Article

Solvation Free Energy of Polar and Nonpolar Molecules in Water: An Extended Interaction Site Integral Equation Theory in Three Dimensions

Qishi Du, Dmitrii Beglov, and Benot Roux

J. Phys. Chem. B, **2000**, 104 (4), 796-805 • DOI: 10.1021/jp992712l

Solvation Free Energy of Polar and Nonpolar Molecules in Water: An Extended Interaction Site Integral Equation Theory in Three Dimensions

Qishi Du, Dmitrii Beglov,† and Benoît Roux*

Centre de Recherche en Calcul Appliqué (CERCA), Départements de physique et de chimie, Université de Montréal, C.P. 6128, succ. Centre-Ville, Montreal, Quebec, Canada H3C 3J7

Received: July 30, 1999; In Final Form: November 2, 1999

The solvation free energy of polar and nonpolar molecules in water is examined using the statistical mechanical integral equation theory which was recently introduced by Beglov and Roux (*J. Phys. Chem.* **1997**, *101*, 7821). The integral equation is an extension of previous theories based on the reference interaction site model (RISM) and the hypernetted chain equation (HNC) to determine the one-dimensional site−site radial pair correlation functions. Since the integral equation provides the average density of the solvent interaction sites in three-dimensional (3d) space around a complex molecular solute of arbitrary shape, it is thereby referred to as the 3d-RISM-HNC integral equation. In the present paper, the accuracy of the 3d-RISM-HNC integral equation is examined by comparing the calculated excess solvation free energy at infinite dilution with experimental data for a set of nonpolar (*n*-alkanes) and polar (*n*-alcohols, *n*-carboxylic acids, and simple amides) molecules. Thermodynamic integration is used to calculate the excess solvation free energy on the basis of the average site densities obtained by the 3d-RISM-HNC equation. Two extensions were made to improve the accuracy of the integral equation. First, the susceptibility of pure liquid water (an input for calculating the excess solvation free energy of molecules in the infinite dilution limit) was obtained by combining information from a molecular dynamics simulation with the RISM-HNC integral equation. Second, empirical water hydrogen and water oxygen bridge functions were introduced and optimized for improving the accuracy of the theory. The calculated solvation free energies are in good accord with experimental data, although further work will be required for quantitatively accurate results. These preliminary results indicate that the 3d-RISM-HNC equation with well-calibrated optimized bridge functions is a promising approach for calculating the hydration free energy of complex molecules.

I. Introduction

Solvation forces are a major determinant of the conformation and association of peptides, proteins, and other biomolecules. For this reason, it is important to develop quantitatively accurate computational approaches enabling one to assess the importance of solvent effects. Current treatments of solvation generally follow one of two very different approaches. Either the solvent molecules are represented *explicitly* for generating detailed atomic trajectories,[1,2] or they are represented *implicitly*, using a continuum approximation.[3,4] Undoubtedly, calculations in which a large number of solvent molecules are explicitly simulated provide the most realistic treatment to examine the solvation of a biomolecular solute.[1,2] However, such an approach has important practical limitations. In particular, the estimation of thermodynamic properties using free energy perturbation simulations relies on the assumption that the configurational space is sampled sufficiently. In practice there remains a significant uncertainty concerning the statistical convergence of such calculations due to the finite length of the simulated trajectories. The need to consider a large number of solvent configurations becomes even more costly when the solute or a subset of the system is treated at a quantum mechanical ab initio level.[5,6] In contrast, continuum solvent models, which represent the solvent in terms of a structureless medium,[4] are computationally inexpensive and do not suffer from such convergence problems. A number of approximate schemes have been developed and used in various contexts, for example, the Kirkwood−Onsager reaction field,[7,8] macroscopic continuum electrostatics,[9] the generalized Born model,[10] and conductor boundary conditions[11,12] (see ref 4 for a recent review of continuum models). Nonetheless, despite their attractive simplicity, continuum models ignore many essential features of a molecular solvent and are thus seriously limited. In particular, continuum macroscopic electrostatics approaches are based on the assumptions that the average polarization density of the solvent $\boldsymbol{P}(\boldsymbol{r})$ is strictly parallel to the total electrostatic field $\boldsymbol{E}^{\mathrm{tot}}(\boldsymbol{r})$ at any point $\boldsymbol{r}$ and that their relative magnitude is governed by the established linear relation $4\pi\boldsymbol{P}(\boldsymbol{r}) = (\epsilon_s - 1)\boldsymbol{E}^{\mathrm{tot}}(\boldsymbol{r})$, where ϵ_s is the dielectric constant of the bulk solvent.[13] Other important features of a molecular solvent are neglected. For example, the ability of water molecules to form hydrogen bonds with a molecular solute is not taken into account and the granularity of the solvent is completely ignored in continuum models. A real liquid is composed of molecules of finite size, which is an important factor influencing the packing of the solvent molecules around a complex solute.

In principle, statistical mechanical theories based on density distribution functions can provide alternative approaches for describing solvation phenomena.[14] Integral equation theories represent the solvent in terms of its average density. Such

* Corresponding author. Telephone: (514) 343-7105. Fax: (514) 343-7586. E-mail: Benoit.Roux@umontreal.ca.

† Current address: Moldyn, 955 Mass Avenue, fifth floor, Cambridge, MA 02139.

10.1021/jp992712l CCC: $19.00
Published on Web 01/05/2000

approaches, which are able to incorporate several aspects of the solvent structure, circumvent the oversimplification of continuum models without having to simulate a large number of solvent molecules explicitly. Recent developments have made it possible to extend standard integral equations to treat the full solvent distribution around solutes of irregular shape.[15−22] In a previous paper, Beglov and Roux introduced a statistical mechanical integral equation theory based on the average solvent interaction site density in three-dimensions (3d) around a complex molecule of arbitrary shape.[17] The integral equation was developed on the basis of the free energy density functional theory formulated by Chandler, McCoy, and Singer[23] and is an extension of previous reference interaction site model (RISM) theories for the one-dimensional radial pair correlation function which were based on the site−site Ornstein−Zernicke (SSOZ) model and the hypernetted chain equation (HNC).[24,25] The SSOZ-RISM theory has been used in solvation studies of monatomic solutes at infinite dilution in liquid water,[26−28] small peptides,[29−33] and a large number of organic molecules.[34] However, the theory is not appropriate for complex molecular solutes, since buried solute atoms are only partially shielded from the solvent, and the dielectric constant of the pure liquid is not described satisfactorily. Because the SSOZ-RISM theory is based on a reduction to site−site solute−solvent radially symmetric distribution functions, there is a loss of information about the three-dimensional spatial organization of the solvent density around a nonspherical molecular solute. In contrast, the 3d-RISM-HNC integral equation makes it possible to take into account the full distribution of the solvent around solutes of irregular shape in three-dimensional space. A similar integral equation has since been examined by Cortis et al.[19] and by Kovalenko and Hirata.[20−22] Although previous tests with the 3d-RISM-HNC theory appear to yield qualitatively reasonable results for complex molecular solutes in water,[17] there were certain problems. For example, it was observed that the solvent density of water hydrogens was overestimated in the vicinity of negatively charged groups (strong hydrogen-bonding acceptor). It is clear that further work is necessary to have a useful theory of quantitative accuracy. In particular, thermodynamic results such as excess solvation free energies need to be examined carefully. The goal of the present paper is thus to pursue our efforts to develop the 3d-RISM-HNC integral equation and to seek methods for its improvement.

A package of programs was developed to calculate distributions of water oxygen and hydrogen densities around a molecular solute, and the nonpolar and electrostatic contributions to the excess solvation free energy using explicit thermodynamic integration. Several steps were specifically taken to extend and improve the 3d-RISM-HNC theory. First, the susceptibility of liquid water (which appears as an input in the integral equation theory, since the solutes are treated in the infinite dilution limit) was obtained by combining the site−site radial pair correlation functions calculated from a molecular dynamics simulation of a box of TIP3P water molecules[35] with the results of the standard SSOZ-RISM using an empirical interpolation method. Second, following the work of Cortis et al.,[19] a water hydrogen bridge function based on a low-order expansion of the short-range solute−water pair correlation function was introduced. This low-order correction is sufficient to remove unreasonably large water hydrogen density peaks in the vicinity of strong hydrogen bond accepting groups, which were observed previously. A water oxygen bridge function is also introduced to improve the thermodynamic accuracy of the theory. Although the concept of water hydrogen and water oxygen bridge functions does not follow from a formal and rigorous statistical mechanical derivation in the 3d-RISM-HNC theory, it is hoped that such bridge functions can be adjusted and calibrated empirically to yield accurate thermodynamic properties.

The accuracy of the extended theory is examined by comparing the calculated solvation free energy with experimental data for a set of polar (*n*-alcohols, *n*-carboxylic acids, and amides) and nonpolar (*n*-alkanes) molecules. The calculated solvation free energies are in good accord with experimental data, although further work will be required for quantitatively accurate results. These preliminary results indicate that the 3d-RISM-HNC equation with well-calibrated optimized bridge functions is a promising approach for calculating the hydration free energies of complex molecules. Directions for future research are discussed.

II. Theoretical Developments

(a) 3d-RISM-HNC Integral Equation and Solvation Free Energy. An isolated molecular solute immersed in a polyatomic liquid of uniform density $\bar{\rho}$ is considered. It is assumed that the solvent molecules are constituted of a number of interaction sites α. Each interaction site is affected by the microscopic solute−solvent potential $U_\alpha(\boldsymbol{r})$, thus perturbing the uniform average solvent density. The interaction potential $U_\alpha(\boldsymbol{r})$ acting on the solvent site α is represented by a sum of radially symmetric Lennard-Jones (LJ) 6−12 and Coulomb electrostatic potential functions centered on the solute interaction site *s*,

$$U_\alpha(\boldsymbol{r}) = \sum_s (U_{\alpha s}^{(\mathrm{LJ})}(\boldsymbol{r}) + U_{\alpha s}^{(\mathrm{elec})}(\boldsymbol{r})) \tag{1}$$

with

$$U_{\alpha s}^{(\mathrm{LJ})}(\boldsymbol{r}) = 4\epsilon_{\alpha s}\left[\left(\frac{\sigma_{\alpha s}}{|\boldsymbol{r}-\boldsymbol{r}_s|}\right)^{12} - \left(\frac{\sigma_{\alpha s}}{|\boldsymbol{r}-\boldsymbol{r}_s|}\right)^{6}\right] \tag{2}$$

and

$$U_{\alpha s}^{(\mathrm{elec})}(\boldsymbol{r}) = \frac{q_\alpha q_s}{|\boldsymbol{r}-\boldsymbol{r}_s|} \tag{3}$$

where $\epsilon_{\alpha s}$, $\sigma_{\alpha s}$, q_α, and q_s are the LJ parameters and atomic charges, respectively (note: in the following, roman s and greek α symbols indicate the solute and solvent interaction sites, respectively). Such a representation, which is commonly used in most force fields for computer simulations of biomolecules (e.g. AMBER,[36] CHARMM,[37] OPLS[38]), is useful for decomposing the reversible work that defines the excess chemical potential into electrostatic contributions and the remaining nonpolar (nonelectrostatic) contributions (see eqs 8 and 9). However, a generalization to different forms of solute−solvent potentials is straightforward (e.g., using a quantum mechanical ab initio description of the solute).

The fundamental quantities for obtaining thermodynamic properties are the reduced average solvent densities $\langle\rho_\alpha(\boldsymbol{r})\rangle$ of the solvent interaction sites. In the present work, we calculate the average densities $\langle\rho_\alpha(\boldsymbol{r})\rangle$ from the 3d-RISM-HNC integral equation

$$h_\alpha(\boldsymbol{r}) = [-\beta U_\alpha(\boldsymbol{r}) + h_\alpha(\boldsymbol{r}) - c_\alpha(\boldsymbol{r}) - b_\alpha(\boldsymbol{r})] - 1 \tag{4}$$

where $h_\alpha(\boldsymbol{r}) = \langle\rho_\alpha(\boldsymbol{r})\rangle/\bar{\rho} - 1$ is the solute−solvent site correlation function and $c_\alpha(\boldsymbol{r})$ is the solute−solvent direct correlation

function defined by

$$\bar{\rho}h_\alpha(\boldsymbol{r}) = \sum_\gamma c_\gamma * \chi_{\gamma\alpha}(\boldsymbol{r}) \tag{5}$$

where $\chi_{\gamma\alpha}(\boldsymbol{r})$ is the solvent susceptibility response function of the pure liquid and the symbol $*$ represents a spatial convolution in three-dimensional space. The function $b_\alpha(\boldsymbol{r})$ is reminiscent of the bridge functions which have a well-defined formal diagrammatic structure in the theory of liquids.[14] Here, the bridge functions play the role of effective potentials which are adjusted empirically to improve the accuracy of the theory. The bridge functions are discussed further in the next section. In the case of a monatomic solute, the 3d-RISM-HNC integral equation with bridge functions reduces to the standard SSOZ-RISM theory when $b_\alpha(\boldsymbol{r}) = 0$.

All solvent effects and thermodynamic properties can be calculated directly from the average densities $\langle\rho_\alpha(\boldsymbol{r})\rangle$. In particular, the average LJ and electrostatic solute−solvent interaction energies are

$$\langle U^{(\mathrm{LJ})}\rangle = \int \mathrm{d}\boldsymbol{r} \sum_\alpha \sum_s U^{(\mathrm{LJ})}_{\alpha s}(\boldsymbol{r})\langle\rho_\alpha(\boldsymbol{r})\rangle \tag{6}$$

and

$$\langle U^{(\mathrm{elec})}\rangle = \int \mathrm{d}\boldsymbol{r} \sum_\alpha \sum_s U^{(\mathrm{elec})}_{\alpha s}(\boldsymbol{r})\langle\rho_\alpha(\boldsymbol{r})\rangle \tag{7}$$

respectively. In addition, the total excess chemical potential or solvation free energy of a solute at infinite dilution can be expressed rigorously as the reversible work (calculated as a thermodynamic integration) for materializing the solute into the solvent in a step-by-step process.[3] In a first step, the nonpolar solute−solvent interactions $U^{(\mathrm{LJ})}{}_{\alpha s}$ are switched "on" in the absence of any solute−solvent electrostatic interactions; in a second step, the solute−solvent electrostatic interactions $U^{(\mathrm{elec})}{}_{\alpha s}$ are switched "on" in the presence of the solute−solvent nonpolar interactions. To calculate the nonpolar and electrostatic free energy contributions, we introduce two thermodynamic coupling parameters λ_1 and λ_2, allowing us to turn the LJ and electrostatic interactions "on" ($\lambda_i = 1$) and "off" ($\lambda_i = 0$), respectively. By construction, the total excess chemical potential is $\Delta\mu_{\mathrm{tot}} = \Delta\mu_{\mathrm{np}} + \Delta\mu_{\mathrm{elec}}$. The quantities $\Delta\mu_{\mathrm{np}}$ and $\Delta\mu_{\mathrm{elec}}$ are often referred to as the "cavity" and "charging" contributions to the solvation free energy, respectively.[3]

For numerical convenience, the coupling parameter λ_1 for the potential $U^{(\mathrm{LJ})}{}_{\alpha s}$ in eq 6 is introduced as a scaling factor of the LJ radius, $\sigma_{\alpha s} \rightarrow \lambda_1\sigma_{\alpha s}$; the coupling parameter λ_2 for the electrostatic potential $U^{(\mathrm{elec})}{}_{\alpha s}$ in eq 7 is introduced as a scaling factor of the solute charges, $q_s \rightarrow \lambda_2 q_s$. It follows that the nonpolar and electrostatic free energy contributions to the excess chemical potential are

$$\Delta\mu_{\mathrm{np}} = \int_0^1 \mathrm{d}\lambda_1 \int \mathrm{d}\boldsymbol{r} \sum_\alpha \sum_s \frac{\partial U^{(\mathrm{LJ})}_{\alpha s}(\boldsymbol{r};\lambda_1)}{\lambda_1}\langle\rho_\alpha(\boldsymbol{r})\rangle_{(\lambda_1,\lambda_2=0)} \tag{8}$$

and

$$\Delta\mu_{\mathrm{elec}} = \int_0^1 \mathrm{d}\lambda_2 \int \mathrm{d}\boldsymbol{r} \sum_\alpha \sum_s \frac{\partial U^{(\mathrm{elec})}_{\alpha s}(\boldsymbol{r};\lambda_2)}{\lambda_2}\langle\rho_\alpha(\boldsymbol{r})\rangle_{(\lambda_1=1,\lambda_2)} \tag{9}$$

respectively. The subscripts (λ_1, λ_2) on the bracket $\langle\ \rangle$ indicate that the average solvent densities are calculated from eqs 4 and 5 with the potential $U_\alpha(\boldsymbol{r};\lambda_1,\lambda_2)$ modified by the thermodynamic coupling parameters λ_1 and λ_2. In eq 8, the solute−solvent electrostatic interactions are turned off, so $\lambda_2 = 0$, and the LJ contribution is calculated by integrating over λ_1 from zero to one. In eq 9, the solute−solvent LJ interactions are turned on, so $\lambda_1 = 1$, and the electrostatic contribution is calculated by integrating over λ_2 from zero to one. In the absence of bridge functions, it is possible to derive closed-form expressions for the thermodynamic integration and, thus, the solvation free energy.[39,40] However, we are not aware of a general expression when there are arbitrary bridge functions.

(b) Empirical Bridge Functions. Previous work showed that the results from the 3d-RISM-HNC equation without $b_\alpha(\boldsymbol{r})$, though qualitatively reasonable, are not quantitatively accurate.[17] In particular, the average solvent density of the hydrogens near negative groups is overestimated. This problem is also observed in application of the standard SSOZ-RISM equation.[25−27] It is caused by a lack of intramolecular correlation in the solvent molecule. The hydrogen−oxygen intramolecular correlations are incorporated only to second order in the free energy density functional expansion which forms the basis of the integral equation.[17,23] In simple words, although some of the intramolecular structure of the water molecule is present via the solvent susceptibility function $\chi_{\alpha\gamma}(r)$, the hydrogen density is not sufficiently coupled to the density of the oxygen to which it is chemically bonded when it is strongly attracted in some region of space.

To address the lack of intramolecular correlation, we follow Cortis et al.[19] and construct a hydrogen bridge function incorporating the short-range hydrogen−oxygen intramolecular correlation to lowest order. Let us consider the distribution of a water molecule interacting with a solute. The solute−water hydrogen pair correlation function at short distances from the solute can be expressed as

$$h_{\mathrm{H}}(\boldsymbol{r}) + 1 = \int \mathrm{d}\boldsymbol{X}\, S(\boldsymbol{X})\delta(\boldsymbol{r} - \boldsymbol{r}_{\mathrm{H1}}) \exp[-\beta W(\boldsymbol{X})] \tag{10}$$

where the integration is carried over all the degress of freedom $\boldsymbol{X}$ of water, $S(\boldsymbol{X})$ is the intramolecular distribution function of one water molecule, and $W(\boldsymbol{X})$ is the total potential of mean force (PMF) between the solute and the water molecule. At a very short distance from the solute, the PMF is expected to be dominated by the direct short-range LJ repulsive interaction between the solute and the water oxygen, that is, $W(\boldsymbol{X}) \approx U^{(\mathrm{LJ})}_{\mathrm{O}}(\boldsymbol{r}_{\mathrm{O}})$. Following Cortis et al.,[19] we separate the total LJ potential acting on the water oxygen into a short-range repulsion $U^{(\mathrm{LJs})}_{\mathrm{O}}$ and an attractive tail $U^{(\mathrm{LJt})}_{\mathrm{O}}$ according to the Weeks−Chandler−Andersen (WCA) perturbation scheme.[41] For a single LJ center, the short-range part has the form

$$U^{(\mathrm{LJs})}(r) = \begin{cases} 4\epsilon[(\sigma/r)^{12} - (\sigma/r)^6] + \epsilon & \text{if } r \le 2^{1/6}\sigma \\ 0 & \text{if } r \ge 2^{1/6}\sigma \end{cases} \tag{11}$$

and the tail has the form

$$U^{(\mathrm{LJt})}(r) = \begin{cases} -\epsilon & \text{if } r \le 2^{1/6}\sigma \\ 4\epsilon[(\sigma/r)^{12} - (\sigma/r)^6] & \text{if } r \ge 2^{1/6}\sigma \end{cases} \tag{12}$$

It follows that the solute−hydrogen pair correlation function is

$$h_{\mathrm{H}}(\boldsymbol{r}) + 1 \approx \int \mathrm{d}\boldsymbol{r}'\, [\int \mathrm{d}\boldsymbol{X}\, S(\boldsymbol{X})\delta(\boldsymbol{r} - \boldsymbol{r}_{\mathrm{H}})\delta(\boldsymbol{r}' - \boldsymbol{r}_{\mathrm{O}})] \times \exp[-\beta U^{(\mathrm{LJs})}_{\mathrm{O}}(\boldsymbol{r}')] \approx \int \mathrm{d}\boldsymbol{r}'\, \omega_{\mathrm{OH}}(\boldsymbol{r} - \boldsymbol{r}') \exp[-\beta U^{(\mathrm{LJs})}_{\mathrm{O}}(\boldsymbol{r}')] \tag{13}$$

where $\omega_{\alpha\gamma}(\boldsymbol{r}) = \delta(|\boldsymbol{r}| - l_{\alpha\gamma})/4\pi l_{\alpha\gamma}^2$ is the site−site intramolecular correlation function.[24] Assuming that the water hydrogens do not carry a LJ interaction center (as in the TIP3 model[35]), the hydrogen bridge function is then simply $b_H(\boldsymbol{r}) \equiv -\ln[h_H(\boldsymbol{r}) + 1]$, with

$$b_H(\boldsymbol{r}) = -\ln[\omega_{OH} * \exp[-\beta U_O^{(LJs)}(\boldsymbol{r})]] \tag{14}$$

The convolution integral can be calculated easily using the transform of the intramolecular correlation function $\hat{\omega}_{\alpha\gamma}(k) = \sin(kl_{\alpha\gamma})/kl_{\alpha\gamma}$. Through this construction, an intramolecular correlation between the hydrogen and oxygen of the water molecule is built-in explicitly for short distances from the solute.

In addition to the hydrogen bridge function, it is also important to introduce a bridge function for the oxygen to obtain accurate solvation free energies. Without the oxygen bridge function the value of the solvent density in contact with the solute is slightly overestimated even close to nonpolar groups, which yields excess free energies $\Delta\mu_{np}$ that are systematically too positive. This problem is not particular to the 3d-RISM theory but is related to the well-known fact that the HNC equation overestimates the pressure.[14] Previous investigations of the bridge function for simple monatomic liquids indicate that it is a rapidly decaying function which plays the role of an effective repulsive potential depending on the radius of the solute.[42−45] For the sake of simplicity, we assume that the water oxygen bridge function consists of a superposition of rapidly decaying radial functions centered on the solute site s,

$$b_O(\boldsymbol{r}) = \sum_s A_s e^{-(|\boldsymbol{r}-\boldsymbol{r}_s| - \lambda_1\sigma_{Os})/y_1} \tag{15}$$

where the amplitude A_s takes the form

$$A_s = a_0 \frac{(\lambda_1\sigma_{Os}/y_0)}{1 + \lambda_1(\sigma_{Os}/y_0)} \tag{16}$$

By construction, the bridge function vanishes when $\lambda_1 = 0$, as it should. The form was chosen so that the importance of the bridge function increases smoothly to a maximum value of a_0 as the radius of the solute particle gets larger. The parameter y_0 controls how much the magnitude of the amplitude A_s varies as a function of the size of the solute particle. It was assigned a value of 2.8 Å, corresponding to the average distance between two oxygen atoms in liquid water. The parameter y_1 represents the decay length of the exponential function. Since bridge functions of simple liquids are effectively short range,[42−45] y_1 was assigned a value of 0.50 Å. The amplitude parameter a_0 controls the overall magnitude of the oxygen bridge function and has a significant influence on the calculated free energies. The optimization of a_0 is described in subsection d below.

(c) Susceptibility of Pure Liquid Water. The susceptibility response function of pure liquid water, $\chi_{\alpha\gamma}(r) = \bar{\rho}\omega_{\alpha\gamma}(r) + \bar{\rho}^2 h_{\alpha\gamma}(r)$ appears as an input in the 3d-RISM-HNC equation. In our previous study,[17] the response function was calculated as using the standard SSOZ-RISM equation[24]

$$h(r) = \omega * \omega(r) + \omega * c * \rho h(r) \tag{17}$$

extended with an HNC closure,[25]

$$h_{\alpha\gamma}(r) = \exp[-\beta u_{\alpha\gamma}^{(LJ)}(r) + A\phi_{\alpha\gamma}(r) + h_{\alpha\gamma}(r) - c_{\alpha\gamma}(r)] - 1 \tag{18}$$

where $\omega_{\alpha\gamma}$ is the intramolecular correlation function (see previous section) and $u_{\alpha\gamma}^{(LJ)}(r)$ and $\phi_{\alpha\gamma}(r) = -\beta q_\alpha q_\gamma/r$ are the site−site LJ and Coulomb potentials, respectively. This is the route that was commonly adopted by several studies of hydration based on the SSOZ-RISM.[26,27] The constant A serves to modify empirically the HNC closure such that the site−site radial distribution functions $h_{\alpha\gamma}(r)$ are consistent with the experimental values of the dielectric constant of water;[46] its value is given by $A = [1 + 1/(\epsilon - 1) - {}^1/_3 y]$, with $3y = 4\pi\beta\rho\mu^2/3$ (μ is the dipole moment of the water molecule in this expression). In the case of the TIP3P water model,[35] the numerical value of A is close to 1 (around 0.98) and has little influence on the pair correlation functions at short range.[26,27]

It is well-known that the average structure of liquid water based on eqs 17 and 18 has very serious deficiencies at short range.[25−27] For example, the first maximum in the oxygen−hydrogen radial distribution occurs at a distance of approximately 1.4 Å, which is much too short. The oxygen−hydrogen peak in the radial distribution is also overestimated. These problems are related to the lack of intramolecular correlation of the RISM equation, as described above. Clearly, it would be desirable to use a better susceptibility response as input in the 3d-RISM-HNC equation. One possibility would be to use the experimental structure of water.[47−50] However, the precision of the oxygen−oxygen, hydrogen−oxygen, and hydrogen−hydrogen pair correlation functions extracted by X-ray or neutron scattering is not sufficient to allow a stable convergence in numerical calculations. Although there is no problem in the case of a nonpolar solute (see the theory of the hydrophobic effect by Pratt and Chandler[51,52]), the integral equation cannot be solved in the case of a polar solute molecule due to the long-range solute−solvent electrostatic interaction. Fundamentally, this is due to the fact that the compressibility and the dielectric response of the pure liquid in a site−site theory correspond to a linear combination of zeroth and second moments of the pair correlation function.[53] A stable numerical theory thus requires that those moments be satisfied with a very high precision. An alternative would be to use the radial distribution function extracted from computer simulations. However, the precision of the pair correlation functions at large distance is again insufficient for constructing an accurate response function for numerical calculations.

We shall extract the water susceptibility function from the TIP3P model[35] rather than from experimental measurements. In so far as a model reproduces the structure of liquid water accurately, one would not expect any qualitative change in the present theory (though the results would be quantitatively different). Nonetheless, the current choice is mainly motivated by concerns about consistency. By extracting the water structure from a model, the solute−solvent interactions are completely consistent with the underlying solvent−solvent interactions that gave rise to the solvent structure. This is particularly important for electrostatic interactions, since the partial charges of the solvent model (TIP3P) are used also to set up the solute−solvent interactions. Furthermore, the current choice shall make it easier to compare the results from the integral equation with those from computer simulations in future work.

To obtain an accurate and dielectrically correct solvent response function, we adopted a simple empirical scheme. We used the radial pair correlation functions calculated from molecular dynamics simulations of a box of TIP3 water molecules in combination with the SSOZ-RISM integral equation with a HNC closure to constrain the short-range solvent structure. In other words, the integral equation is used to extrapolate the information from the MD to a long-range

structure. This was achieved using a modified HNC closure,

$$h_{\alpha\gamma}(r) = \exp[-\beta u_{\alpha\gamma}^{(LJ)}(r) + A\phi_{\alpha\gamma}(r) + h_{\alpha\gamma}(r) - c_{\alpha\gamma}(r) - b_{\alpha\gamma}^{(MD)}(r)] - 1 \quad (19)$$

By construction, the function $b_{\alpha\gamma}^{(MD)}(r)$ constrains the radial distribution functions to follow the results extracted from the MD simulation at the short distances r,

$$b_{\alpha\gamma}^{(MD)}(r) = S(r) \times \{-\ln[g_{\alpha\gamma}^{(MD)}(r)] - \beta u_{\alpha\gamma}^{(LJ)}(r) + h_{\alpha\gamma}(r) - c_{\alpha\gamma}(r)]\} \quad (20)$$

where $g_{\alpha\gamma}^{(MD)}(r)$ is the radial distribution function extracted from the MD simulations, and the function $S(r)$ is an empirical switching function. It is equal to 1 for $r < 12$ Å, going smoothly to 0 at $r = 14$ Å; a simple third-order polynomial was used for the switching. This empirical procedure ensures that the short-range structure matches that of the MD simulation while the long-range electrostatic response is determined by the integral equation. A similar extrapolation procedure was proposed by Verlet to extend the radial distribution function of a simple liquid.[54] A standard iterative scheme with a mixing factor of 0.33 was used to solve eqs 19 and 20. The integral equation was iterated until the difference in the radial distribution was negligible. The $g_{\alpha\gamma}^{(MD)}(r)$ radial distribution functions of liquid water of the O−O, O−H, and H−H pairs were obtained from a 100 ps molecular dynamics simulation of a cubic box of 794 TIP3 water molecules, performed at a constant temperature of 298.15 K and a pressure of 1 atm using the program CHARMM.[37] The long-range electrostatic interactions were treated with the particle mesh Ewald;[55] a cubic grid of 24^3 points with a fourth-order B-spline was used for computing the electrostatic potential. The rigid 3-site TIP3P model water model was used.[35] The charges of the oxygen and hydrogen are $q_O = -0.834$ e and $q_H = 0.417$ e, respectively. The LJ parameters of the oxygen are $\sigma_{OO} = 3.1507$ Å and $e_{OO} = 0.152\,073$ kcal/mol. The length of the OH bond is $l_{OH} = 0.9572$ Å. Following the TIP3 model, no LJ interaction site was assigned to the hydrogens.[35] The radial distribution functions calculated from the trajectories and from the constrained integral equation are shown in Figure 1. They are virtually identical. No abrupt variations occur over the range of the switching region (not shown in Figure 1). By construction, the renormalization of the Coulomb interaction $A\phi_{\alpha\gamma}$ in eq 18 yields a dielectric constant of 78.40, consistent with the electrostatic response of liquid water.

(d) Computational Details. To solve the 3d-RISM-HNC equation, all potentials and correlation functions $f(\mathbf{r})$ defined in 3d space are mapped onto a discrete grid, that is, $f(x,y,z) \rightarrow f(i,j,k)$. A cubic grid with 128^3 points and a spacing of 0.225 Å corresponding to an elementary cell of side 28.8 Å was used in all the calculations. The average density of pure water is 0.0334 $Å^{-3}$, and the temperature is 298.15 K. The LJ parameters and partial atomic charges of the solute were taken directly from the PARAM22 force field of the CHARMM program[37] with the standard combination rule $\epsilon_{\alpha s} = \epsilon_{\alpha\alpha}\epsilon_{ss}$ and $\sigma_{\alpha,s} = (\sigma_{\alpha\alpha} + \sigma_{ss})/2$. Equations 4 and 5 were solved self-consistently using a standard iterative procedure with mixing. Iterations were performed in the real space over the direct correlation function $c_\alpha(r)$. The 3d spatial convolutions $c_\gamma * \chi_{\gamma\alpha}(\mathbf{r})$ in eq 5 were calculated using a 3d fast Fourier transform (3d-FFT).[56] For each interaction cycle there are two forward 3d-FFTs and one inverse 3d-FFT. The iterations are continued until the largest changes in the direct correlation functions $|c_\alpha^{(new)}(i,j,k) - c_\alpha^{(old)}(i,j,k)|$ are $<2.00 \times 10^{-4}$ at all points (i,j,k) on the grid.

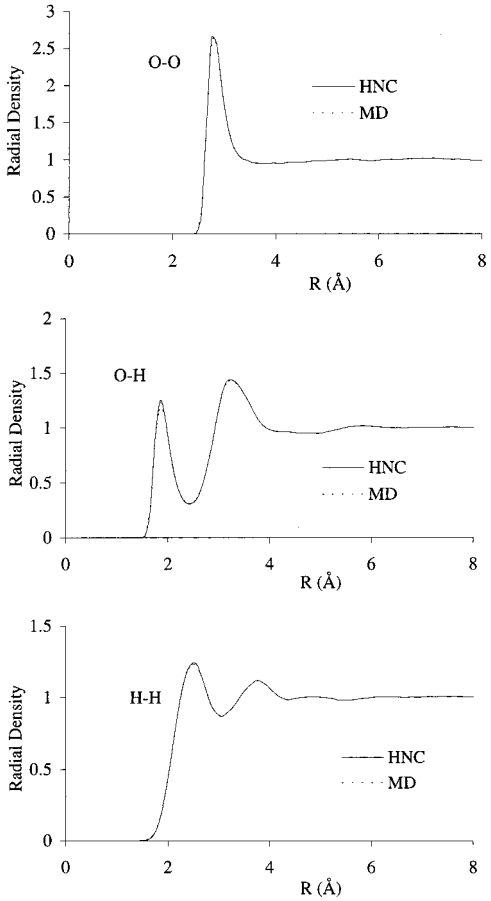

Figure 1. Radial density distribution of pure water obtained from molecular dynamics (MD) simulation (dots) and from the extrapolation scheme based on the radial RISM-HNC integral equation.

TABLE 1: Effects of Steps *n* and Increments of λ in Numerical Integration on Excess Chemical Potentials and Solute−Solvent Interaction Energies (kcal/mol)

	steps n	increment $\Delta\lambda$	$\langle U^{(LJ)}\rangle$	$\Delta\mu$
test 1	10	0.1×10	−1.330	2.671
test 2	$10 + 9$	$(0.01 \times 10) + (0.1 \times 9)$	−1.330	2.667
test 3	$10 + 9 + 9$	$(0.001 \times 10) + (0.01 \times 9) + (0.1 \times 9)$	−1.330	2.667
test 4	$5 + 10$	$(0.1 \times 5) + (0.05 \times 10)$	−1.330	2.656
test 5	$5 + 20$	$(0.1 \times 5) + (0.025 \times 20)$	−1.330	2.652

The thermodynamic integral over the coupling parameters λ_1 and λ_2 in eqs 8 and 9 was calculated using a simple integration scheme based on the trapezoidal rule. The calculation yields the nonpolar and electrostatic contributions to the excess chemical potentials of the solute at infinite dilution, $\Delta\mu_{np}$ and $\Delta\mu_{elec}$. Several values of the coupling parameter λ_1 are needed for an accurate thermodynamic integration, which is a time-consuming process. To use less CPU time, a suitable sequence of increments for λ_1 yielding reliable results was determined. As a test, a neutral spherical LJ particle with $\sigma = 2.7165$ Å and $\epsilon = 0.1521$ kcal/mol was considered. Table 1 gives the results of the test. In each case, a different increment of λ_1 and number of numerical integration steps were used. The calculations show that the resulting excess free energy $\Delta\mu_{np}$ is well-converged with an integration using 10 increments with $\Delta\lambda_1 = 0.1$. The calculations for the electrostatic excess free energy $\Delta\mu_{elec}$ were performed with an integration using 5 increments with $\Delta\lambda_2 = 0.2$.

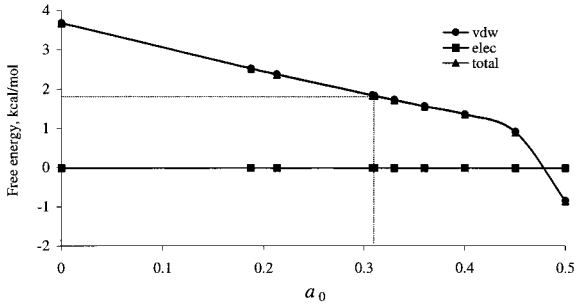

Figure 2. Optimization of the amplitude parameter a_0 of the oxygen bridge function based on the experimental excess chemical potential of ethane (1.83 kcal/mol (ref 57)).

TABLE 2: Solvation of Alkanes (All Energies Are in kcal/mol)

molecule	$\langle U^{(LJ)}\rangle$	$\langle U^{(elec)}\rangle$	$\langle U\rangle$	$\Delta\mu_{np}$	$\Delta\mu_{elec}$	$\Delta\mu_{tot}$	$\Delta\mu_{exp}$
3d-RISM-HNC (no Bh and Bo)							
methane	0.084	0.352	0.436	7.562	0.2530	7.815	2.00[a]
ethane	−4.613	−0.339	−4.952	10.37	−0.1778	10.20	1.83[a]
n-propane	−2.425	1.149	−1.276	13.65	0.7704	14.42	1.96[a]
n-butane	−3.539	1.337	−2.202	16.86	0.9316	17.79	2.08[a]
n-pentane	−4.436	1.626	−2.810	20.04	1.1279	21.17	2.33[a]
3d-RISM-HNC + Bh							
methane	−2.732	−0.075	−2.807	3.265	−0.0210	3.244	2.00[a]
ethane	−4.893	−0.119	−5.012	3.677	−0.0148	3.662	1.83[a]
n-propane	−6.673	−0.108	−6.781	4.522	0.0086	4.531	1.96[a]
n-butane	−8.442	−0.114	−8.556	5.269	0.0212	5.290	2.08[a]
n-pentane	−10.19	−0.133	−10.32	6.063	0.0273	6.090	2.33[a]
3d-RISM-HNC + Bh + Bo							
methane	−2.914	−0.0707	−2.985	2.105	−0.0209	2.084	2.00[a]
ethane	−5.111	−0.1108	−5.222	1.843	−0.0164	1.827	1.83[a]
n-propane	−6.927	−0.1004	−7.027	2.021	0.0031	2.024	1.96[a]
n-butane	−8.727	−0.1070	−8.834	2.123	0.0130	2.136	2.08[a]
n-pentane	−10.51	−0.1250	−10.64	2.274	0.0175	2.292	2.33[a]

[a] From ref 57.

The amplitude parameter a_0 controls the overall magnitude of the oxygen bridge function and has a significant influence on the calculated free energies. The value of the amplitude parameter a_0 in the oxygen bridge function was adjusted to reproduce the experimentally observed hydration free energy of ethane. Its optimization is shown in Figure 2. It is observed that the nonpolar contribution to the solvation free energy $\Delta\mu_{np}$ decreases with increasing a_0, with negligible influence on the electrostatic contribution $\Delta\mu_{elec}$. The optimal value of a_0 is 0.308 according to this procedure. The optimization is discussed further in section III.

III. Results and Discussion

Tables 2−5 give the excess chemical potentials and average solute−solvent interaction energies calculated using the 3d-RISM-HNC integral equation for *n*-alkanes, *n*-alcohols, *n*-carboxylic acids, and amides, respectively. The results obtained with and without hydrogen and oxygen bridge functions are given. All the experimental solvation free energies were taken from references 57 and 58. It is observed that the results obtained with no hydrogen and oxygen bridge functions deviate significantly from the experimental values. In the case of the *n*-alkanes, which are dominated by the nonpolar contribution, the solvation free energies are systematically overestimated. For example, the calculated free energy of methane is +7.8 kcal/mol, whereas the experimental value is +2.0 kcal/mol. The largest error occurs in the case of *n*-pentane, for which the free energy is overestimated by almost 1 order of magnitude (+21.2 kcal/mol instead of +2.3 kcal/mol). In the case of the polar solutes, the results with no bridge function are far worse than those for nonpolar solutes. For example, the calculated free energy of methanol is −108.3 kcal/mol whereas the experimental value is −5.1 kcal/mol. The decomposition of the solvation free energy shows that the largest error arises from the electrostatic contributions, which are unphysically negative for such a moderately polar solute. However, the free energy of decomposition reveals also that the nonpolar contribution is probably overestimated. The molecular size of methanol is similar to that of methane or ethane, and one would expect that the nonpolar contribution to the free energy to be on the order of +2.0 kcal/mol instead of +8.5 kcal/mol. The results with no bridge functions show that the *n*-carboxylic acids and amides follow a similar trend.

Inspection of the three-dimensional solvent distribution function around the solute shows that important problems arise from the solvent hydrogen site. The solvent structure around ethane and ethanol is shown in Figures 3 and 4. The radial distribution functions of the water oxygen, $\langle\rho_O(r)\rangle$, and hydrogen, $\langle\rho_H(r)\rangle$, around the solute sites were calculated from a spherical averaging of the full solvent density $\langle\rho_O(x,y,z)\rangle$ and $\langle\rho_H(x,y,z)\rangle$ obtained by the 3d-RISM-HNC integral equation. The radial distribution functions show that regions of high solvent hydrogen densities are found near centers of the solute such as carbon and oxygen. Figure 3 shows that the water hydrogen density is unphysically high near the carbons even in the case of a nonpolar molecule such as ethane. Such a high hydrogen density in those regions is inconsistent with the geometry of the water molecule because a corresponding oxygen density, which should be located within 1.0 Å, is not observed. Similar problems were observed previously in the case of liquid water[25] and monatomic solutes.[26,27] The significant inaccuracies in the water hydrogen density are related to the lack of intramolecular correlation in the current integral equation. In part, the deficiency of the theory can be qualitatively corrected by modifying the parameter σ in the LJ potential involving the solute interaction centers and the water hydrogen.[17] Nonetheless, it is difficult to generalize such an ad hoc procedure to an arbitrary solute. To cure the worse aspect of the theory, Cortis et al.[19] proposed to introduce a short-range hydrogen bridge function defined by eq 14. In effect, this is equivalent to adding an effective repulsive potential between the solute and the water hydrogens. As shown in eq 14, the effective potential is a direct consequence of the repulsion acting on the water oxygen and the fact that the oxygen and the hydrogen are bonded. The resulting distributing functions calculated with the 3d-RISM-HNC theory augmented by the hydrogen bridge function $B_H(\boldsymbol{r})$ are shown in Figures 3 and 4. In all cases the unreasonable hydrogen solvent densities are corrected. As shown in Tables 2−5, the calculated thermodynamic properties are also in much better agreement with experimental data. In particular, the free energy of ethane has changed from 10.20 kcal/mol, with no bridge functions, to 3.7 kcal/mol, with the simple hydrogen bridge function constructed according to eq 14. Likewise, the free energy of methanol has changed from −108.3 kcal/mol, with no bridge functions, to −2.5 kcal/mol, with a hydrogen bridge function. In the case of all polar solutes, the unphysically large negative electrostatic contributions to the free energy are corrected.

Despite the significant improvement, the solvation free energies of all the solutes are slightly too positive. For example, the solvation free energy of ethane is 3.66 kcal/mol whereas the experimental value is 1.86 kcl/mol; the solvation free energy of ethanol is −1.5 kcal/mol whereas the experimental value is −5.1 kcl/mol. In part, the origin of the problem can be found

TABLE 3: Solvation of Alcohols (All Energies Are in kcal/mol)

molecule	$\langle U_{(LJ)}\rangle$	$\langle U^{(elec)}\rangle$	$\langle U\rangle$	$\Delta\mu_{np}$	$\Delta\mu_{elec}$	$\Delta\mu_{tot}$	$\Delta\mu_{exp}$
			3d-RISM-HNC (with no Bh and Bo)				
methanol	5.612	−274.3	−268.7	8.483	−116.8	−108.3	−5.11[a]
ethanol	4.380	−309.5	−305.1	11.63	−120.7	−109.1	−5.01[a]
n-propanol	3.641	−337.7	−334.1	14.92	−131.2	−116.3	−4.83[a]
n-butanol	2.857	−375.4	−372.5	18.15	−141.6	−123.4	−4.74[a]
n-pentanol	1.850	−399.0	−397.6	21.08	−146.8	−125.7	−4.47[a]
			3d-RISM-HNC + Bh				
methanol	−3.060	−12.99	−16.05	3.067	−5.553	−2.486	−5.11[a]
ethanol	−4.906	−12.78	−17.69	3.816	−5.362	−1.546	−5.01[a]
n-propanol	−6.722	−12.12	−18.84	4.656	−5.080	−0.424	−4.83[a]
n-butanol	−8.456	−12.52	−20.98	5.440	−5.227	0.213	−4.74[a]
n-pentanol	−10.52	−13.55	−24.07	6.005	−5.783	0.222	−4.47[a]
			3d-RISM-HNC + Bh + Bo				
methanol	−3.367	−12.61	−15.98	1.740	−5.374	−3.634	−5.11[a]
ethanol	−5.262	−12.37	−17.64	1.790	−5.165	−3.375	−5.01[a]
n-propanol	−7.111	−11.72	−18.83	1.965	−4.904	−2.939	−4.84[a]
n-butanol	−8.866	−12.12	−20.99	2.111	−5.050	−2.939	−4.74[a]
n-pentanol	−10.79	−11.98	−22.77	2.076	−4.990	−2.914	−4.47[a]

[a] From ref 57.

TABLE 4: Solvation of Carboxylic Acids (All Energies Are in kcal/mol)

molecule	$\langle U^{(LJ)}\rangle$	$\langle U^{(elec)}\rangle$	$\langle U\rangle$	$\Delta\mu_{np}$	$\Delta\mu_{elec}$	$\Delta\mu_{tot}$	$\Delta\mu_{expt}$
			3d-RISM-HNC (with no Bh and Bo)				
acetic acid	6.883	−352.4	−345.5	11.64	−123.1	−111.5	−6.70[a]
propionic acid	6.842	−393.8	−387.0	15.00	−137.8	−122.8	−6.47[a]
butanoic acid	4.853	−408.8	−403.9	18.24	−143.1	−124.8	−6.38[a]
pentanoic acid	4.191	−448.6	−444.4	21.46	−153.1	−131.6	−6.4[b]
hexanoic acid	2.792	−463.4	−460.6	24.76	−157.0	−132.3	−6.2[b]
			3d-RISM-HNC + Bh				
acetic acid	−5.113	−13.91	−19.02	3.533	−5.882	−2.349	−6.70[a]
propionic acid	−6.862	−13.51	−20.37	4.423	−5.812	−1.389	−6.47[a]
butanoic acid	−8.680	−13.10	−21.78	5.220	−5.671	−0.451	−6.38[a]
pentanoic acid	−10.37	−13.33	−23.70	6.065	−5.747	0.318	−6.4[b]
hexanoic acid	−12.14	−13.33	−25.47	6.954	−5.754	1.200	−6.2[b]
			3d-RISM-HNC + Bh + Bo				
acetic acid	−5.539	−12.87	−18.41	1.608	−5.579	−3.971	−6.70[a]
propionic acid	−7.257	−13.03	−20.29	1.822	−5.561	−3.739	−6.47[a]
butanoic acid	−9.106	−12.62	−21.73	1.982	−5.419	−3.437	−6.38[a]
pentanoic acid	−10.83	−12.85	−23.68	2.182	−5.498	−3.316	−6.4[b]
hexanoic acid	−12.63	−12.86	−25.49	2.416	−5.509	−3.093	−6.2[b]

[a] From ref 57. [b] From ref 58.

TABLE 5: Solvation of Amides (All Energies Are in kcal/mol)

molecule	$\langle U^{(LJ)}\rangle$	$\langle U^{(elec)}\rangle$	$\langle U\rangle$	$\Delta\mu_{np}$	$\Delta\mu_{elec}$	$\Delta\mu_{tot}$	$\Delta\mu_{exp}$
			3d-RISM-HNC (with no Bh or Bo)				
ACEM	4.820	−372.6	−367.8	11.62	−130.5	−118.9	−9.75[a]
FORM	7.210	−371.7	−364.5	8.634	−137.8	−129.2	−4.56[a]
DMFA	−2.021	−266.1	−268.1	15.17	−102.2	−87.03	−8.54[a]
trans-NMA	−0.502	−328.0	−328.5	14.33	−114.6	−100.3	−10.07[a]
cis-NMA	−0.305	−322.9	−323.2	14.19	−114.6	−100.4	
			3d-RISM-HNC + Bh				
ACEM	−5.916	−20.12	−26.04	4.232	−8.790	−4.558	−9.75[a]
FORM	−3.862	−18.81	−22.67	3.136	−8.455	−5.319	−4.56[a]
DMFA	−8.496	−21.57	−30.07	5.633	−8.916	−3.283	−8.54[a]
trans-NMA	−8.718	−19.47	−28.19	4.968	−8.276	−3.308	−10.07[a]
cis-NMA	−8.730	−17.50	−26.23	4.830	−7.518	−2.688	
			3d-RISM-HNC + Bh + Bo				
ACEM	−6.357	−19.59	−25.95	1.525	−8.519	−6.994	−9.75[a]
FORM	−4.240	−18.28	−22.52	1.382	−8.194	−6.812	−4.56[a]
DMFA	−8.952	−21.12	−30.07	1.688	−8.689	−7.001	−8.54[a]
trans-NMA	−9.189	−18.96	−28.15	1.142	−8.016	−6.874	−10.07[a]
cis-NMA	−9.176	−17.07	−26.25	1.062	−7.298	−6.236	

[a] From ref 57.

by observing that the nonpolar contribution to the solvation free energy is overestimated in the case of each solute. This error arises from the fact that the HNC closure systematically overestimates the pressure and thus the free energy for creating a cavity.[14] In the formal diagrammatic expansion of the pair correlation of simple liquids, it is shown that the HNC equation

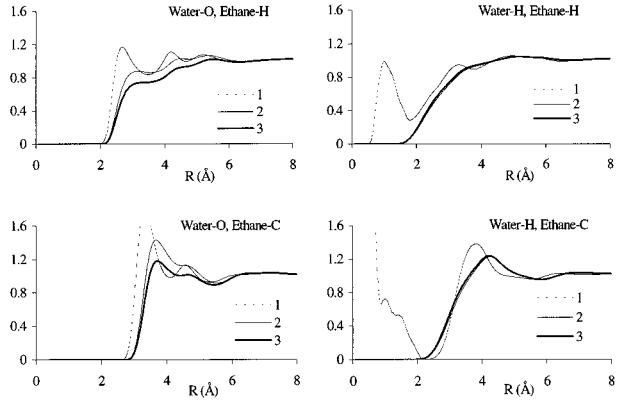

Figure 3. Solvation of ethane. The radial average density distributions of water oxygen $\langle\rho_O(r)\rangle$ and hydrogen $\langle\rho_H(r)\rangle$ around the carbon and hydrogen calculated with the 3d-RSIM-HNC integral equations with and without hydrogen and oxygen bridge functions are shown. 3d-RISM-HNC with no Bh on Bo, dash line; 3d-RISM-HNC + Bh, thin solid line; 3d-RISM-HNC + Bh + Bo, thick solid line.

is obtained by ignoring completely a whole class of diagrams referred to as "bridge diagrams" because of their special highly linked diagrammatic structure.[14] Previous investigations of simple monatomic liquids indicated that those diagrams correspond to a rapidly decaying function which plays the role of an effective short-range repulsive potential.[42−45] The form of the bridge function is generally thought to be rather insensitive to the detail of the potential.[42−45] For example, in the R−HNC theory, the bridge function of a hard sphere liquid (extracted from computer simulations) is used to compute the structures of polar liquids.[59] For the sake of simplicity, we chose a water oxygen bridge function consisting of a superposition of rapidly decaying exponential radial functions centered on the solute interaction site s given by eqs 15 and 16. The amplitude a_0 was optimized to reproduce the experimental solvation free energy of ethane. The optimization procedure is shown in Figure 2. Remarkably, the solvation free energy of all *n*-alkanes calculated from the 3d-RISM-HNC integral equation augmented by the hydrogen and the oxygen bridge functions is now in excellent accord with the experimental values. In particular, the fact that the free energy of methane is slightly larger than that of ethane is reproduced. The energy decomposition given in Table 2 shows that the change in free energy from methane to ethane results from a balance opposing unfavorable free energy for creating a cavity and the favorable van der Waals dispersion interaction. It should be stressed that although the thermodynamic properties are quite sensitive to the value of a_0, the solvent structure around the solute is affected very slightly. As shown in Figures 3 and 4, the solvent density is nearly unchanged with and without the oxygen bridge function. As a further illustration of the influence of the oxygen bridge function, various nonpolar solutes composed of six carbons were considered: hexane, cyclohexane and benzene. Table 6 gives the excess solvation free energies and solute−solvent interaction energies for those solutes. The results are in good accord with the experimental data and the well-known fact that aromatic hydrocarbons and cycloalkanes are less hydrophobic than their equivalent saturated hydrocarbons.[60] However, the calculated excess free energy of benzene has a positive sign, which suggests that further improvement might be required. It is also possible that the discrepancy arises from the lack of induced polarization in the benzene cycle in the microscopic potential function.

The calculations show that the hydrogen and oxygen bridge functions improve significantly the accord with experimental data in the case of all solutes (nonpolar and polar). However,

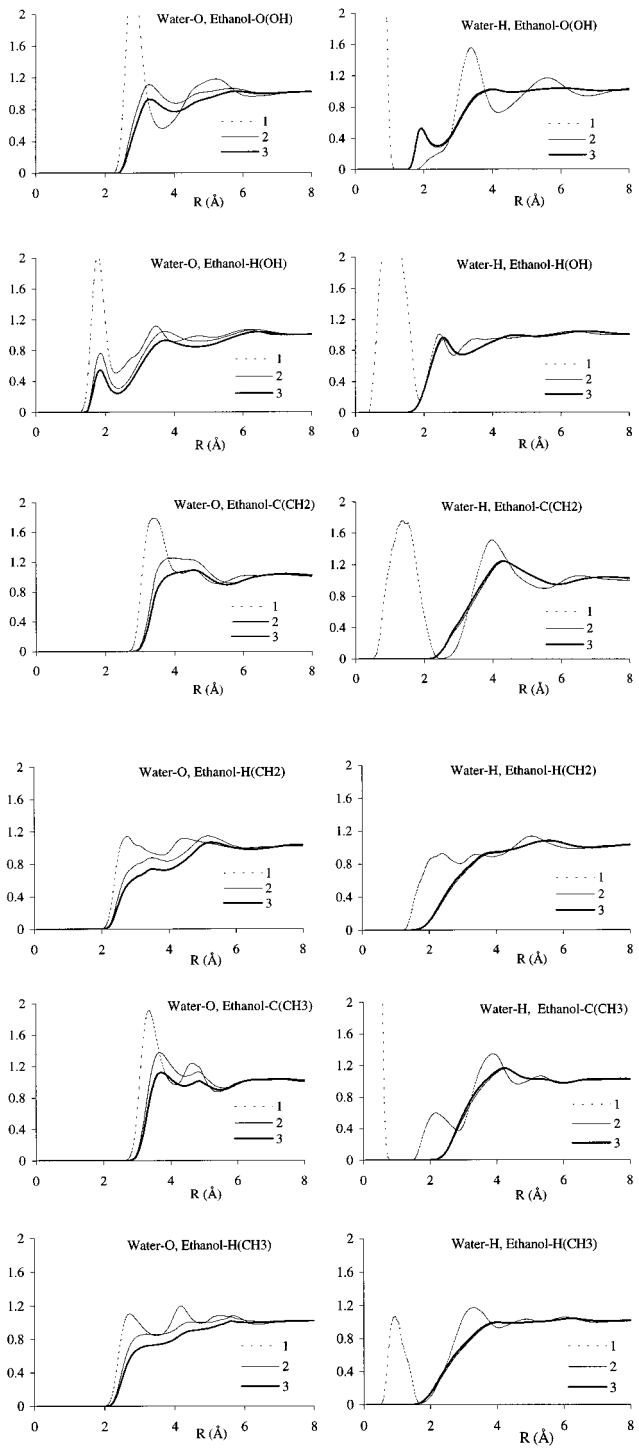

Figure 4. Solvation of ethanol. The radial average density distributions of water oxygen $\langle\rho_O(r)\rangle$ and hydrogen $\langle\rho_H(r)\rangle$ around the carbon and hydrogen calculated with the 3d-RSIM-HNC integral equations with and without hydrogen and oxygen bridge functions are shown. Symbols have the same meaning as in Figure 3.

there remain some differences between the calculated and experimental excess chemical potentials. In particular, the solvation free energies of the polar molecules are not in quantitative agreement with the experimental values. The free energy decomposition suggests that the discrepancies arise from underestimated electrostatic contributions. The origin of the problem is due to the hydrogen bridge function, which ignores the polarities of the solute and of the water molecule (see eq 14 above). For example, the calculated free energy of ethanol is −3.8 kcal/mol, arising from +1.8 kcal/mol from the nonpolar

TABLE 6: Excess Free Energies and Solute−Solvent Interaction Energies of Hexane, Cyclohexane, and Benzene (All Energies Are in kcal/mol)

molecule	$\langle U^{(LJ)}\rangle$	$\langle U^{(elec)}\rangle$	$\langle U\rangle$	$\Delta\mu_{np}$	$\Delta\mu_{elec}$	$\Delta\mu_{tot}$	$\Delta\mu_{exp}$
hexane	−12.29	−0.1378	−12.43	2.564	0.0229	2.587	2.5[a]
cyclohexane	−11.14	−0.0692	−11.21	1.828	0.0430	1.871	1.2[a]
benzene	−9.883	−2.736	−12.22	1.271	−1.0888	0.182	−0.9[a]

[a] From ref 57.

TABLE 7: Solvation with $a_O = 0.397$ and $a_H = 0.862$ (All Energies Are in kcal/mol)

molecule	$\langle U^{(LJ)}\rangle$	$\langle U^{(elec)}\rangle$	$\langle U\rangle$	$\Delta\mu_{np}$	$\Delta\mu_{elec}$	$\Delta\mu_{tot}$	$\Delta\mu_{exp}$
ethane	−5.129	−0.123	−5.252	1.851	−0.019	1.832	1.83[a]
ethanol	−5.066	−17.83	−22.90	2.006	−6.998	−4.992	−5.01[a]

[a] From ref 57.

contribution and −5.2 kcal/mol from the electrostatic contribution. The experimental value is −5.01 kcal/mol, and the error is +1.6 kcal/mol. Because the nonpolar contribution to the free energy of ethanol has a reasonable order of magnitude (relative to ethane, for example), the discrepancy necessarily arises from the lack of electrostatic interaction between the solute and the solvent. To match the experimental value for ethanol, the electrostatic contribution to the free energy should be on the order of −7 kcal/mol. Preliminary calculations indicate that further improvements can be obtained by simultaneously optimizing the hydrogen and oxygen bridge functions. To illustrate the situation, let us introduce an empirical parameter a_H multiplying the hydrogen bridge function $B_H(\boldsymbol{r})$. Simultaneous optimization of the amplitudes a_0 and a_H yields excess free energies which are compatible with the experimental data. The results are shown in Table 7. The amplitude a_H is 0.862, close to unity, and the amplitude for the oxygen bridge function a_0 is 0.397 instead of the original value 0.308, which was used throughout the remaining calculations. However, these values may not be unique and there may be other ways to optimize the bridge functions in order to match the experimental values. The amplitude of the hydrogen bridge function has a significant effect on the nonpolar and electrostatic contributions to the excess chemical potentials. Further adjustments of the parameters of the empirical bridge functions using an extensive "training set" of molecular solutes will be necessary for developing a quantitatively accurate 3d-RISM-HNC integral equation theory.

IV. Summary and Conclusion

In this paper we have presented new developments of the 3d-RISM-HNC integral equation for describing the solvation properties of complex molecules in water. The integral equation provides the average density of the solvent interaction site density in three dimensions (3d) around a complex molecule of arbitrary shape. The excess solvation free energies of a set of nonpolar (*n*-alkanes) as well as polar (*n*-alcohols, *n*-carboxylic acids, and amides) molecules at infinite dilution were calculated using thermodynamic integration and compared with experimental data. Empirical hydrogen and oxygen bridge functions were introduced to improve the accuracy of the 3d-RISM-HNC theory proposed originally.[17] In addition, the susceptibility of pure liquid water (an input in the theory) was constructed by extrapolating the site−site radial pair distribution function calculated from a molecular dynamics simulation using the SSOZ-RISM integral equations. In principle, one could extract the susceptibility of water at different pressures and temperatures in a similar fashion.

Although further work will be required for quantitatively accurate results, the calculated solvation free energies are in good accord with experimental data. The decomposition of the free energy in terms of cavity and electrostatic contributions highlighted the specific problems of the originally proposed 3d-RISM-HNC integral equation. In particular, the lack of intramolecular correlation leads to overestimated water hydrogen density in the neighborhood of negatively charged groups. In addition, the cavity free energy for all the molecules was too positive due to the overestimated pressure in the HNC closure. The worst aspects of those problems can be cured by introducing optimized solvent site bridge functions which act as short-range effective repulsive potentials.

The present results demonstrate that a well-calibrated 3d-RISM-HNC integral equation has the potential for serving as a useful intermediate approach between molecular dynamics free energy simulations with explicit solvent molecules and the traditional Poisson−Boltzmann continuum electrostatics method. Many of the essential aspects of the structure of liquid water are incorporated via the site−site susceptibility response function. The theory avoids the oversimplifications of the continuum electrostatic Poisson−Boltzmann approaches, since it does not represent the solvent as a structureless dielectric medium. On the other hand, despite its approximate nature, the integral equation does not suffer from the statistical convergence problems due to the lack of sampling that occurs with explicit solvent simulations. Nonetheless, the current results are not perfect and further work will be required to have a theory with quantitative prediction capability. At this point, it is necessary to undertake the rather technical task of optimizing the empirical bridge functions using a large "training set" of molecular solutes to develop a numerically useful theory. Our hope is that the 3d-RISM-HNC theory with well calibrated bridge functions will ultimately provide a useful tool for incorporating the influence of solvation in the modelization of biological molecules.

Acknowledgment. This work was supported by a grant from the NSERC of Canada and by Merck-Frosst Canada. B.R. is a Research Fellow of the Medical Research Council of Canada.

References and Notes

(1) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. In *Advances in Chemical Physics*; Prigogine, I., Rice, S. A., Eds.; John Wiley & Sons: New York, 1988; Vol. LXXI.
(2) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford Science Publications: Clarendon Press: Oxford, 1989.
(3) Roux, B.; Simonson, T. *Biophys. Chem.* **1999**, *78*, 1−20.
(4) Cramer, C. J.; Truhlar, D. G. In *Reviews in computational chemistry*; Lipkowitz, K. B., Ed.; VCH: New York, 1995.
(5) Gao, J.; Xia, X. *Science* **1992**, *258*, 631.
(6) Gao, J. *J. Am. Chem. Soc.* **1994**, *116*, 9324.
(7) Kirkwood, J. G. *J. Chem. Phys.* **1934**, *2*, 351.
(8) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1468−1493.
(9) Sharp, K. A.; Honig, B. *Annu. Rev. Biophys. Chem.* **1990**, *19*, 301.
(10) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
(11) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin. Trans.* **1993**, *2*, 799.
(12) York, D. M.; Lee, T. S.; Yang, W. *J. Am. Chem. Soc.* **1996**, *118*, 10940−10941.
(13) Jackson, J. D. John Wiley & Sons: New York, 1962.
(14) Hansen, J. P.; McDonald, I. R. *Theory of Simple Liquids*, 2nd ed.; Academic Press: London, 1986.
(15) Beglov, D.; Roux, B. *J. Chem. Phys.* **1995**, *103*, 360−364.
(16) Beglov, D.; Roux, B. *J. Chem. Phys.* **1996**, *104*, 8678−8689.
(17) Beglov, D.; Roux, B. *J. Phys. Chem. B* **1997**, *101*, 7821.
(18) Ikeguchi, M.; Doi, J. *J. Chem. Phys.* **1995**, *103*, 5011.
(19) Cortis, C. M.; Rossky, P. J.; Friesner, R. A. *J. Chem. Phys.* **1997**, *107*, 6400.
(20) Kovalenko, A.; Hirata, F. *Chem. Phys. Lett.* **1998**, *290*, 237−244.
(21) Kovalenko, A.; Hirata, F. *J. Chem. Phys.* **1999**, *110*, 10095−10112.

(22) Kovalenko, A.; Ten-No, S.; Hirata, F. *J. Comput. Chem.* **1999**, *20*, 928−936.
(23) Chandler, D.; McCoy, J. D.; Singer, S. J. *J. Chem. Phys.* **1986**, *85*, 5971.
(24) Chandler, D.; Andersen, H. C. *J. Chem. Phys.* **1972**, *57*, 1930−1937.
(25) Hirata, F.; Rossky, P. J. *Chem. Phys. Lett.* **1981**, *83*, 329−334.
(26) Yu, H. A.; Karplus, M. *J. Chem. Phys.* **1988**, *89*, 2366.
(27) Yu, H. A.; Roux, B.; Karplus, M. *J. Chem. Phys.* **1990**, *92*, 5020−5033.
(28) Pettitt, M. B.; Karplus, M.; Rossky, P. J. *J. Phys. Chem.* **1986**, *90*, 6335−6345.
(29) Pettitt, M. B.; Karplus, M. *Chem. Phys. Lett.* **1985**, *121*, 194−201.
(30) Karplus, M.; Pettitt, M. B.; Rossky, P. J. *J. Chem. Phys.* **1986**, *90*, 6335.
(31) Pettitt, M. B.; Karplus, M. *Chem. Phys. Lett.* **1987**, *136*, 383−386.
(32) Lau, W. F.; Pettitt, B. M. *Biopol.* **1987**, *26*, 1817−1831.
(33) Ramé, G.; Lau, W. F.; Pettitt, B. M. *Int. J. Peptide Protein Res.* **1990**, *35*, 315−327.
(34) Lee, P. H.; Maggiora, G. M. *J. Phys. Chem.* **1993**, *97*, 10175−10185.
(35) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(36) Weiner, S. T.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*,
(37) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(38) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*,
(39) Morita, T.; Hiroike, K. *Prog. Theor. Phys.* **1960**, *23*, 1003.
(40) Singer, S. J.; Chandler, D. *Mol. Phys.* **1985**, *55*, 621−625.
(41) Weeks, J. D.; Chandler, D.; Andersen, H. C. *J. Chem. Phys.* **1971**, *54*, 5237.
(42) Lado, F. *Phys. Rev. A* **1973**, *8*, 2548.
(43) Lado, F.; Foiles, S. M.; Ashcroft, N. W. *Phys. Rev. A* **1983**, *28*, 2374.
(44) Foiles, S. M.; Ashcroft, N. W.; Reatto, S. M. *J. Chem. Phys.* **1984**, *81*, 6140.
(45) Foiles, S. M.; Ashcroft, N. W.; Reatto, S. M. *J. Chem. Phys.* **1984**, *80*, 4441.
(46) Rossky, P. J.; Pettitt, B. M.; Stell, G. *Mol. Phys.* **1983**, *50*, 1263−1271.
(47) Egelstaff, P. A.; Root, J. H. *Chem. Phys.* **1983**, *76*, 405.
(48) Egelstaff, P. A.; Polo, J. A.; Root, J. H.; Hahn, L. J.; Chen, S. H. *Chem. Phys. Rev. Lett.* **1981**, *47*, 1733.
(49) Stanley, H. E.; Teixeira, J. J. *J. Chem. Phys.* **1980**, *73*, 3404.
(50) Sceats, M. G.; Storola, M.; Rice, S. A. *J. Chem. Phys.* **1979**, *70*, 3927.
(51) Pratt, L. W.; Chandler, D. *J. Chem. Phys.* **1977**, *67*, 3683.
(52) Pratt, L. W.; Chandler, D. *J. Chem. Phys.* **1980**, *73*, 3434.
(53) Chandler, D. *J. Chem. Phys.* **1977**, *67*, 1113−1124.
(54) Verlet, L. *Phys. Rev.* **1968**, *163*, 201.
(55) Ewald, P. *Ann. Phys.* **1921**, *64*, 253.
(56) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. Cambridge University Press: Cambridge, 1990.
(57) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. *J. Solution Chem.* **1981**, *10*, 563.
(58) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 16385.
(59) Fries, P. H.; Patey, G. N. A. *J. Chem. Phys.* **1985**, *82*, 429−440.
(60) Chandler, D. Continuum solvation models: classical and quantum mechanical implementations. In *Reviews in computational chemistry*; Lipkowitz, K. B., Ed.; VCH: New York, 1995.