Potentials of mean force for biomolecular simulations: Theory and test on alanine dipeptide

Matteo Pellegrini
Department of Physics, Stanford University, Stanford, California 94305-4090

Niels Grønbech-Jensen
Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Sebastian Doniach
Departments of Applied Physics and Physics, Stanford University, Stanford, California 94305-4090

(Received 2 October 1995; accepted 28 February 1996)

We describe a technique for generating potentials of mean force (PMF) between solutes in an aqueous solution. We first generate solute–solvent correlation functions (CF) using Monte Carlo (MC) simulations in which we place a single atom solute in a periodic boundary box containing a few hundred water molecules. We then make use of the Kirkwood superposition approximation, where the 3-body correlation function is approximated as the product of 2-body CFs, to describe the mean water density around two solutes. Computing the force generated on the solutes by this average water density allows us to compute potentials of mean force between the two solutes. For charged solutes an additional approximation involving dielectric screening is made, by setting the dielectric constant of water to $\varepsilon = 80$. These potentials account, in an approximate manner, for the average effect of water on the atoms. Following the work of Pettitt and Karplus [Chem. Phys. Lett. **121**, 194 (1985)], we approximate the *n*-body potential of mean force as a sum of the pairwise potentials of mean force. This allows us to run simulations of biomolecules without introducing explicit water, hence gaining several orders of magnitude in efficiency with respect to standard molecular dynamics techniques. We demonstrate the validity of this technique by first comparing the PMFs for methane–methane and sodium–chloride generated with this procedure, with those calculated with a standard Monte Carlo simulation with explicit water. We then compare the results of the free energy profiles between the equilibria of alanine dipeptide generated by the two methods.
 [S0021-9606(96)50321-1]

I. INTRODUCTION

The importance of including explicit water molecules in biomolecular simulations is by now well established.[1,2] Water is essential in creating what is known as the hydrophobic effect: the tendency of hydrophobic groups to cluster together thus minimizing their solvent accessible surface.[3–5] Furthermore water molecules are often involved in the bridging of hydrogen bonds between the solute's hydrogen bond donors and acceptors.[6] Finally the dielectric properties of water dramatically reduce the strength of intra and intermolecular electrostatic interactions.[7–10]

However to fully solvate a biomolecule in a simulation with periodic boundary conditions often requires including many more water molecules than solute atoms. Hence the simulation tends to run much more slowly than a vacuum simulation of the same molecule. It would therefore be very useful, in order to exhaustively sample the conformational space of biomolecules, to develop a technique that implicitly accounts for the effect of water, without the inclusion of explicit water molecules in the simulation. In order to be successful, such an approach should be capable of reproducing the effects of solvation that we have outlined above.

Several techniques have already been developed to allow the exclusion of explicit water molecules in simulations. One of the simplest methods to account for the effect of water on intra-solute electrostatic interactions is to introduce screening functions. A common technique that accounts for the effect of water polarization on electrostatic interactions is to include a distance dependent dielectric constant.[11] Alternately, one could simply set the dielectric constant to 80, the value for bulk water.

Similar ad hoc methods have also been devised to account for the effect of water on van der Waals interactions. When two solute atoms move apart, the intramolecular van der Waals contacts are replaced with solute–solvent contacts. The repulsive component of the solute-solute interaction is unaffected by the presence of water. To account for this behavior, one technique used is to truncate the van der Waals potential at its minimum energy position, and shift it up to zero.[12] Thus one is in effect computing the energy with respect to the solvated state. In a similar approach, sigmoidal potentials have been used to account for the effect of solvation on van der Waals interactions.[13]

Another technique that is widely used to approximate the effects of an aqueous solvent on a solute, is to model the solvation energy as a term proportional to the solvent exposed surface area of the solute. In one such formulation,[14] each atom is assigned a solvation parameter so that the total solvation energy is simply the sum of the products between the atomic solvation parameters and the exposed surface area of the atom in question:

$$\Delta G = \sum_{\text{atom } i} \Delta\sigma_i A_i . \tag{1}$$

A similar technique is Scheraga's hydration shell model,[15] in which each atom in a solute interacts with the volume of water formed by its surrounding shell minus the excluded volume of other solute atoms. These techniques have been shown to adequately reproduce the solvation energy of various conformers of small biomolecules.[15,16] However, they do not account for the screening effects of water on intra-solute electrostatic interactions, which must be treated separately when using these approaches.

A more accurate approach to the dielectric screening problem is to treat the water as a continuum with a dielectric of 80, in which is embedded a solute. In this approach, the van der Waals forces are included in the cavitation energy: the free energy required to form a cavity, with the shape of the solute, in water. Once the cavity is formed, it is possible to solve the Poisson–Boltzmann equation to obtain the charge buildup at the water–solute interface.[17] Once the charge distribution is known, one can deduce the contribution of the electrostatic forces to the free energy of solvation, as well as the screening effects of these charges on intramolecular electrostatic interactions. Another way to estimate the screening effect of water on these interactions is to use a technique developed by Warshel, in which the water dipoles are fixed on lattice points around the solute, and their orientation is self-consistently solved for.[18]

Yet another set of techniques used to approximate solvation, are those that utilize analytic equations to solve for the potential of mean force (PMF) between two solutes. These analytic equations, referred to as RISM type equations,[19] express the 2-body correlation functions (CF) as an integral over 3-body CFs. To solve them one may resort to an approximation, such as that provided by the hypernetted chain equation, to express 3-body CFs in terms of 2-body ones. The interatomic potentials of mean force can then be summed over all the solute atoms to obtain an approximate value of the solute free energy.[20]

Given the vast literature on approximate solutions to the solvation of biomolecules, is there a need for another technique? The first approach outlined above, that empirically screens electrostatic and van der Waals interactions, provides only a qualitative approximation to the effects of water on a solute. It is not possible to quantitatively estimate solvation free energies or interatomic potentials of mean force using such an approximation. The effects of the solvent are far more accurately determined by models that relate the solvation energy to the solvent exposed surface of the solute, although these methods require a further assumption to model the screening effects of water on intra-solute electrostatic interactions. The solution of the Poisson–Boltzmann equation has also been shown to satisfactorily reproduce the solvation energies of small biomolecules. However, these last two approaches are computationally expensive. Both the calculation of surface area, using the algorithm of Lee and Richards,[21] and the solution of the Poisson–Boltzmann equation, using an iterative procedure, require substantially more calculations than merely calculating pairwise interactions. We therefore feel that to be able to explore the conformational landscape of even small biomolecules, these approaches present limited capabilities.

The approach we propose in this work is similar to that described above using RISM equations, where the effect of water is incorporated in pairwise potentials of mean force. The difference lies in the method used for computing the PMFs, which we will outline below. There are two fundamental approximations that one must make when using this approach. The first approximations have to do with the procedure that one uses to calculate the PMFs without having to run a full simulation with explicit water. The second, and probably more important, is related to the fact that we only use pairwise interactions, and not higher order ones (3-body, 4-body etc.), to describe the n-body potential of mean force of the solute, where n is the number of solute atoms. The 2-body potentials treat any pair of solutes as being completely surrounded by water. In a molecule with more than two atoms this will not be the case, only a fraction of the surface area of any pair of atoms will be solvent accessible. Higher order terms in the expansion account for this partial solvation. In this paper we will show that for a small solute, the alanine dipeptide, we obtain reasonable free energies by only considering the 2-body terms. We conclude, therefore, that for small molecules, where a large portion of the surface area is exposed to the solvent, both these approximations yield reasonable results.

Why don't we use the solution of RISM equations to compute intra-solute PMFs? One reason is that when compared to the results of molecular dynamics simulations, the solution to RISM equations may not be very accurate. For instance, it has been shown that the maximum of the correlation function between sodium and chlorine, in an aqueous solution, is off by 50% in a RISM calculation with respect to the molecular dynamics simulation.[22] In contrast to the RISM approach, we compute the solute–solvent correlation functions explicitly from full atom simulations. We will therefore attempt to demonstrate that the procedure described in this paper is both conceptually simpler and more accurate than the RISM approach, especially when applied to small biomolecules such as the alanine dipeptide.

The techniques presented in this work emerge from our previous approach to compute the average water density around biomolecules. We have recently shown[23] (following the work of A. García, G. Hummer and D. Soumpasis on the hydration of biomolecules[24,25]), that it is possible to compute the mean density of water about a solute by using a potential of mean force expansion. We have further demonstrated that this water density correctly reproduces the free energy landscape of various solvated molecules. In this work we use the Kirkwood superposition approximation, to express the mean density of water at position r, with respect to solute atoms $r_1 \ldots . r_n$, as a product of the 2-body correlation functions,

$$\rho^{(r|r_1 \ldots . r_n)} = \rho_0 \prod_{i=1}^{n} g^{(2)}(r, r_i). \tag{2}$$

Unlike our previous work, we neglect the contributions to the density from higher order (3-body, 4-body etc.) correlation functions.

This level of accuracy is sufficient for estimating the mean van der Waals force that the water exerts on the solutes, since these forces are short ranged, decaying as r^{-6} in the description used here. However it is not sufficiently accurate to describe the mean coulombic force that the water exerts on the solutes. To model this effect we introduce a dielectric constant of 80 in the Coulomb term of our potential.

The procedure we adopt to compute PMFs involves the following cycle: we first estimate the free energy to discharge our solutes, we then compute the potential of mean force required to move the uncharged solutes apart, and finally we estimate the free energy to recharge our solutes.

To compute the PMF for uncharged solutes we first generate the solute–solvent 2-body correlation functions between the atoms of interest and water oxygens. This is accomplished using a standard Monte Carlo simulation with explicit water molecules. It is a relatively efficient procedure requiring only a few hours per correlation function. Armed with the CFs we are able to approximate the water density around the solutes as a function of their separation, as shown above. From the average water density we are able to compute the mean force on the two solutes, and hence the potential of mean force.

Using the approximation that the water acts as a homogeneous dielectric with dielectric constant 80, we easily calculate the free energy to discharge and recharge our solutes. Combining these two terms, we then generate symmetric, pairwise potentials of mean force for all the atom species of our solute. Following the work of Pettitt and Karplus,[1] we express the n-body potential of mean force for our solute as a sum of these pairwise potentials:

$$W^{(n)}(\mathbf{r}) = \sum_{ij} W_{ij}^{(2)}(r_i, r_j). \tag{3}$$

To demonstrate the usefulness of this technique, we show that it successfully reproduces the PMF between methanes and sodium-chloride as ions as well as the free energy profiles generated by rotating the alanine dipeptide about its dihedral angles.

II. THEORY

A. Correlation functions and Kirkwood approximation

For a single atom solute in an infinitely dilute solution we can describe the average solvent density around the solute in terms of the 2-body correlation function (CF), $g^{(2)}(r)$. This is a spherically symmetric function that depends only on the radial distance between the solute and the solvent. In the case of an aqueous solution, one would have to compute two separate CFs for the water oxygens and hydrogens, $g_{\mathrm{O}}^{(2)}(r)$ and $g_{\mathrm{H}}^{(2)}(r)$.

In Fig. 1 we show a typical CF for water around a Lennard-Jones particle. We see that both the oxygen and hydrogen CFs have a first hydration shell at 4 Å and a second one at 7 Å from the solute. Beyond that the density of water rapidly approaches its bulk density, which implies a value of the CF of 1. Since the solute in this case is not charged the hydrogens and oxygens of water do not preferentially align themselves so that both the peaks lie at the same distance.

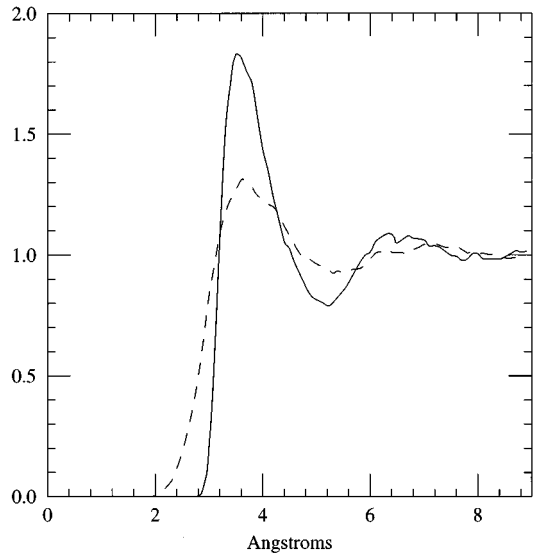

FIG. 1. 2-point correlation function between a united-atom methane and water oxygens (solid) and hydrogens (dashed) obtained from MC simulations.

If instead of a single solute we have two solute atoms, we may wish to describe the water density around the solutes when they are separated by a distance, r_0. In this case we would need a 3-body correlation function, $g^{(3)}(r_0, r_1, r_2)$, where r_1 and r_2 are the distances between a solvent molecule and solutes 1 and 2, respectively. In principle one can compute these functions numerically by running a Monte Carlo (MC) or molecular dynamics (MD) simulations in which one places the two solutes in a boundary box with water molecules.[26] However such a simulation is computationally intensive, requiring approximately a day on a typical workstation. Since for our purposes we need to calculate hundreds of these functions, we must use an approximate technique rather than the full simulation.

The simplest method for approximating 3-point CFs is known as the Kirkwood approximation:[27]

$$g^{(3)}(r_0, r_1, r_2) = g^{(2)}(|r_0 - r_1|) g^{(2)}(|r_0 - r_2). \tag{4}$$

In other words, we treat the 3-body CF as a product of the 2-body CFs. This is not a very accurate approximation, and as a consequence various methods have been designed to attempt to improve it, although with only limited success.[28–30] In particular, it tends to exaggerate the density of the solvent in the areas where the two solutes have overlapping peaks of their CFs. However, as we have shown in previous work,[23] in the case of uncharged Lennard-Jones particles we are still able to correctly calculate PMFs. This is probably due to the fact that van der Waals interactions are short ranged, decaying as r^{-6}. Therefore, even though the

Kirkwood approximation does not work globally, the regions of water density that interact strongly with the solutes are sufficiently well described.

Coulomb interactions, however, decay as r^{-1} and are therefore much more sensitive to errors made by the Kirkwood approximation. Therefore our procedure will not work well for charged particles. Motivated by the need for a computationally efficient algorithm, we propose an approximate method to remedy this in the next sections, by presenting a procedure for calculating the PMFs of charged atoms. Before this, in the next section, we will elaborate on the method by which the Kirkwood approximation may be used to obtain potentials of mean force between two Lennard-Jones particles.

B. Inter-solute mean force

To calculate the free energy change to move two solutes separated by a distance r, to $r+\Delta r$, we compute the mean force on the solutes due to the average water density. The change in free energy may be written as

$$\Delta A = A(r+\Delta r) - A(r) = -\beta^{-1}\langle F_{\mathrm{ax}}(r)\rangle_r \Delta r. \tag{5}$$

Here $\langle\ \rangle$ defines an equilibrium average over all solvent degrees of freedom with the distance between the solutes constrained. F_{ax} is the component of the inter-solute force along the axis connecting the two solutes. All other components of the force average to zero.

To apply this formula we calculate the water shell around the solutes using the above Kirkwood approximation. Since the water density is cylindrically symmetric, we need only calculate it on a plane that intersects the axis of symmetry (defined by the two solutes). On this plane we construct a cubic grid with 0.1 Å spacings extending 9.0 Å from the solute and compute the density of water oxygens and hydrogens at each grid point.

The water density computed in this fashion represents an average water distribution. Therefore we automatically compute the total average force, $\langle F_{\mathrm{ax}}\rangle$, by computing the force between each solute and the mean water distribution and between the solutes.

We can therefore immediately calculate the free energy change in two simple steps: (a) first we compute the density of water oxygens and hydrogens on the grid mentioned above, when the solutes are at a distance r apart, then (b) we sum the total force between the solutes and the water density

$$\langle F_{\mathrm{ax}}(r_{\mathrm{solutes}})\rangle = \left[\sum_{n=1}^{2}\left(2\pi\rho_i \sum_{\rho_i,z_i} F_{ax}(\mathbf{r}_{\mathrm{water}} - \mathbf{r}_{\mathrm{solute}}) D(\rho_i,z_i)\Delta\Omega\right)\right] + F_{\mathrm{ax}}(r_{\mathrm{solutes}}), \tag{6}$$

where the first is the sum over the solutes and the second is the sum over the cylindrical grid coordinates, F_{ax} is the force in the z direction between the solute and the water or the two solutes, D is the density of water oxygens and $\Delta\Omega$ is the area of our grid spacings.

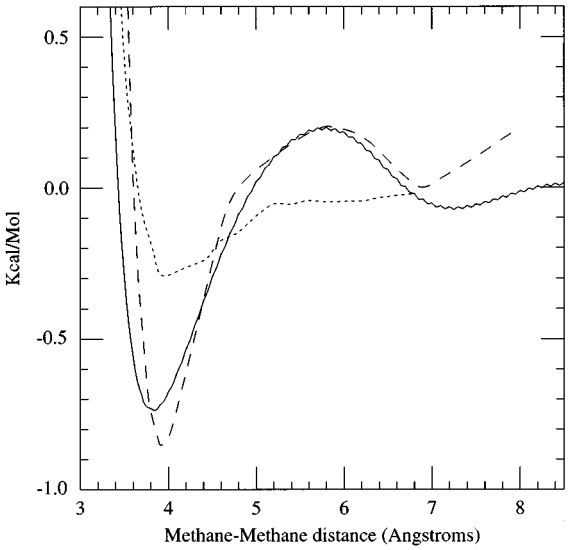

FIG. 2. Potential of mean force as a function of distance between two solvated methanes. The dashed line is from the MD simulation of D. van Belle. The solid line is computed with the methods of this work, with slightly higher accuracy than the same curve of Ref. 23. The dotted line is the potential calculated by the method of Eisenberg and McLachlan where the solvation energy is proportional to the solvent accessible surface.

By integrating the distance between two solutes, in increments of Δr, we can then generate the solute-solute potential of mean force:

$$W^{(2)}(r) = -\int_{r_0}^{r}\langle F_{\mathrm{ax}}(r)\rangle dr + \text{constant}, \tag{7}$$

where the constant is chosen so that the potential goes to zero at infinity. As an example of this technique we show the PMF between two methanes in Fig. 2. In our previous work,[23] we had computed the same potential without taking advantage of the cylindrical symmetry of the problem. Therefore, the potential presented here is in principle more accurate. The details of this calculation will be explained in the next section.

III. METHODS

A. Calculation of correlation functions

To compute the correlation functions (CF) we placed our solute (e.g., united atom methane) in a 18.7 Å box with periodic boundary conditions. We then solvated it with 217 SPC[31] water molecules, representing a density of 1 gm/cm^3.

The potential functions and parameters we used were those developed by Jorgensen and Tirado-Rives, since these are optimized for liquid simulations (hence known as OPLS potential functions).[32,33] The form of the potential function is

$$V(r_1,\ldots.r_n) = \sum_{i,j,i>j}\left(\frac{q_i q_j e^2}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}}\right). \tag{8}$$

TABLE I. OPLS parameters used in simulations.

Atom type	Charge qe	ε	σ
CH_4	0.0	3.73	0.294
C_α	0.2	3.8	0.08
C(O)	0.5	3.75	0.105
O(C)	−0.5	2.96	0.21
N(H)	−0.57	3.25	0.17
H(N)[a]	0.37	1.0	0.01
C_β	0.0	3.91	0.16
$(NH)CH_3$	0.2	3.8	0.17
Na	1.0	1.8974	1.6071
Cl	−1.0	4.4172	0.1178

[a]van der Waals parameters taken from the Levitt parameter set.

In all cases the CH_n groups are treated as united atoms, so that only one position is used to describe these groups (see Table I).

As suggested by Levitt,[34] the Coulomb component of the potential is truncated by multiplying it by the function $S(r)=(1-(r/r_c))^2$, with $r_c=8.0$ Å. This function eliminates spurious forces at the cutoff distance since its derivative also goes to zero there. We have verified that water–water CFs calculated using $S(r)$ are not significantly different than those computed using an unscreened Coulomb potential out to 10 Å, which is set to zero using a switching function[10] beyond that, and that furthermore they reproduce the experimental (obtained from neutron diffraction) water–water CFs well.[35,36]

To select moves (translations and rotations) of molecules we use the standard Monte Carlo Metropolis algorithm.[37] The moves were selected from a flat distribution with ranges ± 0.15 Å for translations and $\pm 15°$ for rotations. At the chosen simulation temperature (100 °C) these moves yield an acceptance rate of approximately 40%.

B. Potentials of mean force

To compute the potentials of mean force we use the Kirkwood approximation to estimate the density around the two solute atoms. We then apply the free energy perturbation equation, as illustrated above, to estimate the free energy change to move the solutes apart.

However, as we have explained above, this procedure works reliably only for uncharged Lennard-Jones particles. Therefore, to generate PMFs for charged atoms we introduce a further approximation, the essence of which is summarized in the thermodynamic cycle of Fig. 3. We begin with the two charged solutes, separated by a distance r, in water. We imagine removing the charge from these solutes, while keeping their distance fixed. We then move them apart by a distance Δr, and finally recharge them to their original values. The free energy to accomplish this cycle is identical to the free energy necessary to move the two charged particles apart by a distance Δr.

We are able to reliably calculate the free energy entailed in moving the two uncharged solutes apart. Therefore the approximation lies in estimating the free energy necessary to discharge and recharge the solutes. The approximation we make along these two portions of the free energy cycle is simply to treat the water as a medium with a dielectric constant of 80. This is, as we stated in the introduction, one of the simplest approximations one can make to account for the screening effect of water on electrostatic interactions. Since we neglect solute–solvent van der Waals interactions when charging and discharging the solutes, the only contribution of these two paths to the free energy cycle is accounted for in the change in solute–solute electrostatic interactions between r and $r+\Delta r$, when screened with a dielectric constant 80.

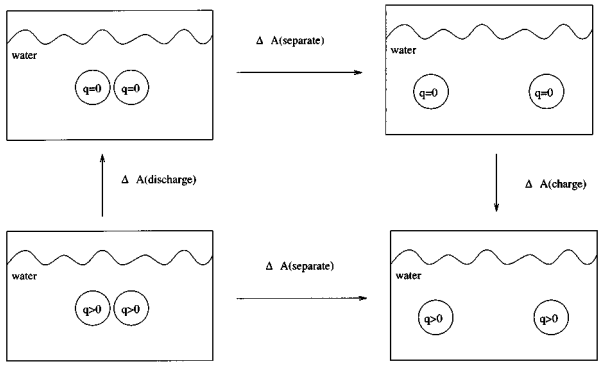

FIG. 3. A schematic of the thermodynamic cycle we use to compute the potentials of mean force. We begin at the bottom left with our two solutes in their charged state in water. We then discharge them (top left), move them apart Δr (top right) and then recharge them (bottom right). This sequence is equivalent, in free energy terms, to separating two charged solutes (bottom left to bottom right).

One might expect that the extent of screening of electrostatic interactions by the rearrangement of water dipoles around the two solutes would depend on the distance between the solutes. For instance, the dielectric constant of water has been modeled as a sigmoidal function that goes from one to 80 in about 10 Å.[38] However, as we will show in the following sections, we find that the correct potentials of mean force are more faithfully reproduced by setting the dielectric to 80 for all distances between the solutes. In particular, distance dependent approximations of the dielectric tend to overestimate the attraction between oppositely charged ions, by underestimating the extent of screening.

To summarize, we first compute the correlation functions for all the atoms of interest while setting their charge to zero. Therefore these particles are only distinguished by their different van der Waals parameters. We then use the procedure outline in Section II to compute the free energy necessary to move them apart. We then add this component of the free energy to that obtained by moving the charged atoms apart, with their electrostatic interactions screened by a dielectric constant of 80. In the following sections we will demonstrate that this technique for approximating PMFs, yields reasonable results when applied to sodium and chlorine ion pairs and the alanine dipeptide.

C. Simulations with explicit water

To determine the accuracy of the potentials of mean force generated by the above method, we will compare the

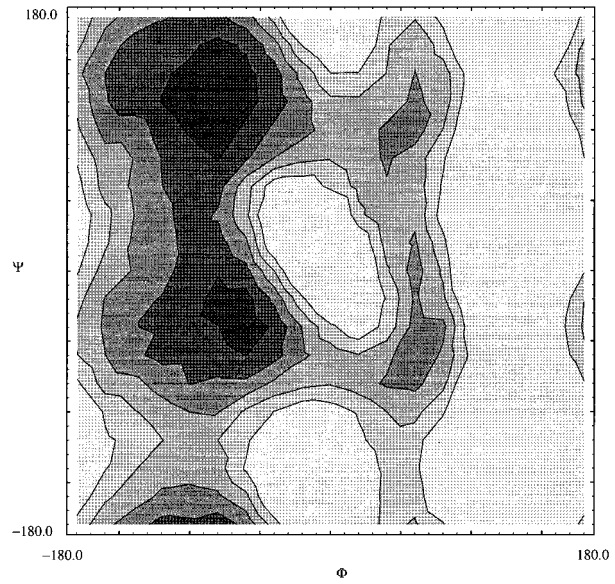

FIG. 4. A contour plot of the ϕ,ψ space of the alanine dipeptide computed using the potentials of mean force described in this work. At 400 points in the ϕ,ψ plane we constrained the dihedral potentials using harmonic constraints and computed the average energy, computed using the pairwise PMFs for non-bonded interactions and the potentials from the original CHARMM Paper (Ref. 10) for the bonded ones, during a 100 000 step simulation. No explicit water molecules were included in the simulation. See Ref. 44 for comparison.

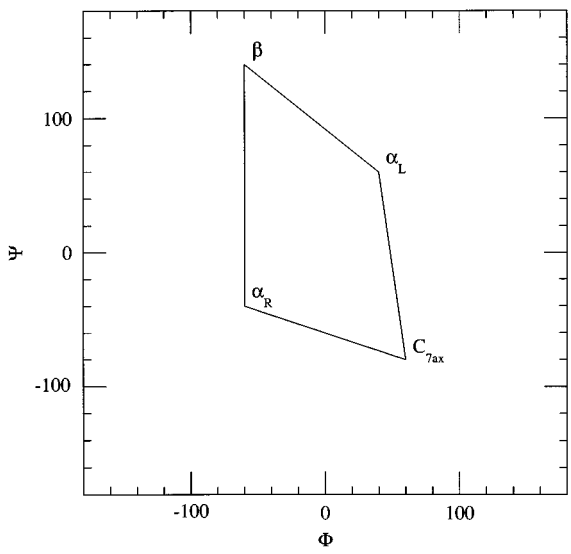

FIG. 5. The paths in the ϕ,ψ space of the alanine dipeptide along which we compute the free energy profiles. We also list the names of the conformations that correspond to the four minima at the end points of the paths.

free energy profile of alanine dipeptide computed with PMFs and with explicit water. Here we will outline the details concerning the simulation with explicit water molecules.

To begin with we must first determine the paths in the dihedral angle space of alanine dipeptide along which to integrate. For simplicity, the four minima in ϕ,ψ space were determined using the PMFs, since this is much faster than determining the real minima of the full simulation with water, and for our purposes of comparison equally good. The ϕ,ψ energy landscape is illustrated in the contour plot of Fig. 4. To obtain this plot we first constrained the dihedral angle using harmonic potentials at 20° intervals and then for each set of constraints averaged the energy, computed using PMFs for non-bonded and CHARMM for bonded interactions, for 100 000 steps. In this paper, CHARMM parameters are exclusively obtained from the tables of the original CHARMM paper of 1983[10] and not from the newer versions. The paths are illustrated in Fig. 5.

The four paths were then divided into 10 segments each, giving us 40 points in ϕ,ψ along which to compute the change in free energy, ΔA. In each interval, the dihedral angles of the alanine dipeptide were constrained with strong harmonic potentials. The molecule was placed in a 19 Å box, with periodic boundary conditions, and 210 water SPC water molecules.

We then conducted a Monte Carlo simulation at the chosen temperature of 100 °C to improve the sampling over a room temperature simulation. The potential we used was that listed in the tables of the original CHARMM paper,[10] for the bonded interactions, and the OPLS parameters for the non-bonded ones.[32] Both parameter sets treat carbons with covalently bonded hydrogens as united atoms, represented by a single position. We did not constrain any atomic degrees of freedom. We first equilibrated the system for one million steps, and then computed the average free energy, $-\beta^{-1}\ln\langle\exp[\Delta U(\phi,\psi)]\rangle$, for an additional four million steps. The calculation was performed on 40 SGI 4400 machines, and required approximately two days per machine.

IV. RESULTS

A. Methane–methane potential of mean force

The potential of mean force (PMF) between two methanes has been studied extensively because it is one of the simplest potentials that describes the hydrophobic effect.[26] As expected from previous work[39] the water becomes more ordered around a hydrophobic solute, thus decreasing its entropy. This effect creates an attraction between two methanes, since reducing their surface area in contact with water minimizes the loss of entropy. Previous studies had also found that a second minimum in the PMF occurs when the two methanes are separated by about 7 Å, allowing a single water molecule to lie between them.

Using the OPLS parameters for a united-atom methane (see Table I) we first calculate the CFs between methane and water oxygens or hydrogens (see Fig. 1). We then determine the PMF by the procedure outlined above.

As was found in Ref. 23, using a slightly less accurate technique, we again find remarkable agreement between our PMF and that computed by MD studies,[26] as can be seen in Fig. 2. The position of the minima, at 4 Å and at 7 Å, are equal to within 0.1 Å in the two curves. The heights of the first barrier are essentially equal. The two curves only begin to show some discrepancy in the second minimum, the water mediated minimum.

In Fig. 2 we have also included a PMF for methane–methane calculated by the technique mentioned above where the solvation energy is taken to be proportional to the solvent accessible area. From the work of Eisenberg and McLachlan,[14] we have taken the atomic solvation parameter of methane to be 16 cal/mol Å^2. We used the program MidasPlus[40] to compute the solvent accessible area using the algorithm of Lee and Richards,[21] with a probe radius of 1.4 Å. The solvation contribution to the free energy was then added to the bare methane–methane OPLS potential. As is clear from Fig. 2, this technique is unable to correctly calculate the PMF. In particular, the minimum in the potential is only 30% of the value computed by MD.

Thus, we conclude that in this very simple case our methodology is effective in reproducing the results of an extensive MD simulation. It should also be noted that to construct our PMF requires only a few seconds, while the evaluation of PMFs using MD requires approximately three orders of magnitude longer.

B. Sodium-chloride potential of mean force

Although we have demonstrated that the PMFs we generate for uncharged particles are accurate, the ones for charged particles are expected to deviate more from the true potentials. This is due to the fact that in the thermodynamic cycle described above, we approximate the free energy contribution from discharging two solutes separated by r and recharging them at $r+\Delta r$, to be the energy required to separate two solutes whose electrostatic interactions are screened by a dielectric of 80. To investigate the magnitude of this error we compare the PMFs for sodium chloride, calculated by our method and by an explicit water simulation.

The explicit water simulation we used to calculate the PMF for Na, Cl is very similar to that described above for the alanine dipeptide. The parameters for sodium and chlorine are listed in Table I. We place the Na, Cl atoms in a periodic boundary box with 217 water molecules and run the simulation at a temperature of 100 °C. At each distance we first equilibrate the system for one million Monte Carlo moves and then compute $-\beta^{-1}\ln\langle\exp[U(r+\Delta r)-U(r)]\rangle$, for another million steps. We repeat this 25 time with increments of 0.2 Å as we separate the two atoms.

The results of the explicit water and PMF calculations are shown in Fig. 6. The first minimum is fairly well reproduced, within 0.1 kcal/mol, by our PMF method, however it is much wider than the minimum computed with explicit water. The first barrier is also lower, and the second minimum is shifted by 1 Å. We have also included the potential of mean force calculated with a distance dependent dielectric. This approximation severely exaggerates the attraction between the two ions, predicting a minimum of -18.0 kcal/mol. Only the portion of the curve above -2.0 kcal/mol appears in the figure.

We conclude that even though the above mentioned approximation reproduces the value of the first minimum in the PMF accurately, it nonetheless introduces significant error in the second minimum and the barrier between the two minima. However, this error is probably maximized in the PMFs between two ions, such as Na, Cl, since it is due to the effects of charging and discharging the atoms. In other atoms, that are modeled as having only a fractional charge, the approximation will be less severe. Since in biomolecules most of the atoms are assigned only a partial charge, we feel the approximation should be sufficient to yield reasonable results, a fact that we will verify in the next section.

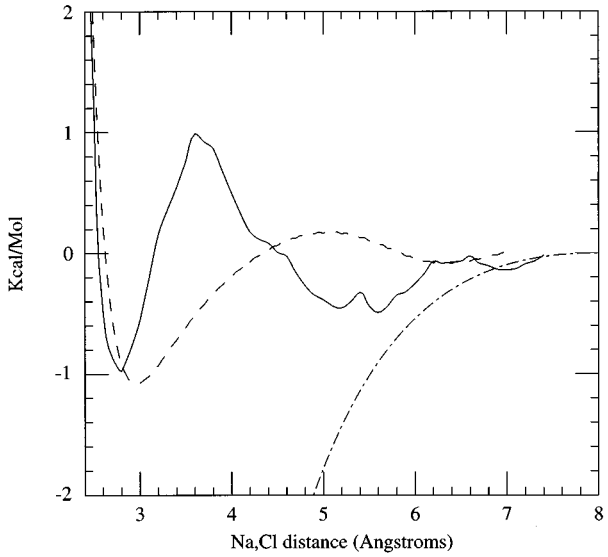

FIG. 6. Potential of mean force as a function of distance between Na and Cl in water. The dashed line is from the MD simulation of D. van Belle. The solid line is from this work. We also include a portion of the curve (dash dot) representing the potential computed with a dielectric constant proportional to the distance ($\varepsilon=r$).

C. Free energy profiles of alanine dipeptide

The effect of water on the free energy landscape of the alanine dipeptide has been extensively studied theoretically.[41–44,16,1] Rather than compare our PMFs for many solute pairs to those generated by simulations with explicit water, we decided to test whether they reproduced the free energy profiles derived from rotating the alanine dipeptide about its dihedral angles. In order to do this we first generated all the necessary PMFs between the atom types that compose the alanine dipeptide, according to the OPLS parametrization. There are seven atom types in all, requiring the calculation of 21 PMFs. The details of this calculation are the same as those of the computations outlined above.

The OPLS parameters we used for the atoms are listed in Table I. Instead of the OPLS parameters for the amide hydrogen, where the van der Waals parameters are zero, we had used the parameters from Levitt *et al.*[34] where this is not the case.

Once we have tabulated all possible PMFs for the alanine dipeptide, we make the approximation that the total potential of mean force for the molecule may simply be expressed as a sum of these 2-body PMFs [see Eq. (3)]. This is the lowest order approximation for such an expansion. To

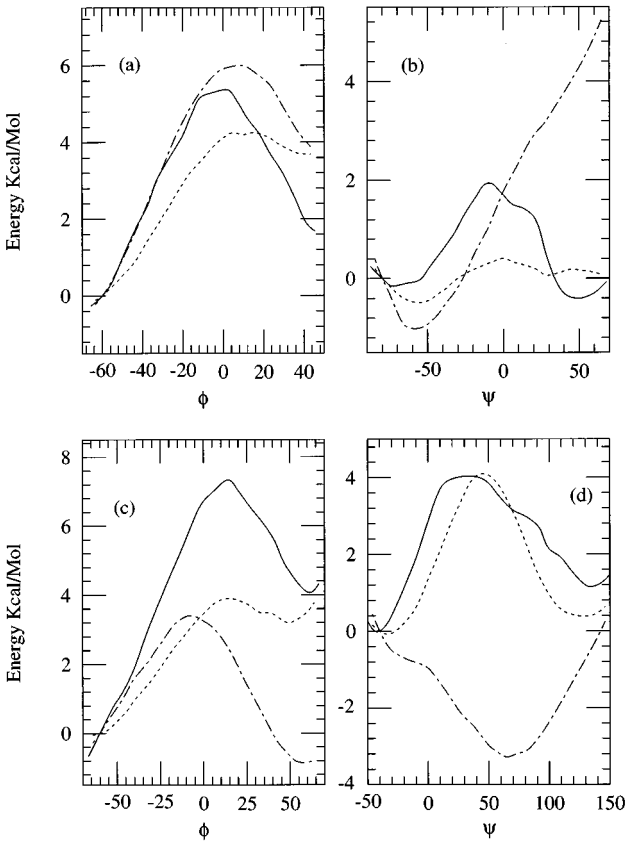

FIG. 7. The four free energy profiles corresponding to the paths of Fig. 4. The solid lines represent the Monte Carlo simulation with explicit water, the dashed lines the simulation with the potentials of mean force and the dash–dot lines correspond to simulations in vacuum (OPLS energy with $\varepsilon=1$). The figures correspond to the following paths: (a) β to α_L, (b) $C_{7\mathrm{ax}}$ to α_L, (c) α_R to $C_{7\mathrm{ax}}$ and (d) α_R to β.

improve the approximation we would add 3-body potentials to the expansion. However, we will now show that even to this order, the n-body PMF is reasonably approximated by the above equation. However, we only expect this agreement for relatively small solutes where the solvent exposed surface area is a large component of the total surface area. We would not, for instance, expect this approach to work for large peptides or proteins, where many of the atoms in the core of the molecule have no solvent exposed surface area.

We computed the free energy profiles by dissecting the paths of Fig. 5 into 15 intervals. At each point we ran a room temperature Monte Carlo simulation for 200 000 steps, and computed $-\beta^{-1}\ln\langle\exp[\Delta U(\phi,\psi)]\rangle$, to obtain the change in free energy. As explained above, we use the parameters from the original CHARMM paper[10] for the bonded interactions and OPLS parameters for the non-bonded ones.[32] Therefore we use the same parameters to describe all interactions in both the full water and mean field simulations, allowing us to make accurate comparisons of the results.

We compare the free energy profiles obtained with the PMFs and explicit water simulations in Fig. 7. We also show the free energy profiles for a simulation performed in vacuum, to show the change in the free energy brought about by including water in the simulation. To perform the vacuum simulation, we use the OPLS parameters, with the dielectric set to 1, for non-bonded interactions, and the parameters from the original CHARMM paper[10] for the bonded ones.

For all four paths the PMFs produce qualitatively correct results, appropriately lowering and raising the energy of the four minima with respect to the vacuum calculations. In all cases, the PMF minima agree with the explicit water simulation within 2.0 kcal/mol. The barrier heights between minima, however, are considerably reduced in the PMF calculation (except in the α_R to β path). Therefore, we conclude that the potentials are adequate for computing relative energies, but inadequate for reproducing the energies of the transition states between them.

In many applications of biomolecular simulations, the goal is to obtain the density distribution of the conformers of the molecule. Since the occupancy of a state is proportional to the Boltzmann factor, only the energetically low lying states are occupied with a high probability. The probability of occupancy of the minimum energy states is fairly insensitive to the energy of the transition states, since these are very sparsely occupied. Therefore, we feel that even though our PMFs inadequately reproduce the heights of the barriers between minima, they should nonetheless reliably reproduce the equilibrium density of states of the solute.

One of the principal advantages of PMF calculations is that they are nearly three orders of magnitude faster than the simulations with explicit water molecules. Furthermore, none of the approximations that we used to generate the PMFs required fitting parameters. Therefore the agreement we find with explicit water calculations only depends on the approximations themselves.

By comparison, other groups[16] have studied how the solvation energy calculated by the method of Eisenberg and McLachlan[14] affects the energy of the alanine dipeptide minima. In their work they compute the solvent accessible surface and multiply it by the atomic solvation parameters to obtain the solvation energy of the α_R and $C_{7\mathrm{ax}}$ conformers. The solvation energy is found to be 0.65 kcal/mol lower for the α_r conformation. In our calculation with explicit water (figure 7(a)) we find that the introduction of water destabilizes the $C_{7\mathrm{ax}}$ conformation by 4.5 kcal/mol. The Eisenberg method does not include the screening effect of water on intra-solute electrostatic interactions, by rearrangement of the water dipoles. Therefore, to explain the discrepancy in the two calculations we conclude that the screening effect accounts for the bulk of the free energy change.

When comparing our results to those computed by the solution of the RISM equations, we find that while our approach seems to correctly reproduce the relative energies of the four minima found in solution, the RISM approach produces results in sharp contrast to ours and other studies. For instance, the work of Pettitt and Karplus,[1] finds that the $C_{7\mathrm{ax}}$ conformation is more stable than the α_R one. In fact, they find that the probability distribution of the alanine dipeptide conformers is heavily peaked in the $C_{7\mathrm{ax}}$ and β region, while most, including our work, finds the α_R and β region to be principally occupied in water. However, it

should be noted that we are not using the same bare potentials as those used in the work of Pettitt and Karplus, and that this may account for some of the differences in our results.

In conclusion, the comparison of the free energy profiles produced with PMFs and explicit water simulations, produces qualitatively correct results. In conjunction with the enormous advantage in computational efficiency of this technique with respect to simulations with explicit water, we feel the results demonstrate that the procedure could be extremely useful in simulations which aim to explore conformational space. As we have seen, with explicit water simulations it is difficult to exhaustively sample the phase space of four paths of the alanine dipeptide, because of the many degrees of freedom associated with 200 water molecules. Therefore, when dealing with even larger molecules, use of the PMFs will allow one to sample molecular conformational spaces that cannot be adequately sampled with conventional simulations.

V. SUMMARY

We have developed a procedure for generating potentials of mean force between any two uncharged atomic solutes in an aqueous solution. This entails the calculation of the solute–solvent correlation functions by running a Monte Carlo simulation with explicit solvent. We next approximate the water oxygen and hydrogen density around our two solutes by making use of the Kirkwood approximation. Finally, by computing the mean force exerted by the water on the solutes, and integrating along the solute separation distance, we are able to generate the PMFs.

We first showed that this procedure reproduces well the PMFs between two methane molecules in water, when compared to a molecular dynamics simulation with explicit water molecules. By introducing the approximation that the effect of water on intra-solute electrostatic interactions may be incorporated into a dielectric constant of 80, we are able to extend the procedure to also produce PMFs between charged molecules.

Finally, by assuming that the n-body potential of mean force, for the molecule alanine dipeptide, may be expressed as the sum of the 2-body PMFs, as in the preceding work of Pettitt and Karplus,[1] we were able to calculate the free energy profiles along four paths in the molecule's ϕ,ψ space. These were in reasonable agreement (within 2.0 kcal/mol) with the same profiles computed in a simulation with 200 water molecules, that used the same intramolecular force field parameters. However, because the calculation with PMFs excludes the water molecules, it is able to generate the profiles three orders of magnitude faster than the full simulation.

Therefore we have shown that without explicitly including water molecules in a simulation of a biomolecule, one is able to reproduce many of the effects of water. Since we screen all electrostatic interactions with a dielectric constant of 80, we include an estimate of the electrostatic screening effects of water. Furthermore we are able to represent explicitly the hydrophobic effect, the attraction between Lennard-Jones particles in water, as seen, for instance, in the PMF between methanes (see Fig. 2).

We cannot yet accurately model water mediated hydrogen bonds. However, in the PMFs between methanes, we notice a second minimum in the potential at 7 Å, when a water molecule exactly fits between the two methanes. Therefore, our technique does account for aspects of water mediated interactions.

We have attempted to make brief comparisons of our results with those computed by some of the techniques mentioned in the introduction. The potential of mean force between methanes is calculated using the approach of Eisenberg and McLachlan[14] and found to produce too shallow a minimum. By contrast, the potential between sodium and chlorine is greatly exaggerated when the dielectric is set to be proportional to the distance ($\varepsilon = r$). We have also shown that the solvation energy of the alanine dipeptide, computed by the method of Eisenberg and McLachlan, represents only a small contribution to the free energy profiles, suggesting that the dominant effect of water arises from its screening of electrostatic interactions. Finally we compared the probability distribution of alanine dipeptide conformers produced by us with those generated by the RISM technique,[1] and found that while our results agreed with other published work, the RISM approach yielded significantly different results.

In conclusion, simulations run with our PMFs, although not as accurate as those with many explicit water molecules, permit a more thorough sampling of conformational space than is possible with other approaches. The primary limitation of the approach presented here is that it is only valid for small molecules, where the solvent accessible surface is close to the total molecular surface. Furthermore the limited accuracy of the Kirkwood superposition approximation does not allow us to accurately compute the electrostatic screening of the intrasolute potentials, which is why we introduce a dielectric constant of 80 in the model.

However, the procedure we have developed has the distinct advantage of using various approximations that do not require any arbitrary fitting parameters. We therefore feel it is a useful initial step in the development of rigorous potentials of mean force, that might include 3-body and higher order terms, for application to biomolecular simulations.

ACKNOWLEDGMENTS

We are grateful to the Advanced Computing Laboratory for access to their computer facilities. Parts of this work were supported by the U.S. Department of Energy and by National Science Foundation Grant No. PHY9512673.

[1] B.M. Pettitt and M. Karplus, Chem. Phys. Lett. **121**, 194 (1985).
[2] V. Madison and K.D. Kopple, J. Am. Chem. Soc. **102**, 4855 (1980).
[3] H.S. Frank and M.W. Evans, J. Chem. Phys. **13**, 507 (1945).
[4] P.L. Privalov and S.J. Gill, Adv. Prot. Chem. **39**, 191 (1988).
[5] C. Simmerling and R. Elber, J. Am. Chem. Soc. **116**, 2534 (1994).
[6] A. Joachimiak, T. E. Haran, and P.B. Sigler, EMBO J. **13**, 367 (1994).
[7] H.L. Friedman, *Ionic Solution Theory* (Interscience, New York, 1962).
[8] L.X. Dang, J. E. Rice, and P.A. Kollman, J. Chem. Phys. **93**, 7528 (1990).
[9] D.J. Lockhart and P.S. Kim, Science **260**, 198 (1993).

[10] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, J.S. States, S. Swaminathan, and M. Karplus, J. Comp. Chem. **2**, 187 (1983).
[11] S.J. Wiener, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Wiener, J. A. Chem. Soc. **106**, 765 (1984).
[12] S.S. Sung, Biophys. J. **66**, 1796 (1994).
[13] J.A. McCammon, S.H. Northrup, M. Karplus, and R.M. Levy, Biopolymers **19**, 2033 (1980).
[14] D. Eisenberg and A.D. McLachlan, Nature **319**, 199 (1986).
[15] G. Nemethy, Z.I. Hodes, and H.A. Scheraga, Proc. Nat. Acad. Sci. USA **75**, 5760 (1978).
[16] C.A. Schiffer, J.W. Caldwell, R.M. Stroud, and P.A. Kollman, Prot. Sci. **1**, 396 (1992).
[17] A. Jean-Charles, A. Nicholls, K. Sharp, B. Honig, A. Tempczyk, T.F. Hendrickson, and W.C. Still, J. Am. Chem. Soc. **113**, 1454 (1993).
[18] V. Luzhkov and A. Warshel, J. Am. Chem. Soc. **113**, 4491 (1991).
[19] D. Chandler, in *The Liquid State of Matter: Fluids, Simple and Complex,* edited by E.W. Montroll and J.L. Lebowitz (North-Holland, Amsterdam, 1982).
[20] B.M. Petitt and M. Karplus, Chem. Phys. Lett. **121**, 194 (1985).
[21] B. Lee and F.M. Richards, J. Mol. Biol. **55**, 379 (1971).
[22] G. Hummer, D.M. Soumpasis, and M. Neumann, Mol. Phys. **77**, 769 (1992).
[23] M. Pellegrini and S. Doniach, J. Chem. Phys. **103**, 2696 (1995).
[24] G. Hummer, A.E. García, and D.M. Soumpasis, Biophys. J. (in press).
[25] G. Hummer, D.M. Soumpasis and A.E. García, in *Non-linear Excitations in Biomolecules,* edited by M. Peyard (Springer, Editions de Physique, Berlin, in press, 1995).
[26] D. van Belle and S.J. Wodak, J. Am. Chem. Soc. **115**, 647 (1993).
[27] J.G. Kirkwood, J. Chem. Phys. **3**, 300 (1935).
[28] G. Hummer and D.M. Soumpasis, J. Chem Phys. **98**, 581 (1993).
[29] R. Abe, Prog. Theo. Phys. **21**, 421 (1959).
[30] P. Attard, Mol. Phys. **74**, 547 (1991).
[31] H.J.C. Berendsen, J.P.M. Postma, W.F. Von Gunsteren, and J. Hermans, *Intermolecular Forces,* edited by B. Pullman (Reidel, Dordrecht, Holland, 1981), p. 331.
[32] W.L. Jorgensen and J. Tirado-Rives, J. Am. Chem. Soc. **110**, 1657 (1988).
[33] W.L. Jorgensen, J.D. Madura, and C.J. Swenson, J. Am. Chem. Soc. **106**, 6638 (1984).
[34] M. Levitt, M. Hirshberg, R. Sharon, and V. Daggett (unpublished).
[35] A.K. Soper and M.G. Phillips, Chem. Phys. **107**, 47 (1987).
[36] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, J. Chem. Phys. **79**, 926 (1983).
[37] N. Metropolis, M.N. Rosenbluth, A. Rosenbluth, H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[38] T. Solmajer and E.L. Mehler, Protein Eng. **4**, 911 (1991).
[39] H.S. Frank and W.J. Evans, Chem. Phys. **13**, 507 (1945).
[40] The solvent accessible surface area was computed using the MidasPlus program from the Computer Graphics Laboratory, University of California, San Francisco (supported by NIH RR-01081).
[41] J.H. Yang, J.K. Shin, and M.S. Jhon, J. Mol. Struct. **268**, 169 (1992).
[42] P.E. Smith, B.M. Pettitt, and M. Karplus, J. Phys. Chem. **97**, 6907 (1993).
[43] D.J. Tobias and C.L. Brooks, J. Phys. Chem. **96**, 3864 (1992).
[44] A.G. Anderson and J. Hermans, Prot.: Struct. Funct. Gen. **3**, 262 (1988).
[45] W.L. Jorgensen, Acc. Chem Res. **22**, 184 (1989).