ELSEVIER

Biophysical Chemistry 78 (1999) 97–111

Biophysical Chemistry

Review

Integral equation theories for predicting water structure around molecules

Yi Liu[1], Toshiko Ichiye*

Departments of Biochemistry / Biophysics and of Chemistry, Washington State University, Pullman, WA 99163-4660, USA

Received 28 July 1998; received in revised form 18 January 1999; accepted 18 January 1999

Abstract

Water plays a crucial role in the structure and function of proteins and other biological macromolecules; thus, theories of aqueous solvation for these molecules are of great importance. However, water is a complex solvent whose properties are still not completely understood. Statistical mechanical integral equation theories predict the density distribution of water molecules around a solute so that all particles are fully represented and thus potentially both molecular and macroscopic properties are included. Here we discuss how several theoretical tools we have developed have been integrated into an integral equation theory designed for globular macromolecular solutes such as proteins. Our approach predicts the three-dimensional spatial and orientational distribution of water molecules around a solute. Beginning with a three-dimensional Ornstein–Zernike equation, a separation is made between a reference part dependent only on the spatial distribution of solvent and a perturbation part dependent also on the orientational distribution of solvent. The spatial part is treated at a molecular level by a modified hypernetted chain closure whereas the orientational part is treated as a Boltzmann prefactor using a quasi-continuum theory we developed for solvation of simple ions. A potential energy function for water molecules is also needed and the sticky dipole models of water, such as our recently developed soft-sticky dipole (SSD) model, are ideal for the proposed separation. Moreover, SSD water is as good as or better than three point models typically used for simulations of biological macromolecules in structural, dielectric and dynamics properties and yet is seven times faster in Monte Carlo and four times faster in molecular dynamics simulations. Since our integral equation theory accurately predicts results from Monte Carlo simulations for solvation of a variety of test cases from a single water or ion to ice-like clusters and ion pairs, the application of this theory to biological macromolecules is promising.

Keywords: Water structure; Statistical mechanical integral equation theory; Computer simulation; Protein solvation

---

* Corresponding author. Fax: +1-509-335-9688
E-mail: ichiye@wsu.edu

[1] Current address: Hewlett-Packard, Mail Stop 410, 1000 NE Circle Blvd., Corvallis, OR 97330, USA.

PII: S0301-4622(99)00008-3

1. Introduction

Water plays an important role in the structure and function of proteins [1]. However, water is a complex solvent whose properties are still not completely understood [2]. At a superficial level, the aqueous solvation of simple solutes can be described simply by the hydrophobic effect for non-polar solutes and by continuum electrostatics for ions. The difference in the solvation free energy of non-polar solutes in water relative to that in hydrocarbons

$$\Delta G(\mathrm{CH_4 \text{ in } H_2O \rightarrow CH_4 \text{ in H.C.}}) = -12\ \mathrm{kcal/mol}$$

is often given as a demonstration of the hydrophobic effect. The solvation free energy of simple ions is predicted with remarkable accuracy by the Born solvation free energy equation, given by

$$\Delta G_{\mathrm{Born}} = -\frac{q^2}{2R}\left(1-\frac{1}{\varepsilon}\right)\rho^2 g(\mathbf{r}_1,\mathbf{r}_2) = N^2\frac{\int \mathrm{d}\mathbf{r}_2 \cdots \mathrm{d}\mathbf{r}_N \exp(-\beta V)}{\int \mathrm{d}\mathbf{r}_1 \cdots \mathrm{d}\mathbf{r}_N \exp(-\beta V)}$$

where q and R are the charge and radius, respectively, of the ion and ε is the dielectric constant of the solvent, and yet it is based on considering water as a dielectric continuum. However, molecular effects, especially hydrogen bonding, are also clearly important. For instance, perhaps the first molecular interpretation of the hydrophobic effect is the formation of enhanced hydrogen-bonded structures of water around non-polar solutes [2], although recent computer simulations give a mixed picture, with some supporting this picture and others supporting the picture that the transfer into water involves an enthalpy gain due enhanced solute–solvent van der Waals interactions and an entropy loss is due to the exclusion of waters from the region occupied by the solute, as discussed in Blokzijl and Engberts [3] and Durell and Wallqvist [4]. Also, hydrogen bonding to polar groups is very important in the solvation energetics of polar solutes. Moreover, the hydrogen bonding properties of water serve to differentiate solvation of negative and positive ions in that first shell waters are in a hydrogen bonding orientation to a negative ion whereas they are in a dipolar orientation to a positive ion [5].

The complexity of aqueous solvation of proteins arises in part because they are globular macromolecules so that neighboring atoms affect the solvation of a given atom. For instance, the water structure around a partially buried non-polar group at the surface of a protein cannot be described simply as a fraction of a clathrate; furthermore, the identity of the neighboring atoms will also influence the structure. In addition, the electrostatic solvation free energy of two atoms adjacent to each other, whether bonded or in van der Waals contact, is not simply the sum of the Born free energies of each atom. The complexity is even greater because proteins are composed not just of one type of atom but of non-polar, polar, and charged groups so that all of the effects mentioned above are important. Moreover, because of the inhomogeneous distribution of these different species, the interplay of these phenomenon at the molecular level is particularly difficult to deconvolute.

Theoretical studies of pure water and aqueous solutions by computer simulation techniques such as molecular dynamics (MD) and Monte Carlo (MC) [6] have provided much insight into their physical and chemical properties. However, an accurate and efficient representation of aqueous solvent in computer simulations of biological macromolecules remains a serious challenge because of the large size of biological macromolecules, so that even a few solvation shells require a very large number of water molecules. For instance, a 55-residue, 6-kDa protein solvated with approximately three shells of water (9 Å) to a box edge requires approximately 1200 water molecules in a rectangular octahedral box [7] or approximately 2800 in a rectangular box. In addition, there are problems associated with boundary conditions. For instance, a droplet boundary can lead to surface tension effects, a stochastic boundary can lead to solvent density anomalies at the boundary and a periodic boundary can lead to periodic artifacts such as standing waves. Furthermore, the use of boundary conditions complicates the calculation of solvation energetics, since the contribution of water beyond 9 Å from a singly charged ion is ~ 20 kcal/mol. Also, proteins and nucleic acids are highly charged, leading to problems in the long range electrostatics. A variety of methods including truncation, Ewald sums, and reaction fields [6] have been proposed, but there are still difficulties associated with each. Furthermore, simulating the salt strength remains a problem since fully equilibrating counterions is difficult [8].

Statistical mechanical integral equation theories, which have been very successful in understanding simple liquids [9], are an alternative approach to providing structural information at a molecular level. Since these theories predict the density distribution of water molecules around a solute so that all particles are fully represented and thus potentially both molecular and macroscopic properties are included. Thus, the problems associated with trying to combine continuum dielectric electrostatic potential calculations with solvation parameters, such as 'over-counting' interactions and the 'correct' decomposition of the interactions, are avoided. In fact, these methods, using the same atomistic potential energy functions as in MD or MC simulations, predict the same type of equilibrium information about the solvent that a MD or MC simulation would. Moreover, they have the added advantage that infinitely large systems can be studied.

This paper describes how several theoretical tools we have developed have been integrated into an integral equation theory that describes the three-dimensional structure of water around globular solutes. First, the theoretical tools are described in the context of the integral equation theory. The integral equation theory predicts the three-dimensional positional and orientational distribution of water molecules around the solute [10,11], and is based on a three-dimensional Ornstein–Zernike (OZ) equation [9] decomposed into a reference part depending only on the position (**r**) of the solvent molecule with respect to the solute and a perturbation part depending also on the orientation (Ω) of the solvent molecule with respect to the solute. The decomposition makes it possible to solve the Ω-independent part by a hypernetted chain based integral equation theory [9] and to treat the Ω-dependent part as a Boltzmann prefactor using a quasi-continuum theory developed by our group [12–14]. This decomposition works best with water models in which a water molecule is represented as a single point with a vector describing its orientation such as the Bratko–Blum–Luzar (BBL) water model [15], which has a hard-sphere, a point dipole and a tetrahedral 'sticky' potential that models the hydrogen bonding properties, and our new soft-sticky dipole (SSD) model of water [16], which is a soft-sphere version of BBL. Remarkably, the SSD model is as good as or better than the SPC [17], SPC/E [18] and TIP3P [19] three-point models typically used in simulations of biological macromolecules for structural [16], dielectric [20] and dynamical [21] properties and yet is about seven times faster in MC and approximately four times faster in MD simulations. After the description of the theoretical methods, the good agreement between our theory and MC simulations for the three-dimensional structure of water around frozen water clusters of various size using BBL water [10] and more recently, around water clusters, single ions and ion pairs using SSD water [11] is described. Finally, the efficiency of the method is discussed.

2. Theoretical methods

2.1. Ornstein–Zernike theories

Statistical mechanical integral equation theories have been a very successful means of studying liquid structure. The Ornstein–Zernike type integral equation theories [9] predict the radial distribution function, $g(r)$, which describes the number of particles at a given distance from a central particle. The radial distribution function is given by

$$\rho^2 g(\mathbf{r}_1,\mathbf{r}_2) = N^2 \frac{\int d\mathbf{r}_2 \cdots d\mathbf{r}_N \exp(-\beta V)}{\int d\mathbf{r}_1 \cdots d\mathbf{r}_N \exp(-\beta V)}$$

where ρ is the density, N is the number of particles, V is the total potential energy, $\beta = 1/k_BT$, k_B is Boltzmann constant and T is the absolute temperature. The radial distribution function can be measured experimentally from X-ray or neutron diffraction experiments and all thermodynamic properties of the system may be obtained from it.

The general form of the OZ integral equation for a simple liquid (i.e. for spherical particles) is given by

$$h(r_{12}) = c(r_{12}) + \int c(r_{13})\rho h(r_{32})\mathrm{d}\mathbf{r}_3 \tag{1}$$

where $h(r) = g(r) - 1$ is the pair correlation function, $c(r)$ is the direct correlation function and ρ is the density. The function $c(r)$ may be described as the direct interaction between the two particles in the field of the other particles. However, the OZ simply relates one unknown, h, to another unknown, c, so 'closure' relations are needed that define c as a function of h and v, the pair potential. Some common closures are the Percus–Yevick (PY), the hypernetted chain (HNC), and the mean spherical approximation (MSA). A general form for a closure is

$$g(r) = \exp[-\beta v(r) + h(r) - c(r) + b(r)] \tag{2}$$

where $b(r)$ is a function that, if equal to the correct bridge function, makes the theory exact whereas if approximated, serves to differentiate between approximate closures [9].

The extension of the OZ equation to molecular fluids has generally followed one of two approaches. One approach is the reference interaction site model (RISM) theory by Chandler and Anderson [22], which is for site models such as the various TIPS and SPC models. One problem encountered with this approach is that the integral equation theory is not exact. The extension to solvation around larger solutes includes studies of solvation of peptides [23,24] and neutral and charged structureless macroparticles [25]. The other approach is to use a three-dimensional OZ, which is suited for moment expansion models of solvent particles. The extension to solvation of larger solutes includes studies with the RHNC closure for the Kusalik–Patey [26] model of water of neutral and charged structureless macroparticles [25] and our studies of the Bratko–Blum–Luzar model of water [15] around ice-clusters [10]. Our approach predicts the three-dimensional positional and orientational distribution of water molecules around the solute, rather than the radial distributions with respect to various sites in the solute predicted by RISM theories.

2.2. Formulation of the integral equation theory for macromolecules

The goal in developing the integral theory described here was to balance accuracy and speed for practical use for macromolecules, in particular, for aqueous solvation of proteins. The integral equation theory has been described in detail elsewhere [10,11] so it is summarized here. The pair distribution function, $g(\mathbf{r}_{sw}\ \Omega_{sw})$, is used to describe the structure of water around the solute, where subscripts s and w denote solute and water, respectively, and $\mathbf{r}_{sw}$ and Ω_{sw} define the position and the orientation, respectively, of a water molecule relative to the solute. A closure is combined with the OZ equation to give

$$\begin{aligned} g(\mathbf{r}_{sw}\Omega_{sw}) = \exp\Big[& -\beta v(\mathbf{r}_{sw}\Omega_{sw}) + b(\mathbf{r}_{sw}\Omega_{sw}) \\ & + \frac{\rho}{8\pi^2}\int c(\mathbf{r}_{ww'}\Omega_{ww'}) \\ & \times h(\mathbf{r}_{sw'}\Omega_{sw'})\mathrm{d}\mathbf{r}_{w'}\mathrm{d}\Omega_{w'}\Big] \end{aligned} \tag{3}$$

The OZ equation written as in Eq. (3) does not involve the solute–water direct correlation function, which is important because the pure solvent direct correlation function $c(\mathbf{r}_{sw}\Omega_{sw})$ can be obtained a priori.

The crux of the present theory is a perturbation-type decomposition of the total distribution function, similar to the work by Kusalik [27] for dipolar soft-sphere fluids. The total solute–water interaction is separated into two parts

$$v(\mathbf{r}_{sw}\Omega_{sw}) = v^{o}(\mathbf{r}_{sw}) + v^{a}(\mathbf{r}_{sw}\Omega_{sw}) \tag{4}$$

where $v^{o}(\mathbf{r}_{sw})$ and $v^{a}(\mathbf{r}_{sw}\Omega_{sw})$ are the reference and perturbation parts of the total potential, respectively. Likewise, $b(\mathbf{r}_{sw}\Omega_{sw})$ and $c(\mathbf{r}_{sw}\Omega_{sw})$ are decomposed in a similar manner. Furthermore, we define

$$g(\mathbf{r}_{sw}\Omega_{sw}) = 8\pi^{2}P(\Omega_{sw};\mathbf{r}_{sw})g^{o}(\mathbf{r}_{sw}) \tag{5}$$

in which $g^{o}(\mathbf{r}_{sw})$ is the probability of finding a water at position $\mathbf{r}_{sw}$ independent of its orientation Ω_{sw} and $P(\Omega_{sw};\mathbf{r}_{sw})$ is the conditional probability that a solvent molecule at position $\mathbf{r}_{sw}$ has the orientation Ω_{sw}. The latter is approximated by

$$P(\Omega_{sw};\mathbf{r}_{sw}) = \frac{\exp[-\beta\omega(\mathbf{r}_{sw}\Omega_{sw})]}{8\pi^{2}k(\mathbf{r}_{sw})} \tag{6}$$

where $\omega(\mathbf{r}_{sw}\Omega_{sw})$ is a potential of mean force and

$$k(\mathbf{r}_{sw}) = \frac{1}{8\pi^{2}}\int \exp[-\beta\omega(\mathbf{r}_{sw}\Omega_{sw})]\mathrm{d}\Omega_{sw} \tag{7}$$

The function $g^{o}(\mathbf{r}_{sw})$ is approximated by

$$g^{o}(\mathbf{r}_{sw}) = k(\mathbf{r}_{sw})\exp\Big[-\beta v^{o}(\mathbf{r}_{sw}) + b^{o}(\mathbf{r}_{sw}) + \rho\int c^{o}(\mathbf{r}_{ww'})h^{o}(\mathbf{r}_{sw'})\mathrm{d}\mathbf{r}_{w'}\Big] \tag{8}$$

Therefore, the difficult solution to Eq. (3) is divided into two steps: the spatial part [Eq. (8)] and the orientational part [Eq. (6)].

2.3. The solution of the spatial part

The solution of the spatial part (Eq. (8)) is further simplified as follows. First, the water–water direct correlation function, $c^{o}(\mathbf{r}_{ww'})$ is approximated by the pure water isotropic function $c^{o}(\mathbf{r}_{ww'}) \approx c^{o}(r_{ww'})$, which can be obtained by solving the HNC–OZ iteratively [28–30] using a radial distribution function from simulation and a spherically-averaged potential. The $c^{o}(\mathrm{r}_{ww'})$ calculation is a one-time task. Second, the unknown bridge function $b^{o}(\mathbf{r}_{sw})$ is handled in two ways. For the BBL calculations, $b^{o}(\mathbf{r}_{sw}) = 0$ is assumed so that Eq. (8) reduces to the familiar HNC-type equation while for the SSD calculations, $b^{o}(\mathbf{r}_{sw})$ is approximated by the bridge function of a reference system (RHNC) [9]. For the latter, the $c^{o}(r_{ww'})$ function is used to calculate the bridge function of pure water $b^{o}(\mathbf{r}_{sw})$ from the simulation, which now serves as the reference for all interactions by assuming b is a function of a variable $r_{sw}/\sigma_{sw} = r_{ww'}/\sigma_{ww'}$ [11]. For a multi-site solute, the total bridge function is approximated by the sum over all sites and is anisotropic (**r**-dependent). Third, the three-dimensional, anisotropic convolution in Eq. (8) is approximated by $g^{o}(\mathbf{r}_{sw}) = g^{o}(r_{sw})|_{\theta,\phi}$, where $g^{o}(r_{sw})|_{\theta,\phi}$ is a separate, isotropic solution for each θ_{sw} and ϕ_{sw}; i.e. a set of one-dimensional solutions.

2.4. The solution of the orientational part

The solution of the orientational part [Eq. (6)] is also simplified as follows. The orientational potential force function $\omega(\mathbf{r}_{sw}\Omega_{sw})$ is approximated using a quasi-continuum theory developed by our group. This theory gives the orientation of water in the field of a charge, and is based on the interaction of a point charge and a single dipole (representing one water) in a cavity in the continuum, allowing for correlation between the dipole and its neighbors [12]. From this, the solvation energy of simple ions in a continuum solvent can be calculated, which approaches the Born solvation energy in the proper limits. The theory using the bulk value for the dielectric constant agrees MD simulation results for explicit ions in explicit water at large r_{sw} for both the orientation and the solvation energy [12,31]. Moreover, it has been useful in studying the origins of the Born radius [13], deviations from Marcus theory for electron transfer [32], and non-linear response in ionic solvation [14].

The previous work has shown that orientation of water for all shells around a singly-charged cation (Fig. 1a,b) [31] and beyond the first one shell around a singly-charged anion (Fig. 1c,d) [12,31] can be simply described as a dipole in a dielectric continuum with a dielectric constant close to the bulk value. Extending this to an arbitrary solute, the functional form for $\omega(\mathbf{r}_{sw}\Omega_{sw})$ is then quite simple

$$\omega(\mathbf{r}_{sw}\Omega_{sw}) = \frac{\mu E(\mathbf{r}_{sw}\Omega_{sw})}{\varepsilon'(\mathbf{r}_{sw})} \tag{9}$$

where μ is the dipole moment of water, $E(\mathbf{r}_{sw}\Omega_{sw})$ is the field due to the solute in a vacuum and $\varepsilon'(r_{sw})$ is a screening factor. The factor $\varepsilon'(r_{sw})$ is small at small r_{sw} but approaches a value that is a function of the bulk dielectric constant at large r_{sw}. An empirical functional form for $\varepsilon'(r_{sw})$ was constructed for both BBL [10] and SSD [11] water. For a single site solute, the theory is used directly but for a multi-site solute, the orientational function is approximated by an infinite dilution superposition, i.e. the dipole orienting electrostatic field due to each individual site ignoring the others is calculated and then summed. The functional form for $\omega(\mathbf{r}_{sw}\Omega_{sw})$ should be quite accurate for positive charges and beyond the first shell around negative charges, but somewhat less accurate for the first shell around negative charges because of hydrogen

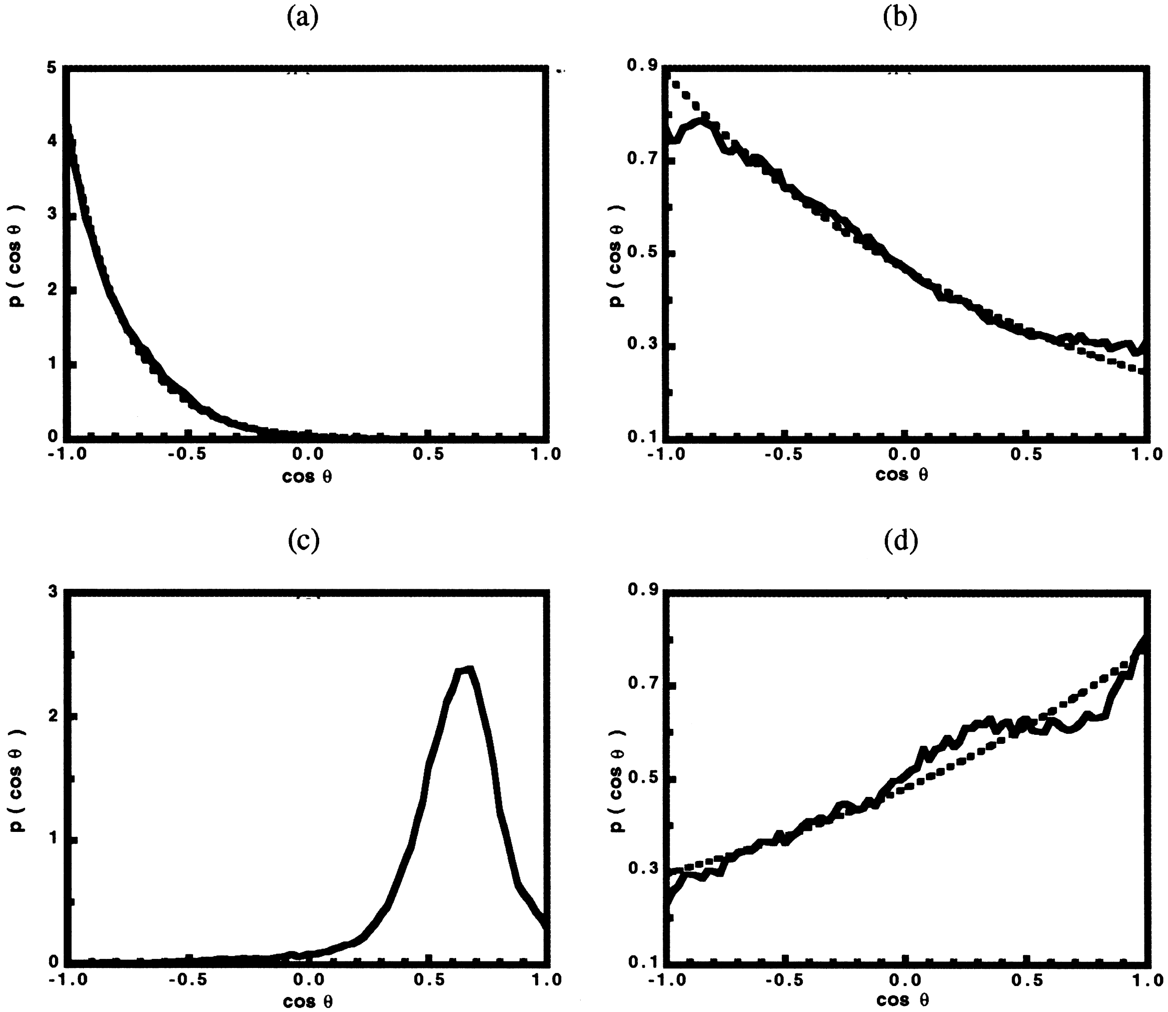


Fig. 1. $P(\cos\theta;r)$ as a function of cos θ for around ions in TIP3P water from MD simulations: (a) the first shell around Na^+, (b) the second shell around Na^+, (a) the first shell around Cl^-, (b) the second shell around Cl^-.

bonding. The latter could be improved by using a different function for the first shell.

2.5. The water models

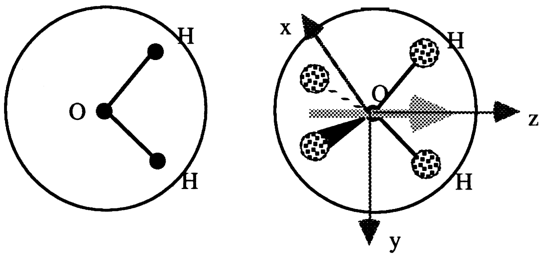
The integral equation theory requires an atomistic potential energy function for the water. The choice of the potential energy function is driven not only by considerations of accuracy of the model but also the constraints of the formulation of the integral equation theory. Electrostatic terms are of course necessary to mimic the ionic solvation properties of water, but the integral equation theory is best suited for single-point models of water with electrostatic interactions rather than multi-site models such as the various TIPS and SPC models because it is a single-site theory. Moreover, these models are readily decomposed into a short-range, orientation-independent part and a long-range, orientation-dependent part dependent on the electrostatics. The earliest models of this sort that account for electrostatic interactions are Stockmayer-type point dipole potentials [33–36], which include no higher than dipolar interactions thus precluding the tetrahedral coordination of the first shell around a water molecule. Of these models, the Surface Constrained Soft Sphere Dipole (SCSSD) model does include additional correction potentials to improve the short range interactions [35,36] but of course hydrogen-bonding properties cannot be reproduced. To introduce the tetrahedral coordination, higher electrostatic moments have been included [33,37–39].

Another approach for including hydrogen bond interactions has been developed by Bratko, Blum and Luzar [15,40]. The BBL model has a hard-sphere embedded with a point dipole of $\mu = 1.85$ D but now a empirical tetrahedral, octupolar ‘sticky’ potential is included for mimicking the hydrogen-bond interactions (Fig. 2). The sticky potential is an attractive square-well potential dependent on the relative orientation of the two interacting molecules, with the point dipole breaking the symmetry between the hydrogens and the lone pairs. A major advantage of this approach over including higher electrostatic order moments is that the sticky potential does not interact with the point dipole directly so that fewer interactions need to be calculated. Recently, our group has developed the SSD model (Fig. 2) [16], which is much like the BBL model but has several important modifications. First, the short range interactions are a Lennard–Jones soft-sphere, which allows it to be used with potential energy functions typically used for computer simulations of biological macromolecules. Second, the point dipole has been increased to the value for TIP3P and SPC/E, $\mu = 2.35$ D, which is based on the estimated dipole moment of a water molecule in liquid water and which has been found to be important for the dielectric proper-

Table 1
Properties of liquid water for SSD and other water models

Property	SSD[a]	TIP3P[a]	TIP4P[a]	SPC[b]	SPC/E[c]	RPOL[c]	Exp
Coord. No. to 3.3 Å (3.5 Å)	4.4 (5.2)	4.5 (5.5)	4.2 (5.0)	– (5.1)	–	–	4.5 (5.3)[d]
No. H-bond to 2.35 Å (2.5 Å)	3.6 (4.0)	3.5 (3.9)	3.5 (3.8)	– (3.9)	–	–	3.6 (4.0)[d]
E_i (kcal/mol)	-9.60	-9.67	-10.02	-10.18	-9.9	-9.9	-9.92[e]
C_v (cal/mol per K)	20.0	15.6[f]	16.0[f]	21.7[f]	–	–	17.9[g]
ε_0	81	82[h]	53–61[ijl]	54–74[ikl]	62–71[klm]	106	78[g]
D (10^{-5} cm^2/s)	2.3	5.1[n], 5.2	3.3[l]	3.6[l], 4.2[o]	2.4	2.4	2.3[p]
τ_1 (ps)	4.8	2.0	2.9[o], 3.6[l]	2.4[o], 3.3[l]	5.1	5.5	–
τ_D (ps)	8.0	5.75	–	11.0[l]	10.0	11.0	8.1[q]

Notes. [a] Ref. [16,20,21] unless otherwise specified; [b] ref. [19] unless otherwise specified; [c] ref. [45] unless otherwise specified; [d] calculated from data in ref. [46]; [e] ref. [47]; [f] recalculated from data in ref. [19]; [g] ref. [48]; [h] ref. [49]; [i] ref. [50–52]; [j] ref. [53]; [k] ref. [45]; [l] ref. [54]; [m] ref. [55]; [n] ref. [56]; [o] ref. [57]; [p] ref. [58]; and [q] ref. [59].



(a) (b)

Fig. 2. Schematic of water models. (a) Three-site models (i.e. SPC, TIP3P) have one point located at the oxygen, which is the center of a van der Waals sphere and a point charge, and two points located at each of the hydrogens, which each are the center of a point charge. (b) Sticky dipole models (i.e. BBL, SSD) have one molecular coordinate system with the origin at the oxygen, which is the center of a hard (BBL) or van der Waals (SSD) sphere, a point dipole (gray arrow), a tetrahedral octupolar sticky potential (dotted circular patches), and an additional sticky angular function (SSD only).

ties [18]. Third, the tetrahedral, octupolar sticky potential is now a smooth function to be compatible with the Lennard–Jones soft sphere. Finally, another angular sticky potential has been added to SSD, which improves the structure of SSD relative to BBL water. The SSD model gives good descriptions of the structural [16], dielectric [20], and dynamical [21] properties of water (Table 1). Moreover, the SSD model is very efficient since it is seven times faster in MC simulations and four times faster in MD simulations than three-site models such as TIP3P and SPC. Because the SSD model is such a good model for computer simulations of biomolecules and works well in the theory, it is the model of choice for the theory. However, earlier studies with the BBL model are also instructive in determining the accuracy of the theory.

3. Comparison of theory with Monte Carlo simulations

Since integral equations predict the structure of liquids from atomistic potential energy functions, the most stringent test of such a theory is to compare results from the theory with those from a MD or MC simulation using the same potential energy function. This is a more stringent test of the integral equation theory itself than comparing the theory directly to experiment since deviation from experiment could be due to deficiencies in either the theory or the potential energy function. In addition, comparison to simulation allows calculation and subsequent comparisons of quantities that may not be easily obtained experimentally. Of course, the accuracy of the overall theoretical treatment, integral equation plus the potential energy function, must be tested by comparison with experiment. However, for SSD water, the potential energy function has already been shown to agree well with experiment, so the important step for this theory is to show agreement with simulations.

The theory has been tested by comparing the three-dimensional spatial structure of water around a solute, $g^o(\mathbf{r}_{sw})$, and the solute–water interaction energy, E_{sw}, computed from theory to results from MC simulations using the identical potential model. Details of the MC simulations are given elsewhere [10,11]. The spatial distribution function $g^o(\mathbf{r}_{sw})$ is the probability of finding a water molecule at $\mathbf{r}_{sw}$ from the solute regardless of the orientation of that molecule and is thus like the standard definition for the pair distribution function for water molecules. For display purposes, $g^o(\mathbf{r}_{sw})$ is averaged to obtain the radial distribution function, $g^o(r_{sw})$ and the spatial distribution of water in the first solvation shell, $g^o(R_{\min} - R_{\max}\,\theta_{sw}\,\phi_{sw})$, where $R_{\min}$ and $R_{\max}$ define the first shell. The solute–water interaction energy E_{sw} is calculated by

$$E_{sw} = \rho \int g^o(\mathbf{r}_{sw})/k(r_{sw})\mathrm{d}\mathbf{r}_{sw} \frac{1}{8\pi^2} \int v(\mathbf{r}_{sw}\Omega_{sw}) \times \exp[-\beta\omega(\mathbf{r}_{sw}\Omega_{sw})]\mathrm{d}\Omega_{sw} \quad (10)$$

3.1. Solvation of fixed ice clusters

The first test systems were clusters of water, which served to test the ability of the theory to predict the solvation both of polar solutes and of globular solutes with buried atoms. The distributions of water around fixed clusters of one, five, or 25 water molecules were studied, in which both

Table 2
Solvent coordination numbers around frozen water clusters calculated from integral equation theory and Monte Carlo simulations

Solute size	One-molecule			Five-molecule		25-Molecule	
Distance (Å)	3.2	4.4	6.5	2.85	5.75	5.5	8.4
Theory	5.08	11.5	38.4	0.12	21.9	1.54	57.6
Simulation	4.58	11.2	37.9	0.07	22.6	3.07	61.6

solute and solvent molecules were either all BBL or all SSD water. Since the solute contains water molecules, there is an uninterrupted three-dimensional hydrogen bonding network between the solute cluster and the solvent water. For all of the clusters, the number of waters in the first few solvation shells given by the theory agree reasonably well with simulation for both BBL (Table 2) and for SSD (Table 3). The solute–solvent interaction energies are also reasonably good for SSD (Table 3).

The monomer solute consists of a single water molecule identical to the rest of the water molecules representing the solvent. The distribution functions around it are thus the same as the water–water distribution functions of the pure liquid. The theory predicts a radial distribution function, $g^o(r_{sw})$, and a spatial distribution of water in the first shell, $g^o(R_{\min} - R_{\max}\theta_{sw}\phi_{sw})$ for both BBL between $R_{\min} = 2.83$ Å and $R_{\max} = 3.0$ Å [10] and SSD between $R_{\min} = 2.7$ Å and $R_{\max} = 2.9$ Å [11] water that are in excellent agreement with MC. The overall good agreement suggests that the theory is able to predict the tetrahedral hydrogen bonding structure of liquid water.

The five-molecule solute consists of five water molecules in an ice-I_h structure: one at the center of the coordinate system (Fig. 3) and the other four at the tetrahedral neighbor sites hydrogen-bonded to the central one. The oxygen atoms form a perfect tetrahedron with an O–O distance of 2.85 Å between the central and each of the corner molecules. The theory predicts a radial distribution function, $g^o(r_{sw})$, that is in excellent agreement with simulation for both BBL (Fig. 4) and SSD (Fig. 5). Remarkably, the theory is able to predict a smaller peak at $r = 2.75$ Å for BBL seen in MC due to penetration into the faces of the tetrahedron formed by the four corner solute water molecules. In addition, the theory predicts ring-like regions encircling each of the 'corner' solute molecules in the spatial distribution function for the first shell, $g^o(R_{\min} - R_{\max}\theta_{sw}\phi_{sw})$, where $\theta_{sw}\phi_{sw}$ are defined in Fig. 3, for both BBL between $R_{\min} = 3.0$ Å and $R_{\max} = 5.75$ Å (Fig. 6) and SSD between $R_{\min} = 4.4$ Å and $R_{\max} = 4.8$ Å (Fig. 7), which are also seen in the simulation.

Table 3
Structural and energetic results of water around various solute calculated from the integral equation theory and MC simulations

Solute	$R_{\max}$[a] (Å)		Coordination number[b]		E_{sw} (kcal/mol)	
	Theory	Simulation	Theory	Simulation	Theory	Simulation
One-water	2.80	2.80	4.7	4.2	−17.5	−19.3 ± 0.5
Five-water	4.70	4.68	23.1	21.6	−53.4	−62.7 ± 2.7
Li^+	2.00	1.98	5.5	5.5	−266	−263 ± 13
Na^+	2.40	2.38	6.5	6.6	−214	−230 ± 23
F^-	2.60	2.58	6.2	7.1	−172	−181 ± 14
Cl^-	3.28	3.33	11.1	9.6	−155	−125 ± 16
Na^+–Cl^- pair (5 Å)					−246	−280 ± 24
Na^+–Cl^- pair (7 Å)					−294	−311 ± 37

[a] $R_{\max}$ is the position of the first peak in $g^o(r_{sw})$.
[b] Calculated by integrating the radial distribution functions out to the first minimum.

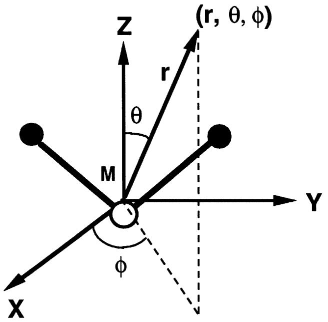


Fig. 3. Schematic of the coordinate system with the origin located at the center-of-mass of a solute water molecule.

Within each ring, there are three peaks that correspond to the three solvent nearest neighbors of the corner molecule, with the fourth nearest neighbor being the water molecule at the center of the solute. For SSD, the ring-like structure is not as apparent in the simulation results. This could be due to the simulation not being long enough to sample the region between the peaks or to approximations in the theory. Although approximations of $c^o(\mathbf{r}_{ww'})$ and $g^o(\mathbf{r}_{sw})$ by isotropic solutions might seem the most likely culprit, the good results for the BBL model using these approximations indicate otherwise. This indicates that further studies of the approximate bridge functions may be necessary. As a whole, these results demonstrate the ability of the theory to predict the tetrahedral bonding structure as well as the excluded volume effects.

The 25-molecule solute consists of 24 BBL water molecules around a central water BBL molecule taken from a single coordinate file of a MC simulation of pure BBL water. Thus, the structure of the cluster is irregular with many atoms buried away from the solvent. The $g^o(r_{sw})$ predicted by theory (Fig. 8) is in reasonable agreement with simulation. Since both show near-zero solvent population at distances within 4.0 Å, the theory predicts excluded volume effects well. The theory also predicts the first major solvation shell approximately at $r = 7.3$ Å, although it misses the shoulder at 6 Å in the MC results. This demonstrates the ability of the theory to handle large solutes efficiently and yet with reasonable accuracy.

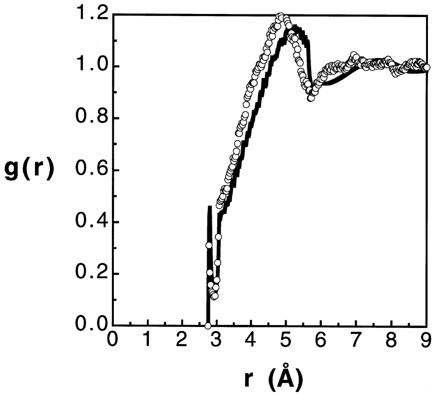


Fig. 4. The radial distribution functions $g^o(r_{sw})$ for BBL water around the five-molecule ice cluster calculated from the theory (line) and from MC simulation (circles).

3.2. Solvation of ions

The distribution of SSD water around a set of single cations and anions of different sizes and an ion pair at two different separations have also been studied. These studies are important tests for the ability of this theory to predict solvation of ionic species. In particular, the study of water near an ion-pair is of biochemical importance, since the association of unlike charges plays an important role in the formation and function of biological molecules [41–43]. For all of the systems studies, the theory is in reasonably good agreement with simulation for the locations of

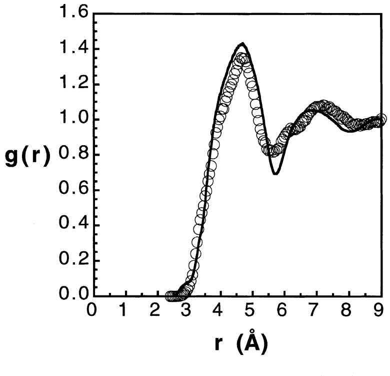


Fig. 5. The radial distribution functions $g^o(r_{sw})$ for SSD water around the five-molecule ice cluster calculated from the theory (line) and from MC simulation (circles).

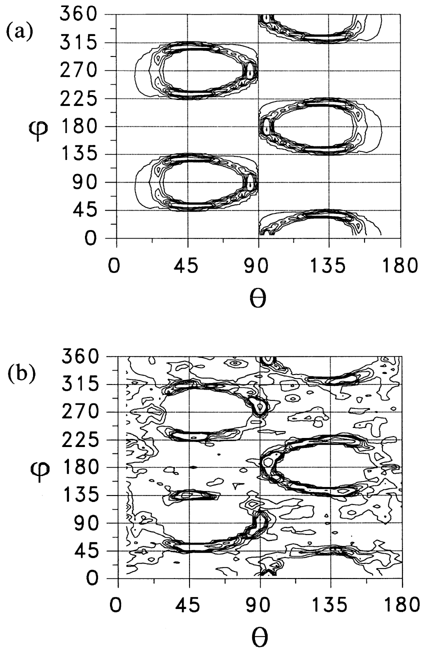


Fig. 6. Contour maps of the spatial distribution $g^o(R_{\min}-R_{\max}\,\theta_{sw}\,\phi_{sw})$ for $R_{\min}=3.0$ Å and $R_{\max}=5.75$ Å of BBL water in the first solvation shell around the five-molecule ice cluster solute calculated from (a) theory, with contour levels at 1, 2, 3, 4, 5, 8; and (b) MC simulation, with contour levels at 1, 2, 3, 4, 5, 10, 20. Contour levels above 5 were chosen to indicate peak positions.

the first peak, the coordination numbers and the ion–water interaction energies (Table 3).

Four simple ions, Li^+, Na^+, F^-, and Cl^- in SSD water were studied [11]. The theory predicts similar $g^o(r_{sw})$ to that from simulations [11]. The overall good agreement is attributed to the bridge functions, despite the fact that they are scaled from the pure water results.

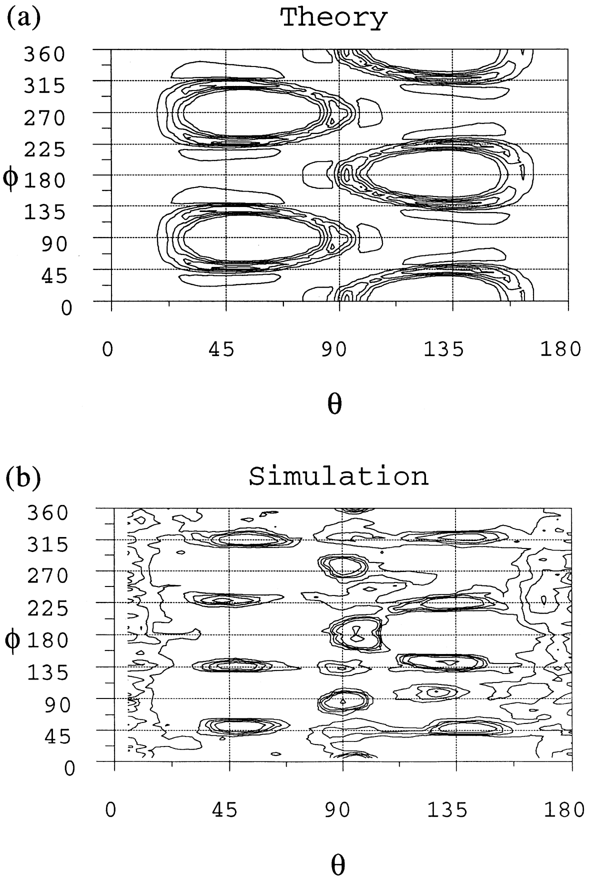


Fig. 7. Contour maps of the spatial distribution $g^o(R_{\min}-R_{\max}\,\theta_{sw}\,\phi_{sw})$ for $R_{\min}=4.4$ Å and $R_{\max}=4.8$ Å of SSD water in the first solvation shell around the five-molecule ice cluster solute calculated from (a) theory, with contour levels at 1, 2, 3, 4, 5, 5.6; and (b) MC simulation, with contour levels at 1, 2, 3, 4, 5, 11, 15, 20, 25. Contour levels above 5 were chosen to indicate peak positions.

Na^+–Cl^- pairs in SSD water at two different separations (5 and 7 Å) were studied. These two distances were chosen because the two separations are likely to either favor (7 Å) or disfavor (5 Å) water intervention between the two. The results are described in terms of a cylindrical coordinate system centered midway between the two ions with the z-axis along the internuclear axis. The theory predicts the water density $g^o(r_{sw}\,\theta_{sw})$ around the ion-pair at 5-Å separation (Fig. 9) and at 7-Å separation (Fig. 10) that are in good agreement with MC. (The ripples in the two main solvation shells are due to the increments in θ_{sw} used for the calculation.) For the 5-Å separation case (Fig. 9), the theory shows a low probability pocket of water located between the ions due to the favorable electrostatic energy when the water dipole aligns with the field from the two ions. Since the MC simulation results do not show this pocket of water, this may reflect an inadequacy of the theory, although it may also be due to insufficient statistics in the simulation since the pocket has a low probabilty. In the 7-Å separation case (Fig. 10), both theory and simulation show two

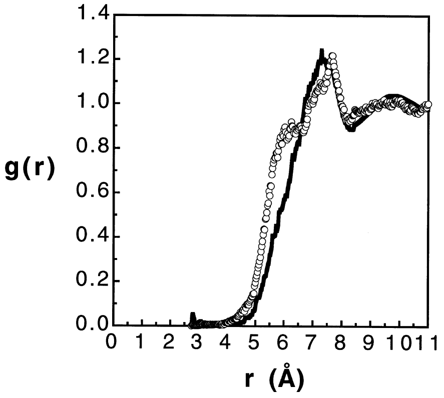


Fig. 8. The radial distribution function $g^o(r_{sw})$ for BBL water around the 25-molecule ice cluster solute calculated from the theory (line) and from MC simulation (circles).

complete solvation shells. These results are similar to other MC simulation results for the Li^+–F^- pair [44].

4. Efficiency

Compared to MC simulations, the theory is extremely efficient. For example, the simulation data around the single water solute required approximately 200 CPU h on an IBM RISC 6000 workstation, while the theory data required < 8 h. Furthermore, the speed could be improved by making the computer code parallel. This is trivial because the approximation $g^o(\mathbf{r}_{sw}) = g^o(r_{sw})|_{\theta,\phi}$;

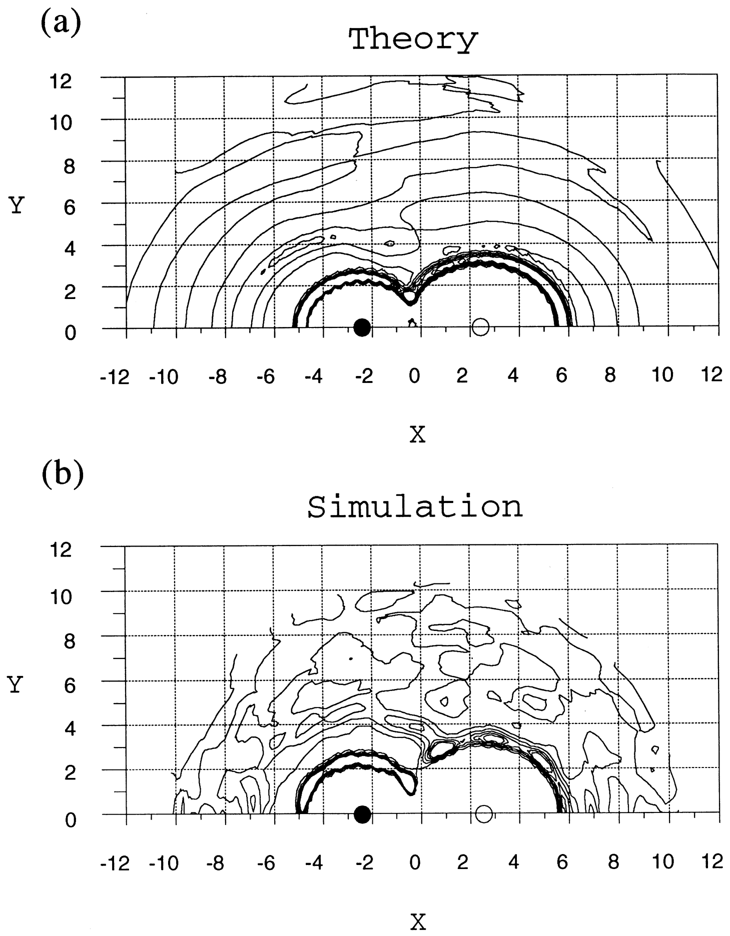
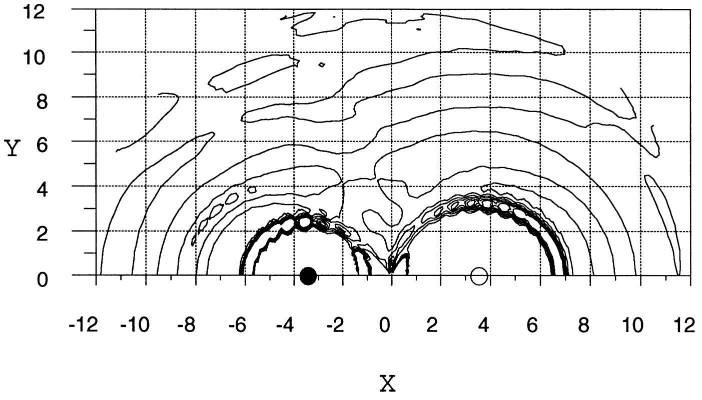


Fig. 9. Contour maps of the SSD water distribution $g^o(r_{sw}\theta_{sw})$ as a function of $X = r_{sw}\cos\theta_{sw}$ and $Y = r_{sw}\sin\theta_{sw}$. around the Na^+–Cl^- ion pair calculated from (a) theory and (b) MC simulation. The interionic separation between the Na^+ (solid circle) and the Cl^- (open circle) is 5 Å. The contour levels are at 0.5, 1, 1.5, 2, 2.5 and 3.

(a) Theory



(b) Simulation

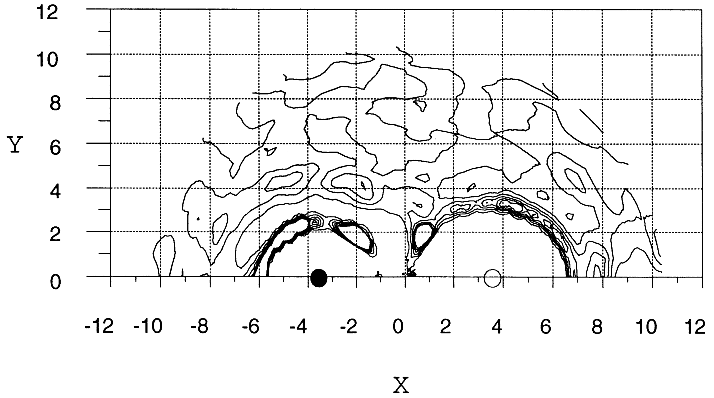


Fig. 10. Same as Fig. 9 except that the interionic separation is 7 Å.

allows the calculations for a given set of θ and ϕ to be done separately on individual nodes of a parallel machine, which is impossible for the more accurate three-dimensional convolution. Thus, the speed of the present theory should scale almost linearly with the use of array processors.

The slowest computational step in the present theory is the numerical evaluation of $k(\mathbf{r}_{sw})$ in Eq. (7), which is an integral over the orientational angles for each spatial position. Thus, the amount of computation is proportional to $N'N_\theta N_\phi$ where N' is the number of non-neutral sites that are near the surface of the solute and $N_\theta N_\phi$ is the total number of grid points. The next slowest step is the iterative solution of Eq. (8), which scales as $N_\theta N_\phi$. The quantity $N_\theta N_\phi$ is proportional to the surface area of the solute and therefore, the overall computational demand of this theory roughly scales as the surface area of the solute times N'. Consequently, the theory should be more efficient in modeling water around globular solutes and less so around solutes with a greater surface/volume ratio, such as linear, planar, or those with very complex surface geometry.

5. Conclusions

The integral equation theory described here is designed for modeling the solvation of large solutes and so a balance of accuracy and computational efficiency is the foremost consideration. The initial results for solutes of various sizes, geometries, net charges and charge distributions are quite accurate and have demonstrated that the theory is quite successful in achieving its initial goal. We believe that the theory is potentially useful for modeling solvation of large, globular solutes such as biological macromolecules.

Acknowledgements

We would like to acknowledge the National Science Foundation for support of this work through grant MCB-9506796 and MCB-9808116. Y.L. thanks Dr Satheesan Babu for helpful discussions. This research was sponsored in part by the Phillips Laboratory, Air Force Material Command, USAF, through the use of the Maui High Performance Computing Center under cooperative agreement number F29601-93-2-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of Phillips Laboratory or the US Government.

References

[1] E. Westhof (Ed.), Water and Biological Macromolecules, 1st ed., CRC Press, Boca Raton, 1993, pp. 466.
[2] F. Franks, Water, a comprehensive treatise, vol. 4, in: F. Franks (Ed.), Aqueous Solutions of Amphiphiles and Macromolecules. The Hydrophobic Interaction, Plenum, New York, 1975.
[3] W. Blokzijl, J.B.F.N. Engberts, Angew. Chem. Int. Ed. 32 (1993) 1545.
[4] S.A. Durell, A. Wallqvist, Biophys. J. 71 (1996) 1695.
[5] H.L. Friedman, C.V. Krischnan, Water: a comprehensive treatise, vol. 3, in: F. Franks (Ed.), Aqueous Solutions of Simple Electrolytes. Thermodynamics of ion hydration, Plenum, New York, 1973, p. 1.
[6] M.P. Allen, D.J. Tildesley, Computer Simulations of Liquids, 1st ed., Clarendon Press, Oxford, 1987.
[7] R.B. Yelle, N.-S. Park, T. Ichiye, Proteins 22 (1995) 154.
[8] V. Mohan, P.E. Smith, B.M. Pettitt, J. Phys. Chem. 97 (1993) 12984.
[9] J.P. Hansen, I.R. McDonald, Theory of Simple Liquids, 2nd ed., Academic Press, London, 1986.
[10] Y. Liu, T. Ichiye, Chem. Phys. Lett. 231 (1994) 380.
[11] Y. Liu, T. Ichiye, J. Chem. Phys. (1999) in preparation.
[12] J.-K. Hyun, C.S. Babu, T. Ichiye, J. Phys. Chem. 99 (1995) 5187.
[13] J.-K. Hyun, T. Ichiye, J. Phys. Chem. B 101 (1997) 3596.
[14] J.-K. Hyun, T. Ichiye, J. Chem. Phys. 109 (1998) 1074.
[15] D. Bratko, L. Blum, A. Luzar, J. Chem. Phys. 83 (1985) 6367.
[16] Y. Liu, T. Ichiye, J. Phys. Chem. 100 (1996) 2723.
[17] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, J. Hermans, Intermolecular forces, in: B. Pullman (Ed.), Interaction Models of Water in Relation to Protein Hydration, Reidel, Dordrecht, Holland, 1981, p. 331.
[18] H.J.C. Berendsen, J.R. Grigera, T.P. Straatsma, J. Phys. Chem. 91 (1987) 6269.
[19] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, J. Chem. Phys. 79 (1983) 926.
[20] Y. Liu, T. Ichiye, Chem. Phys. Lett. 256 (1996) 334.
[21] A. Chandra, T. Ichiye, Chem. Phys. Lett. (1999) in preparation.
[22] D. Chandler, H.C. Andersen, J. Chem. Phys. 57 (1972) 1930.
[23] B.M. Pettitt, M. Karplus, Chem. Phys. Lett. 136 (1987) 383.
[24] M. Kinoshita, Y. Okamoto, F. Hirata, J. Chem. Phys. 107 (1997) 1586.
[25] M. Kinoshita, F. Hirata, J. Chem. Phys. 104 (1996) 8807.
[26] P.G. Kusalik, G. Patey, J. Chem. Phys. 88 (1988) 7714.
[27] P.G. Kusalik, Mol. Phys. 67 (1989) 67.
[28] J.P. Hansen, I.R. McDonald, Phys. Rev. A 11 (1975) 2111.
[29] L. Verlet, Phys. Rev. 165 (1968) 201.
[30] S.M. Foiles, N.W. Ashcroft, J. Chem. Phys. 81 (1984) 6140.
[31] J.-K. Hyun, T. Ichiye, J. Chem. Phys. (1999) in preparation.
[32] T. Ichiye, J. Chem. Phys. 104 (1996) 7561.
[33] P. Barnes, in: C.A. Croxton (Ed.), Progress in Liquid Physics, Wiley-Interscience, New York, 1978.
[34] P. Barnes, J.L. Finney, J.D. Nicholas, J.E. Quinn, Nature 282 (1979) 459.
[35] A. Warshel, Chem. Phys. Lett. 55 (1978) 454.
[36] A. Warshel, J. Phys. Chem. 83 (1979) 1640.
[37] G.N. Patey, S.L. Carnie, J. Chem. Phys. 78 (1983) 5183.
[38] G.M. Torrie, P.G. Kusalik, G.N. Patey, J. Chem. Phys. 88 (1988) 7826.
[39] G.M. Torrie, P.G. Kusalik, G.N. Patey, J. Chem. Phys. 90 (1989) 4513.
[40] L. Blum, F. Vericat, D. Bratko, J. Chem. Phys. 102 (1995) 1461.
[41] B.E. Conway, Ionic Hydration in Chemistry and Biophysics, Elsevier Scientific Publishing, New York, 1981.
[42] M. Berkowitz, O.A. Karim, J.A. McCammon, P.J. Rossky, Chem. Phys. Lett. 105 (1984) 577.
[43] A. Warshel, Acc. Chem. Res. 14 (1981) 284.
[44] R.O. Watts, Mol. Phys. 28 (1974) 1069.
[45] P.E. Smith, W.F. van Gunsteren, J. Chem. Phys. 100 (1994) 3169.

[46] A.K. Soper, M.G. Phillips, Chem. Phys. 107 (1986) 47.
[47] W.L. Jorgensen, J. Am. Chem. Soc. 103 (1981) 335.
[48] D. Eisenberg, W. Kauzmann, The Structure and Properties of Water, 1st ed., Oxford University Press, New York, 1969.
[49] T. Simonson, Chem. Phys. Lett. 250 (1996) 450.
[50] H.E. Alper, R.M. Levy, J. Chem. Phys. 91 (1989) 1242.
[51] H.E. Alper, R.M. Levy, J. Chem. Phys. 99 (1993) 9847.
[52] M. Belhadj, H.E. Alper, R.M. Levy, Chem. Phys. Lett. 179 (1991) 13.
[53] M. Neumann, J. Chem. Phys. 85 (1986) 1567.
[54] K. Watanabe, M.L. Klein, Chem. Phys. 131 (1989) 157.
[55] M.R. Reddy, M. Berkowitz, Chem. Phys. Lett. 155 (1989) 173.
[56] S.E. Feller, R.W. Pastor, A. Rojnuckarin, S. Bogusz, B.R. Brooks, J. Phys. Chem. 100 (1996) 17011.
[57] J.E. Roberts, J. Schnitker, J. Phys. Chem. 99 (1995) 1322.
[58] K. Krynicki, C.D. Green, D.W. Sawyer, Discuss. Faraday Soc. 66 (1978) 199.
[59] R.J. Speedy, C.A. Angel, J. Chem. Phys. 65 (1976) 851.