Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations

Dmitrii Beglov and Benoît Roux
Department of Chemistry, University of Montreal, C.P. 6128, succ. A, Canada H3C 3J7

(Received 28 January 1994; accepted 1 March 1994)

An approach is developed to obtain statistical properties similar to those of an infinite bulk system from computer simulations of a finite cluster. A rigorous theoretical formulation is given for the solvent boundary potential which takes the influence of the surrounding bulk into account. The solvent boundary potential is the configuration-dependent solvation free energy of an effective cluster composed of an arbitrary solute and a finite number of explicit solvent molecules embedded inside a hard sphere of variable radius; the hard sphere does not act directly on the solute or the explicit solvent molecules, and its radius varies according to the instantaneous configurations. The formulation follows from an exact separation of the multidimensional configurational Boltzmann integral in terms of the solvent molecules nearest to the solute and the remaining bulk solvent molecules. An approximation to the solvent boundary potential is constructed for simulations of bulk water at constant pressure, including the influence of van der Waals and electrostatic interactions. The approximation is illustrated with calculations of the solvation free energy of a water molecule and of sodium and potassium ions. The influence of bulk solvent on the conformational equilibrium of molecular solutes is illustrated by performing umbrella sampling calculations of *n*-butane and alanine dipeptide in water. The boundary potential is tested to examine the dependence of the results on the number of water molecules included explicitly in the simulations. It is observed that bulk-like results are obtained, even when only the waters in the first hydration shell are included explicitly.

INTRODUCTION

Computer simulations based on atomic models in which a large number of solvent molecules are treated explicitly with conventional periodic boundary conditions represent one of the most detailed approaches to study the influence of solvation on complex biomolecules.[1–4] Nevertheless, such approaches are very intensive computationally due to the large number of solvent molecules that are required to model a bulk solution. Significant amounts of computational time are spent to calculate a detailed trajectory of a large number of solvent molecules even though the solute of interest may represent a very small fraction of the total number of atoms in the simulation (see Ref. 3, for example). Furthermore, in spite of their computational cost, approaches based on conventional periodic boundary conditions are not exempt from approximations. For example, difficulties arise in thermodynamic perturbation free energy calculations involving charged species when long range electrostatic interactions are truncated[5] or summed over an infinite periodic array using Ewald techniques.[6–8]

Partly due to these difficulties, different approximate schemes have been developed and used to take the dominant effects of solvation into account implicitly, i.e., without including any explicit solvent molecules. Among those are the phenomenological approaches based on a representation of solvation effects in terms of atom or group based solvent-exposed area,[9,10] hydration shell models,[11–13] continuum electrostatic approximations,[14–16] and statistical mechanical theories based on RISM-HNC integral equations.[17] Approximate schemes treating the solvent implicitly can provide useful quantitative estimates and remain computationally inexpensive. In particular, approximations based on continuum electrostatics, in which the solvent is represented as a featureless dielectric material, are remarkably successful in reproducing the electrostatic contribution to the solvation free energy of small solutes.[18] Nevertheless, a description in which all atomic and structural details of the solvent molecules are ignored may not always be appropriate. As an example, direct hydrogen bonding of solvent water molecules to the backbone carbonyls was shown to be important in the unfolding process of an α-helix.[2]

An intermediate approach, considered here, consists of including a small number of explicit solvent molecules in the vicinity of the solute, and representing the influence of the remaining bulk with an effective solvent boundary potential. The approach, which was originally proposed by Stace and Murrel for studying the recombination of radical atoms in the gas phase,[19] has been extended by several authors for the simulation of finite representation of infinite dense bulk systems.[20–26] The first to design a simulation method appropriate for liquids were Berkowitz and McCammon.[20] In their method, the many-body system was divided into three main spherical regions: a central reaction region, a buffer region, and a surrounding static reservoir region. The entire system (reaction, buffer, and reservoir regions) moves with the central solute of interest through a large simulation box. The forces arising from the reservoir region are calculated from fixed atomic centers, thus the influence of the surrounding bulk was not treated implicitly. Instead of using explicit fixed

 0021-9606/94/100(12)/9050/14/$6.00

atomic centers in the bath region, Brooks and Karplus introduced a mean force field approximation (MFFA) to calculate a soft boundary potential representing the average influence of the reservoir region on the reaction region.[21] In the MFFA treatment, the boundary potential was calculated by integrating all contributions to the average force arising from the reservoir region; the density in the reservoir region was approximated by using the radial pair distribution function of the bulk liquid. The MFFA approach was extended by Brooks, Brunger, and Karplus for the simulation of bulk water, although the average contribution of long range electrostatic forces due to the polar nature of the surrounding bulk water was not included.[22] The average electrostatic reaction field was taken into account in the surface constrained all-atom solvent (SCAAS) treatment of King and Warshel,[23,24] and in the reaction field with exclusion (RFE) of Rullmann and van Duijnen[25,26] by using a dielectric continuum approximation to the average electrostatic reaction field originally derived by Kirkwood.[27]

Although the previous treatments of the solvent boundary potential provided many of the essential elements for useful practical approximations, some aspects need additional consideration. In particular, several theoretical formulations of the solvent boundary potential were elaborated by considering the influence of bulk solvent surrounding a spherical region of constant radius (MFFA, SCAAS, and RFE). Such formulation of the solvent boundary potential does not take into account the density and volume fluctuations characteristic of a bulk solvent at constant pressure. For example, this is important when a solvent boundary potential is used to calculate the free energy difference between two solutes occupying significantly different specific volumes.[28] In addition, some difficulties involving the treatment of the electrostatic reaction field remain unresolved. In particular, the presence of point charges near the continuum dielectric interface at the boundary leads to artificial divergences.[26,29] In the Monte Carlo simulations based on the RFE method, this problem was avoided by rejecting the configurations in which the oxygen of a water molecule is closer than a minimum distance from the dielectric discontinuity,[26] i.e., effectively introducing a repulsive wall near the dielectric boundary. In the SCAAS method, phenomenological biasing potentials are introduced to constrain the average density and solvent polarization at the boundary of the finite system around prescribed values.[23,24] However, it was reported that the total dipole of the system was unstable during molecular dynamics simulations of bulk water even though no explanation was provided.[24]

In light of the difficulties raised by the previous treatments, a different route was chosen to formulate and develop a solvent boundary potential for computer simulations of a finite representation of an infinite bulk system. The present theoretical formulation is based on a separation of the multidimensional solute–solvent configurational integral in terms of n "inner" solvent molecules nearest to an arbitrary solute, and the remaining "outer" bulk solvent molecules. Following this formulation, it is shown rigorously that equilibrium statistical properties representative of an infinite bulk system can be obtained from the finite system with a specifically defined "solvent boundary potential." The solvent boundary potential is recognized as the solvation free energy of an effective cluster comprising the solute and n solvent molecules embedded in a large hard sphere. The hard sphere corresponds to a configurational restriction on the outer bulk solvent molecules; its radius is variable such as to include the most distant inner solvent molecule. This formulation, which differs significantly from previous treatments, provides further insight into the statistical mechanical basis of the solvent boundary potential and is helpful in constructing useful approximations for computer simulations in dense liquids.

In the first section, the theoretical formulation of the solvent boundary potential is developed. In the second section, an approximate solvent boundary potential is constructed for simulations of bulk water. In the third section, the boundary potential for water is illustrated with a few applications. Simulations of small systems with different number of explicit waters are used to examine the solvation of a water molecule and of sodium and potassium ions in bulk water and to calculate the intramolecular potential of mean force of n-butane and of the alanine dipeptide. A brief discussion of the results concludes the paper.

FORMAL DEVELOPMENTS

An arbitrary solute immersed in a bulk solvent is considered. It is assumed, without loss of generality, that the center of mass of the solute is fixed at the origin. In the canonical ensemble, all statistical averages to be evaluated involve multidimensional integrals of the form,

$$Z=\int d(\mathbf{X_u})\,\frac{1}{N!}\int d(\mathbf{1})\cdots d(\mathbf{N})e^{-U/k_BT}, \tag{1}$$

where U is the total potential energy of the system and $(\mathbf{X_u},\mathbf{1}\cdots\mathbf{N})$ represent the degrees of freedom of the solute and the N solvent molecules.

At first sight, it appears that the problem of reducing the multidimensional configurational integral Z in terms of $1,\ldots,n$ "explicit" and $n+1,\ldots,N$ effective "implicit" solvent molecules could be simply formulated in the framework of joint density distribution functions as developed in traditional statistical mechanical treatments of liquids,[30,31]

$$\rho^{(n)}(\mathbf{X_u},\mathbf{1},\ldots,\mathbf{n})=\frac{\langle\delta(\mathbf{X_u}-\mathbf{X_u'})\,\delta(\mathbf{1}-\mathbf{1'})\cdots\delta(\mathbf{n}-\mathbf{n'})\rangle}{\langle\delta(\mathbf{1}-\mathbf{1'})\rangle\cdots\langle\delta(\mathbf{n}-\mathbf{n'})\rangle}, \tag{2}$$

where the joint n-particles potential of mean force,

$$\mathscr{W}^{(n)}(\mathbf{X_u},\mathbf{1},\ldots,\mathbf{n})=-k_BT\,\ln[\rho^{(n)}(\mathbf{X_u},\mathbf{1},\ldots,\mathbf{n})], \tag{3}$$

is an explicit function of the coordinates of the solute and n solvent molecules taking implicitly the influence of the $(n+1)$th to Nth molecules into account. Although such treatment is rigorously correct, further consideration reveals that it is inadequate for the development of a theoretical formulation of the solvent boundary potential. The difficulty with Eq. (3) arises from the fact that the n explicit solvent molecules and the $(N-n)$ implicit solvent molecules are not restricted to distinct regions of configurational space that are

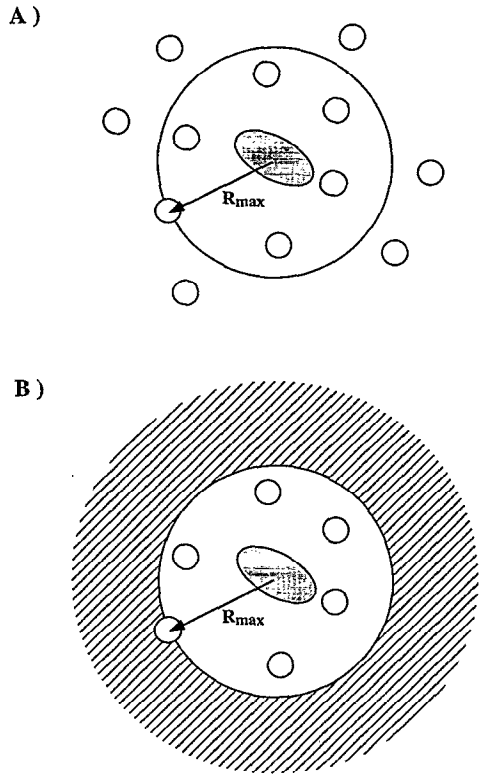

FIG. 1. Schematic representation of the inner and outer regions as described by Eq. (3) in the text. (A) The inner region contains one solute and the n nearest solvent molecules; the nth solvent molecule is at a distance R_{max} from the center of mass of the solute. (B) In the reduced system, only the solute and the n solvent molecules of the inner region are described explicitly. The influence of the $(N-n)$ solvent molecule belonging to the outer region is described through the potential of mean force $\mathscr{W}(\mathbf{X}_u,\mathbf{1},...,\mathbf{n})$ defined by Eq. (5).

near and far away from the solute, respectively. This implies that the potential of mean force $\mathscr{W}^{(n)}(\mathbf{X}_u,\mathbf{1},...,\mathbf{n})$ must be able to take into account the influence of implicit solvent molecules in direct contact with the solute. This is inappropriate to construct a useful simulation method in which all the solvent molecules near the solute are represented with an atomic model. A different formulation of the problem, which was chosen in the MFFA method,[21] is to define the explicit and implicit solvent molecules according to a spherical region fixed in space surrounding the solute. However, in this formulation the number of explicit solvent molecules inside the spherical region must be allowed to vary to account for the density fluctuations present in a constant pressure system. Even though molecular dynamic simulation techniques with a variable number of molecules have been developed,[32] this approach was not chosen here.

To find a useful formulation, the N solvent molecules are separated in terms of two groups: the n inner solvent molecules nearest to the solute and the $(N-n)$ remaining outer solvent molecules. This is represented schematically in Fig. 1. The separation is possible for all configurations contributing to the integral except those for which the nth and $(n+1)$th solvent molecules are exactly at the same distance from the solute. Those configurations were assumed to represent a set of measure zero and were ignored. Exploiting the indistinguishability of the solvent molecules, the solvent molecules are renumbered such that the inner and outer groups are represented by the $(1,...,n)$ first and $(n+1,...,N)$ last solvent molecules, respectively. Based on this separation, the configurational integral in Eq. (1) can be expressed as,

$$Z=\int d(\mathbf{X}_u)\frac{1}{n!}\int d(\mathbf{1})\cdots d(\mathbf{n}) \times\frac{1}{(N-n)!}\int' d(\mathbf{n+1})\cdots d(\mathbf{N})e^{-U/k_BT}, \tag{4}$$

where the prime on the integral symbol indicates that the $(n+1,...,N)$ outer molecules are spatially restricted, i.e., all contribution in which any of the $(n+1,...,N)$ is closer to the solute than any of the $(1,...,n)$ molecules is rejected to avoid multiple counting of identical configurations. The factors of $n!$ and $(N-n)!$ account for the multiple counting of identical configurations of the inner and outer solvent molecules, respectively, and the integral involving the n inner solvent molecules is unrestricted. The exact formal equivalence of the configurational integral of Eq. (4) with the more familiar Eq. (1) is illustrated in Appendix A in the case of the ideal gas partition function.

The separation of the solvent molecule in two groups makes it possible to integrate out the contribution of the outer solvent molecules such that their influence is taken into account implicitly. It follows that statistical properties of the solute and the n nearest solvent molecules representative of an infinite bulk system can be expressed in terms of the Boltzmann average of the finite system interacting with the effective potential energy, $\mathscr{W}(\mathbf{X}_u,\mathbf{1},...,\mathbf{n})$, defined from the restricted configurational integral,

$$e^{-\mathscr{W}(\mathbf{X}_u,\mathbf{1},...,\mathbf{n})/k_BT} \equiv\frac{1}{C}\int' d(\mathbf{n+1})\cdots d(\mathbf{N})e^{-U(\mathbf{X}_u,\mathbf{1},...,\mathbf{N})/k_BT}, \tag{5}$$

where C is an arbitrary constant that can be chosen for convenience (see below). For example, the average of an observable $Q[\mathbf{X}_u]$ depending on the solute degrees of freedom, $\mathbf{X}_u$, is expressed as,

$$\langle Q\rangle=\frac{\int d(\mathbf{X}_u)\frac{1}{N!}\int d(\mathbf{1})\cdots d(\mathbf{n})\cdots d(\mathbf{N})Q[\mathbf{X}_u]e^{-U/k_BT}}{\int d(\mathbf{X}_u)\frac{1}{N!}\int d(\mathbf{1})\cdots d(\mathbf{n})\cdots d(\mathbf{N})e^{-U/k_BT}} =\frac{\int d(\mathbf{X}_u)\frac{1}{n!}\int d(\mathbf{1})\cdots d(\mathbf{n})Q[\mathbf{X}_u]e^{-\mathscr{W}(\mathbf{X}_u,\mathbf{1},...,\mathbf{n})/k_BT}}{\int d(\mathbf{X}_u)\frac{1}{n!}\int d(\mathbf{1})\cdots d(\mathbf{n})e^{-\mathscr{W}(\mathbf{X}_u,\mathbf{1},...,\mathbf{n})/k_BT}}. \tag{6}$$

Similarly, the difference in the free energy of solvation of two solutes can be expressed in terms of the finite system. Following standard free energy perturbation techniques,[5,33] a thermodynamic coupling parameter, λ, is introduced such that the total the solute–solvent interaction energy corre-

sponds to solute 1 or solute 2 when the value of λ is set to λ_1 or λ_2, respectively. The free energy difference between the two solutes with coupling parameters λ_1 and λ_2 can be written as,

$$
\begin{aligned}
e^{-[A(\lambda_1)-A(\lambda_2)]/kBT} &= \frac{\int d(\mathbf{X_u})(1/N!)\int d(\mathbf{1})\cdots d(\mathbf{n})\cdots d(\mathbf{N}) e^{-U(\lambda_1)/k_BT}}{\int d(\mathbf{X_u})(1/N!)\int d(\mathbf{1})\cdots d(\mathbf{n})\cdots d(\mathbf{N}) e^{-U(\lambda_2)/k_BT}} \\
&= \frac{\int d(\mathbf{X_u})(1/n!)\int d(\mathbf{1})\cdots d(\mathbf{n}) e^{-\mathscr{W}(\lambda_1)]/k_BT}}{\int d(\mathbf{X_u})(1/n!)\int d(\mathbf{1})\cdots d(\mathbf{n}) e^{-\mathscr{W}(\lambda_2)/k_BT}} \\
&= \langle e^{-[\mathscr{W}(\lambda_1)-\mathscr{W}(\lambda_2)]/k_BT}\rangle_{(\lambda_2)}, \qquad (7)
\end{aligned}
$$

where the subscript (λ_2) means that the average is performed in an ensemble based on the effective potential $\mathscr{W}(\lambda_2)$ [the potential of mean force $\mathscr{W}(\lambda_i)$ follows from Eq. (5)].

A simple interpretation of the potential of mean force is suggested for one particular choice of the undetermined arbitrary constant C in Eq. (5). For this purpose, the total potential energy of the bulk system is decomposed in terms of the inner–inner (U_{ii}), inner–outer (U_{io}) and outer–outer (U_{oo}) contributions,

$$U = U_{ii} + U_{io} + U_{oo}, \qquad (8)$$

and the configurational restriction on the integral over the outer solvent molecules is be expressed in terms of unrestricted integrals with Heaviside step functions, $H(|\mathbf{r}_i| - R_{max})$,

$$e^{-\mathscr{W}(\mathbf{X_u},\mathbf{1},\ldots,\mathbf{n})/k_BT} = \frac{\int d(\mathbf{n+1}) H(|\mathbf{r}_{n+1}| - R_{max})\cdots\int d(\mathbf{N}) H(|\mathbf{r}_N| - R_{max}) e^{-[U_{ii}+U_{io}+U_{oo}]/k_BT}}{\int d(\mathbf{n+1})\cdots\int d(\mathbf{N}) e^{-U_{oo}/k_BT}}, \qquad (9)$$

where

$$R_{max} = \mathrm{MAX}(|\mathbf{r}_1|,\ldots,|\mathbf{r}_n|), \qquad (10)$$

is defined to include the solvent molecule in the inner region that is the farthest from the solute. In Eq. (9), the Heaviside step functions act effectively as a repulsive hard potential $U_{hsr}(R_{max})$ to prevent the outer molecules from penetrating into a spherical region of radius R_{max}, i.e.,

$$
\begin{aligned}
e^{-U_{hsr}/k_BT} \equiv H(|\mathbf{r}_{n+1}| - R_{max}) H(|\mathbf{r}_{n+2}| - R_{max}) \\
\cdots H(|\mathbf{r}_N| - R_{max}). \qquad (11)
\end{aligned}
$$

The "solvent boundary potential" corresponds to the total influence of the outer solvent molecules on the finite system and is defined as $\Delta\mathscr{W} \equiv \mathscr{W} - U_{ii}$. Expressing the potential of mean force in Eq. (9) in terms of unrestricted averages over the configurations of the outer molecules, the solvent boundary potential can be written as,

$$e^{-\Delta\mathscr{W}(\mathbf{X_u},\mathbf{1},\ldots,\mathbf{n})/k_BT} = \langle e^{-[U_{io}+U_{hsr}]/k_BT}\rangle_{(U_{oo})}, \qquad (12)$$

which can be recognized as the solvation free energy of an effective cluster composed of one solute and n solvent molecules frozen in the $(\mathbf{X_u},\mathbf{1},\ldots,\mathbf{n})$ configuration and embedded inside a hard sphere of radius R_{max}.[34] The hard sphere is introduced as a device to account for the configurational restriction of the outer solvent molecules and does not act directly on the solute or the n inner solvent molecules. The radius R_{max} varies according to the instantaneous configuration of the inner solvent molecules [see Eqs. (10) and (11) above]. This interpretation of the solvent boundary potential, relating the statistical properties representative of an infinite system to that of a finite effective system, is a main result of this paper. The present interpretation is particularly helpful in deriving approximations for the boundary potential $\Delta\mathscr{W}$. Based on this interpretation, an approximate boundary potential for bulk water is developed in the next section.

The separation of the configurational integral in Eq. (1) in terms of two distinct groups of solvent molecules as expressed in Eq. (4) provides a formal link relating the properties of a finite effective system to that of an infinite bulk system. Based on Eq. (6), statistical properties representative of an infinite bulk system can be obtained directly, as a simple Bolztmann average involving the particles of the finite system interacting with an effective potential. Similarly, Eq. (7) is analogous to the familiar expression of free energy perturbation techniques.[5,33] For some observables involving directly the outer solvent molecules, such as the solute–solvent interaction energy or solute–solvent density distribution function at large distances, averages may not be expressed solely in terms of the finite system without further assumptions. Thus to compute averages representative of an infinite system, it is sufficient to generate n-particle configurations obeying a Boltzmann distribution with the potential energy $U_{ii}+\Delta\mathscr{W}$. If an approximation to the solvent boundary potential is available, this task can be accomplished by Monte Carlo, or by stochastic Brownian/Langevin dynamics methods.[6] In generating these configurations only the solvent molecules located inside the sphere are represented explicitly with an atomic model, with the potential U_{ii}; the influence of the outer solvent molecules is taken into account by the boundary potential, $\Delta\mathscr{W}$.

Although Eq. (6) shows that the correct equilibrium average properties representative of an infinite system can be obtained in principle from the finite effective system, a corresponding equivalence for dynamical properties was not demonstrated here. In fact, several dynamical properties cannot be described correctly by the finite system, even if the exact solvent boundary potential is used. For example, the

translational diffusion constant of the solvent molecules in the simulation converges to zero since the mean square displacement of the particles is bound by the finite system. Nevertheless, it may be expected that carefully constructed buffer regions in which the dynamics of the molecules obeys Langevin stochastics algorithms, such as described by Brooks and Karplus,[21] can provide a useful approximation to study dynamical properties involving localized events far away from the boundary.

CONSTRUCTION OF A BOUNDARY POTENTIAL FOR BULK WATER

The configuration-dependent boundary potential due to the surrounding bulk solvent can be obtained by calculating, step by step, the reversible thermodynamical work necessary to assemble the effective cluster composed of the solute and n solvent molecules in the configuration $(\mathbf{X_u},\mathbf{1},...,\mathbf{n})$ embedded in a large hard sphere of radius R_{max} corresponding to the distance from the origin of the outermost inner solvent molecule. Although the final result does not depend on the choice of a particular path, the boundary potential is calculated by assembling the effective complex through the following steps:

(i) A hard sphere of radius R_{max} is inserted in a bulk solvent.
(ii) One by one, the neutral solute and the n solvent molecules are inserted in the bulk solvent away from the origin and translated to their respective positions inside the hard sphere.
(iii) The charges of the solute and of the n solvent molecules located inside the sphere are switched on adiabatically.

Following this path, the free energy contributions at each step are

$$\Delta\mathscr{W}=\Delta\mathscr{W}_{hsr}+\Delta\mathscr{W}_{vdw}+\Delta\mathscr{W}_{elec}, \tag{13}$$

where $\Delta\mathscr{W}_{hsr}$, $\Delta\mathscr{W}_{vdw}$, and $\Delta\mathscr{W}_{elec}$ represent the cavity, the van der Waals, and the electrostatic free energy contributions, respectively. Several approaches based on different level of sophistication can be used to calculate the solvent boundary potential. In principle, the various schemes used to account implicitly for the effects of bulk solvation mentioned in the introduction could be used to approximate the function $\Delta\mathscr{W}$ (see Refs. 9–17). In the present case, analytical functions were chosen to obtain a simple and computationally inexpensive approximation to the boundary potential appropriate for bulk water simulations. One important simplification that is used in the present treatment is based on the assumption that the simulation system is spherical on average. This approximation, which is illustrated schematically on Fig. 2, results in an effective smooth spherical boundary surrounding the finite simulation system. Because this is an approximation, the solvent boundary potential needs to be parametrized and adjusted empirically. Preliminary simulations of a sphere containing 100 water molecules, described below, were used to develop the present boundary potential.

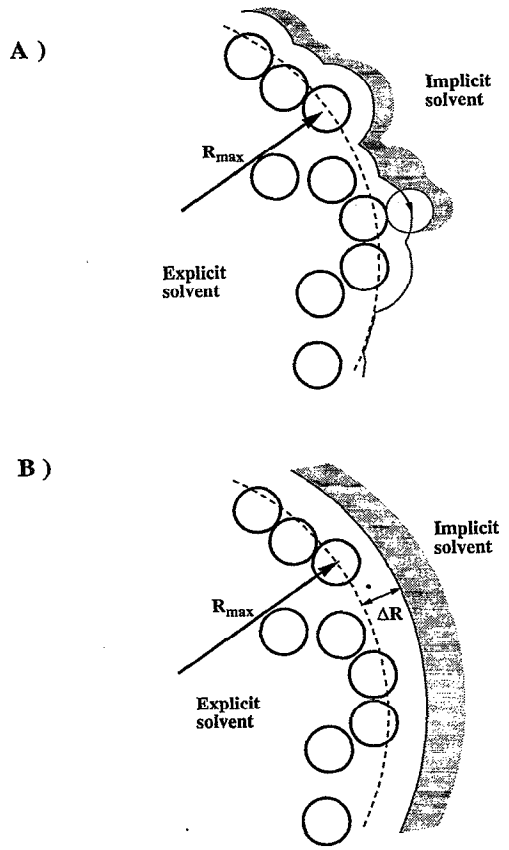

FIG. 2. Schematic representation of an instantaneous configuration of the solvent molecules near the boundary of the simulation system. The hard sphere restriction is represented by a dashed line. (A) Instantaneous configuration of the inner system. The farthest water molecule is located at a radial distance of R_{max} from the center of the simulation system. The region of space from which the outer molecules are excluded is represented by the van der Waals envelop of the inner molecules plus the exclusion hard sphere restriction of radius R_{max} [see Eqs. (8) and (9)] and the instantaneous shape of the van der Waals envelop of the system is not perfectly spherical. (B) The assumption of the average effective spherical symmetry of the simulation system results in a smooth effective interface located around $R_{max}+\Delta R$ due to the repulsion between the inner and outer molecule, where ΔR is related to the diameter of the solvent molecules. The effective interface for the dielectric reaction field and the van der Waals interactions were adjusted empirically; the effective dielectric interface was placed at a distance $R_{max}+\Delta R_{diel}$; the van der Waals interactions were modeled, based on HNC-RISM calculations, as a smooth sphere of radius $R_{max}+\Delta R_{vdw}$.

Contribution from the hard sphere restriction

Because the radius of the hard sphere is expected to be much larger than the size of a solvent molecule, a simple asymptotic form for $\Delta\mathscr{W}_{hsr}$ should provide an adequate approximation, i.e.,

$$\Delta\mathscr{W}_{hsr}\simeq pV+\sigma S, \tag{14}$$

where V and S are the volume and the surface corresponding to the hard sphere restriction of radius R_{max}, and p and σ are the pressure and surface tension of the bulk liquid.

Contribution from van der Waals interactions

For a pure solvent, the total free energy contribution from core repulsion and van der Waals interactions can be described as the total reversible work necessary to assemble the system of electrically neutral molecules embedded in the effective hard sphere restriction potential. In the case of water, the interaction of a neutral solvent molecule with the surroundings is usually represented by a simple short range radial potential centered on the oxygen. In particular, the oxygen interaction center is a Lennard-Jones potential $4\epsilon[(\sigma/r)^{12}-(\sigma/r)^{6}]$ for several simple water models such as TIP3P[35] or SPC.[36] The total reversible work is obtained by transporting the solute and the n Lennard-Jones oxygens, one by one, to the interior of the hard sphere restriction, that is,

$$\Delta\mathscr{W}_{\rm vdw}=\Delta w^{(1)}(\mathbf{r}_1;R_{\rm max})+\Delta w^{(2)}(\mathbf{r}_2;R_{\rm max},\mathbf{r}_1)+\cdots +\Delta w^{(n)}(\mathbf{r}_n;R_{\rm max},\mathbf{r}_1,\ldots,\mathbf{r}_{n-1}), \quad (15)$$

where the functions $\Delta w^{(i)}(\mathbf{r}_i;R_{\rm max},\mathbf{r}_1,\ldots,\mathbf{r}_{i-1})$ are the solvent-induced cavity potentials of mean force between the ith Lennard-Jones oxygen and the partly assembled system with $(i-1)$ Lennard-Jones oxygens embedded in the hard sphere of radius $R_{\rm max}$. The constant corresponding to the free energy necessary to insert one neutral water molecule in the bulk, away from the origin, has been ignored with no consequence since the boundary potential can be defined relative to an arbitrary constant according to Eq. (5).

The first term in Eq. (15), $\Delta w^{(1)}(\mathbf{r}_1;R_{\rm max})$, is the bulk solvent cavity potential of mean force between one hard sphere of radius $R_{\rm max}$ located at the origin and one Lennard-Jones oxygen located at $\mathbf{r}_1$. The subsequent terms represent the bulk solvent cavity potential of mean force between one Lennard-Jones center and the incompletely assembled system. The functions $\Delta w^{(i)}$ do not depend on the direct interactions between the system of $(i-1)$ neutral solvent molecules embedded in the hard sphere and the ith solvent molecule but are influenced by their interactions with the surrounding bulk water. To find a simple approximation, it is assumed that the system of hard sphere and the embedded neutral solvent molecules is on average spherical. The influence of the spherical system with the embedded neutral water oxygens is represented by an effective potential, $u_{\rm eff}(|\mathbf{r}|,R_{\rm max})$, calculated from an integral of Lennard-Jones centers distributed uniformly over the spherical volume $V_{\rm max}$ corresponding to the radius $R_{\rm max}$ with the bulk density of water $\rho_{\rm bulk}$,

$$u_{\rm eff}(|\mathbf{r}|,R_{\rm max})=4\pi\int_{V_{\rm max}} d\mathbf{r}'\ \rho_{\rm bulk}\left[4\epsilon\left(\frac{\sigma}{|\mathbf{r}-\mathbf{r}'|}\right)^{12}-\left(\frac{\sigma}{|\mathbf{r}-\mathbf{r}'|}\right)^{6}\right]. \quad (16)$$

The average cavity potential is then,

$$\langle\Delta w^{(n)}(\mathbf{r}_n;R_{\rm max},\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_{n-1})\rangle \simeq\Delta w^{(1)}_{\rm eff}(\mathbf{r}_i;[u_{\rm eff}(|\mathbf{r}|,R_{\rm max}]). \quad (17)$$

The approximation is illustrated schematically in Fig. 2. The effective cavity potential was calculated in the case of a simple hard sphere potential $u_{\rm hs}(r;R_{\rm max})$ and in the case of the effective potential $u_{\rm eff}(r;R_{\rm max})$, using the HNC-RISM integral equation,[17,37,38]

$$\Delta w^{(1)}(r;[u_{\rm eff}])=c_{{\rm LJ},\alpha}\star\chi_{\alpha\gamma}\star c_{\gamma,{\rm hs}}(r;[u_{\rm eff}]), \quad (18)$$

where the symbol $\star$ represents a convolution, $\chi_{\alpha\gamma}$ is the site–site solvent density succeptibility and $c_{{\rm LJ},\alpha}$ and $c_{\gamma,{\rm hs}}$ are the site–site LJ–solvent and solvent–hard sphere direct correlation functions, respectively. The site–site direct correlation functions were obtained from the RISM-HNC integral equations for a solute at infinite dilution using standard numerical techniques (see Ref. 39 for more details). The calculated RISM-HNC cavity potentials are shown in Fig. 3 for a hard sphere with a radius $R_{\rm max}$ of 12 Å. The result of the calculations indicates that the cavity potential of the effective uniformly filled sphere is qualitatively similar to that of a hard sphere with an effective van der Waals radius $R_{\rm vdw}=R_{\rm max}+\Delta R_{\rm vdw}$, i.e.,

$$\Delta w^{(1)}_{\rm eff}(\mathbf{r}_i;[u_{\rm eff}])\simeq\Delta w^{(1)}(\mathbf{r}_1;R_{\rm vdw}). \quad (19)$$

This is due to the repulsion between the outer solvent molecules and the inner solvent molecules lying near the edge of the hard sphere region (see also Fig. 2). The contribution of the van der Waals interaction is very similar to the solvent boundary potential used in the MFFA approximation.[21]

Further calculations of the cavity potential of a Lennard-Jones oxygen with a hard sphere were done using the RISM-HNC integral equation with $R_{\rm max}$ varying from 2.4 to 20.0 Å and an analytical function was constructed to obtain an approximation valid for systems of various size. The total free energy contribution from the van der Waals potential is,

$$\Delta\mathscr{W}_{\rm vdw}\simeq\sum_{i=1,n}\Delta w^{(1)}(\mathbf{r}_i;R_{\rm vdw}), \quad (20)$$

with the analytical functions,

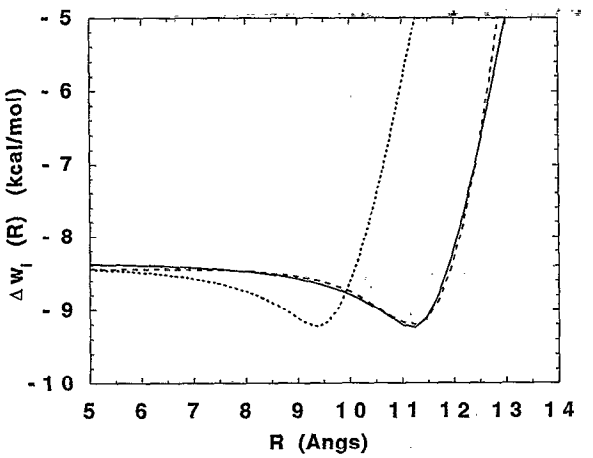

FIG. 3. Effective cavity potential $\Delta w^{(1)}(r)$ calculated by the RISM-HNC equation with Eq. (18). The result for the potential $u_{\rm eff}(r)$ (solid line) and the simple hard sphere $u_{\rm hs}(r)$ (dotted line) are shown. In both cases, the radius of the hard sphere restriction is 12 Å. The cavity potential calculated from the analytic approximation with Eqs. (22) and (23) is also shown in dashed line (a value of $R_{\rm vdw}=11.25$ Å was used).

$$\Delta w^{(1)}(\mathbf{r}_i;R_{\rm vdw})=A(R_{\rm vdw})+B(r_i-R_{\rm vdw}), \tag{21}$$

where,

$$A(R_{\rm vdw})=\begin{cases} a_0+a_1R_{\rm vdw}+a_2R_{\rm vdw}^2+a_3R_{\rm vdw}^3+a_4R_{\rm vdw}^4 & \text{if } R_{\rm vdw}<X_c \\ a_c \quad \text{if } R_{\rm vdw}\geqslant R_c \end{cases} \tag{22}$$

with the coefficients $a_0=-1.665$, $a_1=0.562$, $a_2=-0.072\,8$, $a_3=0.004\,26$, $a_4=-0.000\,092\,5$, $a_c=0.084$, $X_c=15.393$, and

$$B(x)=\begin{cases} b_2(1+x^2/b_1)^{-1}+b_3x^2+b_4 & \text{if } -5\leqslant x\leqslant 0 \\ b_c \quad \text{if } x\leqslant -5 \\ b_5+b_6x^2 \quad \text{if } x>0 \end{cases}, \tag{23}$$

where $x=r_i-R_{\rm vdw}$ and the coefficients are $b_1=1.320$, $b_2=-0.841$, $b_3=-0.001\,60$, $b_4=-8.393$, $b_c=-8.475$, $b_5=-9.234$, $b_6=1.6$. A value of 2.6 Å was taken for the radius increment $\Delta R_{\rm vdw}$. The analytical approximation is shown in Fig. 3.

Contribution from the electrostatic charging free energy

The last contribution to the solvent boundary potential of mean force corresponds to the free energy of adiabatically switching on the charges of the hard sphere system with the embedded solute and the n solvent molecules,

$$\Delta\mathscr{W}_{\rm elec}=\int_0^1 \sum_i d\lambda\ Q_i\psi_i(\lambda Q_1,\ldots,\lambda Q_i,\ldots), \tag{24}$$

where ψ_i is the electrostatic potential due the outer polar solvent molecules at the position of charge Q_i and induced in reaction to the charges $(\lambda Q_1,\ldots,\lambda Q_i,\ldots)$ inside the subsystem. If the dimension of the simulation system is large compared to the size of the solvent molecules, the dominant contribution to the reaction field may be approximated by representing the outer region as a dielectric continuum. To calculate the free energy of charging it is necessary to first solve the Poisson differential equation for each configuration of the inner system,

$$\nabla\cdot[\epsilon(\mathbf{r})\nabla\psi(\mathbf{r})]=-4\pi\rho_{\rm elec}(\mathbf{r}), \tag{25}$$

where $\rho_{\rm elec}(\mathbf{r})$ is the charge density in the system and $\epsilon(\mathbf{r})$ is a spatially varying dielectric constant $\epsilon(\mathbf{r})$ defined from the following prescription:[40] $\epsilon(\mathbf{r})=1$ at all points in space falling inside the hard sphere restriction or within the core exclusion of any of the inner solvent molecules; $\epsilon(\mathbf{r})=\epsilon_{\rm bulk}$ otherwise [Fig. 2(A)]. Although it is possible to solve the continuum electrostatic problem with numerical methods,[14,40] it is more advantageous computationally to use an analytical approximate expression. One possible approximation is obtained by assuming that, on average, the dielectric boundary is located at the surface of a sphere of radius $R_{\rm diel}$. The nature of the approximation involved in replacing the instantaneous dielectric boundary by a smooth spherical dielectric boundary with an effective radius $R_{\rm diel}$ is illustrated schematically in Fig. 2. The free energy of a distribution of point charges inside a spherical cavity enclosed in a dielectric continuum with relative dielectric constant $\epsilon_{\rm bulk}$ is given by,[27,41]

$$\Delta\mathscr{W}_{\rm elec}\simeq-\frac{1}{2}\sum_{lm}\frac{4\pi|Q_{lm}|^2}{(2l+1)}\frac{1}{R_{\rm diel}^{2l+1}}\left[\frac{\epsilon_{\rm bulk}-1}{\epsilon_{\rm bulk}+l/(l+1)}\right], \tag{26}$$

where the $|Q_{lm}|^2$ are the square of the moments of the charge distribution,

$$Q_{lm}=\sum_i Y_{lm}^*(\theta_i,\phi_i)|\mathbf{r}_i|^l, \tag{27}$$

where the Y_{lm} are the spherical harmonics, defined in terms of the Legendre polynomials,

$$Y_{lm}(\theta,\phi)=\frac{1}{2}\sqrt{\frac{2l+1}{\pi}\frac{(l-m)!}{(l+m)!}}P_{lm}[\cos(\theta)]e^{im\phi}. \tag{28}$$

To account for the finite size of the inner solvent molecules lying near the surface of the hard sphere restriction, it is expected that the dielectric boundary will be at some distance from the radius of the hard sphere restriction, i.e., $R_{\rm diel}=R_{\rm max}+\Delta R_{\rm diel}$. The average dielectric interface was chosen 2.8 Å away from the farthest solvent molecule (located at $R_{\rm max}$) according to the position of the first peak in the oxygen–oxygen radial distribution function of bulk water. This construction of the dielectric interface avoids the well-known divergences due to the presence of point charges near the boundary of the system.[26,29] In the reaction field with exclusion model (RFE), Rullmann and van Duijnen resolved this difficulty by imposing a minimum distance between the oxygen of the water molecules and the dielectric boundary.[26] They tested values of 1.58 and 3.16 Å for the so-called "exclusion radius," which plays a role similar to that of $\Delta R_{\rm diel}$ in the present work. One important difference with the RFE method is that the radius of the spherical dielectric cavity is determined from Eq. (10) and, therefore, is variable.

To accurately reproduce the absolute solvation free energy of ions in simulations in which only the primary hydration shell is included explicitly, a variable parameter $\Delta R_{\rm diel}$ was used

$$\Delta R_{\rm diel}=2.8-1.6|Q_{\rm tot}|e^{-R_{\rm max}/2}, \tag{29}$$

where $Q_{\rm tot}$ is the total charge of the simulation system. This empirical relation may be interpreted as a manifestation of the electrostriction effect, i.e., the dielectric boundary is attracted closer to the simulation system when the total charge

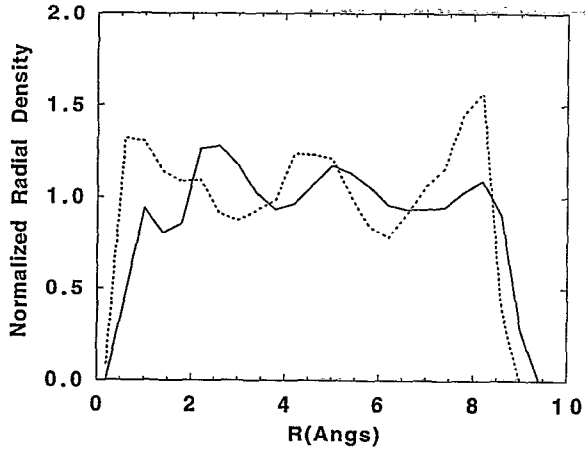

FIG. 4. Normalized radial distribution of the water oxygens from the center of a sphere containing 100 explicit TIP3P waters (Ref. 35). The density distribution function was divided by 0.0334/Å^3 for a better comparison with bulk water. The results with (solid line) and without (dotted line) the angular potential given by Eq. (32) are shown. In the absence of angular potential, the density has a maximum near the boundary of the system. In the presence of the angular potential, the density distribution function is more uniform throughout the simulation system.

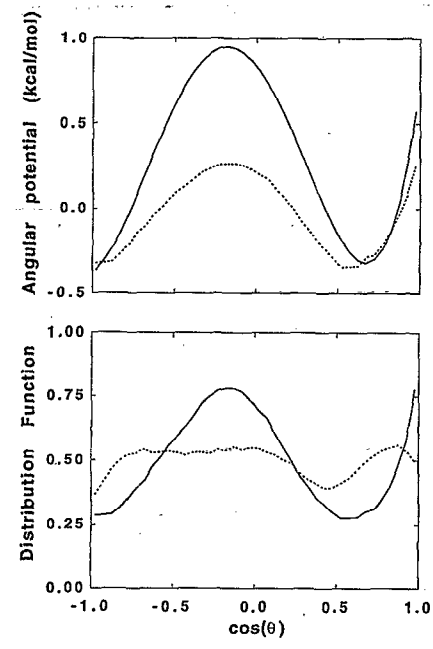

FIG. 5. Angular distribution function of the O–H bonds of the water molecules located in a 1 Å shell at the boundary of a sphere containing 100 explicit TIP3P waters (bottom) and angular potentials (top). The initial guess (dotted line), extracted from $\langle\rho(\cos)\rangle$ using Eq. (30), and the empirically fitted angular potentials (solid line) given by Eq. (32), are shown at the top. For clarity, the angular potentials are plotted with the switching function $\mathrm{SW}(r_O)$ equal to one. The distribution functions with (dotted line) and without (solid line) the angular potential are shown at the bottom. In the absence of the angular potential, the distribution is anisotropic; one O–H bond is pointing outward in average. In the presence of the angular potential, the resulting distribution is more isotropic, though some structure remains.

of the inner region is increased. The dielectric boundary is not affected by Eq. (29) for neutral solutes and is located at 2.8 Å from $R_{\max}$.

TEST SIMULATIONS WITH A WATER SPHERE

Preliminary simulations were performed with a sphere of 100 TIP3P water molecules[35] to test the solvent boundary potential described by Eqs. (13), (14), (20), and (26). The simulations were done with a modified version of the program CHARMM[42] incorporating the boundary potential. The configurational sampling was achieved with by generating five Langevin dynamics trajectories of 10 ps at 300 K with a friction constant corresponding to a relaxation time of 25 ps^{-1}. For each trajectory, the system was first heated to 400 K for 2 ps followed by 5 ps of equilibration at 300 K. The averages structural properties were calculated by combining the configurations of the five trajectories. For all the calculated trajectories, the bonds involving hydrogens atoms were kept fixed using SHAKE[43] and an integration step of 0.001 ps was used. The electrostatic reaction field was calculated using Eq. (26) with 16 moments of the charge distribution.

The normalized density distribution of water oxygens around the center of the simulation sphere is shown in Fig. 4. It is observed that the density distribution is close to the bulk value of 0.0334/Å^3 though it is slightly too structured near the edge of the simulation system. The normalized orientational distribution function of the O–H bonds of the water molecules relative to the radial director in a shell of 1 Å near the boundary shown in Fig. 5. The distribution function has a maximum near $\cos(\theta)=-0.2$, corresponding to an O–H bond lying parallel to the boundary, and a maximum near $\cos(\theta)=1$, corresponding to an O–H bond pointing away from the bulk. The anisotropy is due to the approximations involved in constructing the solvent boundary potential. Because the approximations to $\Delta\mathscr{W}_{\mathrm{vdw}}$ and $\Delta\mathscr{W}_{\mathrm{elec}}$ were developed based on the assumption that the simulation system is on average spherical, the present implementation of the boundary potential has the characteristics of a smooth restraining boundary. It is well known that the water molecules near smooth spherical and planar boundaries do not have an isotropic orientational distribution.[24,44,45] To correct the spurious anisotropy introduced by the average spherical approximation, an empirical angular potential acting on the orientation of the waters located near the boundary was introduced. As a first guess, the angular potential necessary to compensate the edge effect in the system was constructed based on the the average orientational distribution function of the O–H bonds of the water molecules in a shell of 1 Å near the boundary,

$$f_i(\cos_i) \approx +k_B T \ln[\langle\rho(\cos)\rangle/\rho_{\mathrm{iso}}]\times \mathrm{SW}(r_O), \qquad (30)$$

where ρ_{iso} is the value of the distribution function in an isotropic system (equal to 0.5) and $\mathrm{SW}(r_O)$ is a half-harmonic switching function,

$$\mathrm{SW}(r_O)=\begin{cases}(r_O-R_{\max}+1.0)^2 & \text{if } r_O>R_{\max}-1.0\\ 0 \quad \text{otherwise}\end{cases}, \qquad (31)$$

where r_O is the distance of the water oxygen from the center of the sphere (in Å). The initial guess, extracted from $\langle\rho(\cos)\rangle$ using Eq. (30) is shown in Fig. 5. The initial form of the angular potential was further refined empirically by trial

and error, using several simulations of a sphere of 100 water molecules. The resulting angular potential was represented in terms of a polynomial,

$$f(\cos_i)=\mathrm{SW}(r_O)(c_1\cos_i^4+c_2\cos_i^3+c_3\cos_i^2+c_4\cos_i+c_5) \quad (32)$$

with $c_1=2.409$, $c_2=1.767$, $c_3=-3.067$, $c_4=-1.201$, $c_5=0.841$. The empirically fitted angular potential is shown in Fig. 5. It can be observed that the fitted potential is similar to the initial guess bu differ in amplitude. This is due to the switching function Eq. (31) which is less than one in the 1 Å shell at the boundary of the system.

The structural properties of the pure water, recalculated from simulations including the angular correction Eq. (32) in the boundary potential, are shown in Figs. 4 and 5. The average structural properties were calculated by combining the configurations of five trajectories of 10 ps at 300 K generated with the same procedure as described above. In the presence of the angular potential, it observed that the density distribution is improved towards the normal value of 0.0334/Å^3. The orientation distribution function of the O–H bonds is also more isotropic, though some structure remains near $\cos(\theta)=-1$ and $\cos(\theta)=0.5$. Attempts to further refine the angular potential were unsuccessful.

APPLICATION OF THE METHOD

To illustrate the present approach, the solvation properties of charged and polar solutes and the intramolecular conformational equilibrium of flexible molecular solute in bulk water were examined. The absolute free energy of solvation of a single solute–water molecule and single sodium and potassium ions was calculated with free energy perturbation techniques.[5,33] The potential of mean force (PMF) along the dihedral angle of n-butane and the relative free energy of the C_{7ax} and α_L conformations of the alanine dipeptide were calculated using umbrella sampling techniques.[46]

The purpose of these calculations is primarily to examine the performance of the boundary potential developed for bulk water simulations. The ability of the solvent boundary potential to yield accurate estimates of the absolute free energy of solvation of charged and polar solutes with a small number of explicit water molecules is very important and must be verified. The absolute solvation free energy of charged and polar solutes is difficult to calculate accurately with simulation based on conventional periodic boundary conditions.[5] In addition, the ability of the boundary potential to reproduce the influence of bulk solvation on the intramolecular conformational equilibrium of flexible molecular solutes must be examined. The n-butane and the alanine dipeptide were chosen because they are well-known systems for which similar calculations have been reported in the literature. The solvent-induced potential of mean force around the dihedral angle of n-butane has been used to study the hydrophobic effect,[47–49] and the alanine dipeptide molecule has been used to study the influence of solvation on the conformational equilibrium of polypeptides and proteins.[17,50]

In all cases, the configurational sampling was performed by generating Langevin dynamics trajectories with a friction constant corresponding to a relaxation time of 5 ps^{-1} applied to the oxygens; an integration time step of 1 ps was used and the bonds involving hydrogen atoms were kept fixed using SHAKE.[43] The center of mass of the solute was constrained at the center of the simulation system. The solvent boundary potential including the angular potential correction was used. The electrostatic reaction field was calculated with 16 multipoles. The Lennard-Jones parameters used in calculations are given in Table I. More specific details about the simulation are given below. The performance of the solvent boundary potential was tested by considering a number of explicit water molecules varying from 4 to 100.

TABLE I. Lennard-Jones parameters.

Particles	Parameters	
	ϵ (kcal/mol)	R_{min} (Å)
K^+	−0.0870	3.5275
Na^+	−0.0469	2.7275
O^a	−0.1521	3.5365

[a]Parameters corresponding to the TIP3P water model (Ref. 35).

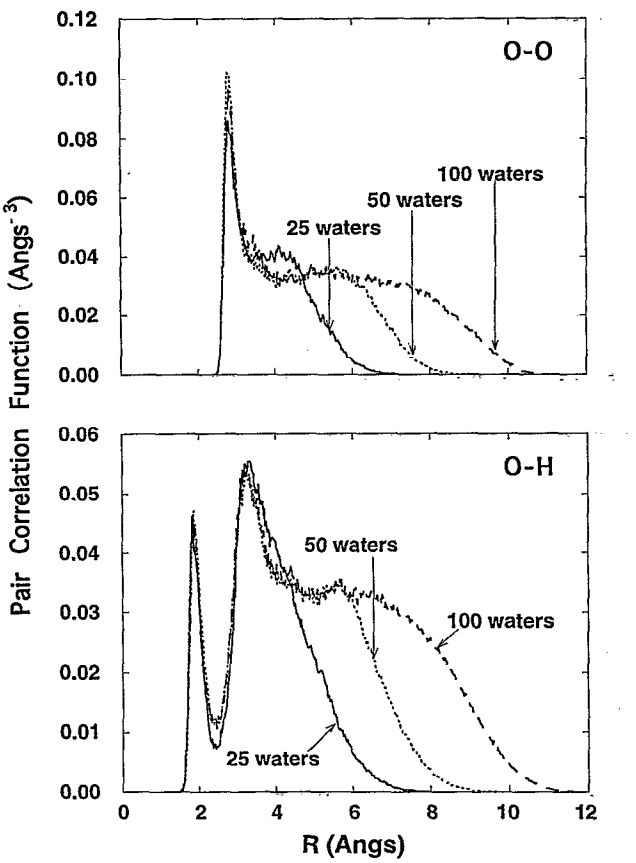

FIG. 6. Solute water molecule. Radial distribution function of the solvent water oxygen (top) and of the solvent water hydrogen (bottom) around the oxygen of one solute water molecule kept in the center of a sphere of 25, 50, and 100 explicit waters. The average was calculated from a trajectory of Langevin dynamics of 50 ps.

TABLE II. Solvation free energy of a solute water molecule (kcal/mol).

Systems[a]	Water molecule		
	Charging	Cavity	Total
4	−6.6	3.9	−2.7
6	−6.5	3.2	−3.3
25	−8.7	2.7	−6.0
50	−8.8	3.0	−5.8
100	−8.8	2.4	−6.4

[a]Number of explicit waters around the central solute water.

Solvation of one water molecule

The solvation structure around a solute water molecule constrained at the center of a water sphere was calculated for simulation systems including different numbers of explicit solvent molecules. Simulation systems containing 25, 50, and 100 explicit solvent waters were examined. For each case, the averages were calculated from a Langevin trajectory of 50 ps. The oxygen–oxygen and oxygen–hydrogen radial pair distribution functions are shown in Fig. 6. It is observed that the solvent structure around the solute water molecule is well reproduced by the solvent boundary potential when 50 or 100 waters are included explicitly, though the edge effects are stronger with 25 waters. The positions of the strong first and weak second peaks in the oxygen-oxygen radial distribution function (near 3.0 and 5.5 Å) are in good agreement with previous simulations of TIP3P water.[35] The positions and amplitude of the first and second peaks of the oxygen–hydrogen radial distribution function (near 1.9 and 3.3 Å) are also in good agreement with previous simulations.[35]

The free energy of solvation of a solute water molecule was calculated using standard free energy perturbation techniques.[5,33] The total free energy of solvation of the solute was calculated as a sum of two contributions: the free energy to insert a neutral solute in the solvent ("cavity formation") and the free energy to switch-on the charge of the solute ("charging"). The free energy for cavity formation was calculated using ten windows with the scaled Lennard-Jones parameters $\lambda_i\sigma_{\text{ion}}$ and $\lambda_i\epsilon_{\text{ion}}$. Langevin dynamic trajectories of 10 ps each were generated for each window, varying λ_i from 0.05 to 0.95 with a step increment of 0.1. Each trajectory was started from the last configuration of the previous window and equilibrated during 10 ps. Free energy perturbations were calculated for values of $\lambda_i \pm 0.05$ according to Eq. (7). An identical procedure was used to calculate the charging free energy. The results of the calculations for the solvation free energy of a solute water molecule are given in Table II. The results obtained with 25 to 100 waters are very similar. The total solvation free energy is around −6 kcal/mol, with contributions of +2.8 kcal/mol from the cavity formation and −8.8 kcal/mol from charging. The free energy calculated with four and six waters is underestimated due to a combination of errors. The charging free energy is smaller (−6.6 and −6.5 kcal/mol) and the cavity formation is larger (+3.9 and +3.2 kcal/mol). As observed in other calculations, the relative error is dominated by the free energy for cavity formation.[33] Based on separate calculations done with different starting conditions, the uncertainty is estimated to be on the order of 0.5 kcal/mol.

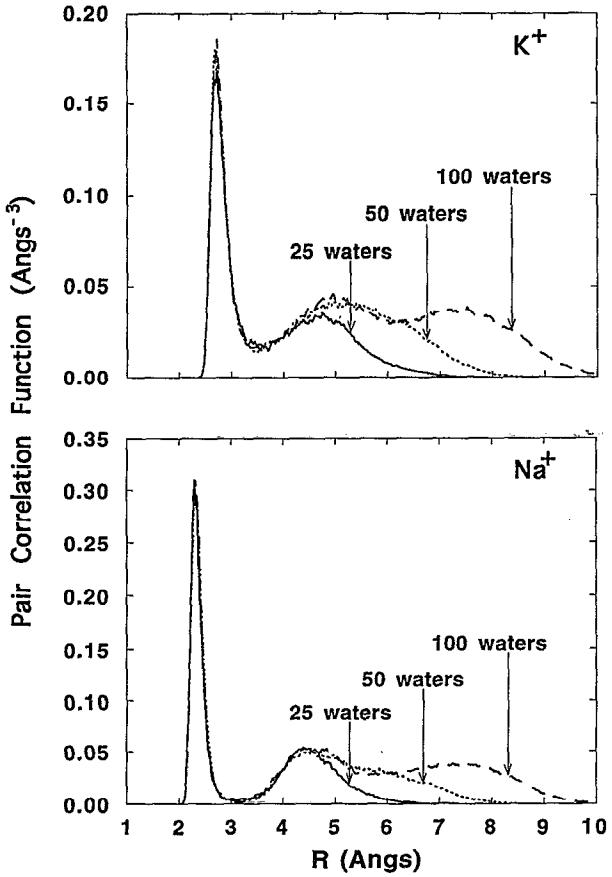

FIG. 7. Radial distribution function of the solvent water oxygen around one solute K^+ (top) and Na^+ (bottom) ion constrained in the center of a sphere of 25, 50, and 100 explicit waters. The average was calculated from trajectories of Langevin dynamics of 50 ps.

Properties of solvated ions

The solvation structure around one ion constrained at the center of a water sphere was calculated for simulation systems including 25, 50, and 100 explicit waters. For both ions, the averages were calculated from a single 50 ps Langevin trajectory. The ion–water oxygen radial pair correlation functions for Na^+ and K^+ are shown in Fig. 7. It is observed that the first peak, representing the primary hydration shell, is well reproduced in the three simulation systems. The second hydration shell is reproduced with 50 and 100 explicit waters. The number of waters is not sufficient to provide a second hydration shell in the system with 25 waters. The calculated free energy of solvation and the relative free energy difference are given in Table III for Na^+ and K^+. The free energy simulation procedure was similar to that used in the water solute calculations (see above). The relative free energy of Na^+ and K^+ was calculated using Eq. (7) with linearly interpolated Lennard-Jones parameters $0.5(\sigma_{\text{Na}}+\sigma_{\text{K}})$ and $0.5(\epsilon_{\text{Na}}+\epsilon_{\text{K}})$ from 50 ps Langevin dynamics trajectories of the ions.

TABLE III. Solvation free energy of ions (kcal/mol).

	K^+			Na^+			Relative	
Systems[a]	Charging	Cavity	Total	Charging	Cavity	Total	Difference[c]	Perturbation[d]
expt[b]			−80.6			−98.2	−17.6	
6				−106.2	2.2	−104.0		
8	−82.0	3.3	−78.7					
25	−82.3	3.2	−79.1	−103.5	2.8	−100.7	−21.6	−21.5
50	−83.9	3.5	−80.4	−103.4	2.6	−100.8	−20.4	−21.5
100	−84.3	2.8	−81.5	−105.1	2.5	−102.6	−21.1	−20.5

[a]Number of explicit water molecules around the ions.
[b]Experimental chemical potentials taken from Ref. 51.
[c]Calculated directly from the difference between the absolute free energy (column 3 and 6).
[d]Calculated from a free energy perturbation approach.

The free energy of Na^+ and K^+ calculated with different numbers of water molecules are very similar. To provide more realistic models, the Lennard-Jones parameters used for the ions were adjusted to reproduce the experimentally determined chemical potentials with the free energy calculations.[51] The charging free energy of K^+ is around −83.5 kcal/mol and the free energy for cavity formation is +3.2 kcal/mol (averaged over the 25, 50, and 100 systems). Due to the smaller radius of Na^+, the charging free energy is larger (−104 kcal/mol) and the free energy for cavity formation is slightly smaller (+2.6 kcal/mol). The relative free energy difference of K^+ to Na^+, calculated with the free energy perturbation is consistent with the difference in absolute solvation free energy. The free energy of the ions calculated with only the first hydration shell (six waters for Na^+ and eight waters for K^+) is very similar to the values obtained with the larger systems. The relation Eq. (29), determining the location of the dielectric boundary ΔR_{diel}, was adjusted empirically to yield the good agreement. In the systems with 25–100 waters, the value of ΔR_{diel} is nearly constant and around 2.7–2.8 Å. In the systems including only the first shell, the value of ΔR_{diel} is 2.33 Å for Na^+ and 2.47 Å for K^+. Because the free energy calculations with these small systems are computationally inexpensive, this approximation useful in the early stage of parametrization of new ion models.

Conformational equilibrium

The potential of mean force around the dihedral angle of *n*-butane was calculated using the umbrella sampling technique.[46] The butane was modeled in the extended atom representation, in which the nonpolar hydrogens are included in the carbon to which they are attached. The parameters of the butane model were taken from previous studies.[49] A harmonic window potential $V^i_{\text{umb}} = 0.5k(\phi - \phi_i)^2$ with k=4.0 kcal/(mol rad^2) was used. The configurational sampling was performed from Langevin dynamics trajectories of 100 ps for each window successively centered at −270°,−240°,...,90°. Each window was equilibrated during 20 ps starting from the last configuration of the previous window. The intrinsic torsional potential of butane was removed to obtain the solvent-induced contribution. The histograms were unbiased and combined using a statistical weighting method.[52] The symmetry of the PMF around 0° and 180° was used to improve the statistical convergence of the calculation. The PMF was calculated with 25 and 100 explicit water molecules. The results are shown in Fig. 8. In both the cases, solvation favors the *cis* conformation relative to the *trans* conformation by approximately 0.5 kcal/mol. It is remarkable that the correct free energy difference between the *cis* and *trans* conformers is reproduced even when only 25 water molecules are included explicitly. As shown in Fig. 9, this number of waters is barely sufficient to provide a first hydration shell around the *n*-butane molecule.

The relative free energy of the C_{7ax} (ϕ=60, ψ=−60) and α_L (ϕ=60, ψ=60) conformations of the alanine dipeptide was calculated from the PMF along the dihedral ψ with a constant value of 60° for the dihedral ϕ. The simulation procedure was similar to that used in the *n*-butane umbrella sampling calculations (described above). The potential energy function of the system was biased by the torsional umbrella potential $V^n_{\text{umb}} = 0.5k(\psi - \psi_n)^2$ with k=4.0

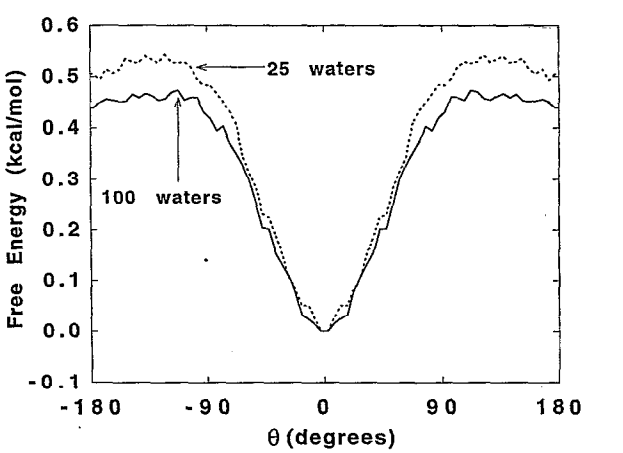

FIG. 8. Solvent-induced *n*-butane dihedral PMF calculated by umbrella sampling in a cluster of 100 (solid line) and 25 (dotted line) explicit water molecules. The calculation was performed in the absence of the intramolecular dihedral potential of *n*-butane for the sake of clarity. The PMF was extracted from 19 windows histograms calculated from 100 ps trajectory of Langevin dynamics in the presence of a dihedral harmonic potential of 2 kcal/(mol rad^2). The histograms were binned with 5 degrees intervales and were symmetrized with respect to ϕ=0 and ϕ=180 to increase the statistical convergence.

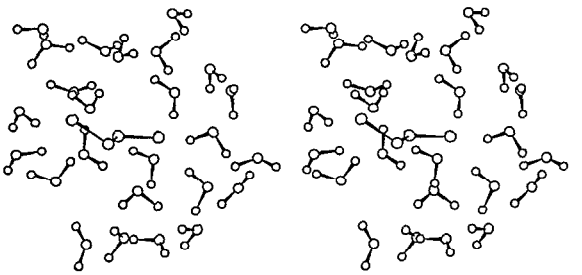

FIG. 9. Stereo representation of the system with n-butane and 25 waters. The value of R_{max} is 5.5 Å for this configuration.

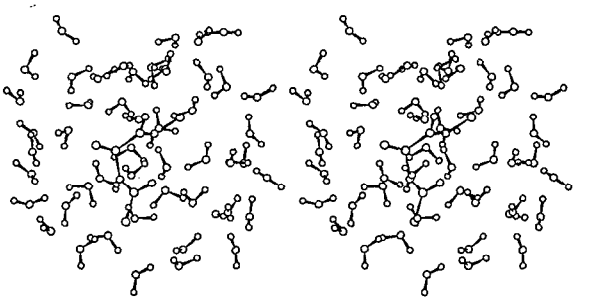

FIG. 11. Stereo representation of the system with the alanine dipeptide and 50 waters. The value of R_{max} is 7.2 Å for this configuration.

kcal/(mol rad^2); the value of the dihedral ϕ was restrained around 60° with a strong harmonic potential. Successive windows with $\psi_i = -100°$, $-80°$,..., 80° were generated. The polar hydrogen potential function of CHARMM[42] was used to model the alanine dipeptide (see also Ref. 50). The results of the calculations in the system with 50 and 100 waters are shown in Fig. 10. For comparison, the PMF calculated in vacuum is also shown. In both cases, the free energy difference between the α_L and C_{7ax} conformations is only 0.1 kcal/mol, while it is 8 kcal/mol in vacuum. As in the case on n-butane, it is remarkable that the PMF is reproduced accurately, even when only 50 water molecules are included explicitly. It can be seen in Fig. 11 that this number of waters corresponds to a first hydration shell around the alanine dipeptide. The PMF determined by Tobias and Brooks has been replotted (Fig. 5 from Ref. 50). Their result was obtained from molecular dynamics simulations of the alanine dipeptide molecule surrounded by 202–207 explicit water molecules treated with conventional periodic boundary conditions; a method based on holonomic internal coordinate constraints with free energy perturbation was used. The same potential function was used. Despite the differences in simulation systems and in free energy methodologies, the results obtained with the boundary potential are in good agreement with their calculations. The largest discrepancy is at the barrier between the C_{7ax} and α_L conformations. In the present calculations it is lower by about 1.0 kcal/mol relative to their calculations.

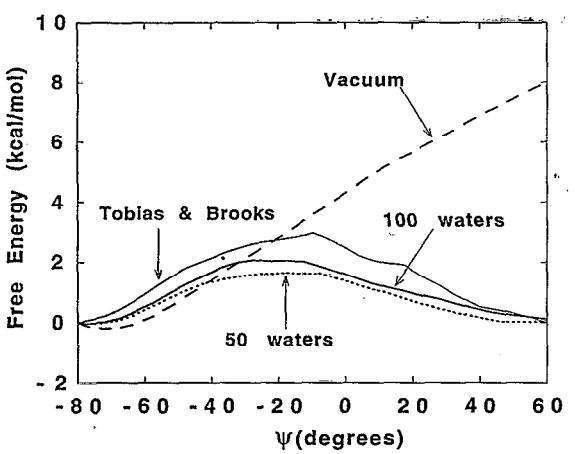

FIG. 10. Alanine dipeptide relative free energy difference for conformational transition C_{7ax} to α_L in a gas phase (vacuum) and in presence of 50 and 100 explicit water molecules. The histograms were binned with 5 degrees intervales. The result of the study by Tobias and Brooks (Ref. 50) is also shown for comparison.

CONCLUDING DISCUSSION

An approach to approximate the effect of bulk solvation from molecular dynamics simulations of a finite cluster has been described. The approximation follows from a formal separation of the multidimensional configurational integral in terms of n inner solvent molecules nearest to the solute and the remaining outer bulk solvent molecules. Using this formulation, it was shown that averages can be expressed in terms of a finite cluster with one arbitrary solute and n explicit solvent molecules under the influence of a solvent boundary potential. The solvent boundary potential corresponds to the solvation free energy of an effective object composed of one hard sphere of radius R_{max}, one solute and n solvent molecules in a frozen configuration. This interpretation is helpful in constructing a boundary potential to approximate the influence of the surrounding bulk solvent.

Following the theoretical formulation, an approximate effective potential of mean force was constructed to reproduce the main feature of solvation in bulk water. The water boundary potential is practically realized as a part of a modified version of the CHARMM program[42] but could be easily implemented in other biomolecular simulation programs. The water boundary potential was adjusted empirically for simulations of the TIP3P model,[35] although it might also be appropriate for other similar water models such as SPC.[36] In the present implementation, many features of the boundary potential are closely related to others developed previously. For example, the van der Waals contribution, $\Delta\mathscr{W}_{vdw}$ is very similar to the MFFA potential calculated by Brooks and Karplus.[21] Similarly, the position of the dielectric interface used in the calculation of the Kirkwood reaction field calculation is analogous to the exclusion radius of Rullmann and van Duijnen.[26] One important difference here is that the present boundary potential was derived from the exact expression given by Eq. (12). We consider this equation a main result of this article. The approximate boundary potential developed here is not unique and could be improved further. For example, improvements of the angular correction poten-

tial are possible. Several difficulties may be resolved by going beyond the simplifying assumption that the boundary is on average spherical.

The boundary potential was tested with standard systems. The solvation free energy of a water molecule and of sodium and potassium ions was calculated. The influence of bulk solvent on the conformational equilibrium of molecular solutes is illustrated by performing umbrella sampling calculations of n-butane and alanine dipeptide in water. The boundary potential was tested to examine the dependence of the results on the number of water molecules included explicitly in the simulations. It was observed that bulklike results were obtained, even when only the waters of the first hydration shell were included. The boundary potential is flexible for any number of solvent molecule such that the density corresponding to a constant pressure is adjusted automatically. This can be useful in alchemical free energy calculations in which the difference in molecular size between two solutes is significant.[28] The formal separation of the configurational integral was developed for the case of a spherical inner region. The formalism can also be developed for other geometries, e.g., extensions to a planar boundary should be straightforward. However, it may be more difficult to find satisfactory approximations in the case of a nonspherical inner region such as ellipsoids. The solvent boundary potential approach may also be useful to provide a primary hydration around a small subset of atoms described with *ab initio* quantum chemical potentials.

APPENDIX

In this Appendix, the formal equivalence of Eqs. (1) and (4) is demonstrated in the case of an ideal monoatomic gas of N particles in a volume V. The classical ideal gas configuration integral, expressed as in Eq. (4) in terms of n inner particles and $(N-n)$ outer particles, is,

$$Z_{\text{ideal}}(N,V)=\frac{1}{n!}\int d\mathbf{r}_1\int d\mathbf{r}_2\cdots d\mathbf{r}_n\,\frac{(V-V_R)^{(N-n)}}{(N-n)!}, \tag{A1}$$

where the outer particles have been integrated over the full space except the volume V_R, occupied by the sphere of radius R defined by the distance of the farthest of the inner particle with $R=\text{MAX}\,(r_1,\ldots,r_2)$.

To solve the multidimensional integral, the inner particles are renumbered from 1 to n, starting from the center of the system. In the ordered integral, the $n!$ is removed. Assuming that the ideal gas in contained in a sphere of radius R_V and volume V, the partition function is expressed in spherical coordinates,

$$Z_{\text{ideal}}(N,V)=\int_0^{R_V}4\pi r_1^2\,dr_1\int_{r_1}^{R_V}4\pi r_2^2\,dr_2\cdots \times\int_{r_{n-1}}^{R_V}4\pi r_n^2\,dr_n\,\frac{(V-v_n)^{(N-n)}}{(N-n)!}, \tag{A2}$$

making the change of variables, $v_i=4\pi r_i^3/3$, the integral is

$$Z_{\text{ideal}}(N,V)=\int_0^V dv_1\int_{v_1}^V dv_2\cdots\int_{v_{n-1}}^V dv_n\,\frac{(V-v_n)^{(N-n)}}{(N-n)!}. \tag{A3}$$

Performing the integral over the nth inner particle,

$$Z_{\text{ideal}}(N,V)=\int_0^V dv_1\int_{v_1}^V dv_2\cdots\int_{v_{n-2}}^V dv_{n-1} \times\frac{(V-v_{n-1})^{(N-n+1)}}{(N-n+1)!}, \tag{A4}$$

yields a similar expression for $(n-1)$ inner particles and $(N-n+1)$ outer particles. Repeating the process n times yields the familiar $V^N/N!$ of the ideal gas.

[1] C. L. Brooks III, Martin Karplus, and B. M. Pettitt, in *Advances in Chemical Physics*, Vol. LXXI, edited by I. Prigogine and S. A. Rice (Wiley, New York, 1988).
[2] J. Tirado-Rives and W. L. Jorgensen, Biochem. **30**, 3864 (1991).
[3] D. J. Tobias and C. L. Brooks III, Biochem. **30**, 6059 (1991).
[4] A. R. Van Buuren and H. C. Berendsen, Biopol. **33**, 1159 (1993).
[5] T. P. Straatsma, H. J. C. Berendsen, and J. P. M. Postma, J. Chem. Phys. **89**, 5876 (1988).
[6] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford Science Publications (Clarendon, Oxford, 1989).
[7] J. S. Bader and D. Chandler, J. Phys. Chem. **96**, 6423 (1992).
[8] J. P. Valleau and A. A. Gardner, J. Chem. Phys. **86**, 4162 (1987).
[9] D. Eisenberg and A. McLachlan, Nature (London) **319**, 199 (1986).
[10] A. A. Lipanov, D. B. Beglov, and V. P. Chuprina, J. Mol. Biol. **210**, 399 (1989).
[11] Y. K. Kang, G. Nemethy, and H. A. Scheraga, J. Phys. Chem. **91**, 4105 (1987).
[12] F. Colonna-Cesari and C. Sander, Biophys. J. **57**, 1103 (1990).
[13] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, J. Am. Chem. Soc. **112**, 6127 (1990).
[14] K. Sharp, J. Comput. Chem. **12**, 454 (1990).
[15] H. Sklenar, F. Eisenhaber, M. Poncin, and R. Lavery, in *Theoretical Biochemistry and Molecular Biophysics*, edited by D. L. Beveridge and R. Laver (Adenine, Schenectady, NY, 1990).
[16] R. J. Zauhar and S. Morgan, J. Mol. Biol. **186**, 815 (1985).
[17] M. B. Pettitt and M. Karplus, Chem. Phys. Lett. **121**, 194 (1985).
[18] A. Jean-Charles, A. Nicholls, K. Sharp, B. Honig, A. Tempczyk, T. Hendrickson, and W. C. Still, J. Am. Chem. Soc. **113**, 1454 (1991).
[19] A. J. Stace and J. N. Murrel, Mol. Phys. **33**, 1 (1977).
[20] M. Berkowitz and J. A. Mc.Cammon, Chem. Phys. Lett. **90**, 215 (1982).
[21] C. L. Brooks III and M. Karplus, J. Chem. Phys. **79**, 6312 (1983).
[22] A. Brunger, C. L. Brooks III, and M. Karplus, Chem. Phys. Lett. **105**, 495 (1984).
[23] A. Warshel and G. King, Chem. Phys. Lett. **121**, 124 (1985).
[24] G. King and A. Warshel, J. Chem. Phys. **91**, 3647 (1989).
[25] J. A. C. Rullmann, Ph.D. thesis, University of Utrecht, The Netherlands (1988).
[26] J. A. Rullmann and P. Th. van Duijnen, Mol. Phys. **61**, 293 (1987).
[27] J. G. Kirkwood, Chem. Rev. **19**, 275 (1936).
[28] M. Prevost, S. J. Wodak, B. Tidor, and M. Karplus, Proc. Natl. Acad. Sci. USA **88**, 10880 (1991).
[29] H. L. Friedman, Mol. Phys. **29**, 1533 (1975).
[30] D. A. McQuarrie, *Statistical Mechanics* (Harper and Row, New York, 1976).
[31] J. P. Hansen and I. R. McDonald, *Theory of Simple Liquids* (Academic, London, 1976).
[32] T. Çağin and B. M. Pettitt, Mol. Phys. **72**, 169 (1991).
[33] T. P. Straatsma, H. J. C. Berendsen, and J. P. M. Postma, J. Chem. Phys. **85**, 6720 (1986).
[34] R. W. Zwanzig, J. Chem. Phys. **22**, 1420 (1954).
[35] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[36] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, in *Intermolecular Forces*, edited by B. Pullman (Reidel, Dordrecht, 1981).
[37] D. Chandler and H. C. Andersen, J. Chem. Phys. **57**, 1930 (1972).

[38] F. Hirata and P. J. Rossky, Chem. Phys. Lett. **83**, 329 (1981).
[39] B. M. Pettitt and P. J. Rossky, J. Chem. Phys. **84**, 5836 (1986).
[40] K. A. Sharp and B. Honig, Annu. Rev. Biophys. Chem. **19**, 301 (1990).
[41] J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1962).
[42] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).
[43] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, J. Comput. Chem. **23**, 327 (1977).
[44] A. C. Belch and M. Berkowitz, Chem. Phys. Lett. **113**, 278 (1985).
[45] C. Y. Lee, J. A. McCammon, and P. J. Rossky, J. Chem. Phys. **80**, 4448 (1984).
[46] G. M. Torrie and J. P. Valleau, Chem Phys. Lett. **28**, 578 (1974).
[47] W. L. Jorgensen, J. Chem. Phys. **77**, 5757 (1982).
[48] W. L. Jorgensen and J. K. Buckner, J. Phys. Chem. **91**, 6083 (1987).
[49] D. J. Tobias and C. L. Brooks III, J. Chem. Phys. **92**, 2582 (1990).
[50] D. J. Tobias and C. L. Brooks III, J. Phys. Chem. **96**, 3864 (1992).
[51] M. A. Burgess, *Metal Ions in Solution* (Horwood, Chichester, 1978).
[52] T. B. Woolf and B. Roux, J. Am. Chem. Soc. (in press).