemical potentials μ_a^* and μ_{ab}^* can be obtained from the above expressions as $\mu_a + kT \ln \langle V \rangle_{N,a}$ and $\mu_{ab} + kT \ln \langle V \rangle_{N,ab}$, where

$$\langle V \rangle_{N,a} = \frac{\int dV V \int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N e^{-\beta U_N} e^{-\beta U_a} e^{-\beta PV}}{\int dV \int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N e^{-\beta U_N} e^{-\beta U_a} e^{-\beta PV}}, \quad (19)$$

$$\langle V \rangle_{N,ab} = \frac{\int dV V \int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N e^{-\beta U_N} e^{-\beta U_a} e^{-\beta U_b} e^{-\beta PV}}{\int dV \int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N e^{-\beta U_N} e^{-\beta U_a} e^{-\beta U_b} e^{-\beta PV}},$$

are the average solution volumes with one and two solutes, respectively.⁸² It follows that

$$e^{\beta\mu_a^*} = \frac{\int dV \int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N e^{-\beta U_N} e^{-\beta PV}}{\int dV \int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N e^{-\beta U_N} e^{-\beta U_a} e^{-\beta PV}}, \quad (20)$$

$$e^{\beta(\mu_{ab}^* - U_{ab})} = \frac{\int dV \int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N e^{-\beta U_N} e^{-\beta U_a} e^{-\beta U_b} e^{-\beta PV}}{\int dV \int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N e^{-\beta U_N} e^{-\beta U_a} e^{-\beta U_b} e^{-\beta PV}}.$$

Naturally, the *NPT* version of the potential distributions are defined as

$$g_a(\epsilon) = \frac{\int dV \int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N e^{-\beta U_N} e^{-\beta U_a} e^{-\beta PV} \delta(\epsilon - U_a)}{\int dV \int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N e^{-\beta U_N} e^{-\beta U_a} e^{-\beta PV}}, \quad (21)$$

$$g_{ab}(\epsilon; \xi) = \frac{\int dV \int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N e^{-\beta U_N} e^{-\beta U_a} e^{-\beta U_b} e^{-\beta PV} \delta(\epsilon - U_a)}{\int dV \int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N e^{-\beta U_N} e^{-\beta U_a} e^{-\beta U_b} e^{-\beta PV}}.$$

Equations (12) and (14), which form the basis of the formal development above, remain valid in the present *NPT* formulation.

3. Shape of potential energy distributions

Now that a connection has been established between PMF and potential distribution functions, it is useful to know to what extent these functions can be approximated by Gaussians. The motivation is simple: If these g functions are well approximated by Gaussians, the first two terms in the cumulant expansion

$$\Delta G_h(\xi) = (\langle U_a \rangle_{N,ab} - \langle U_a \rangle_{N,a}) + \frac{\beta}{2} (\langle \Delta U_a^2 \rangle_{N,ab} - \langle \Delta U_a^2 \rangle_{N,a}) + \dots \quad (22)$$

of Eq. (15) would be sufficient to reproduce key features of the hydration part of PMF. The average solute–water energies in this expansion are given by

$$\langle U_a \rangle_{N,ab} \equiv \int d\epsilon [\epsilon g_{ab}(\epsilon; \xi)], \quad (23)$$

$$\langle U_a \rangle_{N,a} \equiv \int d\epsilon [\epsilon g_a(\epsilon)];$$

the corresponding variances in energy are

$$\langle \Delta U_a^2 \rangle_{N,ab} \equiv \int d\epsilon [\epsilon - \langle U_a \rangle_{N,ab}]^2 g_{ab}(\epsilon; \xi), \quad (24)$$

$$\langle \Delta U_a^2 \rangle_{N,a} \equiv \int d\epsilon [\epsilon - \langle U_a \rangle_{N,a}]^2 g_a(\epsilon).$$

In Sec. VIIC, we will use these formulas to examine the potential distributions we obtain from simulations.

C. Connection between local water density and water–water hydrogen bonding

We have thus far connected hydration mean forces to local water density, and the PMF to potential distribution functions. These relations are useful as a first step toward a microscopic molecular understanding of hydrophobic interactions. We now take a further step by introducing formulations to relate local water density to hydrogen bonding properties of water. To do this, we first derive a relation between local water density and pairwise water interaction energies. This is motivated by the fact that the latter can be compared against certain established energetic criteria for interwater hydrogen bonding.⁸³

Consider the local density of water $\rho(\mathbf{r}_1)$ at position \mathbf{r}_1 when a solute is present at \mathbf{r}_a

$$\rho(\mathbf{r}_1) = \frac{N \int d\mathbf{r}_2 d\mathbf{r}_3 \cdots d\mathbf{r}_N e^{-\beta U_N(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)} e^{-\beta U_a}}{\int d\mathbf{r}'_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N e^{-\beta U_N(\mathbf{r}'_1, \mathbf{r}_2, \dots, \mathbf{r}_N)} e^{-\beta U_a}}. \quad (25)$$

This expression is identical to that for ρ in Eq. (4) except that for simplicity we consider only one solute here. It is straightforward to generalize the present discussion to cases with multiple solutes. In Eq. (25), the position of water 1 is integrated in the denominator but not in the numerator (i.e., \mathbf{r}'_1 is a dummy integration variable in the denominator), $U_N = U_1(\mathbf{r}_1) + U_{N-1}$, where $U_1(\mathbf{r}_1)$ is the interaction of the water 1 with the rest of the water molecules and U_{N-1} is the

interaction energy among the other $N-1$ water molecules. (To simplify notation, U_1 and U_{N-1} 's dependences on $\mathbf{r}_2, \mathbf{r}_3 \dots, \mathbf{r}_N$ are not indicated.) Similarly, the interaction U_a between solute a and the N water molecules can be written as

$U_{a,N-1} + u_{a1}$, where $u_{a1} = u_{a1}(\mathbf{r}_a, \mathbf{r}_1)$ is the potential energy between water 1 and solute a , and $U_{a,N-1}$ is the interaction between the solute and the other $N-1$ water molecules. Using these decompositions, Eq. (25) becomes

$$\rho(\mathbf{r}_1) = \frac{N \int d\mathbf{r}_2 d\mathbf{r}_3 \dots d\mathbf{r}_N e^{-\beta U_{N-1}} e^{-\beta U_1(\mathbf{r}_1)} e^{-\beta U_{a,N-1}} e^{-\beta u_{a1}}}{\int d\mathbf{r}_2 d\mathbf{r}_3 \dots d\mathbf{r}_N e^{-\beta U_{N-1}} e^{-\beta U_{a,N-1}}} \left[\frac{\int d\mathbf{r}_1 d\mathbf{r}_2 \dots d\mathbf{r}_N e^{-\beta U_N} e^{-\beta U_a}}{\int d\mathbf{r}_2 d\mathbf{r}_3 \dots d\mathbf{r}_N e^{-\beta U_{N-1}} e^{-\beta U_{a,N-1}}} \right]^{-1}$$

$$\propto \frac{N}{e^{-\beta \mu_{\text{water}}}} \frac{1}{\langle e^{\beta U_1} e^{\beta u_{a1}} \rangle_{a, \mathbf{r}_1}}, \quad (26)$$

where μ_{water} is the chemical potential of water in the system, and $\langle \dots \rangle_{a, \mathbf{r}_1}$ is the Boltzmann average with the position of water 1 fixed at \mathbf{r}_1 , and the position of solute fixed at \mathbf{r}_a . This equation was obtained by Widom.^{75,76} Since u_{a1} depends only on the position of water 1 and the solute, it is not affected by the $\langle \dots \rangle_{a, \mathbf{r}_1}$ average. Hence

$$\rho(\mathbf{r}_1) e^{\beta u_{a1}(\mathbf{r}_a, \mathbf{r}_1)} \propto \frac{N}{e^{-\beta \mu_{\text{water}}}} \frac{1}{\langle e^{\beta U_1} \rangle_{a, \mathbf{r}_1}}. \quad (27)$$

Equation (27) says that, besides the contribution from the single-pair solute–water interaction $\exp[\beta u_{a1}(\mathbf{r}_a, \mathbf{r}_1)]$, local water density is determined by two factors: (i) The contribution $\exp(\beta \mu_{\text{water}})$ which is independent of \mathbf{r}_1 . When the solute concentration is sufficiently dilute, this quantity is essentially equal to the chemical potential of pure water. (ii) The contribution $(\langle \exp(\beta U_1) \rangle_{a, \mathbf{r}_1})^{-1}$ that depends on \mathbf{r}_1 [because $U_1 = U_1(\mathbf{r}_1)$]. This implies that fewer occurrences of high U_1 values are concomitant with a larger local water density.

Therefore, $\langle \exp(\beta U_1) \rangle_{a, \mathbf{r}_1}$ emerges as a determining factor of local water density. It is useful for subsequent analyses to define a distribution function

$$g_{\text{water}}(E_{\text{ww}}, \mathbf{r}_1) = \frac{\int d\mathbf{r}_2 d\mathbf{r}_3 \dots d\mathbf{r}_N \delta(U_1 - E_{\text{ww}}) e^{-\beta U_N} e^{-\beta U_a}}{\int d\mathbf{r}_2 d\mathbf{r}_3 \dots d\mathbf{r}_N e^{-\beta U_N} e^{-\beta U_a}} \quad (28)$$

of the interaction (binding) energy E_{ww} between water molecule 1 with the other $N-1$ water molecules, such that

$$\langle e^{\beta U_1} \rangle_{a, \mathbf{r}_1} = \int dE_{\text{ww}} g_{\text{water}}(E_{\text{ww}}, \mathbf{r}_1) e^{\beta E_{\text{ww}}}. \quad (29)$$

Because of the exponential factor $\exp(\beta E_{\text{ww}})$, water configurations with high binding (i.e., unfavorable) energies are expected to have dominant effects on local water density.

In this paper, we adopt an energetic criterion for hydrogen bonding,⁸³ in which a pair of water molecules having an interaction energy lower than -2.238 kcal/mol (-9.364 kJ/mol) is defined to be hydrogen bonded. The binding energy distribution thus has a direct bearing on the number of inter-water hydrogen bonds. It follows that the relationship between the effects of solutes on local water density and inter-water hydrogen bonding can be studied in this framework.

III. COMPUTATIONAL METHODS

All results in this work are produced by *NPT* Monte Carlo simulations of 396 TIP4P water molecules in a box with periodic boundary conditions. We use the software BOSS version 4.1.⁸⁴ The cutoff distance of water–water and methane–water interactions is set to 9.0 Å. Energetic parameters and potential functions are given in Fig. 1. A Monte Carlo step represents an attempted move on one molecule. For a solvent (water) molecule, such a step consists of an attempted translational displacement as well as an attempted rotational displacement. In our simulations, the ranges of translational and rotational displacements are ± 0.15 Å (components of the linear displacement along the three axes of a fixed Cartesian frame) and $\pm 15^\circ$, respectively. The rotational displacement of a given water molecule is made around an axis randomly chosen among the three axes of a fixed Cartesian frame, whereas the translational displacement is made in a randomly chosen direction, not restricted to the three Cartesian axes.⁸⁴ The solute is spherically symmetric. The solute translational move (± 0.08 Å) and volume move (± 190.0 Å³) are attempted every 90 and 2,375 steps, respectively. (When there are a pair of solutes in the box, the solutes move in tandem so that their separation ξ is unchanged by the move.) Acceptance ratios of attempted moves are about 40%. Simulations are performed on DEC-alpha's (533 MHz) and Pentium (350 and 450 MHz) machines. Two independent methods are employed to compute PMFs.

A. PMF computation

1. Free energy perturbation

The first method uses the standard free energy perturbation technique.⁷⁸ We calculate the Gibbs free energy difference

$$\Delta G(\xi_i \rightarrow \xi_i + \Delta \xi)$$

$$\equiv G(\xi_i + \Delta \xi) - G(\xi_i)$$

$$= -kT \ln \langle \exp\{\beta[U(\xi_i) - U(\xi_i + \Delta \xi)]\} \rangle_{\xi} \quad (30)$$

between the systems of intersolute separations ξ_i and $\xi_i + \Delta \xi$, then we integrate the Gibbs free energy difference over ξ to obtain the full PMF curve. In Eq. (30), the average $\langle \dots \rangle_{\xi}$ is over the (unperturbed) ensemble with intersolute separation ξ . In our simulation, the perturbation step $\Delta \xi$

$=0.05 \text{ \AA}$, and double-wide sampling is employed to increase sampling efficiency. For each of the perturbation steps, the initial equilibration run takes $1.26\text{--}2.50 \times 10^4$ passes and generates $5\text{--}10 \times 10^6$ configurations, which are discarded. (In the present simulations, a pass equals 396 Monte Carlo steps.) This is followed by a production phase of 2×10^5 passes that generates over 8×10^7 configurations or “snapshots” for ensemble averaging. Statistical errors are evaluated by the standard block average method,⁷⁸ which is less than 0.003 kcal/mol for each step.

After these simulations, a PMF between a $\xi = 3.5 \text{ \AA}$ and $\xi = 12 \text{ \AA}$ is obtained by discretized integration

$$\text{PMF} = \Delta G(\xi_j) = \sum_{i=0}^j \Delta G(\xi_i \rightarrow \xi_i + \Delta \xi) + (\text{constant}), \quad (31)$$

where $\xi_j = \xi_0 + j\Delta\xi$. Ideally, the constant in the above equation should be determined from the boundary condition

$$\Delta G(\xi \rightarrow \infty) = 0. \quad (32)$$

Practically, however, it is impossible to extend the perturbation calculation to $\xi \rightarrow \infty$. In previous studies

$$\Delta G(\xi = \xi_{\text{large}}) = 0 \quad (33)$$

was assumed, where ξ_{large} is the largest intersolute separation in a given perturbation calculation: Dang⁴⁶ and Rick and Berne⁵³ used $\xi_{\text{large}} = 8.5 \text{ \AA}$, whereas Lüdemann *et al.*^{50,51} and Rank and Baker⁵¹ used $\xi_{\text{large}} = 12 \text{ \AA}$. However, it is possible for the true PMF at $\xi = \xi_{\text{large}}$ to deviate from zero if the proper baseline condition [Eq. (32)] were used. Therefore, instead of relying on the assumption in Eq. (33), we determine zero-PMF baselines using the test-particle method below.

2. Test-particle insertion

The second method we use to calculate PMF is based on Eq. (8). Here we employ the test-particle insertion technique⁴⁷ to compute μ_a^* and $\mu_{ab}^*(\xi)$. To calculate

$$\mu_a^* = -kT \ln \left[\frac{\langle V \exp(-\beta U_a) \rangle_N}{\langle V \rangle_N} \right], \quad (34)$$

configurations of pure water molecules are generated using *NPT* Monte Carlo simulations, then solute *a* is inserted into random positions to compute the required averages. This recipe is provided by Shing and Gubbins,⁸⁵ and it accounts for volume fluctuations in the *NPT* ensemble.^{82,86} In these simulations, the discarded initial equilibration runs take at least 1.3×10^5 passes. These are followed by production runs of at least 1.3×10^6 passes (4×10^6 passes for 278 K), during which configurations of pure water are recorded at every 100 passes. 10 000 attempted solute insertions per snapshot are used to estimate the averages in the above equation. For

$$\mu_{ab}^*(\xi) = -kT \ln \left[\frac{\langle V \exp\{-\beta[U_a + U_{ab}(\xi)]\} \rangle_{N,b}}{\langle V \rangle_{N,b}} \right], \quad (35)$$

configurations of one methane plus water molecules are first generated. As before, 1.3×10^5 initial passes are discarded, then coordinates of the one-methane solution are collected every 100 passes for at least 2.6×10^6 passes (8×10^6 passes

for 298 K and 1.4×10^7 passes for 278 K). 10 000–30 000 insertions per snapshot are attempted to estimate the ensemble averages in Eq. (35).

3. Accuracy of PMFs

We use the following heuristic procedure to optimize the numerical reliability of the simulated PMFs by comparing results obtained by the two independent methods described above: (1) Insofar as it is feasible, long simulation times are used until a good agreement between the two methods over a wide range of ξ is reached. In other words, we take the degree of agreement between the two independent methods to be a measure of accuracy, which is intuitive. (2) If good agreement cannot be obtained over the entire range of ξ despite long simulation runs (this is likely to occur for low temperatures), we simply adopt the test-particle result, including its zero-PMF baseline. This is motivated by the fact that the PMF at every ξ calculated from test-particle insertion is independent of each other. On the other hand, free energy perturbation entails a summation over results at many values of ξ [Eq. (31)], so there is a chance for errors to accumulate. Hence, for the same given run time, we regard test-particle PMFs at large ξ as more trustworthy than those generated by free energy perturbation and assumption (33). Therefore, in these situations, we do not use Eq. (33). Instead, we allow the free-energy perturbation $\Delta G(\xi_{\text{large}})$ to vary [which is equivalent to redefining the constant in Eq. (31), $\xi_{\text{large}} = 12 \text{ \AA}$ in our simulations] and determine whether a decent match with the PMF obtained by test-particle insertion is possible by setting $\Delta G(\xi_{\text{large}})$ equal to a small offset.

B. Potential distribution calculation

The accurate calculation of $g_a(\epsilon)$ and $g_{ab}(\epsilon; \xi)$ for large ϵ is indispensable for the accurate evaluation of Eqs. (12) and (14), because both of these expressions contain the multiplicative factor $\exp(+\beta\epsilon)$ in their integrands. Numerically this would not be straightforward if these quantities were to be estimated by averaging over the (water + solute *a*) system as prescribed by Eq. (11), because the presence of the Boltzmann factor $\exp(-\beta U_a)$ in the numerator implies that high- U_a configurations are rarely sampled. However, this difficulty can be overcome by using a special test-particle insertion technique due to Shing and Gubbins,⁸⁵ who noted that

$$\begin{aligned} g_a(\epsilon) &= \frac{\int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N \delta(\epsilon - U_a) e^{-\beta U_N} e^{-\beta U_a}}{\int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N e^{-\beta U_N} e^{-\beta U_a}} \\ &= \frac{\int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N e^{-\beta U_N}}{\int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N e^{-\beta U_N} e^{-\beta U_a}} \\ &\quad \times \frac{\int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N \delta(\epsilon - U_a) e^{-\beta U_N}}{\int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N e^{-\beta U_N}} \\ &\quad \times \frac{\int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N \delta(\epsilon - U_a) e^{-\beta U_N} e^{-\beta U_a}}{\int d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N \delta(\epsilon - U_a) e^{-\beta U_N}} \\ &= e^{\beta \mu_a^*} g_a^{\text{tp}}(\epsilon) e^{-\beta \epsilon}, \end{aligned} \quad (36)$$

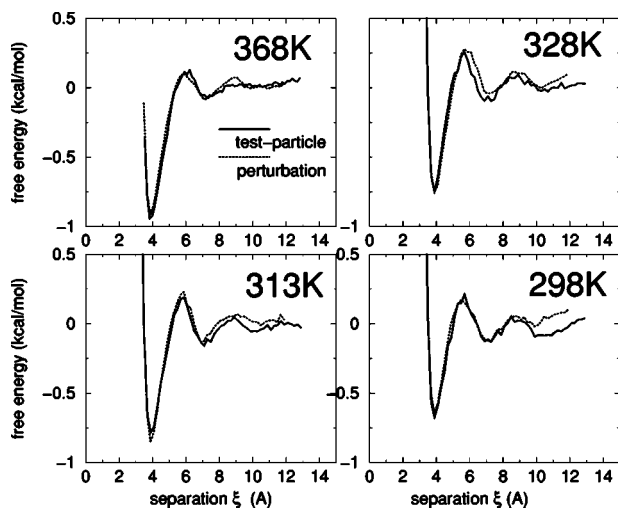


FIG. 2. Two-methane PMFs under the same (1 atm) pressure at four different temperatures are calculated by two independent methods: test-particle insertion (solid curves) and free energy perturbation (dotted curves). An offset is allowed for each dotted curve to achieve optimal matching with the solid curve. See the text for details.

where $g_a^{\text{tp}}(\epsilon)$ is the distribution of interaction energy between a randomly inserted test particle (solute a) and pure solvent. Since the test particle is allowed to overlap freely with the water molecules [there is no $\exp(-\beta U_a)$ factor in g_a^{tp}], large ϵ values can be sampled sufficiently by this technique to give an accurate determination of $g_a(\epsilon)$. A similar consideration leads to the corresponding relation

$$g_{ab}(\epsilon) = e^{\beta(\mu_{ab}^* - U_{ab})} g_{ab}^{\text{tp}}(\epsilon) e^{-\beta\epsilon} \quad (37)$$

for the two-solute potential distribution in Eq. (13). We use these two equations and test-particle insertion techniques to obtain $g_a(\epsilon)$ and $g_{ab}(\epsilon)$ from our simulations.

IV. RESULTS: TEMPERATURE AND DENSITY EFFECTS

A. Temperature dependence of the potential of mean force

Figure 2 shows PMFs at four different temperatures. The offset used to match free energy perturbation results to test-particle insertion results are (in units of kcal/mol) 0.05 (368 K), 0.1 (328 K), 0.05 (313 K), and 0.1 (298 K). For 278 K (Fig. 3), only test-particle insertion has been performed. The agreement of the results in Fig. 2 is reasonably good. As discussed in Sec. III A 3, discrepancy increases for lower temperatures and/or at larger separation. Nevertheless, the good agreement between the two methods for separations $\xi < 9.0$ Å (Fig. 2) lends credence to the overall numerical reliability of the test-particle results (Fig. 3). The depth of the contact pair (0.67 kcal/mol, which equals 2.80 kJ/mol) and the height of the desolvation barrier (0.88 kcal/mol, 3.68 kJ/mol) at 298 K agree with previous studies.^{31,46,50,53}

To establish a basis for our analysis, we first simulated the insertion of a single methane into water, and obtained an hydration Gibbs free energy of 2.34 ± 0.05 kcal/mol (9.79 ± 0.21 kJ/mol) at 298 K [Fig. 4(A)]. This is consistent with the value of 2.5 ± 0.4 kcal/mol from a previous *NPT* calcu-

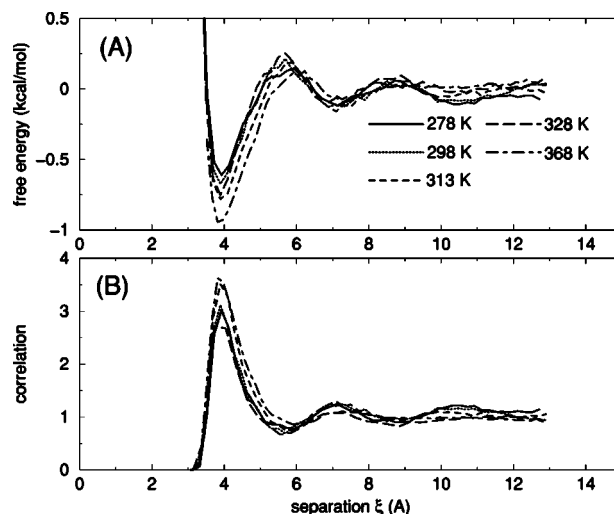


FIG. 3. Temperature dependence of the association of two methanes, simulated by the test-particle insertion technique. (A) PMFs [$\Delta G(\xi)$] at five different temperatures and 1 atm. (B) The corresponding two-methane correlation functions are given by $\exp[-\beta\Delta G(\xi)]$.

lation that also used TIP4P water.⁸⁷ Figure 4(A) indicates that the general trend of temperature dependence of the one-methane hydration Gibbs free energy (chemical potential) is in fair agreement with experimental data. However, the latter shows a steeper dependence on temperature. This is most likely related to a well-known artifact of the TIP4P water model, that its density changes more gradually with temperature than real water.⁸³ This can have an impact on the simu-

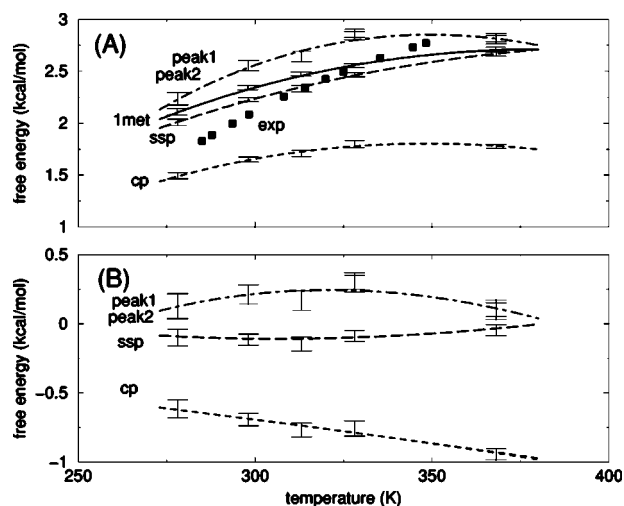


FIG. 4. Temperature dependence of hydrophobic hydration and hydrophobic interactions under 1 atm. (A) Simulated Gibbs free energy change upon insertion of a single methane into pure water (1met), into a water plus one methane system at the contact-pair (cp) minimum, desolvation barrier (peak1, peak2), and solvent-separated minimum (ssp) positions. 1met is the single methane hydration Gibbs free energy. The corresponding experimental data (exp, black squares) is from Morrison and Billett (Ref. 89). (B) Temperature dependence of PMF values. The error bar for each data point in (A) and (B) shows the range of variation between the minimum and maximum values observed among 50–300 different cumulative averages taken at regular intervals during the summation of contributions from configurational snapshots to compute the overall (final) average. The curves are optimal fits of the overall averages (as reported in Fig. 3) to Eq. (38).

TABLE I. Temperature dependence of hydrophobic hydration and interaction. The unbracketed quantities are calculated by least-square fitting Eq. (38), with $T_0=298.15$ K, to the simulated single methane (1met) hydration Gibbs free energy in Fig. 4(A) and the two-methane results at the PMF barrier and minima in Fig. 4(B). Included in square brackets are estimates from the commonly employed solvent accessible surface area (Ref. 103) (SASA) method (Ref. 104).^a

Solute configuration	ΔH_0^b (kcal/mol)	ΔS_0 (cal/mol/K)	ΔC_p (cal/mol/K)
Single-methane hydration	-0.562	-9.72	40.1
Contact formation ^c ($\xi=3.8$ Å)	0.295 0.292 [0.128 (0.239)]	3.29 3.28 [4.12]	1.59 2.93 [-17.0]
Desolvation peak ($\xi\approx 5.7$ Å)	-0.667 [-8.39 $\times 10^{-3}$ (0.0767)]	-2.92 [1.33]	40.4 [-5.46]
Solvent-separated minimum ($\xi=7.0$ Å)	-1.51 $\times 10^{-2}$ [-0.0263 (0)]	0.347 [0]	-13.3 [0]

^aThese estimates assume that the ΔH_0 , ΔS_0 , and ΔC_p of the solvent (water)-mediated or indirect part ΔG_h of PMF are proportional to the SASA lost by bringing two methanes to close proximity with each other. The proportionality constants used in the present comparison are set by the simulated single-methane hydration data listed in the table. The purely enthalpic direct Lennard-Jones interaction between the two methanes affects only the ΔH_0 estimates.

^bThe first entry inside a square bracket in this column is the total estimate that includes both the direct Lennard-Jones interaction between the methanes and the hydration contributions estimated using SASA. The latter is tabulated separately as the second entry in parentheses.

^cThe first and second rows of contact formation data correspond, respectively, to the thick and thin fitted dashed curves for cp in Fig. 4(B).

lated hydration Gibbs free energy.⁶⁰ The agreement between simulation and experiment is much better for the hydration heat capacity (see below).

To a first approximation, any Gibbs free energy change ΔG may be described by

$$\Delta G = \Delta H_0 + \Delta C_p(T - T_0) - T\Delta S_0 - T\Delta C_p \ln \frac{T}{T_0}, \quad (38)$$

where ΔH_0 and ΔS_0 are enthalpy and entropy changes, respectively, at a given reference temperature T_0 , and the constant-pressure heat capacity change ΔC_p is assumed to be temperature independent. Equation (38) has been useful for analyzing experimental transfer data.^{10,12} Here we use this equation to analyze our simulation results by least-square fitting it to the Gibbs free energy data in Fig. 4. The resulting thermodynamic parameters are given in Table I. The assumption here that ΔC_p is temperature independent is sensible in view of the fact that simulation data are available only at five temperatures, notwithstanding that ΔC_p s of some hydrophobic processes are weakly temperature dependent.^{10,88}

We first consider the hydration heat capacity of a single methane. The best fit of our simulation result gives $\Delta C_p = 40.1$ cal/mol/K (168 J/mol/K). This is in fair agreement with the 52.0 cal/mol/K (217 J/mol/K) value obtained by using the same procedure to fit the experimental solubility data⁸⁹ in Fig. 4(A), as well as Naghibi *et al.*⁸⁸ direct calorimetric measurements of 53.4 cal/mol/K (223 J/mol/K) at 298 K and 47.4 cal/mol/K (198 J/mol/K) at 313 K.

Our simulation data indicate that the entropy and heat capacity effects of two-methane association are intricate and dependent on the intersolute distance (Table I and Fig. 4). For completeness, we use two definitions for the desolvation peak: peak1 is defined by the local PMF maximum around $\xi \approx 5.9 \pm 0.2$ Å, whereas peak2 denotes PMF at $\xi = 5.7$ Å. When numerical uncertainties are taken into account, the two definitions give essentially identical results. The data points and their fitted curves in Fig. 4(B) are given by their corresponding plots in Fig. 4(A) minus the single-methane hydration Gibbs free energy [1met in Fig. 4(A)]. For the contact pair (cp) position, we also provide an independent fit of the PMF values themselves (thin dashed curve), but it practically overlaps with the difference [thick dashed curve for cp in Fig. 4(B)] between the cp and 1met curves in Fig. 4(A).

The hydration of a single methane decreases entropy and increases heat capacity (see above). Therefore, one would expect the entropy change upon the formation of a two-methane contact to be positive, and the heat capacity change negative. This is because the spatial arrangement of two methanes in contact minimizes their geometric exposure to the aqueous environment, and thus contact formation may be viewed as a desolvation process. The large positive ΔS_0 from simulation results at the contact minimum is consistent with this expectation. However, contrary to expectation, the corresponding ΔC_p has a surprisingly small magnitude ($\sim 1.59-2.93$ cal/mol/K, i.e., $\sim 6.65-12.3$ J/mol/K), amounting to only $\sim 7\%$ of the hydration heat capacity of a single methane; and its positive sign is opposite to what had been expected.

When the methanes are farther apart at the desolvation barrier, the heat capacity is large and positive ($\sim +40.4$ cal/mol/K = +169 J/mol/K). It has the same sign and essentially the same magnitude as the *hydration* heat capacity of an entire methane. From the ΔS_0 fits, it is also clear that at 25 °C, substantial entropic costs have to be paid to reach the top of this barrier (Table I): 2.92 cal/mol/K (12.2 J/mol/K) if the two methanes are approaching each other from an initially large separation, and a much higher value of 6.20 cal/mol/K (25.9 J/mol/K) if the two methanes are dissociating from the contact minimum. Thus, remarkably, our results indicate that the thermodynamic signatures of the two-methane desolvation barrier are *opposite* to that of hydrophobic desolvation itself.

Moving to a larger separation, we find that the PMF at the solvent-separated minimum has a heat capacity of $\sim -1.33 \times 10^{-2}$ cal/mol/K (-5.56×10^{-2} J/mol/K) and a positive ΔS_0 . In this case, both thermodynamic signatures have the signs expected of a desolvation process.

We next turn to the relative favorability of different two-methane configurations. A few salient features are noteworthy in Figs. 3 and 4: (i) The contact minimum deepens as temperature is raised. The increase in stability is almost linear in temperature. This trend is qualitatively consistent with the fixed-density *NVT* simulations by Lüdemann *et al.*⁵⁰ (ii) The desolvation barrier exhibits a Gibbs free energy maximum (i.e., a stability minimum) around 313–328 K. The top of the barrier is seen to shift slightly toward larger ξ as temperature is raised. This trend is similar to that observed in

Lüdemann *et al.*'s *NVT* simulations.⁵⁰ However, the temperature dependence of their PMF desolvation barrier (their Figs. 3 and 4) indicates that it has a negative constant-volume heat capacity (ΔC_V). This is of particular interest in light of the present results, because it suggests that the ΔC_P and ΔC_V of the desolvation barrier may have opposite signs. (iii) The solvent-separated pair exhibits a shallow minimum around 313 K. (iv) As a consequence of (i) and (iii), the solvent-separated pair (ssp) to contact pair (cp) population ratio

$$K_{\text{eq}} = \frac{[\text{ssp}]}{[\text{cp}]} = \frac{\int_{\xi_1}^{\xi_2} d\xi 4\pi\xi^2 e^{-\beta\Delta G(\xi)}}{\int_0^{\xi_1} d\xi 4\pi\xi^2 e^{-\beta\Delta G(\xi)}} \quad (39)$$

decreases with increasing temperature. Here $\xi_1 = 5.5$ Å and $\xi_2 = 8.5$ Å, as in Rick and Berne,⁵³ except $\xi_1 = 6.0$ Å is used for 368 K. The shapes of the correlation functions in Fig. 3(B) indicate that these definitions are intuitively reasonable. For the five temperatures studied here, $K_{\text{eq}} = 2.2, 2.2, 2.0, 1.6,$ and 1.4 for 278, 298, 313, 328, and 368 K, respectively. K_{eq} 's tendency to decrease with increasing temperature is consistent with Rick and Berne's recent *NPT* simulation using a polarizable water model.⁵³

B. Effects of average water density

Previous investigations have shown that variation in overall water density has significant effects on hydrophobic hydration,^{60,90} especially under high pressure.^{54,55} Can the variation of water density under atmospheric pressure (1 atm) over the temperature range from 0 to 100 °C have significant effects on PMF properties? To address this question, we compare two PMFs computed using the test-particle insertion method at the same temperature 368 K: one with the 1 atm model water density at 368 K, the other with a higher density that equals to 1 atm TIP4P model water densities at 298 K. Figure 5 shows that the two PMFs are significantly different: The higher water density raises the desolvation barrier dramatically and stabilizes the contact pair.

Our results contradict Lüdemann *et al.*,⁵⁰ but are consistent with that of Dang,⁴⁶ who has reported a deepening of both the contact and solvent-separated minima at higher water density. Dang's results indicate that a higher water density leads to a larger Gibbs free energy difference between the desolvation barrier and the contact minimum. This is qualitatively consistent with our results in Fig. 5.

V. RAMIFICATIONS FOR PROTEIN FOLDING

The present simulation results are relevant to protein folding.

Typically, protein folding kinetics are non-Arrhenius; folding rates are often observed to be fastest at some intermediate temperatures.⁹¹⁻⁹⁷ This means that the effective free energy barrier to folding (i.e., the Gibbs free energy difference between the "folding transition state" and the unfolded state) has a minimum as a function of temperature, and therefore has an effective negative heat capacity. Recently, Rank

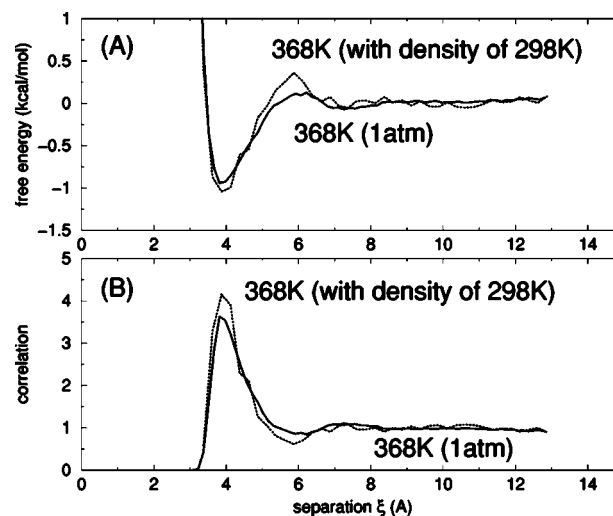


FIG. 5. Density effects on methane association are significant. (A) PMF, and (B) two-methane correlation function at 368 K at two different densities, computed by test-particle insertions. Imposing a density (equal to that at 298 K under 1 atm) higher than that at 368 K under 1 atm significantly increases the free energy difference between the contact minimum and the desolvation barrier. The increase in the desolvation barrier height is particularly noteworthy. It is almost twice the range of variation in barrier height observed at constant pressure (1 atm) from 278 to 368 K (see Figs. 3 and 4). The PMF at 298 K density is computed here using an *NVT* ensemble.

and Baker³¹ suggested that hydrophobic PMF desolvation barriers may be important in determining folding rates. However, results in this paper indicate that the two-methane desolvation barrier has a positive heat capacity [Fig. 4(A) and Table I]. Thus its sign is opposite to that of folding free energy barriers. This implies that crossing two-methane-like hydrophobic desolvation barriers may not be rate limiting in protein folding.

Some recent molecular dynamics studies of protein kinetics⁹⁸⁻¹⁰⁰ use an extremely high temperature of 498 K to induce unfolding, and the simulations are conducted at an average water density of 0.829 g/cm³. This particular water density corresponds to the density of a *hypothetical* unstable state of water at 498 K under atmospheric pressure (the stable state under these conditions is the vapor phase.) This density was apparently calculated by long extrapolation from data below 100 °C using the R_{51} formula provided by Kell.¹⁰¹ Practically, however, a pressure of ~26 atm is required¹⁰⁰ to maintain a water density of 0.829 g/cm³ at 498 K. These studies have provided useful insight. However, results presented in this paper demonstrate that hydrophobic interactions are sensitive to density and pressure at a given temperature (Fig. 5), and depend on temperature at a given pressure (Fig. 3). Figures 4, 5, and Table I show that the relative Gibbs free energy differences between key points along the PMF curve can vary significantly with temperature and pressure. Therefore, inasmuch as hydrophobic interactions are important to protein folding, caution should be used in connecting results of high temperature, high pressure unfolding simulations to folding and unfolding transition states^{100,102} under ambient conditions.

Our results may also be used to examine to what extent protein folding potentials with implicit solvent effects can

capture microscopic details of hydrophobic interactions. One popular approach has been to assume that the aqueous solvation contributions to heat capacity and entropy of a collection of molecules are proportional to their total solvent (water) accessible surface area (SASA).^{103,104} Table I compares SASA predictions with results from simulated PMFs. At the PMF contact minimum, the ΔS_0 predicted by SASA (4.12 cal/mol/K, 17.2 J/mol/K) compares reasonably well with the simulated value of 3.29 cal/mol/K (13.8 J/mol/K), confirming the conventional view that hydrophobic contacts at 25 °C are favored by entropy. However, the ΔC_P predicted by SASA (-17.0 cal/mol/K, -71.1 J/mol/K) is ~ 7 times larger in magnitude and has a sign opposite to the one estimated from simulation (~ 2.3 cal/mol/K, 9.6 J/mol/K). Interestingly, Lazaridis and Karplus's recent solvent exclusion approach, which is based on the volume integration of a Gaussian function,²³ predicts $\Delta C_P = -4.59$ cal/mol/K (-19.20 J/mol/K) and $\Delta S_0 = 1.11$ cal/mol/K (4.64 J/mol/K). Their volume-based model's error in ΔC_P at the contact minimum is less appreciable than that of the SASA method, but its ΔS_0 error is much larger. Apparently, both models are not capable of offering accurate representations of both ΔC_P and ΔS_0 at the two-methane contact minimum. The discrepancies at the desolvation peak are even more serious, here the ΔC_P and ΔS_0 predicted by both models have signs opposite to the simulated values. These comparisons suggest that any implicit-solvent model that relies on monotonic variations of thermodynamic signatures with distance would fail to predict the thermodynamic signatures of the desolvation peak. Detailed implications of this finding on the accuracy of implicit-solvent simulations of much larger systems such as entire proteins^{23,24,105} remain to be elucidated. The observations here point to the fact that nonadditive effects¹⁰⁶ are significant in hydrophobic interactions, and must be tackled before a coherent physical picture of protein energetics can emerge.

VI. A PHYSICAL PICTURE IN TERMS OF HYDRATION MEAN FORCES

Based on the formulation developed in Sec. II A for hydration mean forces, we now analyze in more detail the hydration part ΔG_h of the two-methane PMF.

A. Dependence on solute–solute distance ξ

We begin by considering the PMF at 298 K. Figure 6 shows the spatial decomposition of the hydration mean force at three intersolute distances. In this figure, positive forces push the two methanes apart (dissociative, i.e., tend to increase ξ), while negative forces are associative. Interestingly, the dominant mean force contributions come from water molecules closer than the first hydration shell at $r = 3.7$ Å, which corresponds to the solute–water correlation peak. While the top row in Fig. 6 shows that local water density increases monotonically from $r = 2.9$ Å to $r = 3.7$ Å, force contributions (middle row) peak at $r = 3.3$ Å and decreases at larger r . This is because $-(\partial u_{a1}/\partial r)$ diminishes rapidly as r increases. Results in this figure shows that dominant hydration mean force contributions come from water molecules in the “inner” shell, which is defined here by $r \leq 3.9$ Å. Within

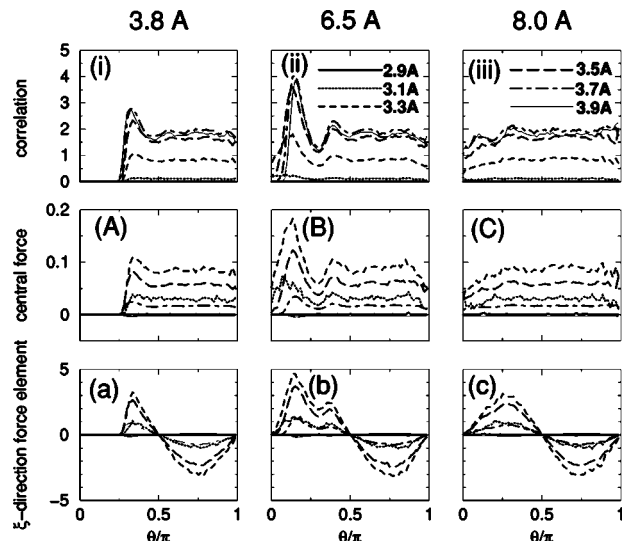


FIG. 6. Understanding PMFs in terms of the hydration mean force. The first, second, and third columns describe results simulated by test-particle insertions at 298 K and 1 atm for $\xi = 3.8, 6.5,$ and 8.0 Å, respectively. Angular dependences (horizontal axes) of mean forces contributed by water molecules situated at different distances r from solute a (cf. Fig. 1) are indicated by six different line styles defined in (ii) and (iii), from $r = 2.9$ Å (thick solid line) to $r = 3.9$ Å (thin solid line). Top row (i)–(iii) The water–solute correlation function is the local water density $\rho(\xi, r, \theta)$ divided by the average water density in the bulk at the given temperature and pressure [c.f. Eq. (7)]. We use the volume element within $r \pm 0.05$ Å and $\theta \pm \pi/120$ to compute $\rho(\xi, r, \theta)$. Middle row (A)–(C) Force per unit volume directed towards the center of solute a in units of kcal/mol/Å⁴. This is equal to the contribution $-\rho(\xi, r, \theta)(\partial u/\partial r)$ from water molecules at an r, θ position. Bottom row (a)–(c) Component of force in the increasing ξ direction per unit solute–water distance (r), applied by all water molecules at r, θ . This is given by $-\pi r^2 \sin 2\theta \rho(\xi, r, \theta)(\partial u/\partial r)$, and plotted in units of kcal/mol/Å².

this shell, water–solute interactions are repulsive. Forces exerted by water molecules outside this shell are attractive [$-(\partial u_{a1}/\partial r) < 0$], but their contributions to the mean force are significantly smaller than the repulsive forces originating from the inner shell (data not shown). Hence we focus only on the inner shell in the discussion below.

The angular (θ) dependence of the hydration mean force offers additional physical insight. Since the inner-shell solute–water interactions are repulsive, inner-shell water molecules located between the solutes ($\theta \leq \pi/2$, “interior” region) contribute to a dissociative force, whereas those on the other side ($\theta \geq \pi/2$, “exterior” region) contribute to an associative force (see Fig. 6, bottom row). Comparison between Figs. 6(A), 6(B), and 6(C) shows that associative forces change little with ξ , but dissociative forces are strongly ξ dependent. This is because inner-shell local water density is very sensitive to ξ in the $\theta \leq \pi/2$ interior region, but local water density at any given r is not that sensitive to ξ in the exterior $\theta \geq \pi/2$ region (Fig. 6, top row).

The two-methane PMF contact minimum is at $\xi \approx 3.8$ Å. This represents a balance between an attractive hydration mean force and a repulsive direct interaction between the two solutes. [In the current model, the methane–methane Lennard-Jones force $-(\partial U_{ab}/\partial \xi)$ is repulsive for $\xi < 4.2$ Å.] Figure 6(A) shows that at $\xi = 3.8$ Å, no hydration mean force comes from the region $0 < \theta \leq \pi/4$, because the volume of solute b has excluded water molecules from this region

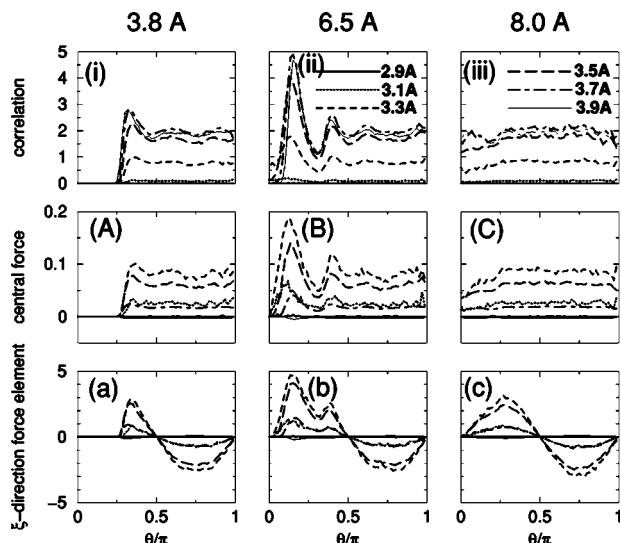


FIG. 7. Same as Fig. 6, also at 1 atm, but for 278 K.

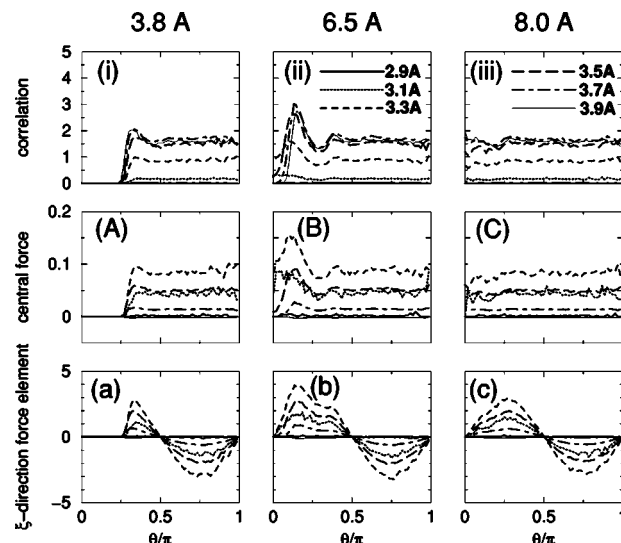


FIG. 8. Same as Fig. 6, also at 1 atm, but for 368 K.

[Fig. 6(i)]. This implies a decrease in dissociative force [Fig. 6(a)], the net result is an attractive mean force.

On the other hand, the mean force is repulsive (negative PMF slope) at $\xi=6.5$ Å, when the two methanes are a little farther apart than the desolvation barrier. This results from a balance between a weak attractive direct intersolute force and a stronger repulsive hydration mean force. Figure 6(b) shows that dissociative hydration mean forces become larger when the two-methane separation is increased from $\xi=3.8$ Å to $\xi=6.5$ Å. This is clearly related to the increased inner-shell local water density in the interior region at around $\theta = \pi/8$ [Fig. 6(ii)], which leads to a maximum repulsion from water molecules located in the same region [Fig. 6(c)].

Further increasing ξ to 8.0 Å weakens the dissociation force [Fig. 6(c)], making the net hydration mean force once again attractive. Figure 6(iii) shows that this is primarily caused by a decrease in inner-shell local water density in the interior region relative to that at $\xi=6.5$ Å.

B. Temperature dependence

Following the above reasoning, insight into the temperature dependence of the hydration mean force may also be gained by examining how the difference in local water density between the exterior and interior parts of the inner shell depends on temperature. This can shed some light on the microscopic molecular basis of PMF temperature dependence, although the hydration mean force, being the hydration part of the PMF slope (not PMF itself), only accounts for PMF indirectly.

When temperature is lowered from 298 to 278 K, the contact minimum at $\xi \approx 3.8$ Å becomes a bit more shallow. Concomitantly, the magnitude of the PMF slope around the contact minimum becomes less steep (Fig. 3). Because the direct interaction between the two methanes is temperature independent, this implies that the attractive hydration mean force at this configuration is weakened at 278 K relative to that at 298 K. Figure 7(i) shows that this is caused by a slight decrease in inner-shell local water density at $r \approx 3.3$ Å over a

broad range of angles in the exterior region ($\theta > \pi/2$), resulting in a weaker associative force from this region at 278 K [Fig. 7(c)] than that at 298 K (Figure 6c).

On the other hand, when temperature is raised from 298 to 368 K, the contact minimum deepens substantially and the solvent-separated minimum becomes more shallow (Figs. 3 and 4). Concomitantly, the contact minimum is shifted to a slightly smaller ξ , which implies a stronger attractive hydration mean force at $\xi \approx 3.8$ Å; and more gradual slopes on both sides of the solvent-separated minimum, at $\xi \approx 6.5$ and 8.0 Å. These more gradual slopes imply that the hydration mean force at 368 K is not as repulsive at $\xi=6.5$ Å and less attractive at $\xi=8.0$ Å than their counterparts at 298 K.

Again, these trends correlate with that of inner-shell local water density. Specifically, comparison between Figs. 6 and 8 shows that at $\xi=3.8$ Å (i) and 6.5 Å (ii), water density is lower in the interior region at 368 K than that at 298 K. However, at $\xi=8.0$ Å (iii), water density over a broad angular range in the exterior region at 368 K is lower relative to that at 298 K. Correspondingly, dissociative components of the hydration mean force are reduced at $\xi=3.8$ Å (a) and 6.5 Å (b), whereas the associate component is reduced at $\xi=8.0$ Å (c). The net effects of these are consistent with the trends of hydration mean forces noted above.

C. Effects of average (overall) water density

Figures 9(i)–9(iii) shows that when overall water density is increased at constant temperature (which requires an increase in pressure), there is a general increase in inner-shell local water density in excess of the already increased overall water density [c.f. Figs. 8(i)–8(iii)]. This leads to much stronger associative as well as dissociative hydration mean forces (Fig. 9, middle and bottom rows), which translates into steeper PMF slopes. This general trend is consistent with a deeper contact minimum and a higher desolvation barrier (Fig. 5).

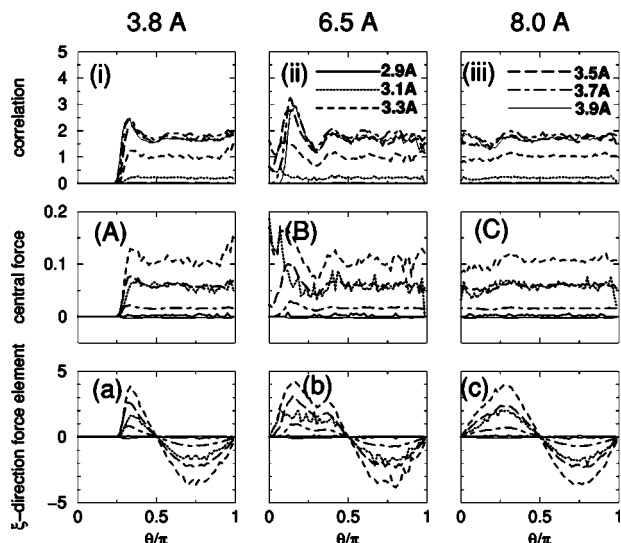


FIG. 9. Same as Fig. 6, now at 368 K, but with the 1 atm water density at 298 K. (The pressure is therefore higher than 1 atm; c.f. Fig. 5.)

D. Temperature-induced shift of the desolvation peak

The peak of the PMF desolvation barrier around $\xi = 5.7\text{--}6.1\text{Å}$ results from a balance between an attractive direct intersolute force and a dissociative (repulsive) hydration mean force. As temperature is raised at constant pressure from 298 to 368 K, the peak of this barrier shifts toward a larger ξ ; see Fig. 3. Comparing the local water density at $\xi = 6.5\text{Å}$ in Figs. 6(ii) and 8(ii) provides a rationalization of this behavior: Inner-shell local water density in the interior region is lower at 368 K than that at 298 K. This leads to a weaker dissociative force at 368 K [Figs. 6(b) and 8(b)], resulting in the balance point with the attractive intersolute force being shifted to a larger ξ relative to that at 298 K.

VII. A PHYSICAL PICTURE IN TERMS OF POTENTIAL DISTRIBUTIONS

The hydration mean force analysis above accounts for the slope of PMF. Here we provide a complementary perspective that provides rationalizations of the hydration part of PMF itself. The analysis in this section is based on the potential distribution formulation in Secs. II B and III B.

A. The role of high potential energy states in hydrophobic hydration

We first consider the hydration of a single methane. As discussed in Sec. II B, the hydration Gibbs free energy is given by

$$\mu_a^* = kT \ln \int d\epsilon g_a(\epsilon) e^{\beta\epsilon}, \quad (40)$$

where ϵ is the solute–water interaction energy. This relation is equivalent to Eq. (12). It follows that, for a solute to be “hydrophobic,” i.e., to have low solubility and hence a large and positive μ_a^* , the right hand side of Eq. (40) must be large and positive. What functional form of $g_a(\epsilon)$ can lead to such a result? By definition [Eq. (11)], $g_a(\epsilon)$ has to satisfy the normalization condition

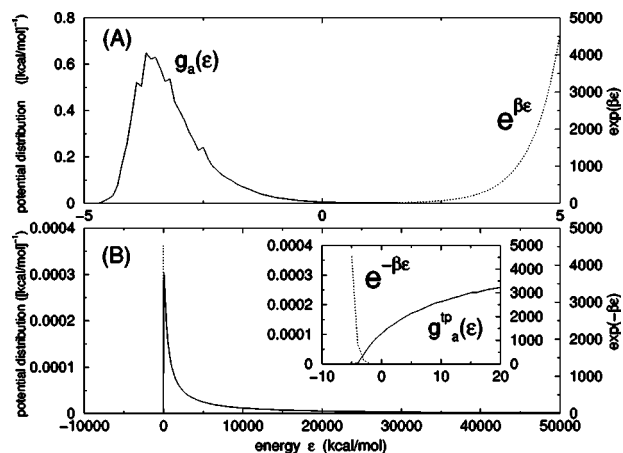


FIG. 10. Two equivalent perspectives of hydrophobic hydration. (A) Potential distribution theory viewpoint [Eq. (40)]: g_a (solid curve) and $\exp(\beta\epsilon)$ (dotted curve) are given by the left and right vertical scales, respectively. (B) Test-particle insertion viewpoint [Eq. (42)]: g_a^{tp} (solid curve) and $\exp(-\beta\epsilon)$ (dotted curve) are given by the left and right vertical scales, respectively. The inset in (B) shows the variation of these two functions over a smaller range of energies, from -10 to $+20$ kcal/mol. Results in this figure are obtained by test-particle insertion simulation of methane hydration at 298 K and 1 atm.

$$\int d\epsilon g_a(\epsilon) = 1. \quad (41)$$

This means that for μ_a^* to be large and positive, one has to rely on large contributions from the $\exp(\beta\epsilon)$ factor in the integrand of Eq. (40). As can be seen from Eqs. (13) and (14), a similar condition applies to the two-methane chemical potential μ_{ab}^* and the potential distribution $g_{ab}(\epsilon)$.

Figure 10 is an analysis of our simulation data in this formulation. The potential distribution $g_a(\epsilon)$ peaks at $\epsilon \approx -3.8$ kcal/mol (15.9 kJ/mol), as seen in Fig. 10(A), but the integration of $g_a(\epsilon)$ from $\epsilon = -5.0$ to 2.0 kcal/mol yields only 0.46 [c.f. Eq. (41)]. Hence the rest of the distribution (0.54) must come from higher values of ϵ . This is consistent with the condition above for a large and positive μ_a^* . A larger hydration Gibbs free energy (i.e., the solute being more hydrophobic) is associated with a distribution g_a of solute–water interaction energies that has higher weights at high interaction energies ϵ , where it overlaps with large values for $\exp(\beta\epsilon)$. Therefore, potential distribution theory stipulates that hydrophobicity has its origin in the high probabilities of solution configurations with strongly unfavorable solute–water interactions. Conversely, when probabilities of solution configurations with strongly attractive solute–water attraction are increased, the normalization condition Eq. (41) implies that μ_a^* would have to be smaller. A case in point is the favorable hydration of ions, whose interactions with water are concentrated in far lower energies than that of methane.^{107–109}

This perspective is consistent with that of Pratt and Pohorille,^{110,111} even though they have used a very different approach, which is based on hard-sphere solutes instead of the “soft” Lennard-Jones solutes studied here. [It may be more appropriate to call these “tough” spheres, because the $(\sigma/r)^{12}$ term is highly repulsive at small r .] They relate hy-

drophobicity to the structure of *pure* water; whereas the present approach relates hydrophobicity to the structure of water *around the solute*. The conceptual difference of the two approaches parallels that between the two different computational routes represented by Eqs. (9) and (10). Pratt and Pohorille^{110,111} note that “free volume is distributed in smaller packets” in pure water, in spite of its larger free volume fraction than organic liquids; and they identify this as the origin of hydrophobicity. Using Eq. (36), we now provide a specific connection between their perspective and the potential distribution formulation.

It follows from the normalization condition Eq. (41) that Eq. (36) can be re-written as

$$\mu_a^* = -kT \ln \int d\epsilon g_a^{\text{tp}}(\epsilon) e^{-\beta\epsilon}. \quad (42)$$

This equation represents the test-particle approach to hydration. It describes the insertion of a solute into pure water, and states that hydration Gibbs free energy decreases monotonically with increasing value of the integral of $g_a^{\text{tp}}(\epsilon) \exp(-\beta\epsilon)$ over all ϵ . In other words, because of the factor $\exp(-\beta\epsilon)$, a large $g_a^{\text{tp}}(\epsilon)$ at high values of ϵ serves to reduce the integral on the right-hand side [Fig. 10(B)]. This in turn implies a large positive μ_a^* , i.e., low solubility, or equivalently, hydrophobicity. Physically, large $g_a^{\text{tp}}(\epsilon)$ at high ϵ values must originate from significant overlaps between the test particle and water molecules, presumably because of the fact that free volume in water is distributed in small packets. Thus this offers a physical picture essentially equivalent to Pratt and Pohorille's, and may be viewed as a potential distribution reinterpretation of their perspective.

As discussed above, an alternate and equivalent perspective is represented by Eq. (40). This view is based on a potential distribution $g_a(\epsilon)$ obtained from averaging over an ensemble with the solute already in solution [Eq. (11) and Fig. 10(A)] instead of the test-particle potential distribution $g_a^{\text{tp}}(\epsilon)$ computed from pure water. In either views, hydrophobicity requires long, high- ϵ tails in both g_a and g_a^{tp} , as the two distributions are intimately related through Eq. (36).

B. Dependence of PMFs on the two-methane separation ξ

We now explore further the relation between potential distribution and the ξ dependence of PMF. We consider as an example the situation at 298 K. Figure 11 shows the change in potential distribution (g_{ab} vs g_a) when solute b is introduced at different distances from where solute a is inserted. As has been shown in Eq. (17), the sign of the hydration part of PMF is identical to the sign of the integral of $[g_{ab}(\epsilon) - g_a(\epsilon)] \exp(\beta\epsilon)$ over ϵ .

PMF is large and negative at $\xi = 3.8 \text{ \AA}$ (Fig. 3). By the above consideration, this means that the introduction of the second solute b must have led to a potential distribution more heavily weighted at lower values of ϵ . This is not easy to discern from Fig. 11(A), since the peaks of g_a and g_{ab} are close to each other. The situation becomes perfectly clear in Fig. 11(D), however. This example shows that the resultant

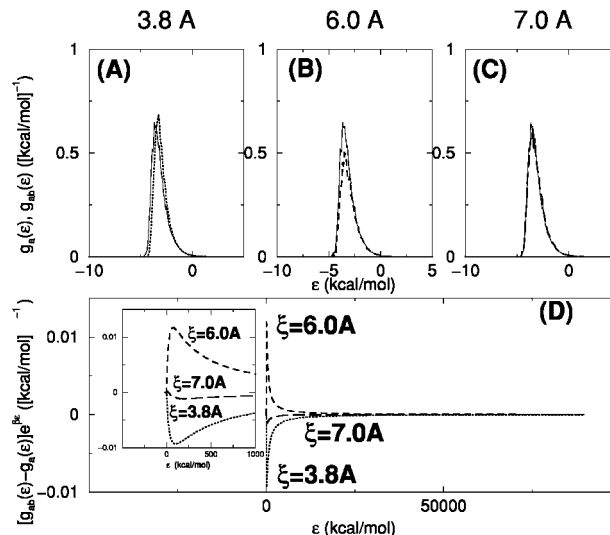


FIG. 11. Using potential distributions to account for the hydration (indirect part of PMF) at three solute-solute separations ξ . Results are from test-particle insertion simulations at 298 K and 1 atm. (A)–(C) $g_{ab}(\epsilon)$ for $\xi = 3.8, 6.0,$ and 7.0 \AA are plotted by dotted, dashed, and long dashed lines, respectively, whereas the corresponding $g_a(\epsilon)$ are represented by solid curves. (D) Plots for $[g_{ab}(\epsilon) - g_a(\epsilon)] \exp(\beta\epsilon)$. Line styles for different values of ξ follow that in (A)–(C). The inset shows the same plot over a smaller energy range.

effect of potential distributions on chemical potentials is often not fully apparent until the factor $\exp(\beta\epsilon)$ is included explicitly in the analysis.

Figures 11(B) and 11(C) indicate that differences between potential distributions g_{ab} and g_a are also difficult to spot for $\xi = 6.0 \text{ \AA}$ (close to the desolvation barrier) and $\xi = 7.0 \text{ \AA}$ (solvent-separated minimum), let alone the *difference* in value of $g_{ab} - g_a$ between $\xi = 6.0$ versus 7.0 \AA . However, their very different effects on the ΔG_h (positive at $\xi = 6.0 \text{ \AA}$ but negative at $\xi = 7.0 \text{ \AA}$) are apparent in Fig. 11(D). As discussed above, because of the $\exp(\beta\epsilon)$ factor, these effects arise from differences in probabilities, depending on intersolute separation, of very rare configurations with very high solute–water interaction energies. Figure 11(D) shows that $[g_{ab}(\epsilon) - g_a(\epsilon)] \exp(\beta\epsilon)$ peaks around $\epsilon = 100 \text{ kcal/mol}$, and decays very slowly with increasing ϵ , with substantial contributions from $\epsilon = 1000 \text{ kcal/mol}$ and beyond. Therefore, according to this view, whether a given separation ξ is favorable hinges on subtle shifts caused by the presence of a second solute in the probabilities of highly unfavorable configurations.

We have conducted similar analyses to study how the temperature dependence of PMF is related to changes in potential distribution. As in the cases considered above, dominant contributions originate from the small but nonzero population of configurations with very high solute–water interaction energies (data not shown).

C. Non-Gaussian nature of the potential distribution functions

Following Sec. II B 3, we test the accuracy of cumulant expansion predictions [Eq. (22)]. This is to ascertain whether Gaussian approximations¹¹² are viable for the potential dis-

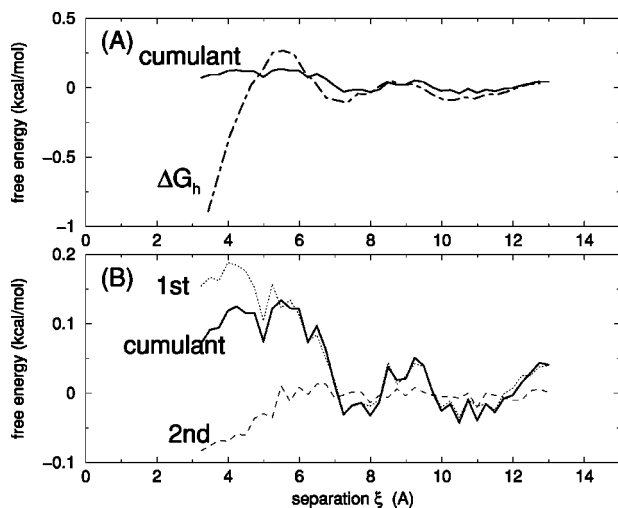


FIG. 12. Potential distributions are non-Gaussian. (A) Comparing the hydration (indirect) part (ΔG_h) of the two-methane PMF simulated at 298 K and 1 atm by test-particle insertions with the corresponding cumulant expansion consisting of the two terms shown in Eq. (22). Contributions from the first (1st) and second (2nd) terms in the cumulant expansion are plotted separately in (B). [Note that the vertical scale in (B) covers a smaller Gibbs free energy range than that in (A).]

tributions studied here. Figure 12 shows that the expansion is apparently successful in qualitatively predicting the solvent-separated minimum. However, Gaussian expansions fail to predict the contact minimum. Instead, it predicts destabilization at small ξ . Figure 12(B) shows that this arises from the fact that at small ξ the two-term expansion is dominated by the first cumulant term. This contribution is equal to the change in average solute–water interaction energy upon association of two methanes [Eq. (22)]. It is unfavorable (positive, ≈ 0.17 kcal/mol = 0.71 kJ/mol) at the contact minimum ($\xi = 3.8$ Å). By itself, this is not inconsistent with the observation that the total change in the hydration part of the enthalpy upon contact formation is also positive (≈ 0.24 kcal/mol = 1.00 kJ/mol, see Table I), although the latter contains not only solute–water interactions but also water–water interactions and PV contributions. However, the fact that the two-term cumulant expansion predicts neither the desolvation barrier nor the contact minimum implies that its utility in PMF studies is very limited.

VIII. HYDROGEN BONDING AND LOCAL WATER DENSITY

It is intuitively obvious that many peculiar properties of water and hydrophobic interactions must arise from the hydrogen bonding properties of water. Previous studies have elucidated the balance between entropic and enthalpic contributions to hydrophobic effects and their relationship with the structure of water, by focusing on clathrate structures^{64,65} and hydrogen bond number distributions in the solvation shell around a nonpolar solute versus that in the bulk.^{41,66} However, a quantitative relationship between water hydrogen bonding pattern and PMF has yet to be established. Here we take a preliminary step in that direction by exploring the correlation between hydrogen bond numbers and local water density. Our hope is that, by utilizing the formulation devel-

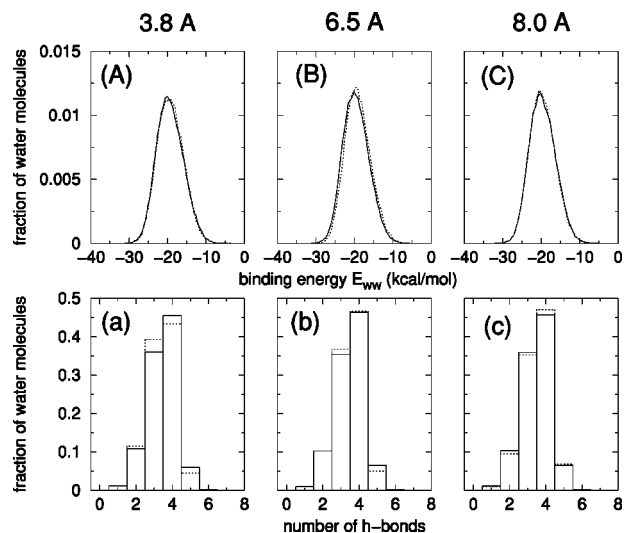


FIG. 13. Water–water binding energies and hydrogen bonds. (A)–(C): Distribution of binding energy E_{ww} among water molecules residing within the inner shell surrounding a methane solute in the interior (dotted curve, $\theta \leq \pi/2$) and exterior (solid curve, $\theta \geq \pi/2$) regions, for a two-methane system at 298 K and 1 atm, at the three intermethane distances ξ indicated. (a)–(c): Distribution of number of hydrogen bonds (h -bonds) per water molecule within the inner-shell volume. Dotted and solid lines represent the interior and exterior regions, respectively. Hydrogen bonds are defined here by an upper bound on a pair of water molecule's interaction energy (see Ref. 83). See the text for details.

oped above that relates local water density to the hydration mean force as an intermediate link in our theoretical development, this line of inquiry would ultimately lead to a more quantitative and microscopic account of PMF in terms of hydrogen bonding patterns.

Using the formulation in Sec. II C, the upper panels of Fig. 13 shows water–water binding energies at three intermethane separations. The distributions of binding energy E_{ww} among water molecules in the interior and exterior regions of the inner shell around a methane are given by

$$2\pi \int_0^{\pi/2} d\theta \sin \theta \int_{r_{i0}}^{r_{i1}} dr r^2 g_{\text{water}}(E_{ww}, \mathbf{r}_1) \quad (43)$$

and

$$2\pi \int_{\pi/2}^{\pi} d\theta \sin \theta \int_{r_{i0}}^{r_{i1}} dr r^2 g_{\text{water}}(E_{ww}, \mathbf{r}_1), \quad (44)$$

respectively, where $g_{\text{water}}(E_{ww}, \mathbf{r}_1)$ is the position-dependent binding energy distribution in Eq. (28), and the integrations from $r_{i0} = 2.9$ Å to $r_{i1} = 3.9$ Å cover the volume of the inner shell. Based on these binding energy distributions, numbers of hydrogen bonds per water are calculated (lower panels of Fig. 13).

At $\xi = 3.8$ and 6.5 Å, inner-shell binding energy distribution for the interior region is shifted slightly toward higher energies relative to that for the exterior region. At $\xi = 8.0$ Å, however, an opposite shift is observed. The present observation for $\xi = 6.5$ Å is consistent with the $\xi = 7.0$ Å results of Zichi and Rossky,⁶⁶ who considered both the inner and outer shells as a whole. They reported that water binding energy distribution in the interior region was shifted slightly toward higher energies than that in the exterior region, and attributed

this to the reduction of number of neighboring water molecules. Geiger *et al.*⁶⁴ and Pangali *et al.*⁴¹ have also calculated binding energy distributions. They focused on a comparison between shell and bulk, rather than between interior and exterior regions. However, as suggested by the above local water density considerations, we expect a comparison between the interior and exterior regions would be more directly relevant to PMFs.

Figure 13 shows that the relationship between hydrogen bond numbers and local water density in the inner shell is subtle; and there is no simple direct correlation between the two properties. It shows that at $\xi=8.0$ Å, each water molecule on average forms a slightly smaller number of hydrogen bonds in the interior than the exterior region, whereas the opposite holds for $\xi=6.5$ Å. When compared with the inner-shell water density for the same ξ in Figs. 6(ii) and 6(iii), these observations seem to suggest a correlation between a higher number of hydrogen bonds and a higher local water density. However, inner-shell water density in the interior region is lower at $\xi=3.8$ Å than that at $\xi=6.5$ Å [Figs. 6(i) and 6(ii)], yet Fig. 13(a) shows that the average number of hydrogen bonds per water molecule in the interior region is higher at $\xi=3.8$ Å than that at $\xi=6.5$ Å.

At first sight, these observations appear to be inconsistent with the general relation between local water density and water-water binding energy derived above in Sec. II C, which states that local water density $\rho(\mathbf{r}_1)$ is proportional to $(\langle \exp(\beta U_1) \rangle_{a,r_1})^{-1}$, where U_1 is the water-water binding energy of a given water molecule [Eq. (27)]. This means that a higher local water density implies a lower average for $\exp(\beta U_1)$; and since hydrogen bonds are defined by low U_1 in the present treatment, one might expect that a higher local water density would automatically lead to a larger number of hydrogen bonds per water molecule.

However, this simple-minded picture is incomplete. Equation (29) implies that, because of the $\exp(\beta E_{ww})$ factor in the integral, configurations with high (repulsive) E_{ww} s have more dominant effects on $\langle \exp(\beta U_1) \rangle_{a,r_1}$, and therefore local water density $\rho(\mathbf{r}_1)$, than configurations with low E_{ww} . High E_{ww} configurations correspond to collisions among water molecules. But hydrogen bond number counts only stable low- E_{ww} configurations. By definition, they do not account for rare collisions that have high E_{ww} s. Therefore, no direct correlation between hydrogen bond count and local water density should be expected.

This points to where future efforts should be directed to decipher the relationship between hydrogen bonding and local water density around hydrophobic solutes. It is evident from the present result that merely counting the number of hydrogen bonds is insufficient. More fundamentally, one should seek to understand how hydrogen bond network or clathrate-like structures^{64,65} affect probabilities of water-water and water-solute collisions, as these collisions are the principal determinants of local water density and the mean force exerted on the solutes. For instance, Fig. 6(a) indicates that the inner-shell waters in the interior region at the contact minimum ($\xi=3.8$ Å) are more ordered than those in the same region at $\xi=6.5$ Å. This follows from the fact that the former has a larger number of hydrogen bonds per water and

also a lower entropy (Table I). One may then speculate that a higher degree of water ordering could lead to lower water mobility, which may then be translated into less water-water and water-solute collisions. If this argument is valid, it would offer a microscopic explanation for the observed lower local water density and smaller dissociate mean force component at $\xi=3.8$ Å relative to that at $\xi=6.5$ Å in this region. We intend to put these conjectures to the test in future work.

The above discussion does not belittle the important role of hydrogen bonding in water. As the most essential property of water, the *possibility* that water molecules can form energetically favorable hydrogen bonds has to be a dominant determinant of water properties and hydrophobic effects. However, this intuitive logic does not imply that any given water property can be related to a *simple* hydrogen bond count. The reason is that statistical mechanical averages take into account all configurations, both with and without a preponderance of hydrogen bonds, even though it is likely that the former are energetically more favored than the latter. But the number of hydrogen bonds accounts mainly for low-energy but not high-energy configurations. It follows that the relationship between a given water property and the number of hydrogen bonds can be complex, as we have observed above for local water density.

IX. CONCLUDING REMARKS

Extensive *NPT* simulations have been conducted to determine the temperature dependence of the two-methane PMF under atmospheric pressure. Temperature effects are found to be significant. Contrary to previous assertions, we have demonstrated that the PMF at a given temperature depends sensitively on overall (average) water density (and therefore pressure). Hence it is important to employ an overall water density that is appropriate for the experimental situation one aims to model.

Thermodynamic signatures—enthalpy, entropy, and heat capacity—at the contact minimum, desolvation barrier, and solvent-separated minimum were estimated. We observed nonadditivity effects. The heat capacity of the desolvation barrier is found to be large and positive; opposite in sign to the negative heat capacity change expected of the association of nonpolar groups in water. Solvent accessible surface area gives reasonable prediction for the entropy increase upon formation of a methane contact pair, but it drastically overestimates the concomitant decrease in heat capacity. Similar comparisons indicated that another implicit-solvent model also fails to provide accurate predictions for the PMF thermodynamic signatures reported here.

We have presented a formulation to relate the hydration mean force exerted by water molecules on a nonpolar solute, and found that the strength of this force correlates well with the local water density residing within the inner shell surrounding the solute. A potential distribution formalism for dissecting energetic contributions to PMFs has also been developed. Both of these perspectives emphasize the importance of high-energy molecular collisions as determining factors for both the hydration mean force and the PMF itself. This view has allowed us to make the observation that no

simple correlation should be expected *a priori* between local water density and the average number of hydrogen bonds made per water molecule. This is because hydrogen bond number by itself is only concerned with stable, low-energy configurations. Therefore this count is oblivious to high-energy configurations that are rare yet critical in controlling local water density.

Note added in Proof. After this paper had been accepted for publication, a study by Rick¹¹⁴ on the entropy and heat capacity effects of two-methane association appeared, the conclusions of which are quite different from ours. In particular, Rick reported a $\Delta C_P \approx -600$ cal/mol/K when the two methanes are in contact, and observed no positive ΔC_P at the desolvation barrier at 1 atm. However, the reliability of his results is questionable because the magnitude of his reported ΔC_P is more than one order of magnitude larger than the experimental hydration heat capacity of a single methane.^{88,89} Furthermore, while the accuracy of zero-PMF baselines is crucial for the reliability of heat capacity estimated by comparing PMFs at different temperatures, zero-PMF baselines may not have been properly determined in Ref. 114. This is illustrated by the fact that although Rick set PMFs to zero at the point when the two methanes are 8.0 Å apart, this data (Fig. 1 of Ref. 114) shows that there are substantial changes in PMF values ranging from ≈ -0.2 to -0.4 kcal/mol when the two-methane separation is further increased to ≥ 9 Å.

ACKNOWLEDGMENTS

We are very grateful to Nobuyuki Matubayasi for helpful discussions on the method of test-particle insertion, to William Jorgensen for his generous help with BOSS, and to Danny Heap for his tireless effort in maintaining our computing system. This work was supported in part by the Connaught Fund.

- ¹W. Kauzmann, *Adv. Protein Chem.* **14**, 1 (1959).
- ²K. A. Dill, *Biochemistry* **29**, 7133 (1990).
- ³W. Blokzijl and J. B. F. N. Engberts, *Angew. Chem. Int. Ed. Engl.* **32**, 1545 (1993).
- ⁴G. I. Makhatadze and P. L. Privalov, *J. Mol. Biol.* **213**, 375 (1990).
- ⁵R. S. Spolar, J. R. Livingstone, and M. T. Record, *Biochemistry* **31**, 3947 (1992).
- ⁶K. P. Murphy and E. Freire, *Adv. Protein Chem.* **43**, 313 (1992).
- ⁷J. Gomez, V. J. Hilsner, D. Xie, and E. Freire, *Proteins* **22**, 404 (1995).
- ⁸J. M. Sturtevant, *Proc. Natl. Acad. Sci. USA* **74**, 2236 (1977).
- ⁹R. L. Baldwin, *Proc. Natl. Acad. Sci. USA* **83**, 8069 (1986).
- ¹⁰P. L. Privalov and S. J. Gill, *Adv. Protein Chem.* **39**, 191 (1988).
- ¹¹H. S. Chan and K. A. Dill, *Annu. Rev. Biophys. Biomol. Struct.* **26**, 425 (1997).
- ¹²D. R. DeVido, J. G. Dorsey, H. S. Chan, and K. A. Dill, *J. Phys. Chem. B* **102**, 7272 (1998).
- ¹³B.-L. Chen and J. A. Schellman, *Biochemistry* **28**, 685 (1989).
- ¹⁴K. A. Dill, D. O. V. Alonso, and K. Hutchinson, *Biochemistry* **28**, 5439 (1989).
- ¹⁵F. Franks, *Adv. Protein Chem.* **46**, 105 (1995).
- ¹⁶G. Graziano, F. Catanzano, A. Riccio, and G. Barone, *J. Biochem. (Tokyo)* **122**, 395 (1997).
- ¹⁷G. Panick, G. J. Vidugiris, R. Malessa, G. Rapp, R. Winter, and C. A. Royer, *Biochemistry* **38**, 4175 (1999).
- ¹⁸G. I. Makhatadze and P. L. Privalov, *Adv. Protein Chem.* **47**, 307 (1995).
- ¹⁹D. Shortle, *FASEB J.* **10**, 27 (1996).
- ²⁰T. R. Sosnick and J. Trehwella, *Biochemistry* **31**, 8329 (1992).
- ²¹Y. Hagihara, M. Hoshino, D. Hamada, M. Kataoka, and Y. Goto, *Folding Des.* **3**, 195 (1998).
- ²²D. Yang, Y. K. Mok, J. D. Forman-Kay, N. A. Farrow, and L. E. Kay, *J. Mol. Biol.* **272**, 790 (1997).
- ²³T. Lazaridis and M. Karplus, *Proteins* **35**, 133 (1999).
- ²⁴T. Lazaridis and M. Karplus, *Biophys. Chem.* **78**, 207 (1999).
- ²⁵H. S. Chan, *Proteins* **40**, 543 (2000).
- ²⁶H. Kaya and H. S. Chan, *Proteins* **40**, 637 (2000).
- ²⁷B. Roux and T. Simonson, *Biophys. Chem.* **78**, 1 (1999).
- ²⁸R. H. Wood and P. T. Thompson, *Proc. Natl. Acad. Sci. USA* **87**, 946 (1990).
- ²⁹R. M. Jackson and M. J. E. Sternberg, *Protein Eng.* **7**, 371 (1994).
- ³⁰R. B. Hermann, *J. Comput. Chem.* **18**, 115 (1997).
- ³¹J. A. Rank and D. Baker, *Protein Sci.* **6**, 347 (1997).
- ³²C. Y. Lee, J. A. McCammon, and P. J. Rossky, *J. Chem. Phys.* **80**, 4448 (1984).
- ³³Y.-K. Cheng and P. J. Rossky, *Nature (London)* **392**, 696 (1998).
- ³⁴K. A. T. Silverstein, A. D. J. Haymet, and K. A. Dill, *J. Am. Chem. Soc.* **120**, 3166 (1998).
- ³⁵K. A. T. Silverstein, A. D. J. Haymet, and K. A. Dill, *J. Chem. Phys.* **111**, 8000 (1999).
- ³⁶N. T. Southall and K. A. Dill, *J. Phys. Chem. B* **104**, 1326 (2000).
- ³⁷K. Lum, D. Chandler, and J. D. Weeks, *J. Phys. Chem. B* **103**, 4570 (1999).
- ³⁸A. Sacco, A. Ascioffa, E. Matteoli, and M. Holz, *J. Chem. Soc., Faraday Trans.* **92**, 35 (1996).
- ³⁹A. Sacco and M. Holz, *J. Chem. Soc., Faraday Trans.* **93**, 1101 (1997).
- ⁴⁰M. Mayele, M. Holz, and A. Sacco, *Phys. Chem. Chem. Phys.* **1**, 4615 (1999).
- ⁴¹C. Pangali, M. Rao, and B. J. Berne, *J. Chem. Phys.* **71**, 2982 (1979).
- ⁴²W. L. Jorgensen, J. K. Buckner, S. Boudon, and J. Tirado-Rives, *J. Chem. Phys.* **89**, 3742 (1988).
- ⁴³D. E. Smith, L. Zhang, and A. D. J. Haymet, *J. Am. Chem. Soc.* **114**, 5875 (1992).
- ⁴⁴D. E. Smith and A. D. J. Haymet, *J. Chem. Phys.* **98**, 6445 (1993).
- ⁴⁵D. van Belle and S. J. Wodak, *J. Am. Chem. Soc.* **115**, 647 (1993).
- ⁴⁶L. X. Dang, *J. Chem. Phys.* **100**, 9032 (1994).
- ⁴⁷J. Forsman and B. Jönsson, *J. Chem. Phys.* **101**, 5116 (1994).
- ⁴⁸P. Jungwirth and R. Zahradnik, *Chem. Phys. Lett.* **217**, 319 (1994).
- ⁴⁹G. Hummer, S. Garde, A. E. Garcia, A. Pohorille, and L. R. Pratt, *Proc. Natl. Acad. Sci. USA* **93**, 8951 (1996).
- ⁵⁰S. Lüdemann, H. Schreiber, R. Abseher, and O. Steinhauser, *J. Chem. Phys.* **104**, 286 (1996).
- ⁵¹S. Lüdemann, R. Abseher, H. Schreiber, and O. Steinhauser, *J. Am. Chem. Soc.* **119**, 4206 (1997).
- ⁵²W. S. Young and C. L. Brooks III, *J. Chem. Phys.* **106**, 9265 (1997).
- ⁵³S. W. Rick and B. J. Berne, *J. Phys. Chem. B* **101**, 10488 (1997).
- ⁵⁴V. A. Payne, N. Matubayasi, L. R. Murphy, and R. M. Levy, *J. Phys. Chem. B* **101**, 2054 (1997).
- ⁵⁵G. Hummer, S. Garde, A. E. Garcia, M. E. Paulatis, and L. R. Pratt, *Proc. Natl. Acad. Sci. USA* **95**, 1552 (1998).
- ⁵⁶N. T. Skipper, *Chem. Phys. Lett.* **207**, 424 (1993).
- ⁵⁷R. L. Mancera and A. E. Buckingham, *Chem. Phys. Lett.* **234**, 296 (1995).
- ⁵⁸N. T. Skipper, C.H. Bridgeman, A. D. Buckingham, and R. L. Mancera, *Faraday Discuss.* **103**, 141 (1996).
- ⁵⁹R. L. Mancera, A. D. Buckingham, and N. T. Skipper, *J. Chem. Soc., Faraday Trans.* **93**, 2263 (1997).
- ⁶⁰M. Ikeguchi, S. Shimizu, S. Nakamura, and K. Shimizu, *J. Phys. Chem. B* **102**, 5891 (1998).
- ⁶¹N. Matubayasi, L. H. Reed, and R. M. Levy, *J. Phys. Chem.* **98**, 10640 (1994).
- ⁶²R. Brem, H. S. Chan, and K. A. Dill, *J. Phys. Chem. B* **104**, 7471 (2000).
- ⁶³L. R. Pratt and D. Chandler, *J. Chem. Phys.* **67**, 3683 (1977).
- ⁶⁴A. Geiger, A. Rahman, and F. H. Stillinger, *J. Chem. Phys.* **70**, 263 (1979).
- ⁶⁵T. Head-Gordon, *Proc. Natl. Acad. Sci. USA* **92**, 8308 (1995).
- ⁶⁶D. A. Zichi and P. J. Rossky, *J. Chem. Phys.* **83**, 797 (1985).
- ⁶⁷B. J. Berne, *Proc. Natl. Acad. Sci. USA* **93**, 8800 (1996).
- ⁶⁸H. Reiss, H. L. Frisch, and J. L. Lebowitz, *J. Chem. Phys.* **31**, 369 (1959).
- ⁶⁹R. A. Pierotti, *J. Phys. Chem.* **69**, 281 (1965).
- ⁷⁰R. A. Pierotti, *Chem. Rev.* **76**, 717 (1976).
- ⁷¹T. L. Hill, *Statistical Mechanics* (McGraw-Hill, New York, 1956).
- ⁷²N. Matubayasi, *J. Am. Chem. Soc.* **116**, 1450 (1994).

- ⁷³M. Pellegrini, N. Gronbech-Jensen, and S. Doniach, *J. Chem. Phys.* **104**, 8639 (1996).
- ⁷⁴H. S. Ashbaugh, S. Garde, G. Hummer, E. W. Kaler, and M. E. Paulaitis, *Biophys. J.* **77**, 645 (1999).
- ⁷⁵B. Widom, *J. Chem. Phys.* **39**, 2808 (1963).
- ⁷⁶B. Widom, *J. Phys. Chem.* **86**, 869 (1982).
- ⁷⁷A. Ben-Naim, *Solvation Thermodynamics* (Plenum, New York, 1987).
- ⁷⁸M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, Oxford, 1987).
- ⁷⁹H. S. Chan and K. A. Dill, *J. Chem. Phys.* **101**, 7007 (1994).
- ⁸⁰S. Shimizu, M. Ikeguchi, S. Nakamura, and K. Shimizu, *J. Chem. Phys.* **110**, 2971 (1999).
- ⁸¹K. S. Shing and S. T. Chung, *J. Phys. Chem.* **91**, 1674 (1987).
- ⁸²These expressions are the liberation term *after* the insertion of solute *a* into the system. Because the system is sufficiently large, the differences between using these expressions vis-à-vis the *pre*-insertion $-kT \ln \langle V \rangle_N$ and $-kT \ln \langle V \rangle_{N,b}$ are negligible.
- ⁸³W. J. Jorgensen and J. D. Madura, *Mol. Phys.* **56**, 1381 (1983).
- ⁸⁴W. L. Jorgensen, boss, Version 4.1 Yale University, New Haven, CT (1999).
- ⁸⁵K. S. Shing and K. E. Gubbins, *Mol. Phys.* **46**, 1109 (1982).
- ⁸⁶In the test-particle insertion approach, $\langle V \rangle_N$ and $\langle V \rangle_{N,b}$ are readily calculated because they entail averaging over the generated configurations. Because the test-particle method averages over ensemble *before* but not after solute insertion, it is not straightforward to use it to calculate $\langle V \rangle_{N,a}$ or $\langle V \rangle_{N,ab}$.
- ⁸⁷W. J. Jorgensen, J. F. Blake, and J. K. Buckner, *Chem. Phys.* **129**, 193 (1989).
- ⁸⁸H. Naghibi, S. F. Dec, and S. J. Gill, *J. Phys. Chem.* **90**, 4621 (1986).
- ⁸⁹T. J. Morrison and F. Billett, *J. Chem. Soc.* 3819 (1952).
- ⁹⁰B. Lee, *Biopolymers* **31**, 993 (1991).
- ⁹¹S.-I. Segawa and M. Sugihara, *Biopolymers* **23**, 2473 (1984).
- ⁹²B.-L. Chen, W. A. Baase, and J. A. Schellman, *Biochemistry* **28**, 691 (1989).
- ⁹³S. E. Jackson and A. R. Fersht, *Biochemistry* **30**, 10428 (1991).
- ⁹⁴T. R. Sosnick, L. Mayne, R. Hiller, and S. W. Englander, *Nat. Struct. Biol.* **1**, 149 (1994).
- ⁹⁵T. Schindler and F. X. Schmid, *Biochemistry* **35**, 16833 (1996).
- ⁹⁶H. S. Chan, in *Monte Carlo Approach to Biopolymers and Protein Folding*, edited by P. Grassberger, G. T. Barkema, and W. Nadler (World Scientific, Singapore, 1998).
- ⁹⁷H. S. Chan and K. A. Dill, *Proteins* **30**, 2 (1998).
- ⁹⁸V. Daggett and M. Levitt, *Proc. Natl. Acad. Sci. USA* **89**, 5142 (1992).
- ⁹⁹D. O. V. Alonso and V. Daggett, *J. Mol. Biol.* **247**, 501 (1995).
- ¹⁰⁰A. Li and V. Daggett, *J. Mol. Biol.* **257**, 412 (1996).
- ¹⁰¹G. S. Kell, *J. Chem. Eng. Data* **12**, 66 (1967).
- ¹⁰²V. Daggett, A. Li, L. S. Itzhaki, D. E. Otzen, and A. R. Fersht, *J. Mol. Biol.* **257**, 430 (1996).
- ¹⁰³B. Lee and F. M. Richards, *J. Mol. Biol.* **55**, 379 (1971).
- ¹⁰⁴T. Ooi, M. Oobatake, G. Nemethy, and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **84**, 3086 (1987).
- ¹⁰⁵T. Lazaridis and M. Karplus, *Science* **278**, 1928 (1997).
- ¹⁰⁶K. A. Dill, *J. Biol. Chem.* **272**, 701 (1997).
- ¹⁰⁷J. K. Buckner and W. L. Jorgensen, *J. Am. Chem. Soc.* **111**, 2507 (1989).
- ¹⁰⁸J. Chandrasekhar, D. C. Spellmeyer, and W. L. Jorgensen, *J. Am. Chem. Soc.* **106**, 903 (1984).
- ¹⁰⁹J. Chandrasekhar and W. L. Jorgensen, *J. Chem. Phys.* **77**, 5080 (1982).
- ¹¹⁰A. Pohorille and L. R. Pratt, *J. Am. Chem. Soc.* **112**, 5066 (1990).
- ¹¹¹L. R. Pratt and A. Pohorille, *Proc. Natl. Acad. Sci. USA* **89**, 2995 (1992).
- ¹¹²B. Lee, *Biophys. Chem.* **51**, 271 (1994).
- ¹¹³W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.* **79**, 926 (1983).
- ¹¹⁴S. W. Rick, *J. Phys. Chem. B* **104**, 6884 (2000).