Molecular simulation study of cooperativity in hydrophobic association

CEZARY CZAPLEWSKI,^{1,2} SYLWIA RODZIEWICZ-MOTOWIDŁO,² ADAM LIWO,^{1,2}
DANIEL R. RIPOLL,³ RYSZARD J. WAWAK,¹ AND HAROLD A. SCHERAGA¹

¹Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301

²Faculty of Chemistry, University of Gdańsk, ul. Sobieskiego 18, 80-952 Gdańsk, Poland

³Cornell Theory Center, Ithaca, New York 14853-3801

(RECEIVED December 22, 1999; FINAL REVISION April 11, 2000; ACCEPTED April 18, 2000)

Abstract

To investigate the cooperativity of hydrophobic interactions, the potential of mean force of two- and three-molecule methane clusters in water was determined by molecular dynamics simulations using two methods: umbrella-sampling with the weighted histogram analysis method and thermodynamic integration. Two water models, TIP3P and TIP4P, were used, while each methane molecule was modeled as a united atom. It was found that the three-body potential of mean force is not additive, i.e., it cannot be calculated as a sum of two-body contributions, but requires an additional three-body *cooperative* term. The cooperative term, which amounts to only about 10% of the total hydrophobic association free energy, was found to increase the strength of hydrophobic association; this finding differs from the results of earlier Monte Carlo studies with the free energy perturbation method of Rank and Baker (1997). As in the work of Rank and Baker, the solvent contribution to the potential of mean force was found to be well approximated by the molecular surface of two methane molecules. Moreover, we also found that the cooperative term is well represented by the difference between the molecular surface of the three-methane cluster and those of all three pairs of methane molecules. In addition, it was found that, while there is a cooperative contribution to the hydrophobic association free energy albeit a small one, the errors associated with the use of pairwise potentials are comparable to or larger than this contribution.

Keywords: cooperativity; hydrophobic association; molecular surface; potential of mean force; solvation models

Hydrophobic effects play an extremely important role in many chemical phenomena in aqueous solution (Némethy & Scheraga, 1962a, 1962b, 1962c; Ben-Naim, 1980). These include the folding and stabilization of proteins (Kauzmann, 1959; Scheraga, 1998), molecular recognition, stability of molecular membranes, surfactant aggregation, coagulation, complexation, detergency, and the formation of gas clathrates. In particular, any attempt to model the folding of proteins or to develop an *ab initio* method for protein structure prediction depends on a proper description of hydrophobic interactions. Therefore, all force fields used in the simulation of protein conformation must include water explicitly or must contain terms representing the free energy of hydrophobic association (Gibson & Scheraga, 1969; Gö & Scheraga, 1976; Ravishanker et al., 1982; Cramer & Truhlar, 1999; No et al., 1999; Roux & Simonson, 1999). Mean-field approaches should be derived and parameterized on the basis of all-atom studies of hydrophobic association of model systems.

Godzik et al. (1993) have shown that some cooperative terms must be included, but they derived them on a heuristic basis. Recently, based on the cumulant expansion of the restricted free energy of a polypeptide chain, we derived analytical formulas for the cooperative contribution to the energy of hydrogen-bonding interactions of united peptide groups (Liwo et al., 1998) and for mixed cooperative terms accounting for the coupling between backbone hydrogen-bonding and local interactions (Liwo et al., 1999a). The hydrogen-bonding cooperative terms were implemented in our united-residue UNRES force field (Liwo et al., 1999c), which enabled us to carry out *de novo* predictions of predominantly helical proteins with considerable success, as demonstrated during the CASP3 experiment (Liwo et al., 1999b). Without the cooperative terms, the UNRES force field was capable of recognizing only the native structure among the decoys derived from the Protein Data Bank (PDB) (Liwo et al., 1997, 1999c).

A single methane molecule as well as a methane dimer or small methane clusters in water have been the most commonly used systems to model hydrophobic hydration (Owicki & Scheraga, 1977; Swaminathan et al., 1978; Rapaport & Scheraga, 1982) and hydrophobic association (Swaminathan & Beveridge, 1979; Lüde-

Reprint requests to: Harold A. Scheraga, Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301; e-mail: has5@cornell.edu.

mann et al., 1997; Rank & Baker, 1997; Tsai et al., 1997; Young & Brooks, 1997) theoretically. Cooperativity effects in solvent-induced forces [the negative gradient of the potential of mean force (PMF)] have been the focus of recent investigations (San Biagio et al., 1998). Recently, Rank and Baker (1997) determined the PMF of methane clusters in water and found that it cannot be decomposed into pairwise terms. They have shown that the solvent contribution to the hydrophobic-association free energy is related to the molecular surface of the hydrophobic cluster. However, these authors did not analyze a possible relation between the *cooperativity* term and the molecular surface of clusters of methanes. Therefore, in our present study, we focus attention on this issue.

To determine the cooperative part of the hydrophobic-association free energy quantitatively, we have determined the PMF of two- and three-molecule clusters of methanes in water at different configurations, using the molecular dynamics (MD) procedure with explicit water molecules to obtain statistical ensembles. To evaluate the PMF, we tried various methods: free-energy perturbation (FEP) (Tembe & McCammon, 1984; Jorgensen & Ravimohan, 1985), thermodynamic integration (TI) (Pearlman, 1993), and umbrella sampling (Torrie & Valleau, 1977) with the Weighted Histogram Analysis Method (WHAM) (Kumar et al., 1992, 1995). We found it necessary here to carry out a comparison of the performance of these methods to calculate the PMF because the free-energy changes considered in this study are very small.

This paper is organized as follows. In Results and discussion, we present the performance of various approaches and the selection of the appropriate method. Then, we present our results for the determination of the PMF of methane pairs and triplets in water. Finally, we discuss the similarity between the cavity formation free energy and the molecular surface of the hydrophobic cluster. In Materials and methods, we describe the methods used to calculate the PMF and the model systems.

Results and discussion

Comparison of the accuracy of TI and umbrella sampling/WHAM methods

Figure 1 shows the results of TI calculations of the PMF for a two-methane system ($m + m$) and TIP3P water model for different simulation times in each of 80 windows (defined as the range of the reaction coordinate). The PMF has a characteristic shape with one deep minimum at 3.9 Å, corresponding to the methane-methane van der Waals (VDW) contact distance (hereafter referred to as the *contact minimum*) and the second (shallow) minimum at about 7 Å, corresponding to the distance at which precisely one water molecule can enter the space between the two solutes (Geiger et al., 1979). This second minimum will hereafter be referred to as the *solvent-separated minimum*. To assess the error propagation, one series of runs was carried out starting from the lower boundary of the distance and increasing it and another series of runs starting from the upper boundary of the distance and decreasing the distance. As shown, the curves from the forward and backward runs differ significantly (by ~ 0.2 kcal/mol at the contact minimum) even for the longest MD runs. The differences do not become smaller by increasing the simulation time per window. It can therefore be concluded that accumulation of errors during the Euler integration rather than insufficient sampling is the main source of error.

Fig. 1. PMF curves for the $m + m$ system and TIP3P water model obtained from the TI calculations using a box of 20 Å and 80 windows. Different equilibration times (the first number following a graph legend) and data collection times (the second number following a graph legend) were used, as well as backward or forward direction of sweeping the reaction coordinate (either starting at 3.1 Å and ending at 9.1 Å or starting at 9.1 Å and ending at 3.1 Å). The reaction coordinate corresponds to the distance between the methane molecules. Calculations were carried out at 298 K.

To overcome this, we would have to decrease the integration intervals (increase the number of windows) or use a better method than the Euler integration. Other evidence of the problem of error propagation is shown in Figure 2 where the PMF obtained using a 30 Å box and backward integration is compared with the PMF obtained using a 20 Å box, with the same simulation time per window and backward integration. It can be seen that the values at the contact minimum differ significantly. Therefore, the difference

Fig. 2. PMF curves for the $m + m$ system and TIP3P water model obtained from the TI calculations with a large (30 Å) box using 200 windows compared with results for a smaller (20 Å) box and 80 windows. The simulations were carried out assuming a temperature of 298 K.

depends not only on the direction of integration, but also on the position of the starting point, which indicates that the dominant source of error is associated with an insufficiently small integration step that is generally used in this type of simulations. The magnitude of the error found in the TI calculations described above is comparable to that reported in other PMF calculations using the TI method (Pearlman, 1993; Chipot et al., 1996).

Figure 3 shows the dependence of the $m + m$ PMF (20 Å box) obtained using the umbrella sampling/WHAM method on the amount of data collected from each window using the same water model as in the TI calculations. Unlike the behavior of TI, convergence is achieved as the number of data increases. However, for this system there is no systematic dependence of the results on the number of windows (convergence is achieved after 500 ps for 11 windows separated by 0.5 Å and 900 ps for 6 windows separated by 1 Å); the differences are well accommodated within the random error inherent in the simulation (it must be kept in mind that the magnitude of this error decreases only as the square root of the simulation time). Even this random error can be estimated to be within 0.05 kcal/mol, which is sufficient accuracy for the purpose of this study, as opposed to the accuracy of the TI method obtained with similar computational effort (a total of ~6 ns simulation time). As will be shown later, the magnitude of the cooperative effect is ~0.1 kcal/mol per pair of methane molecules in a cluster and, therefore, the error inherent in TI with the control parameters that we used is unacceptably large for studying cooperativity.

Figures 4A and 4B show the convergence of the PMFs computed by using 6 and 12 windows, respectively, for a three-methane system in which an equilateral-triangle geometry was assumed ($m + m + m$). The PMFs were obtained by using the umbrella sampling/WHAM method, and the figures indicate that the convergence is as good as that corresponding to the methane-dimer simulations, especially for 12 windows. For the above reasons, we decided to use the umbrella sampling/WHAM method, rather than trying to fine tune the TI control parameters to improve the accuracy; the latter probably would have multiplied the required computational effort.

Fig. 3. PMF curves for the $m + m$ system obtained from the umbrella sampling/WHAM method using the TIP3P water model and a 20 Å box with increasing simulation time and 11 windows. The simulations were carried out assuming a temperature of 298 K.

Fig. 4. Convergence of the three-body PMF in umbrella sampling/WHAM simulations of the $m + m + m$ system using the TIP4P water model and a 20 Å box with increasing simulation time in each window for (A) 6 and (B) 12 windows. The PMF curves have been normalized to show the free energy per unconstrained pair of methanes [$W(\xi)/3$]. The simulations were carried out assuming a temperature of 298 K.

Dependence of the PMF on nonbonded and electrostatic interaction energy cut-off and on the method to compute the electrostatic energy

To determine possible sources of systematic error, we carried out test calculations of the $m + m$ system using different nonbonded and electrostatic interaction energy cut-offs. We also carried out one calculation using particle-mesh Ewald summation (Darden et al., 1993) to compute the electrostatic interaction energy. As shown in Figure 5, the PMFs obtained with the 8.5 or 9.5 Å cut-offs for both nonbonded and electrostatic, as well as that obtained using the Ewald summation differ only within the random error. Moreover, the simulations with the Ewald summation are three times more time consuming, and the TIP3P and TIP4P water models had been parameterized for calculations with the use of an electrostatic interaction energy cut-off rather than the Ewald summation. Therefore, the main calculations were carried out by using a nonbonded and an electrostatic interaction energy cut-off and not the Ewald summation.

Fig. 5. Dependence of the PMF on the procedure to compute the electrostatic interaction energy for the $m + m$ system with the TIP3P water model and on the size of the box (20 and 30 Å). The simulations were carried out assuming a temperature of 298 K.

Determination of the cooperative term

As explained in Materials and methods, to compute the cooperative terms, $\delta F^{(3)}$ (Equation 15), we used the PMFs from simulations of two and three methane molecules. Two different geometries for the three methane molecules were considered: (1) an equilateral triangle ($m + m + m$), with the dimension of the triangle being the reaction coordinate and (2) the three molecules on the vertices of an isosceles triangle ($2m + m$), with the distance between molecules 1 (m_1) and 2 (m_2) set to 3.9 Å (the contact-minimum distance between two methane molecules in water) and the distance, $\xi_{3,1}$, from the third methane molecule (m_3) to m_1 (same as $\xi_{3,2}$) being the reaction coordinate (Fig. 6).

The cooperative terms were calculated from the umbrella sampling/WHAM PMF curves of the $m + m + m$ and $m + m$ or $2m + m$ and $m + m$ simulations, respectively, using Equation 15,

Fig. 6. Definition of the reaction coordinate (ξ) for studying the distance dependence of the $m + m$, $m + m + m$ and $2m + m$ systems. For the $m + m$ system, it corresponds to the methane-methane distance, for $m + m + m$ it corresponds to the side of an equilateral triangle and for $2m + m$ system it corresponds to the side of an isosceles triangle.

i.e., by subtracting the sum of normalized two-body PMFs from the normalized three-body PMF. The total PMF and corresponding cooperative-term curves for the $m + m + m$ system obtained with the TIP3P and TIP4P water models are shown in Figures 7A and 7B, respectively. The corresponding curves for the $2m + m$ model in which two alternative methods to constrain/restrain molecules m_1 and m_2 are shown in the two sets of Figures 8A,B and 8C,D. In contrast to the determination of the pairwise PMF, the coverage of the distances has to be quite dense; when using only six windows (separated by 1 Å), the shape of $\delta F^{(3)}$ almost disappears within the noise, and the cooperative contribution appears negligible (Figs. 7A, 8A, 8C, 8D). The shape of $\delta F^{(3)}$ becomes well defined in all cases only when 12 windows are used.

For the $m + m + m$ system the cooperative term is negative for short distances (Fig. 7A,B). With the TIP4P model, the region where the cooperative term is negative is narrower (up to 6.5 Å); it corresponds to the region of the contact minimum of the total PMF. For the TIP3P water model, the region of negative cooper-

Fig. 7. Curves for the PMF and the cooperative term for the $m + m + m$ system obtained with the (A) TIP3P and the (B) TIP4P water model and the umbrella-sampling/WHAM method, using 6, 11, or 12 windows and a 20 Å box. The PMF curves have been normalized to show the free energy per unconstrained pairs of methanes [$W(\xi)/3$]. The simulations were carried out assuming a temperature of 298 K.

Fig. 8. Curves for the PMF and the cooperative term for the $2m + m$ system obtained with the umbrella-sampling/WHAM method, using 6, 11, or 12 windows, and (A) the TIP3P water model and 100 kcal/mol/Å² restraint on the $m_1 - m_2$ distance, (B) the TIP4P water model with distance restraints as in (A); (C) the TIP4P water model with angular and distance restraints to force the isosceles triangle geometry, as specified in Figure 9B; (D) the TIP4P water model with angular and distance restraints plus exact constraints on the distance between m_1 and m_2 as specified in Figure 9C. The 2 kcal/mol/Å² distance restraints on the reaction coordinates were imposed in each case to keep it within the assumed distance window. The PMF curves have been normalized to show the free energy per unconstrained pairs of methanes [$W(\xi)/2$]. The simulations were carried out assuming a temperature of 298 K.

activity extends to 7 Å to overlap with the region of the maximum of the total PMF. It should be noted that the maximum of the PMF for the $m + m$ system is remarkably higher with the TIP3P water model (~ 0.3 kcal/mol), compared to the TIP4P water model (~ 0.1 kcal/mol); for the $m + m + m$ system, the difference between the maximum values of the *total* PMF for the two water models is smaller than for the $m + m$ system.

As described in Materials and methods, three types of constraints or restraints were used to carry out simulations for the $2m + m$ system. In the first case, restraints were imposed only on the distances and not on the angles (Fig. 9A), which allowed some deviation from the isosceles-triangle geometry during the simulations [however, the data gathered for further processing corresponded to almost exact (within the dimensions of the distance bins) isosceles-triangle geometries]. This could cause some inefficiency of the sampling. In the second case, additional symmetry restraints were imposed to keep the geometry at that of an isosceles triangle (Fig. 9B). Within the error limits, the resulting PMFs did

not depend on the type of restraints used (Fig. 8B,C). The cooperative contribution to the PMF is negative for short distances, but the minimum is, however, narrower compared to the $m + m + m$ system and its center is shifted to the right (4.9 Å). It can also be seen that the negative region is broader for the TIP3P water model (Fig. 8A compared to Fig. 8B).

The above results differ from the findings of Rank and Baker (1997). Their simulations were carried out by using exact constraints (not harmonic restraints) on one of the methane-methane distances (corresponding to our $2m + m$ system). To check whether the imposition of *exact* constraints in the $2m + m$ trimer simulations changes the results, we carried out additional runs applying the SHAKE algorithm to constraint the $m_1 - m_2$ distance (Fig. 9C). For these simulations, we used the TIP4P water model, which was employed by Rank and Baker (1997). The results of these additional simulations are given in Figure 8D. For 12 windows for which better convergence is achieved, a negative cooperative term is obtained consistent with our previous results.

Fig. 9. The topologies of the $2m + m$ systems used in umbrella-sampling simulations. **A:** 2 kcal/mol/Å² force constant (k) is imposed on the reaction coordinate (the m_1 - m_3 and m_2 - m_3 distances) and a strong 100 kcal/mol/Å² force constant (K) is imposed on the m_1 - m_2 distance. **B:** Additional restraints for the angles α , β , and γ with 100 kcal/mol/rad² force constant are included to maintain the geometry of the isosceles triangle. **C:** The SHAKE algorithm is applied to constrain the distance between m_1 and m_2 , and angular restraints on the angles α and β with 100 kcal/mol/rad² force constant are used.

Relationship between the change of molecular surface in formation of hydrophobic clusters and the three-body contribution to the PMF

Recent results (Tuñón et al., 1992) indicate that the surface area of the solute is the only factor in determining the hydrophobic effect and that the molecular surface (Richards, 1977) is the most adequate one to be used for quantitative evaluation of this effect. As shown by Rank and Baker (1997), the dependence of the *solvent contribution to the PMF* (defined as the difference between the PMF and the solute-solute interaction energy of a hydrophobic pair) on the distance between the solutes is qualitatively similar to the change of the molecular surface of the solute pair. We have pursued this idea further and compared the distance dependence of the cooperative part of the three-body PMF, $\delta F^{(3)}$, with the change of molecular surface.

For the solute dimer and a three-solute cluster, the changes of the molecular surface $\delta S^{(2)}$ and $\delta S^{(3)}$ are defined by Equations 1 and 2, respectively.

$$\delta S^{(2)}(r) = [S^{(2)}(r) - 2S^{(1)}] \quad (1)$$

$$\delta S^{(3)}(r_{12}, r_{13}, r_{23}) = [S^{(3)}(r_{12}, r_{13}, r_{23}) - 3S^{(1)}] - [\delta S^{(2)}(r_{12}) + \delta S^{(2)}(r_{13}) + \delta S^{(2)}(r_{23})] \quad (2)$$

where $S^{(3)}(r_{12}, r_{13}, r_{23})$ is the surface area of the cluster with distances between the components equal to r_{12} , r_{13} , and r_{23} , respectively, $S^{(1)} = 4\pi R_s^2$ is the molecular surface area of an isolated solute, R_s being solute radius, and $S^{(2)}(r_{ij})$ is the molecular surface area of solute pair ij . The surface area of the isolated solute is subtracted in Equations 1 and 2 to establish a zero reference value at large distances, consistent with the reference value of the PMF (Equation 15).

The distance dependence of $S^{(3)}$, $\delta S^{(2)}$, and $\delta S^{(3)}$, for the $m + m + m$ and $2m + m$ systems is shown in Figure 10. As in the case of the corresponding $\delta F^{(3)}$, $\delta S^{(3)}$ exhibits a minimum between the contact and the solvent-separated distance and is negative in the region of this minimum. The minimum occurs at a greater distance for the $2m + m$ system (4.5 Å), compared to the $m + m + m$ system in the case of the $\delta F^{(3)}$ (4.2 Å). Therefore, it can be concluded that the three-body contribution to the hydrophobic-association PMF can be accounted for, at least qualitatively, by the difference between the molecular surface area of the cluster and that of the sum of the surface areas of all the constituent dimers.

This result can be explained as follows. The molecular surface area of the dimer consists of the convex area and the saddle area (Fig. 11A). In the case of a trimer configuration, a portion of the original surface areas of the dimers (facing toward the inside of the trimer) disappears (provided that all the constituents are closer than the solvent-separated distance). Instead, a concave surface area is formed to cover the top and the bottom of the trimer. This concave surface area, however, is smaller than the saddle surface area that vanished and, consequently, $S^{(3)} - 3S^{(1)}$ is smaller than the sum of the three $\delta S^{(2)}$ s (Fig. 11B).

Fig. 10. Distance dependence of the molecular surface area of the methane dimer $S^{(2)}$ and the methane trimers ($m + m + m$ and $2m + m$) $S^{(3)}$ together with corresponding differences $\delta S^{(3)}$. The curves have been normalized for comparison purposes. The curve corresponding to the $m + m + m$ system was normalized by subtracting from the $S^{(3)}$ term the surface areas of the three isolated monomers $[S^{(1)}]$ and then dividing the total by 3. The curve corresponding to the $2m + m$ system was normalized by subtracting from the $S^{(3)}$ term the surface areas of a monomer $[S^{(1)}]$ and the constrained-dimer with monomers separated by 3.9 Å $[S^{(2)}(3.9)]$ and dividing the total by 2. The molecular surfaces of the $m + m$, $m + m + m$, and $2m + m$ systems were calculated numerically using a standard triangulation method implemented in the MOLMOL program (Koradi et al., 1996).

Fig. 11. Partition of the molecular surface of the dimer into convex and saddle parts, and the molecular surface of the trimer into convex, saddle, and concave parts. These can be envisaged as the surfaces in contact with a probe ball rolling over the trimer.

Conclusions

As follows from this study, the cooperative three-body contribution to the hydrophobic-association PMF is negative for distances between the contact minimum and the solvent-separated minimum. This result has been obtained using two different types of cluster geometry (an equilateral and an isosceles triangle) and two different water models (TIP3P and TIP4P). Moreover, the results do not depend on the restraint/constraint scheme applied in umbrella sampling, which indicates that the coverage of the configurational space has been sufficient. Therefore, the results obtained in this study may be considered as reliable.

The cooperative contribution accounts for only about 10% of the total PMF at the contact minimum. The distance dependence of the cooperative term is qualitatively similar to the distance dependence of the difference between the molecular surface area of the cluster and that of the sum of the surface areas of the isolated constituent pairs. Together with the similarity between the shape of the distance dependence of the solvent contribution to the PMF of hydrophobic dimer formation (pointed out by Rank & Baker, 1997), this suggests that surface area models based on molecular surface area are reasonable ones to describe the energetics of hydrophobic association (Tuñón et al., 1992; Rank & Baker, 1997), even up to such fine details as multibody contributions to the PMF of a hydrophobic cluster.

Regarding the cooperative contribution to the free energy of hydrophobic association, our results are different from those of Rank and Baker (1997) obtained by using Monte Carlo simulations and the free energy perturbation method, in which they obtained a positive contribution to the free energy of hydrophobic association. The difference between our and their results is not caused by the manner in which constraints or restraints are imposed on the cluster geometry, because our results did not change after using exactly the same constraint scheme as in their calculations. Thus, the difference could arise only from the use of different techniques for determining the PMF. As indicated in the literature, the FEP method is liable to induce instability and hysteresis and therefore poor convergence (Pearlman, 1993; Radmer & Kollman, 1997). On the other hand, umbrella sampling with use of the WHAM provides much more stable results, if the sampling is sufficiently extensive

(Radmer & Kollman, 1997). The tests carried out in this study have also shown that the TI method [which has been reported to be superior to FEP (Pearlman, 1993)] resulted in much poorer convergence compared to umbrella sampling/WHAM within the same amount of computational effort. In addition, Rank and Baker (1997) indicated that the errors in their simulations from zeroing the PMFs may be as much as 0.3 kcal/mol, a value that is larger than the cooperative contribution that we measured. We have used a different technique to normalize the PMFs that minimizes the errors from the superposition of two- and three-body PMFs. Our method is based on the assumption that the cooperative term vanishes faster with distance than the PMFs.

Although the multibody term favors more pronounced hydrophobic association than would result from a pairwise model, its net contribution is small (about 10%). This implies that simple coarse-grain potentials of side-chain–side-chain interactions can be fairly well approximated as sums of pairwise terms. The inclusion of the cooperative terms in coarse-grain potentials could be justified only if the functional form of these potentials is more sophisticated than the usually-applied Lennard–Jones-like form (Seetharamulu & Crippen, 1991; Liwo et al., 1997). Consequently, the first improvement of a coarse-grain potential should be the incorporation of the maximum between the contact and the solvent-separated minima. Additional refinements may be carried out by including cooperative terms.

Materials and methods

For a general classical statistical mechanical treatment of a solute in water, references may be made to papers by Gibson and Scheraga (1969) and Gō and Scheraga (1976). The following sections summarize the theoretical treatment of methane clusters in water.

Calculation of the potential of mean force (PMF)

Let the configurational space of the system be defined by the vector of variables $\mathbf{x} = (x_1, x_2, \dots, x_m)$ (e.g., the Cartesian coordinates of all the atoms of the solute). Let us also define the variables of interest by the vector $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_n)$; in the present case, $\boldsymbol{\xi}$ will consist of all the distances between the methane molecules. The coordinates of $\boldsymbol{\xi}$ are functions of $\mathbf{x}$, and $n \ll m$. The potential of mean force (PMF), $W(\boldsymbol{\xi})$, is defined by the following equation:

$$W(\boldsymbol{\xi}) = -RT \ln P(\boldsymbol{\xi}, p, T, V) \quad (3)$$

where $P(\boldsymbol{\xi})$ is the conditional probability density of the variables of interest at a given value of $\boldsymbol{\xi}$; p , T , and V are the pressure, temperature, and volume of the system, and

$$P(\boldsymbol{\xi}, p, T, V) = \frac{\int_{V_{\boldsymbol{\xi}}} \exp(-H(\mathbf{x})) dV}{\int_V \exp(-H(\mathbf{x})) dV} \quad (4)$$

where H is the Hamiltonian of the system (solute plus solvent), V is the volume of the whole configurational space of the system, and $V_{\boldsymbol{\xi}}$ is the volume of the fraction of the configurational space restricted to $\boldsymbol{\xi}$. If P is evaluated at constant pressure and temperature, the potential of mean force W corresponds to the restricted Gibbs free energy of the system, while for constant volume and temper-

ature it corresponds to the restricted Helmholtz free energy. Throughout this paper, we identify the PMF with the restricted Helmholtz free energy.

Although many different methods have been developed to evaluate the free energy of interaction or changes in free energy, only a few of them are accurate enough to compute the potential of mean force. For example, the free-energy perturbation (FEP) method suffers from the “orthogonality” problem and requires a large number of very small windows in each of which an equilibrated Monte Carlo (MC) or MD ensemble must be generated (Pearlman, 1993). For this reason, we did not use the FEP method in our study.

The TI method is free from the “orthogonality” problem and has been recommended (Pearlman, 1993; Chipot et al., 1996) as a useful tool to compute the PMF. It is based on the following master equation (in this case ξ can only be a one-dimensional (1D) reaction coordinate denoted as ξ):

$$W(\xi) = \int_0^\xi \left\langle \frac{\partial H(\mathbf{x}, \lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \quad (5)$$

where λ defines the normalized reaction coordinate along which the PMF is calculated, related to the variables of interest ξ , by $\xi(\lambda) = \xi_0(1 - \lambda) + \xi_1\lambda$, ξ_0 and ξ_1 define the initial and final state, respectively, $\mathbf{x}$ denotes the degrees of freedom of the system, H is the Hamiltonian, and $\langle \dots \rangle_\lambda$ denotes averaging over the degrees of freedom $\mathbf{x}$ for which the normalized reaction coordinate is fixed at λ ; in the case of the dimer calculations, $\partial H / \partial \lambda$ is given by

$$\frac{\partial H}{\partial \lambda} = [\mathbf{f}(\mathbf{r}_1) - \mathbf{f}(\mathbf{r}_2)] \cdot \frac{\mathbf{r}_{12}}{r_{12}} \frac{\partial r_{12}}{\partial \lambda} \quad (6)$$

where $\mathbf{r}_i$ is the position vector of the solute i , $\mathbf{r}_{12}$ is the vector pointing from solute 1 to 2, and $\mathbf{f}(\mathbf{r}_i)$ is the force (from solvent molecules and the other solute molecule) acting on solute i (adapted from Pearlman, 1993).

The integration of Equation 5 has to be carried out numerically:

$$W(\xi) \approx \sum_{i=1}^{N_\xi} \left[\left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{\lambda(i)} + \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{\lambda(i+1)} \right] \left[\frac{\lambda(i+1) - \lambda(i)}{2} \right] \quad (7)$$

where N_ξ is the number of intervals (windows) from ξ_0 to ξ . The number of windows has to be large enough to minimize errors caused by discretization of the integrand. On the other hand, the errors accumulate during the integration, and increasing the number of windows too much would increase the error. In each window, the average (Equation 7) is computed by MC or MD sampling.

In the umbrella-sampling method (Torrie & Valleau, 1977; Frenkel & Smit, 1996), a series of restraints (usually harmonic) are imposed on the reaction coordinate to assure that all regions are sampled sufficiently (see Molecular dynamics simulations below). Each restraint defines a sampling window. [The meaning of a *window* is different in the TI and umbrella-sampling methods. For TI (and FEP), the window is defined as an interval of λ while, in the umbrella-sampling method, the center of the window, but not the boundaries, is well defined.] In general, the total Hamiltonian of the system (including the biasing potential) is defined by

$$H_{\{\lambda\}}(\xi) = V_0(\xi) + \sum_{i=1}^L \lambda_i V_i(\xi) = \sum_{i=0}^L \lambda_i V_i(\xi) \quad (8)$$

where L is the number of windows, V_0 is the energy function of the system consisting of the solute and the solvent, $\lambda_0 = 1$, and $V_1, V_2, \dots, V_L$ are the biasing restraining potentials; $\lambda_1, \lambda_2, \dots, \lambda_L$ are the coupling parameters. Each simulation at a given window is characterized by a different set of values of the coupling parameters. For example, in the case of our $m + m$ system, for the simulation at the i^{th} window $\lambda_i = 1$ and $\lambda_j = 0$, $j = 1, 2, \dots, i - 1, i + 1, \dots, L$. The regions covered by each window typically overlap to enhance sampling in the boundary regions.

To calculate the PMF from umbrella-sampling simulation data, the data from all simulation windows must be combined, and the contribution due to the restraining potential must be eliminated. The most advanced approach to this problem is the Weighted Histogram Analysis Method (WHAM) (Kumar et al., 1992, 1995). In this method, the *unbiased* probability density corresponding to a given ξ , $P(\xi)$ is computed as a weighted average of the counts for this ξ , $N_k(\{V\}, \xi)$, over all simulations (with different restraining potentials) (Equations 9, 10). Thus, the final probability density is constructed using *all* data, including those of the overlapping regions. The formulas for the weights (Kumar et al., 1992, 1995) were derived based on the requirement that the sum of the squares of the errors over all ξ in the calculated distribution be minimized.

$$P(\xi) = \sum_{k=1}^R N_k(\{V\}, \xi) w(\{V\}) \quad (9)$$

$$w(\{V\}) = \frac{1}{\sum_{l=1}^R n_l \exp \left[F_l - \beta \sum_{j=1}^L \lambda_j^l V_j \right]} \quad (10)$$

where R is the number of simulations (windows); usually $R = L$, which means that the number of restraints is equal to the number of simulations, n_l , $l = 1, 2, \dots, R$ is the total number of counts in simulation l :

$$n_l = \sum_{\{V\}, \xi} N_k(\{V\}, \xi), \quad l = 1, 2, \dots, R \quad (11)$$

and the quantities $F_1, F_2, \dots, F_R$ can be considered *dimensionless free energies* corresponding to the windows $1, 2, \dots, R$, respectively (Kumar et al., 1992, 1995). They are determined by solving the system of nonlinear equations defined by Equations 12 and 13; these equations arise directly from the minimization of the sum of the squares of the errors in the calculated probability density.

$$\exp(-F_l) = \sum_{\{V\}, \xi} P_{\lambda_l}(\xi), \quad l = 1, 2, \dots, R \quad (12)$$

$$P_{\{\lambda\}}(\xi) = \frac{\sum_{k=1}^R N_k(\{V\}, \xi) \exp \left(-\beta \sum_{j=1}^L \lambda_j^k V_j \right)}{\sum_{l=1}^R n_l \exp \left[F_l - \beta \sum_{j=1}^L \lambda_j^l V_j \right]}. \quad (13)$$

The sets of Equations 12 and 13 are iterated, until the dimensionless free energies $F_1, F_2, \dots, F_R$ attain self-consistency (Kumar et al., 1992, 1995). The unbiased probability density function $P(\xi)$ is then calculated from Equation 9 and, finally, the PMF along the reaction path $[W(\xi)]$ is calculated from $P(\xi)$ using Equation 3. For

a detailed description of the method, the reader is referred to the original papers (Kumar et al., 1992, 1995).

The great advantage of this approach is that, as opposed to TI and FEP, the errors do not propagate with the amount of the configurational space that is covered by the computations. Also, in contrast to FEP or TI, distributions can be determined for a multidimensional reaction coordinate ξ . Even if only a one-dimensional section of a multidimensional PMF is requested by symmetry constraints (as for the trimer in this study), the multidimensional distribution in the neighborhood of this section has to be calculated to take into account the volume factor corresponding to different points on the reaction coordinate.

Definition of cooperativity of hydrophobic hydration and model systems

The PMF defined by Equation 3 is not necessarily pairwise although, in our simulation, all the interactions that contribute to it are assumed to be pairwise. Generally, the PMF of an n -particle interaction can be decomposed into single-body, pairwise, and multi-body terms:

$$\begin{aligned} W(1,2,\dots,n) &= F(1,2,\dots,n) - \sum_i^n F^{(1)}(i) \\ &= \sum_{i<j}^n \delta F^{(2)}(i,j) + \sum_{i<j<k}^n \delta F^{(3)}(i,j,k) \\ &\quad + \dots + \sum_{i<j<\dots}^n \delta F^{(l)}(i,j,\dots) + \delta F^{(n)}(i,j,\dots,n) \end{aligned} \quad (14)$$

where $f(1,2,\dots,n)$ is the shorthand for $f(\xi_1, \xi_2, \dots, \xi_{3n-6})$, f being any function, and $n > 1$. The normalization term $\sum_i^n F^{(1)}(i)$ serves to define the reference state.

In the case of hydrophobic phenomena, the single-body contributions $F^{(1)}$ correspond to hydrophobic hydration, and the two-body terms $\delta F^{(2)}$ correspond to hydrophobic association. This term is defined by using Equation 14 for $n = 2$. The higher-order terms correspond to the cooperativity in hydrophobic association and can be derived recursively from Equation 14. In this work, we investigate the lowest-order cooperative terms $\delta F^{(3)}$.

Consider a cluster of three methane molecules in water. The relative geometry of the cluster is described by three site–site distances that define our reaction coordinate ξ , viz., $\xi_1 = r_{12}$, $\xi_2 = r_{13}$, and $\xi_3 = r_{23}$. Given the unnormalized total PMF, $F(r_{12}, r_{13}, r_{23})$, (i.e., the function obtained from the $m + m + m$ or $2m + m$ simulations), the cooperativity terms $\delta F^{(3)}(r_{12}, r_{13}, r_{23})$ can be expressed as

$$\begin{aligned} \delta F^{(3)}(r_{12}, r_{13}, r_{23}) &= F(r_{12}, r_{13}, r_{23}) - \delta F^{(2)}(r_{12}) \\ &\quad - \delta F^{(2)}(r_{13}) - \delta F^{(2)}(r_{23}) - 3F^{(1)} \\ &= F(r_{12}, r_{13}, r_{23}) - 3F^{(1)} \\ &\quad - [F(r_{12}) - 2F^{(1)}] - [F(r_{13}) - 2F^{(1)}] \\ &\quad - [F(r_{23}) - 2F^{(1)}] \\ &= W(r_{12}, r_{13}, r_{23}) \\ &\quad - [W(r_{12}) + W(r_{13}) + W(r_{23})] \end{aligned} \quad (15)$$

where $F(r_{ij})$ is the unnormalized total PMF for an isolated methane pair in water and, as mentioned above, $F^{(1)}$ is the hydrophobic hydration free energy of a single methane molecule, which is an unknown constant. Computing the differences $F(r_{12}, r_{13}, r_{23}) - 3F^{(1)}$ and $F(r_{ij}) - 2F^{(1)}$ is equivalent to shifting the respective total PMFs to zero values at infinite (practically sufficiently large) distances between the methane molecules and provides the *normalized* PMF, $W(r_{12}, r_{13}, r_{23})$, and $W(r_{ij})$, respectively. Thus, $\delta F^{(3)}(r_{12}, r_{13}, r_{23})$ can be determined from the difference between the three-body and the sum of two-body PMFs.

The determination of the PMF of two methane molecules in water is a 1D problem; we define the respective system as $m + m$ (Fig. 6), while that of the PMF of three methane molecules is a 3D one. In the latter case, it is reasonable to study the PMF in a chosen 1D subspace. Therefore, we defined two systems for a three methane cluster: (1) $m + m + m$, in which the three methane molecules are on the vertices of an equilateral triangle, with the side of the triangle being the reaction coordinate and (2) $2m + m$, in which the three molecules are on the vertices of an isosceles triangle, with the distance between molecules 1 and 2 fixed at 3.9 Å (the contact-minimum distance between two methane molecules in water) and the distance of the third methane molecule from the first and the second one, respectively, being the reaction coordinate (Fig. 6). The first system corresponds to uniform expansion or contraction of a hydrophobic cluster and the second system to the approach of an additional nonpolar molecule to an already formed hydrophobic-contact pair. It should be noted that the methane molecules were constrained exactly to these geometries in the TI calculations, while only restrained by a harmonic potential in the umbrella-sampling calculations.

Molecular dynamics simulations

Calculations were carried out by using a cubic periodic box with a 20 Å side, and a 9.5 Å cut-off for both the nonbonded and electrostatic interactions in most of the runs. To determine the long-distance behavior of the PMF, some of the calculations were carried out with a cubic box with a 30 Å side and an 8.5 Å cut-off. To check the influence of the cut-off on the accuracy of the electrostatic energy calculations, one series of simulations was carried out by using the Ewald summation.

Each methane molecule was modeled by a single interaction site, with VDW parameters taken from Jorgensen et al. (1984) (Table 1). The TIP3P and TIP4P models were used for the water molecules, with parameters (Table 1) taken from Jorgensen et al. (1983). These two water models were used to assess the dependence of the results on the representation of the water molecule. These are also the models most commonly used in macromolecular simulations. It is known that the TIP4P model produces a more ordered water structure than the TIP3P model (Levesque & Weis, 1995; Young & Brooks, 1997). The VDW parameters of the oxygen–methane interaction were calculated by using the standard Lorentz–Berthelot mixing rules (Allen & Tildesley, 1987), given by

$$\sigma_{ij} = (\sigma_{ii} + \sigma_{jj})/2 \quad (16)$$

$$\epsilon_{ij} = \sqrt{\epsilon_{ii} \epsilon_{jj}} \quad (17)$$

where i, j are two different type of atoms, and σ and ϵ corresponds to VDW parameters associated with the contact distance and the

Table 1. Force field parameters for TIP3P and TIP4P water models, and the methane model (Jorgensen et al., 1983, 1984) used in this study^a

Atom type	σ (Å)	ϵ (kcal/mol)	q (eu)
Methane			
C4	3.730	0.294	0.0
TIP3P			
HW	0.0	0.0	+0.417
OW	3.1507	0.1521	−0.834
TIP4P			
HW	0.0	0.0	+0.52
OW	3.1535	0.1550	0.0
M	0.0	0.0	−1.04

^aC4 represents the methane molecule, HW and OW represent hydrogen and oxygen atoms of the water molecule, respectively, and M represents a dummy atom in the TIP4P water model.

depth of the well, respectively. All VDW parameters and charges are summarized in Table 1. The 20 Å periodic box contained 265 and 264 water molecules in the simulations for two- and three-methane clusters, respectively. The simulations were carried out in the NVT ensemble with temperature 298 K. The time step in the MD simulations was 2 fs, and a modified SHAKE algorithm was applied to keep the water molecules rigid (van Gunsteren & Berendsen, 1977). The simulations were carried out by using both the AMBER (Pearlman et al., 1995) and the GROMACS (Berendsen et al., 1995) programs. Simulations with the TIP4P water were about four times faster using GROMACS, because of the more efficient way to achieve the rigidity of the four-point water model, compared to AMBER.

The TI calculations were carried out by using only the TIP3P water, with the GIBBS module of the AMBER package. The reaction coordinates used for all systems studied are shown in Figure 6. TI calculations were carried out with 80 windows for the 20 Å box, which corresponded to the change of the methane–methane distance from 3.1 to 9.1 Å. For each window, the equilibration period varied from 5 to 40 ps, and data from 15 to 40 ps were collected. To evaluate the effect of error accumulation, in the case of the methane dimer, integration was carried out twice: first, by starting from the lowest distance and increasing it (the forward integration), and second, by starting from the largest distance and decreasing it (the backward integration). Two backward integration runs (Fig. 2) were also carried out with the 30 Å box with the distance varying from 14.0 to 3.4 Å, using 200 windows. In the first of these, the equilibration period was 5 ps and in the second one it was 15 ps. The data were collected for 15 ps in both backward runs.

The MD umbrella-sampling simulations were carried out by using Equation 8. Calculations with the TIP3P water model were carried out with the program AMBER and those with the TIP4P water models were carried out with the program GROMACS. The simplest choice for the set of coupling parameters was made, for a simulation at the i^{th} window, viz., $\lambda_i = 1$ and $\lambda_j = 0$, for $j \neq i$. For two methane molecules, the umbrella-sampling simulations were carried out in 6 or 11 windows, respectively, with the following restrained potentials:

$$V(\xi) = k(\xi - d_0)^2 \quad (18)$$

with the force constant $k = 2$ kcal/mol/Å² and the “equilibrium” distance for a given window d_0 equal to 3.5, 4.5, ..., 8.5 Å for windows 1 to 6 or $d_0 = 3.5, 4.0, \dots, 8.5$ Å for windows 1 to 11. The reaction coordinate ξ was chosen as the methane–methane distance. In the case of the $m + m + m$ cluster, each distance was restrained by Equation 18 with the same force constant and the same values of d_0 as above. Additionally, the internal angles of the equilateral-triangle geometry were restrained to 60° with the force constant of 100 kcal/mol/rad².

In the case of the $2m + m$ system, one series of umbrella-sampling runs was carried out by imposing 2 kcal/mol/Å² restraints on the m_1 – m_3 and m_2 – m_3 distances with d_0 values varying as described above, and a 100 kcal/mol/Å² restraint (K) on the m_1 – m_2 distance with d_0 fixed at 3.9 Å (the methane–methane contact-minimum distance in water), as shown in Figure 9A. However, because the geometry was not explicitly restrained to that of an isosceles triangle, the sampling was not very efficient and we carried out some runs by applying additional restraints or constraints. First, we restrained the three angles α , β , and γ described in Figure 9B to 120° by a force constant of 100 kcal/mol/rad², keeping the restraints described above. In another series of runs, the distance between m_1 and m_2 was constrained exactly at 3.9 Å, by applying the SHAKE algorithm (Fig. 9C). In all the umbrella-sampling simulations, the Cartesian coordinates of the methane molecules were stored at every time step. In TI simulations there was no need to store coordinates, since the program GIBBS reports free energy directly.

Calculation of the PMF

The bin dimension applied in the WHAM calculations of the PMF was equal to 0.05 Å for the $m + m$ system (a 1D bin) and 0.1 Å for the $m + m + m$ and $2m + m$ systems (3D bins in all cases except for $2m + m$ with a frozen $m_1 - m_2$ distance where 2D bins were used). For the $m + m + m$ system, the histogram values of the bins whose center positions satisfied the condition $\xi = r_{12} = r_{13} = r_{23}$ were used for the calculation of PMF. For the $2m + m$ system, the conditions were $r_{12} = 3.9$ Å and $\xi = r_{13} = r_{23}$. It should be noted that, because of the finite bin dimensions, the PMFs are not exactly those defined by Equation 3, but each calculated value is an average over the bin volume.

The calculated PMFs can be identified with $W(r_{12})$ or $W(r_{12}, r_{13}, r_{23})$ of Equation 15 and should therefore tend to zero with increasing distance (after subtracting the constant factor accounting for the hydrophobic-hydration free energy of an *isolated* solute molecule). We noticed that, in the production runs (i.e., box size 20 Å, cut-off 9.5 Å, with the last window centered at 8.5 or 9 Å), the PMF calculated for distances beyond 8.5 Å exhibited some increasing oscillations. A test performed with an increased box size to 30 Å and a larger number of windows indicated that these oscillations disappeared and, therefore, they must be caused by the fact that distances beyond 8.5 Å are already close to the box’s limits for the last window in the previous simulations. Therefore, to calculate the cooperativity term $\delta F^{(3)}(r_{12}, r_{13}, r_{23})$, we made the assumption that this term vanishes faster than $W(r_{12})$ and $W(r_{12}, r_{13}, r_{23})$ with distance, i.e., $\delta F^{(3)}(r_{12}, r_{13}, r_{23})$ is assumed nearly zero for distances greater than 7.5 Å. Consequently, the region 7.5–8.5 Å (far region of the PMF but well behaving) was used to superimpose the PMFs of the trimer (normalized to show the free energy per unconstrained pairs of methanes) onto the PMFs of the dimer. To plot the PMFs, all curves were shifted so that the $W(r_{12})$ curves around 9 Å are zero (based on the results from the simulation with a box size of 30 Å, Fig. 5).

This shifting of all the PMFs is immaterial to the computation of the cooperativity term.

Acknowledgments

We thank J. Skolnick for helpful discussions that motivated this investigation. This work was supported by grants from the National Science Foundation (MCB95-13167), the National Institutes of Health (GM-14312 and R03 TW1064-1), the NIH National Center for Research Resources (P41RR-04293), and the Polish State Committee for Scientific Research, KBN (3 T09A 111 17). This research was conducted using the resources of (1) the Cornell Theory Center, which receives funding from Cornell University, New York State, the NIH National Center for Research Resources, and members of the Theory Center's Corporate Partnership Program; (2) the National Partnership for Advanced Computational Infrastructure at the San Diego Supercomputer Center supported in part by NSF cooperative agreement ACI-9619020; (3) the Informatics Center of the Metropolitan Academic Network (IC MAN) in Gdańsk, Poland; and (4) the Interdisciplinary Center for Mathematical Modeling (ICM) in Warsaw, Poland.

References

- Allen MP, Tildesley DG. 1987. *Computer simulation of liquids*, Chapter 1.4.2. New York: Oxford University Press. p 21.
- Ben-Naim A. 1980. *Hydrophobic interactions*. New York: Plenum Press.
- Berendsen HJC, van der Spoel D, van Drunen R. 1995. GROMACS: A message-passing parallel molecular dynamics implementation. *Comp Phys Commun* 91:43–56.
- Chipot C, Kollman PA, Pearlman DA. 1996. Alternative approaches to potential of mean force calculations: Free energy perturbation versus thermodynamic integration. Case study of some representative nonpolar interactions. *J Comput Chem* 17:1112–1131.
- Cramer CJ, Truhlar DG. 1999. Implicit solvation models: Equilibria, structure, spectra, and dynamics. *Chem Rev* 99:2161–2200.
- Darden T, York D, Pedersen L. 1993. Particle mesh Ewald: An $n \log(n)$ method for Ewald sums in large systems. *J Chem Phys* 98:10089–10092.
- Frenkel D, Smit B. 1996. *Understanding molecular simulation—From algorithms to applications*, Chapter 7. San Diego: Academic Press. pp 176–181.
- Geiger A, Rahman A, Stillinger FH. 1979. Molecular dynamics study of the hydration of Lennard-Jones solutes. *J Chem Phys* 70:263–276.
- Gibson KD, Scheraga HA. 1969. Minimization of polypeptide energy. V. Theoretical aspects. *Physiol Chem Phys* 1:109–126.
- Gö N, Scheraga HA. 1976. On the use of classical statistical mechanics in the treatment of polymer chain conformation. *Macromolecules* 9:535–542.
- Godzik A, Koliński A, Skolnick J. 1993. De novo and inverse folding predictions of protein structure and dynamics. *J Comput Aid Mol Des* 7:397–438.
- Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. 1983. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935.
- Jorgensen WL, Madura JD, Swenson CJ. 1984. Optimized intermolecular potential functions for liquid hydrocarbons. *J Am Chem Soc* 106:6638–6646.
- Jorgensen WL, Ravimohan C. 1985. Monte Carlo simulation of differences in free energies of hydration. *J Chem Phys* 83:3050–3054.
- Kauzmann W. 1959. Some factors in the interpretation of protein denaturation. *Adv Protein Chem* 14:1–63.
- Koradi R, Billeter M, Wüthrich K. 1996. MOLMOL: A program for display and analysis of macromolecular structures. *J Mol Graph* 14:51–55.
- Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. 1992. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem* 13:1011–1021.
- Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. 1995. Multi-dimensional free-energy calculations using the weighted histogram analysis method. *J Comput Chem* 16:1339–1350.
- Levesque D, Weis JJ. 1995. Recent progress in the simulation of classical fluids. In: Binder K, ed. *Topics in applied physics. The Monte Carlo method in condensed matter physics*, Volume 71, Chapter 6.6.6. Berlin: Springer. pp 172–177.
- Liwo A, Czaplewski C, Pillardy J, Lee J, Ripoll DR, Scheraga HA. 1999a. Origin and role of cooperative terms coming from local and long-range interactions in united-residue potentials. 5th World Congress of theoretically oriented chemists, WATOC'99. London: Imperial College.
- Liwo A, Kaźmierkiewicz R, Czaplewski C, Groth M, Oldziej S, Wawak RJ, Rackovsky S, Pincus MR, Scheraga HA. 1998. United-residue force field for off-lattice protein-structure simulations; III. Origin of backbone hydrogen-bonding cooperativity in united-residue potentials. *J Comput Chem* 19:259–276.
- Liwo A, Lee J, Ripoll DR, Pillardy J, Scheraga HA. 1999b. Protein structure prediction by global optimization of a potential energy function. *Proc Natl Acad Sci USA* 96:5482–5485.
- Liwo A, Oldziej S, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA. 1997. A united-residue force field for off-lattice protein-structure simulations. I. Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data. *J Comput Chem* 18:849–873.
- Liwo A, Pillardy J, Kaźmierkiewicz R, Wawak RJ, Groth M, Czaplewski C, Oldziej S, Scheraga HA. 1999c. Prediction of protein structure using a knowledge-based off-lattice united-residue force field and global optimization methods. *Theor Chem Accounts* 101:16–20.
- Lüdemann S, Abseher R, Schreiber H, Steinhäuser O. 1997. The temperature-dependence of hydrophobic association in water. Pair versus bulk hydrophobic interactions. *J Am Chem Soc* 119:4206–4213.
- Némethy G, Scheraga HA. 1962a. The structure of water and hydrophobic bonding in proteins. I. A model for the thermodynamic properties of liquid water. *J Chem Phys* 36:3382–3400.
- Némethy G, Scheraga HA. 1962b. The structure of water and hydrophobic bonding in proteins. II. A model for the thermodynamic properties of aqueous solutions of hydrocarbons. *J Chem Phys* 36:3401–3417.
- Némethy G, Scheraga HA. 1962c. The structure of water and hydrophobic bonding in proteins. III. The thermodynamic properties of hydrophobic bonds in proteins. *J Phys Chem* 66:1773–1789. [Erratum. *J Phys Chem* 67:2888.]
- No KT, Kim SG, Cho KH, Scheraga HA. 1999. Description of hydration free energy density as a function of molecular physical properties. *Biophys Chem* 78:127–145.
- Owicki JC, Scheraga HA. 1977. Monte Carlo calculations in the isothermal-isobaric ensemble. 2. Dilute aqueous solution of methane. *J Am Chem Soc* 99:7413–7418.
- Pearlman DA. 1993. Determining the contributions of constraints in free energy calculations: Development, characterization, and recommendations. *J Chem Phys* 98:8946–8957.
- Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE III, DeBolt S, Ferguson D, Seibel G, Kollman P. 1995. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comp Phys Commun* 91:1–41.
- Radmer RJ, Kollman PA. 1997. Free energy calculation methods: A theoretical and empirical comparison of numerical errors and a new method for qualitative estimates of free energy changes. *J Comput Chem* 18:902–919.
- Rank JA, Baker D. 1997. A desolvation barrier to hydrophobic cluster formation may contribute to the rate-limiting step in protein folding. *Protein Sci* 6:347–354.
- Rapaport DC, Scheraga HA. 1982. Hydration of inert solutes. A molecular dynamics study. *J Phys Chem* 86:873–880.
- Ravishanker G, Mezei M, Beveridge DL. 1982. Monte Carlo computer simulation study of the hydrophobic effect. Potential of mean force for $[(CH_4)_2]_{aq}$ at 25 and 50 °C. *Faraday Symp Chem Soc* 17:79–91.
- Richards FM. 1977. Areas, volumes, packing, and protein structure. *Ann Rev Biophys Bioeng* 6:151–176.
- Roux B, Simonson T. 1999. Implicit solvent models. *Biophys Chem* 78:1–20.
- San Biagio PL, Bulone D, Martorana V, Palma-Vittorelli MB, Palma MU. 1998. Physics and biophysics of solvent induced forces: Hydrophobic interactions and context-dependent hydration. *Eur Biophys J* 27:183–196.
- Scheraga HA. 1998. Theory of hydrophobic interactions. *J Biomol Struct Dyn* 16:447–460.
- Seetharamulu P, Crippen GM. 1991. A potential function for protein folding. *J Math Chem* 6:91–110.
- Swaminathan S, Beveridge DL. 1979. Monte Carlo computer simulation of hydrophobic bonding. *J Am Chem Soc* 101:5832–5833.
- Swaminathan S, Harrison SW, Beveridge DL. 1978. Monte Carlo studies on the structure of a dilute aqueous solution of methane. *J Am Chem Soc* 100:5705–5712.
- Tembe BL, McCammon JA. 1984. Ligand-receptor interactions. *Comput Chem* 8:281–283.
- Torrie GM, Valleau JP. 1977. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J Comp Phys* 23:187–199.
- Tsai J, Gerstein M, Levitt M. 1997. Simulating the minimum core for hydrophobic collapse in globular proteins. *Protein Sci* 6:2606–2616.
- Tuñón I, Silla E, Pascual-Ahuir JL. 1992. Molecular surface area and hydrophobic effect. *Protein Eng* 5:715–716.
- van Gunsteren WF, Berendsen HJC. 1977. Algorithms for macromolecular dynamics and constraint dynamics. *Mol Phys* 34:1311–1327.
- Young WS, Brooks CL III. 1997. A reexamination of the hydrophobic effect: Exploring the role of the solvent model in computing the methane-methane potential of mean force. *J Chem Phys* 106:9265–9269.