

Coupled reference interaction site model/simulation approach for thermochemistry of solvation: Theory and prospects

Holly Freedman and Thanh N. Truong^{a)}

Department of Chemistry, Henry Eyring Center for Theoretical Chemistry, University of Utah, Salt Lake City, Utah 84112

(Received 17 November 2003; accepted 20 April 2004)

We present a new methodology for computing solvation free energy, which is based upon the reference interaction site model (RISM)/hypernetted chain (HNC) solvation free energy expression, but which substitutes radial distribution functions taken from simulations for those calculated by simultaneous solution of the RISM and HNC equations. Consequently, solvation free energy can be obtained from a single molecular dynamics or Monte Carlo simulation. Here we describe in detail the coupled RISM/simulation approach, and offer some error analysis. Finally we give the results of its application to a set of small test molecules in aqueous solution. The success shown in some of our results demonstrates that the coupled RISM/simulation approach is worth considering further as a potentially useful tool in studies of solvated systems, such as aqueous molecular biosystems.

© 2004 American Institute of Physics. [DOI: 10.1063/1.1760741]

I. INTRODUCTION

The reference interaction site model (RISM) integral equation theory has received a large amount of interest over the past three decades as a means to calculate the solvation structure surrounding a solute molecule, as described by radial distribution functions.¹⁻⁴ Still more importantly, by employing the hypernetted chain (HNC) closure relation, an expression has been derived for the solvation free energy as a function of these solvent distributions.⁵⁻⁷ Solvation free energy constitutes an important, but difficult-to-find quantity, posing a challenge to computational chemists. This approach provides an alternative to the more laborious free-energy simulation methodology, and moreover one which, like these simulations, depicts the solvent structure, thus presenting an advantage over the popular continuum solvation models, which instead require that the solvent be described by an extensive set of parameters.⁸ However unlike free-energy simulations, RISM calculations are relatively computationally cheap.⁹⁻¹¹

The determination of a solvation free-energy difference by free-energy simulation requires that many simulations be run as one solvated molecule is gradually “mutated” into another, with explicit modeling of all solute and solvent atoms. A value between 0 and 1 of the coupling parameter, which describes the fractional character of the first molecule at each point along the path and is denoted by λ , is associated with each simulation. Free-energy differences associated with the concomitant displacement of water molecules effected during this transformation can be used to calculate the relative solvation free energies of the two molecules as an integral over the coupling parameter in either the thermodynamic integration or the free-energy perturbation approach. Reasonable, and sometimes quite accurate, relative solvation free energies can be determined by this method. However the process is quite complex, thus making it rather

impractical.^{8,11} On the other hand, the RISM/HNC formalism allows the integration over the coupling parameter to be performed analytically by expressing $d\Delta\mu/d\lambda$ as the exact differential of a function that is independent of λ .⁶ In this way, the complexity of a solvation free-energy calculation is reduced tremendously.

Although RISM appears to be a promising alternative to free-energy simulations, difficulties are encountered during its application. In the first place, it can be difficult to achieve convergence to a simultaneous solution of the RISM and closure relations, especially for molecules with many atomic sites such as proteins.¹² In addition, the accuracy of solvation free energies calculated from RISM can be somewhat discouraging.¹³

With these problems in mind, we have proposed that a simple alternative to the RISM/HNC solvation free-energy methodology might be to find accurate solvent distributions using a simulation, from these to solve the RISM and HNC equations for the direct correlation functions, which are also needed in a RISM free-energy calculation, and lastly to combine these functions as input to the RISM solvation free-energy expression. That is, radial distribution functions entering the solvation free-energy expression would come from simulation results, and the direct correlation functions utilized would be derived from these distributions by making use of the RISM and HNC equations.¹⁴

Other studies have been carried out mixing information obtained from RISM theory with that derived from simulations, but in a different framework and for a different purpose. These studies have utilized the RISM approach, rather, for its cost-efficient depiction of solvation structure within simulations modeling a solute’s dynamics. For example, a stochastic dynamics simulation was performed by Pettitt and co-workers in which RISM/calculated solute-solute correlation functions were compiled in a look-up table, which could be referenced to find solvent contributions to potentials of mean force that were combined with nonsolvated internal

^{a)} Author to whom correspondence should be addressed.

energy terms.¹⁵ Likewise both molecular dynamics (MD) and Monte Carlo (MC) simulations have been used to explore the configuration spaces of peptides, in which solvation free energies calculated by RISM were summed with conformational energies to obtain total free-energy profiles.^{12,16} However, in the approach we take here, RISM, instead of being used to supply the solvation structure, is taken advantage of for its simple expression for solvation free energy, while solvent distributions are taken from simulation trajectories. Kast has explored some ideas similar to our own.¹⁷

Here we clarify the procedure used to carry out a coupled RISM/simulation solvation free-energy calculation, focusing on solutes dissolved in water. In addition, we discuss the errors occurring as these calculations are performed. We also present the results of applying coupled RISM/MC to a few small neutral test molecules.

II. METHODOLOGY

In our approach, a simulation is used to determine the solvation structure, described by radial distribution functions around each solute site. This data is combined with information gained from the RISM and HNC equations to find the solvation free energy within the RISM formalism. Besides the radial distribution function, g , the second function needed as input in the RISM formulation of solvation free energy is the direct correlation function, c . Because g is determined from simulation results, we no longer need to apply an iterative strategy to solve the RISM and HNC equations simultaneously. Instead both of these equations can be employed separately for the determination of c .¹⁴

The RISM and HNC equations are given as

$$\hat{\mathbf{h}}(k) = \hat{\omega}(k)\hat{\mathbf{c}}(k)\hat{\omega}(k) + \rho\hat{\omega}(k)\hat{\mathbf{c}}(k)\hat{\mathbf{h}}(k), \quad (1)$$

$$g_{\alpha\gamma}(r) = \exp[-\beta u_{\alpha\gamma}(r) + h_{\alpha\gamma}(r) - c_{\alpha\gamma}(r)], \quad (2)$$

where $h \approx g - 1$ is the correlation function, the carets indicate spatial Fourier transforms, the boldface denotes matrices, $\hat{\omega}$ is the intramolecular correlation function, and u is the intermolecular pair potential. More precisely, if $l_{\alpha\beta}$ denotes the intramolecular separation between sites α and β , then $\hat{\omega}_{\alpha\beta}(k) = \delta_{\alpha\beta} + (1 - \delta_{\alpha\beta})(\sin kl_{\alpha\beta}/kl_{\alpha\beta})$.

When the HNC equation is solved for c , the resulting expression is

$$c_{\alpha\gamma}(r) = -\beta u_{\alpha\gamma}(r) + h_{\alpha\gamma}(r) - \ln[h_{\alpha\gamma}(r) + 1]. \quad (3)$$

Note, however, that c cannot be determined by this equation at small distances where g vanishes and thus the natural logarithm in the above expression loses meaning.

Using the RISM equation, c can be expressed in terms of h as

$$\hat{c}_{uv} = [\hat{\omega}_u]^{-1}[\hat{h}_{uv}][\hat{\omega}_v + \rho_v \hat{h}_{vv}]^{-1}. \quad (4)$$

For large \mathbf{r} , c is known to decay as \mathbf{r}^{-1} ; in particular,

$$c_{ij}(r) \rightarrow -\beta u_{ij}^c(r) \text{ as } r \rightarrow \infty, \quad (5)$$

where u^c is the Coulomb potential given by

$$u_{ij}^c(r) = \frac{q_i q_j}{r},$$

with q_i denoting the charge on site i . As a result of this slow decay of c , \hat{c} becomes unbounded as $k \rightarrow 0$, and thus c does not lend itself well to numerical Fourier transformation. This difficulty was formerly encountered when iterative techniques for solving the RISM/HNC equations were extended to polar molecules and ions. In the extended RISM approach (XRISM) the long-range Coulomb asymptotics of c are separated out and “resummed,” using analytic Fourier transformation, to yield their contribution to h .^{1,2} This process has been referred to as renormalization. We make use of the particular renormalization scheme proposed by Hirata *et al.*,¹⁸ in which c is expressed as the sum of long- and short-range components as

$$c_{uv} = c_{uv}^{sr} + \phi_{uv}, \quad (6)$$

where $\phi(r) = -\beta u^C(r)$.

Because, unlike c , the above function is divergent at $r = 0$, it is convenient to make the modification

$$\phi(r) = -\beta u^C(r) \operatorname{erf}\left(\frac{-r}{\tau}\right), \quad (7)$$

where τ is a constant best selected from the range -0.5 to -1.0 ; we used a value of $\tau = -0.8$.¹⁹

The resultant decomposition of \hat{h} is

$$\hat{h}_{uv} = \hat{\gamma}_{uv} + \hat{\omega}_u \hat{c}_{uv}^{sr} \{\hat{\omega}_v + \rho_v \hat{h}_{vv}\} \quad (8)$$

with

$$\hat{\gamma}_{uv} = \hat{\omega}_u \hat{\phi}_{uv} \{\hat{\omega}_v + \rho_v \hat{h}_{vv}\}. \quad (9)$$

Solving for c^{sr} we obtain

$$\hat{c}_{uv}^{sr} = [\hat{\omega}_u]^{-1} [\hat{h}_{uv} - \hat{\gamma}_{uv}] [\hat{\omega}_v + \rho_v \hat{h}_{vv}]^{-1}. \quad (10)$$

Note that two matrices must be inverted in the above expression for \hat{c} , namely the matrix $\hat{\omega}$ with dimensions equal to the number of sites on the solute, and the matrix, $\hat{\chi}_v = \hat{\omega}_v + \rho_v \hat{h}_{vv}$, with dimensions equal to the number of solvent atoms. Both of these matrices become ill conditioned near $k=0$, where all elements of each take on an identical value, leading to difficulties when evaluating Eq. (10).

The HNC expression for solvation free energy is given by^{6,7}

$$\Delta\mu^{(\text{HNC})} = \frac{\rho}{2\beta} \sum_{\alpha\gamma} \int 4\pi r^2 [-2c_{\alpha\gamma}(r) + h_{\alpha\gamma}^2(r) - h_{\alpha\gamma}(r)c_{\alpha\gamma}(r)] dr. \quad (11)$$

Note that the integral of the first term is a multiple of $\hat{c}(0)$. Although this term cannot be obtained from either the HNC or the RISM equation, fortunately, as we will describe below, the complete solvation free-energy expression may be evaluated by appropriately combining terms dependent upon two components of c , which are calculated separately from Eqs. (1) and (2).

A. Solvation free energy

In this section we formulate an improved expression for solvation free energy, $\Delta\mu$. In doing so it is recognized that the HNC and RISM equations are, in fact, inherently incompatible and thus modifications must be made to incorporate

the differences between their two separate solution sets. We will also discuss our handling of long-range boundary conditions.

We first considered how we could combine the solutions c to the RISM and HNC equations in an adaptation of the solvation free-energy expression that would allow us to calculate energies most accurately. To arrive at this adaptation, we considered the invariance principle satisfied by the free-energy expression, pointed out by Singer and Chandler.⁶ An invariance of the solvation free-energy expression about the simultaneous solution to the RISM and HNC equations was shown when this expression was rewritten as

$$\begin{aligned} \beta\Delta\mu = & \rho \sum_{\alpha,\gamma} \int dr \left\{ -\exp[-\beta u_{\alpha\gamma}(r) + h_{\alpha\gamma}(r) \right. \\ & \left. - c_{\alpha\gamma}(r)] + h_{\alpha\gamma}(r) + 1 - c_{\alpha\gamma}(r) - h_{\alpha\gamma}(r)c_{\alpha\gamma}(r) \right. \\ & \left. + \frac{1}{2} h_{\alpha\gamma}^2(r) \right\} \\ & + (2\pi)^{-3} \rho \int d\mathbf{k} \left\{ \sum_{\alpha,\gamma} \hat{c}_{\alpha\gamma}(k) \hat{h}_{\alpha\gamma}(k) \right. \\ & \left. - \frac{1}{2} \sum_{\alpha,\gamma,s,s'} \hat{c}_{\alpha\gamma}(k) \hat{c}_{ss'}(k) \hat{\omega}_{\alpha s}(k) \hat{\chi}_{\gamma s'}(k) \right\}. \quad (12) \end{aligned}$$

The differential of this equation with respect to c is⁶

$$\begin{aligned} \beta\delta\Delta\mu = & \rho \sum_{\alpha,\gamma} \int dr \left\{ \exp[-\beta u_{\alpha\gamma}(r) + h_{\alpha\gamma}(r) - c_{\alpha\gamma}(r)] \right. \\ & \left. - 1 - h_{\alpha\gamma}(r) \right\} \delta c_{\alpha\gamma}(r) \\ & + (2\pi)^{-3} \rho \int d\mathbf{k} \sum_{\alpha,\gamma} \left\{ \hat{h}_{\alpha\gamma}(k) \right. \\ & \left. - \sum_{s,s'} \hat{c}_{ss'}(k) \hat{\omega}_{\alpha s}(k) \hat{\chi}_{\gamma s'}(k) \right\} \delta \hat{c}_{\alpha\gamma}(k). \quad (13) \end{aligned}$$

Note that the terms in the brackets, respectively, vanish when the HNC and RISM equations are satisfied by h and c . A natural decomposition of $\Delta\mu$ is now evident. Namely, to preserve the invariant form of Eq. (13), we substitute the solutions c to the HNC and RISM equations, respectively, into the two brackets in Eq. (12) and condense terms to write

$$\begin{aligned} \Delta\mu = & \frac{\rho}{2\beta} \sum_{\alpha,\gamma} \int 4\pi r^2 \left\{ -2c_{\alpha\gamma'}^{\text{HNC}}(r) [1 + h_{\alpha\gamma}(r)] \right. \\ & \left. + [h_{\alpha\gamma}(r)]^2 + h_{\alpha\gamma}(r) c_{\alpha\gamma}^{\text{RISM}}(r) \right\} dr, \quad (14) \end{aligned}$$

where c^{HNC} satisfies the HNC equation and c^{RISM} satisfies the RISM equation. The invariance satisfied by the expression for the solvation free energy implies that although the direct correlation function being used may be approximate, yet our approximation can lead to only a minor error in our calculated solvation free energy.

Upon examination of this expression for solvation free energy, its import becomes apparent: this decomposition allows the data available for the direct correlation functions from the RISM and HNC equations to complement each

other. The range of r for which c^{HNC} is unknown matches the condition $1+h=0$, and thus contributes nothing to the first term in the free-energy expression. Moreover, the last term, given by a multiple of the integral of the product of h and c^{RISM} , is proportional to the integral of the product of the Fourier transforms \hat{h} and \hat{c}^{RISM} over k . Integration over three-dimensional space contributes a factor of k^2 to this term, which reduces the error associated with the uncertainty in \hat{c} at low k .

The Gaussian fluctuation (GF) method does not depend upon the HNC equation for its derivation for solvation free energy. Instead the direct correlation function c is treated as the linear response function of the probability distribution to a perturbation. The probability distribution is assumed to be Gaussian. This expression can be derived either by fixing one solute molecule and one solvent molecule and letting the rest of the solvent perturb the system, or by representing the pure solvent as perturbed in a linear response to the solute molecule. Either way, the expression obtained is²⁰

$$\begin{aligned} \Delta\mu^{(\text{GF})} = & \frac{\rho}{2\beta} \sum_{\alpha,\gamma} \int 4\pi r^2 \left[-2c_{\alpha\gamma}(r) \right. \\ & \left. - h_{\alpha\gamma}(r) c_{\alpha\gamma}(r) \right] dr. \quad (15) \end{aligned}$$

Recently Ten-no has shown that there is a serious flaw in the HNC expression (11) for solvation free energy; namely, its value has an artificial dependence upon the number of dummy solute sites added.²¹ The HNC approach approximates the solvation free energy as a sum over all pairs of sites of excess chemical potential terms due to each site. Each such term is found from the product of the particular site-site correlation function with the pair potential, with the positions of all other sites being averaged over

$$-\beta\Delta\mu^{\text{HNC}} = \rho \sum_{\alpha,s} \int_0^1 d\lambda \int d\mathbf{r} \left[-\beta u_{\alpha s}(r) g_{\alpha s}(r;\lambda) \right]. \quad (16)$$

However although we may assume that the potential itself is pair decomposable, yet the radial density of a particular site is correlated to the density of other sites in the same molecule. When the angular dependence of this density is preaveraged in forming $g_{\alpha s}$, the accuracy of calculated free energies is compromised. In an alternate solvation free-energy expression, Ten-no treats the solvation free energy as arising from the potential energy, rather, of the entire solute molecule interacting with solvent molecules

$$-\beta\Delta\mu^{\text{PW}} = \rho \int d\mathbf{r}_{1,2} \left\langle \int_0^1 d\lambda \left[-\beta U(\mathbf{r}_{1,2}) G(\mathbf{r}_{1,2};\lambda) \right] \right\rangle_{\hat{\Omega}_1 \hat{\Omega}_2}, \quad (17)$$

where the average is taken over all angular orientations of solute and solvent molecules. Integration over the coupling parameter is performed prior to decomposition into site-site contributions. "Partial wave" relations are then used to obtain the solvation free-energy expression,^{21,22}

$$\Delta\mu^{(\text{PW})} = \frac{\rho}{2\beta} \sum_{\alpha\gamma} \int 4\pi r^2 \{ -2c_{\alpha\gamma}^{\text{HNC}}(r)[1+h_{\alpha\gamma}(r)] + h_{\alpha\gamma}(r)h_{\alpha\gamma}^{\text{PW}}(r) + h_{\alpha\gamma}(r)c_{\alpha\gamma}^{\text{RISM}}(r) \} dr,$$

where

$$\hat{h}_{sv}^{\text{PW}} = [\hat{\omega}_s]^{-1} [\hat{h}_{sv}] [\hat{\omega}_v]^{-1}. \quad (18)$$

Note that the second term in this integral substitutes for the term

$$\frac{\rho}{2\beta} \sum_{\alpha\gamma} \int 4\pi r^2 [h_{\alpha\gamma}(r)]^2 dr$$

in the expression for $\Delta\mu^{(\text{HNC})}$ [see Eq. (11)] and incorporates, through the matrices $\hat{\omega}$, structural information about both solute and solvent molecules. This second term is absent altogether in the Gaussian fluctuation expression for $\Delta\mu$.

Because of the thoroughness shown in the derivation of the partial wave (PW) free-energy expression when compared to the derivations of the HNC and GF expressions, which make rather broad assumptions, the PW expression is expected to yield the most accurate energies. We will confirm this prediction with a comparison of these three methods in the results section, Sec. IV. For this reason, we take advantage of the PW free-energy formulation for the results reported in this paper.

1. Bridge-function corrections

It has been pointed out that it is in fact impossible to simultaneously satisfy both the HNC and RISM equations. The HNC equation is based upon topological reduction of the diagrammatic expansion for the correlation function, h . However starting with the RISM equation, an attempt to impose a diagrammatic reduction leads to the diagrammatic expansion of c as the sum of those diagrams in h that have no nodal circles. This, however, is not consistent with the RISM equation, but rather leads to more complex systems of equations, such as the Chandler–Silbey–Ladanyi integral equations.²³ It follows that no such topological reduction of the RISM equation can exist. Thus, in order to achieve accurate results, it is essential to make an appropriate correction to the solvation free-energy expression using the bridge function to account for the differences between solutions to the HNC and RISM equations.

By going about the solution of the RISM equation starting from a prior knowledge of h , we obtain insight into where the discrepancy between c^{RISM} and c^{HNC} is most significant. At low k , the matrices in the RISM equation are not only inexact in relation to the HNC equation, but also their associated condition numbers converge to ∞ as $k \rightarrow 0$. It follows that the solution \hat{c} to the RISM equation is not at all accurate at low k , relative to the HNC relation. To see this, consider the standard result of numerical analysis stating that an upper bound to the relative error Δx resulting from evaluating the solution $x = A^{-1}b$ to the matrix equation $Ax = b$ is given by $\text{cond}(A) \|E\| / \|A\|$, where the matrix E represents the error in the matrix A .²⁴ Now let $\hat{\omega}^{\text{exact}}$ represent the matrix which, when substituted for $\hat{\omega}$ in the RISM equation, would provide consistency with the HNC equation. Then, in multi-

plying by the matrix inverse $\hat{\omega}^{-1}$ when solving the RISM equation [see Eq. (10)], the incongruity between the two matrices $\hat{\omega}^{\text{exact}}$ and $\hat{\omega}$ is amplified to an error in the solution c^{RISM} , which can be as large as

$$\text{cond}(\hat{\omega}) \left(\frac{\|\hat{\omega}^{\text{RISM}} - \hat{\omega}^{\text{exact}}\|}{\|\hat{\omega}^{\text{RISM}}\|} \right) \quad (19)$$

and can be large enough to dominate at low k . Thus it is essential to correct for this discrepancy, or bridge function, $B_{\alpha\gamma}(r) = c_{\alpha\gamma}^{\text{RISM}}(r) - c_{\alpha\gamma}^{\text{HNC}}(r)$, by using an appropriate term to incorporate it into the expression being used for solvation free energy. Note that we expect this bridge function, B , to have the following property: It should be largest at $k=0$, and decrease rapidly to 0 at larger k , where the relevant matrices are well conditioned.

Duh and Haymet have used empirical methods to determine this bridge correction as a function of the variable $s = h - c + u^{(2)}$, where the potential energy function u has been decomposed as the sum $u = u^{(1)} + u^{(2)}$ of short- and long-range components, $u^{(1)}$ and $u^{(2)}$, respectively. The following expression was used for this separation:

$$u_2^{\alpha\gamma}(r) = -4\varepsilon_{\alpha\gamma} \left(\frac{\sigma_{\alpha\gamma}}{r} \right)^6 \exp \left[\frac{-1}{\rho^*} \left(\frac{\sigma_{\alpha\gamma}}{r} \right)^{6\rho^*} \right] - \frac{\phi_{\alpha\gamma}(r)}{\beta},$$

where

$$\rho^* = \sum_{\alpha \in \text{solvent}} \rho_{\alpha} \sigma_{\alpha\alpha}^3. \quad (20)$$

By making the approximation that both s and h are proportional to the coupling parameter λ , they modified the argument used by Singer and Chandler⁶ in deriving their original expression for solvation free energy to obtain the following expression for the bridge-function-corrected solvation free energy.^{25,26}

$$\Delta\mu^{(\text{HNC})} = \frac{\rho}{2\beta} \sum_{\alpha\gamma} \int 4\pi r^2 \left[-2c_{\alpha\gamma}^{\text{HNC}}(r)[1+h_{\alpha\gamma}(r)] + h_{\alpha\gamma}^2(r) + h_{\alpha\gamma}(r)c_{\alpha\gamma}^{\text{RISM}}(r) - \frac{2h_{\alpha\gamma}(r)}{s_{\alpha\gamma}(r)} \int_0^{s_{\alpha\gamma}} ds' B_{\alpha\gamma}(s') \right] dr. \quad (21)$$

By following the arguments used by Ten-no in deriving the partial wave solvation free-energy expression,²¹ it is observed that the bridge-function-dependent term in the expression above remains correct in the partial wave formalism.

Duh and Haymet obtained $B = c^{\text{RISM}} - c^{\text{HNC}}$ by running simulations and RISM calculations, and then graphed B against s .²⁶ An equation was then fit to the data to express B as a function of s , which was later improved by Duh and Henderson.²⁷ The integral of this expression is^{26,27}

$$\int_0^s ds' B(s') = \begin{cases} \frac{81}{50}s - \frac{7}{20}s^2 - \frac{9}{2}\ln\left(1 + \frac{s}{3}\right) - \frac{9}{125}\ln\left(1 + \frac{5s}{3}\right), & \text{for } s \geq 0 \\ -\frac{1}{6}s^3, & \text{for } s < 0 \end{cases} \quad (22)$$

The mean spherical approximation (MSA) closure,

$$g(r) = [1 - e^{\beta u_1(r)}]^{-1} [c(r) + \beta u_2(r)], \quad (23)$$

was used for the RISM calculations in performing the fit just described. We calculated $s(r)$ by using the MSA to write

$$\begin{aligned} s_{\alpha\gamma}(r) &= h_{\alpha\gamma}(r) - c_{\alpha\gamma}^{\text{MSA}}(r) - \beta u_{\alpha\gamma}^{(2)}(r) \\ &= g_{\alpha\gamma}(r) e^{\beta u_{\alpha\gamma}^{(1)}(r)} - 1, \end{aligned} \quad (24)$$

and evaluating s using the last expression in this equality.

Our final solvation free-energy expression is thus given by

$$\begin{aligned} \Delta\mu &= \frac{\rho}{2\beta} \sum_{\alpha\gamma} \int 4\pi r^2 [-2c_{\alpha\gamma}^{\text{HNC}}(r)[1 + h_{\alpha\gamma}(r)] \\ &\quad + h_{\alpha\gamma}(r)\phi_{\alpha\gamma}(r)] dr \\ &\quad + (2\pi)^{-3} \frac{\rho}{2\beta} \sum_{\alpha\gamma} \int 4\pi k^2 [\hat{h}_{\alpha\gamma}(k)\hat{h}_{\alpha\gamma}^{\text{PW}}(k) \\ &\quad + \hat{h}_{\alpha\gamma}(k)\hat{c}_{\alpha\gamma}^{\text{RISMsr}}(k) - 2\hat{h}_{\alpha\gamma}(k)\hat{\beta}_{\alpha\gamma}(k)] dk, \end{aligned} \quad (25)$$

where

$$\beta_{\alpha\gamma}(r) = \frac{1}{s_{\alpha\gamma}(r)} \int_0^{s_{\alpha\gamma}(r)} ds' B_{\alpha\gamma}(s')$$

and $\hat{c}_{\alpha\gamma}^{\text{RISMsr}} = \hat{c}_{\alpha\gamma}^{\text{RISM}} - \phi_{\alpha\gamma}$ [see Eq. (6)].

Because the function of k given by

$$\sum_{\alpha,\gamma} \hat{h}_{\alpha\gamma}(k)\hat{c}_{\alpha\gamma}^{\text{RISMsr}}(k)$$

is invariant with respect to small changes in \hat{c}^{RISMsr} and this term is precisely a multiple of the contribution made to the solvation free-energy integrand by \hat{c}^{RISMsr} , we chose to show this function in Fig. 1. We graphed this function using radial distribution functions generated by the BOSS Monte Carlo simulation program,²⁸ for methanol in water. Also shown in the same diagram is the contribution made to the solvation free-energy integrand by the bridge function B . This contribution is given by

$$\sum_{\alpha,\gamma} 2\hat{h}_{\alpha\gamma}(k)\hat{\beta}_{\alpha\gamma}(k).$$

Note the shape of the latter function, starting out at $k=0$ with a large value comparable to that of the contribution from \hat{c}^{RISMsr} and declining rapidly to zero. The rapid decrease in the bridge function coincides with and represents the effect of the rapid improvement of conditioning of the matrices in the RISM equation as k is increased.

To investigate the significance of the resemblance between the two functions plotted in Fig. 1, we analyze in more detail their difference. We first express $\beta(r)$ as the integral over the coupling parameter,

$$\begin{aligned} \beta(r) &= \int_0^1 B_{\alpha\gamma}(r,\lambda) d\lambda \\ &= \int_0^1 [c_{\alpha\gamma}^{\text{RISM}}(r,\lambda) - c_{\alpha\gamma}^{\text{HNC}}(r,\lambda)] d\lambda \\ &= \int_0^1 [c_{\alpha\gamma}^{\text{RISMsr}}(r,\lambda) + \phi_{\alpha\gamma}(r,\lambda) - c_{\alpha\gamma}^{\text{HNC}}(r,\lambda)] d\lambda \\ &\equiv \frac{1}{2} c_{\alpha\gamma}^{\text{RISMsr}}(r) + \int_0^1 [\phi_{\alpha\gamma}(r,\lambda) - c_{\alpha\gamma}^{\text{HNC}}(r,\lambda)] d\lambda, \end{aligned} \quad (26)$$

where the last approximation is justified by the assumption made above that both s and h are proportional to the coupling parameter λ ,²⁶ from which it can be inferred that $c^{\text{RISMsr}} \approx s - h$ is also approximately linear in the variable λ . Thus the difference at $k=0$ between the two functions graphed in Fig. 1 can be approximated as the product of the $k=0$ Fourier transform of the sum of site-site components,

$$\begin{aligned} &-\sum_{\alpha\gamma} \int_0^1 [\phi_{\alpha\gamma}(r,\lambda) - c_{\alpha\gamma}^{\text{HNC}}(r,\lambda)] d\lambda \\ &= \sum_{\alpha\gamma} \int_0^1 c_{\alpha\gamma}^{\text{HNCsr}}(r,\lambda) d\lambda, \end{aligned} \quad (27)$$

with $\hat{h}(0)$. The lower value taken on at $k=0$ by the difference between these two functions compared to the $k=0$

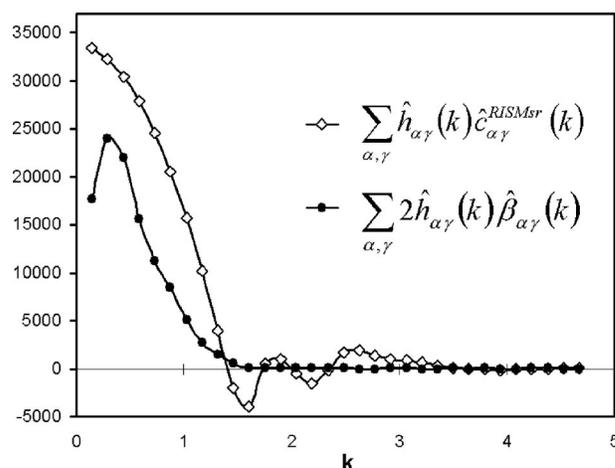


FIG. 1. Contributions to solvation free-energy integrand from c^{RISMsr} and B for methanol in water.

TABLE I. Free energies of solvation (kcal/mol) of methanol using different cutoff procedures.

Cutoff distance (Å)	Hard-sphere cutoff	Soft-sphere cutoff
10.4	-4.94	-4.44
10.6	-1.06	-4.55
10.8	0.23	-4.45
11.0	1.86	-4.36
11.2	0.08	-4.46
11.4	-4.93	-4.69
11.6	-7.51	-5.07

value taken on by the function $\sum_{\alpha,\gamma} \hat{h}_{\alpha\gamma}(k) \hat{c}_{\alpha\gamma}^{\text{RISMsr}}(k)$ may therefore signify that the solution c^{HNC} to the HNC equation approaches the asymptotic limit of ϕ somewhat faster than c^{RISM} does.

2. "Soft" sphere boundary condition

Because the sphere centered around a solute site being integrated over to obtain the solvation free energy may not contain stoichiometric numbers of water O and H atoms, our calculated solvation free energy may depend upon the total charge of atoms included within this sphere. This lack of charge neutrality will affect our calculated solvation free energy, especially the integral over the outermost shell of water molecules of the solute-site components,

$$\begin{aligned} & \frac{\rho}{2\beta} \sum_{\gamma} 4\pi r^2 \{-2\phi_{\alpha\gamma}(r)[1+h_{\alpha\gamma}(r)]+h_{\alpha\gamma}(r)\phi_{\alpha\gamma}(r)\} \\ &= -\frac{\rho}{2\beta} \sum_{\gamma} 4\pi r^2 g_{\alpha\gamma}(r)\phi_{\alpha\gamma}, \end{aligned} \quad (28)$$

of the term

$$\frac{\rho}{2\beta} \sum_{\alpha\gamma} 4\pi r^2 \{-2c_{\alpha\gamma}^{\text{HNC}}(r)[1+h_{\alpha\gamma}(r)]+h_{\alpha\gamma}(r)\phi_{\alpha\gamma}(r)\}$$

occurring in the integrand of the solvation free-energy expression (25). Such a component represents the total Coulombic potential energy exercised on site α by the solvent molecules occupying this shell, and, in effect, all shells beyond this outermost one, and should equal 0 for an uncharged solvent. If this component is nonzero, then the effect

is, basically, to introduce a charge separation spanning the distance R between the solute atom and the sphere's boundary (Table I).

To correct for this error, a term must be added on, given by

$$-\frac{1}{2\beta} \sum_{\alpha} \frac{q_{\alpha} q_{\text{O,H}_2\text{O}}}{R} \Delta n_{\alpha}(R), \quad (29)$$

where

$$\Delta n_{\alpha}(R) = 4\pi\rho \int_0^R g_{\alpha\text{O}}(r)r^2 dr - 4\pi\rho \int_0^R g_{\alpha\text{H}}(r)r^2 dr.$$

Δn_{α} represents the difference in coordination numbers for O and H atoms of water about solute site α . By making this adjustment, we are subtracting off the effect of extra charge from the boundary of the sphere surrounding each solute atom. Equivalently, we can think of each sphere as being replaced by a "soft" sphere containing the entire water molecules within its inner boundary, without truncation of these molecules at the boundary. We found that by adding this term to the free-energy expression reasonable independence of the free energy from the integration limit can be obtained. This is seen in Table I, where, in the first column, the integration limit is given, in the second column, the calculated solvation free energy before any adjustment has been made, and in the third, the results of adding our adjustment to the values in the second column.

B. Low- k extrapolation

Having addressed the issue of the exact form of the solvation free-energy expression to be evaluated, we now turn to the problem of solving the renormalized RISM equation, (10), for the solution c^{RISMsr} .

We have discussed how bridge functions can be used to correct for the inexact nature of the matrices occurring in the RISM equation at low k . However the relative error in the solution \hat{c}^{RISMsr} to the RISM equation resulting from the inaccuracies in \hat{h} must also be considered. The full numerical error in evaluating the solution $x=A^{-1}b$ to the linear equation $Ax=b$ has been shown to satisfy²⁴

$$\frac{\|\Delta x\|}{\|x\|} \leq \text{cond}(A) \left(\frac{\|E\|}{\|A\|} + \frac{\|\Delta b\|}{\|b\|} \right), \quad (30)$$

TABLE II. Free energies of solvation (kcal/mol).

Solute	RISM/MC	Bridge-corrected RISM/MC	RISM/MC ^a	Bridge corrected RISM/MC ^a	Expt.
Acetic acid	2.92	-1.92	-4.84	-6.27	-6.7
Ammonia	-1.94	-3.44	-2.48	-3.16	-4.3
Ethane	11.2	6.37	2.67	1.30	1.8
Methylamine	3.76	0.59	-1.50	-2.65	-4.6
Methane	0.95	-1.61	-0.16	-1.08	2.0
Methanol	0.66	-4.31	-3.34	-4.36	-5.1
Methanethiol	6.46	2.80	-1.23	-2.24	-1.2
TIP3P water	-4.50	-5.34	-5.88	-6.27	-6.3

^aWe used a low- k integration cutoff of c^{RISMsr} up to $5\Delta k$.

where, as above, the matrix E represents the error in the matrix A . Therefore there is a second term in the upper bound for the relative error in \hat{c} that arises when multiplying by the matrix inverse $\hat{\omega}^{-1}$ in solving the RISM equation [again, see Eq. (10)], which must also be considered,

$$\frac{\|\Delta\hat{c}\|}{\|\hat{c}\|} \leq \text{cond}(\hat{\omega}) \left(\frac{\|\Delta\hat{\omega}\|}{\|\hat{\omega}\|} \right) + \text{cond}(\hat{\omega}) \left(\frac{\|\Delta\hat{h}\|}{\|\hat{h}\|} \right). \quad (31)$$

At low k , we can expect, again, that this error term will dominate our solution \hat{c}^{RISMsr} . To address this problem, we must modify $\hat{\omega}$ to a better-conditioned matrix.

We first note that our expression for solvation free energy, Eq. (25), can be approximated without knowing individual site-site direct correlation functions at small k . In fact at $k=0$, for all sites α and γ , $\rho\hat{h}_{\alpha\gamma}(k)$ represents the total fractional change in the number of solvent molecules in the entire three-dimensional space surrounding a solute molecule compared to in pure solvent, and thus is independent of the individual sites α and γ . For small positive k , we can make the approximation that values of $\hat{h}(k)$ concur over a range of proximal sites. In that case all that is needed to find the solvation free energy are the terms,

$$\sum_{\gamma} \hat{h}_{\alpha\gamma}(k) \hat{c}_{\alpha\gamma}^{\text{RISMsr}}(k) \equiv \langle \hat{h}_{\alpha\gamma}(k) \rangle_{\gamma} \left(\sum_{\gamma} \hat{c}_{\alpha\gamma}^{\text{RISMsr}}(k) \right), \quad (32)$$

where the sum is taken over those sites for which values of $\hat{h}(k)$ approximately agree. The ill conditioning of the matrix $\hat{\omega}$ results from the slow decay from the value of 1 of matrix elements proceeding away from the diagonal either along a row or a column. If sites are ordered so that consecutive sites are neighboring to each other within the solute molecule, then adjacent entries of the matrix $\hat{\omega}$ are close in value. These two aspects of the low- k behavior of \hat{c} and $\hat{\omega}$, first, the replacement of separate site-site values of \hat{c} with a sum of such values in the free-energy expression, and, second, the similarity of neighboring entries of the matrix $\hat{\omega}$, allow the dimensionality of the original RISM equation to be reduced, as the ill-conditioned matrix $\hat{\omega}$ is replaced by a better-conditioned matrix with only one row and one column for each group of solute sites being incorporated into the sum. This process is given in detail in the Appendix.

Thus although for most k the evaluation of the term $\sum_{\gamma} \hat{h}_{\alpha\gamma}(k) \hat{c}_{\alpha\gamma}^{\text{RISMsr}}(k)$ in the free-energy expression relies upon the entire molecular geometry as given by $\hat{\omega}$, as k approaches 0 this geometry becomes decreasingly important, and the focus on the solute molecule for this calculation can be broadened to encompass only groups of sites, so that adjacent pairs, and at still smaller k -values groups of more than two sites, can be merged together and regarded as a single site. Finally at $k=0$, $\hat{\omega}$ becomes completely irrelevant to our calculation of this term and we have

$$\sum_{\alpha\gamma} \hat{c}_{\alpha\gamma}(0) = \frac{\hat{h}_{\text{solute}}(0)}{1 + \rho\hat{h}_{\text{solvent}}(0)}. \quad (33)$$

It has been shown that the denominator in the above ratio is equal to $\rho_{\text{solvent}}kT\chi_T^0$, where χ_T^0 denotes the isothermal compressibility of the pure solvent.⁴

The procedure just described can be applied only as long as k is small enough so that adjacent rows and columns of $\hat{\omega}$ are roughly equal. For larger k for which $\hat{\omega}$ is still ill conditioned, some elements of $\hat{\omega}$ may be small enough so that their replacement by 0 will produce better-conditioned matrices, which can be used to replace the original matrices. For k very close to zero, rather than solving the RISM equation we can use a quadratic interpolation of the form $\sum_{\alpha\gamma} \hat{c}_{\alpha\gamma}(k) = (\sum_{\alpha\gamma} \hat{c}_{\alpha\gamma}(0)) + Bk^2$, which is justified by the low- k Taylor approximation of the sine function used in the spatial Fourier transformation.

In practice, we must address the problem of what limit to set to the condition numbers occurring in the RISM equation before applying one of the procedures just described. The number of digits of accuracy in \hat{c} is less than that in \hat{h} by as much as the sum of the orders of magnitude of these condition numbers. Therefore a condition number of 100 may seem rather high, especially considering that there are two ill-conditioned matrices occurring in the RISM equation. However the invariance of its expression allows the solvation free energy to maintain stability versus the ill conditioning of the solution, and thus we found that the value of 100 is a reasonable maximum for permissible condition numbers of these matrices.

Note that the same invariance satisfied by the term $\sum_{\alpha\gamma} \hat{h}_{\alpha\gamma}(k) \hat{c}_{\alpha\gamma}^{\text{RISMsr}}(k)$ in the integrand of $\Delta\mu$ with respect to the solution to the RISM equation is also satisfied by the term $\sum_{\alpha\gamma} \hat{h}_{\alpha\gamma}(k) \hat{h}_{\alpha\gamma}^{\text{PW}}(k)$, with respect to variations in \hat{h}^{PW} . We can therefore apply methods similar to those just described to the evaluation of $\hat{h}_{sv}^{\text{PW}} = [\hat{\omega}_s]^{-1} [\hat{h}_{sv}] [\hat{\omega}_v]^{-1}$.

To test our inversion process, we first started with radial distribution functions calculated by stand-alone RISM/HNC and then used these as input in the program we designed for determining the direct correlation functions, and from these the solvation free energy. In Fig. 2 we graph the same function $\sum_{\alpha\gamma} \hat{h}_{\alpha\gamma}(k) \hat{c}_{\alpha\gamma}^{\text{RISMsr}}(k)$ as was shown previously in Fig. 1. The figure compares the function evaluated using \hat{c} coming from self-consistent solution of the RISM and HNC equations to the same function where \hat{c} is derived using our method with \hat{h} obtained from RISM/HNC. Three different approximations in the low- k range were used. For the higher spectrum of values of k for which $\hat{\omega}$ is ill conditioned, we use the method of zeroing out entries of $\hat{\omega}$ which are close to zero. This set of points is labeled by the series (1) in the chart. At a somewhat lower range of k , because the smallest values occurring in $\hat{\omega}$ are now closer to 1, it becomes more appropriate to use the matrix reduction procedure described above. These values of k are labeled by the series (2). Finally for those k values closest to 0, we used a quadratic fit of the form $A + Bk^2$, labeled by the series (3). Although we hope to eliminate the large error term caused by the enlargement of small errors in \hat{h} by huge condition numbers, yet we introduce by our modification of $\hat{\omega}$ another smaller error term. If this error term is adequately minor, then the two functions

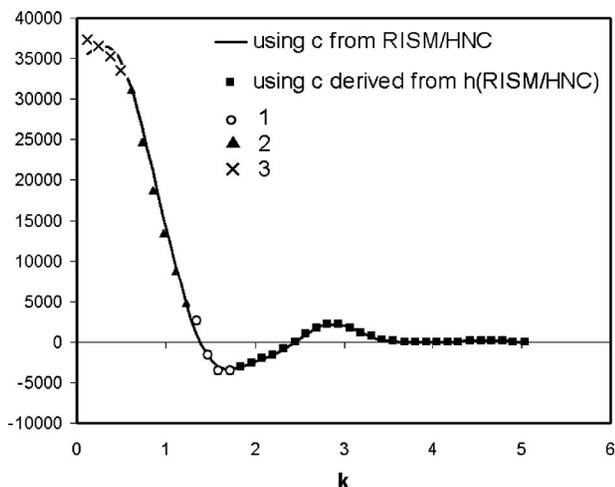


FIG. 2. Contributions to solvation free-energy integrand, for methanol in water, from (a) the solution c^{sr} to the RISM/HNC equations, and (b) the solution c^{RISMsr} derived from h , where h has been solved for using RISM/HNC. Better-conditioned matrices $\hat{\omega}$ were obtained by (1) zeroing out all elements of the matrix $\hat{\omega}$ that were sufficiently close to zero, and (2) reducing matrix dimensionality. For series (3), we used a quadratic fit of the form $A+Bk^2$.

shown in Fig. 2 should almost overlap. In fact, we observe a close match between the two plots.

III. COMPUTATIONAL DETAILS

In this study, distribution functions were calculated by Monte Carlo simulations at 298 K using the BOSS package.²⁸ The dimensions of solvent boxes used were large enough to contain solvent water molecules at distances of up to ~ 15 Å from the center of any solute atom, and a periodic boundary condition was used. OPLS all-atom parameters were used for solutes and the TIP3P model was used for water in MC simulations. Except for the case of solvated water for which convergence is reached much more quickly, solvated systems were equilibrated for 12×10^6 configurations separated into three batches of 4×10^6 configurations each, followed by averaging over 28×10^6 configurations run in seven batches of 4×10^6 configurations each. For water, batches were of 250 000 configurations. Calculated radial distribution functions were sampled at intervals of 0.04 Å and were gradually smoothed to zero at distances exceeding 11 Å.

The OPLS force field was also used to obtain the pair potential when solving for c from the HNC equation. This force field represents the pair potential as a Coulomb potential with the usual 12-6 van der Waals interactions, and uses the Lorentz–Berthelot mixing rules.

We used a water density of $0.033\ 34/\text{Å}^3$. The Fourier transformations for the RISM calculations were carried out on a linear grid of 537 points, with zero aliasing to twice this range,² meaning that an additional 537 points were zero filled. This gave a total interval length of 43 Å, which was used to calculate values of k . Three kinds of approximations used when solving the RISM equation for c^{RISMsr} have been described above. For $k > 10\Delta k = 1.46$, the problem of ill conditioning was addressed by setting elements of $\hat{\omega}$ less than 0.1 equal to 0, and checking for the validity of the condition

$\{1/\text{cond}(\hat{\omega}) > 0.011\}$. In case this condition was not met, all entries of the matrix $\hat{\omega}$ less than 0.2 were set equal to zero, the condition number was again checked, and this process was continued until all elements of ill-conditioned matrices $\hat{\omega}$ less than 0.4 had been made equal to 0. The next step was to reduce by a factor of 2 the dimensionality of all remaining ill-conditioned matrices $\hat{\omega}$. We used only one such reduction, although for larger molecules, further reduction by factors of larger powers of 2 would most likely be advantageous. In the final stage, at small enough k so that either $\hat{\omega}$ remained ill conditioned or else the matrix $\hat{\chi}$ was ill conditioned, we used a quadratic fit such as was described above. To obtain $\Sigma_{\alpha\gamma} \hat{c}_{\alpha\gamma}(0)$ in Eq. (33), we used the reference value²⁹ for the isothermal compressibility χ_T^0 of water at 298 K equal to $4.524 \times 10^{-10}/\text{Pa}$, or $1 + \rho \hat{h}_{\text{solvent}}(0) = \rho_{\text{solvent}} kT \chi_T^0 = 0.062$. A similar technique was employed in solving for \hat{h}^{PW} as we used to find c^{RISMsr} .

When calculating the bridge function, the function $s = h - c + u^{(2)}$ was calculated using Eq. (24). The function s seemed, for some intervals close to $r=0$, to assume values unreasonably large for our purposes in using this function to determine the bridge function. We therefore reassigned a value of 1 to this function whenever its value exceeded 1.

In carrying out the numerical integration for the solvation free energy, radial distribution functions were cut off at 11 Å. Because of the problem with low accuracy at small k values, we started the integration over k from a value of $k = 6\Delta k = 0.876$.

Stand-alone RISM/HNC equations were performed using the modified direct inversion of iterative subspace convergence technique.³⁰ For stand-alone RISM/HNC calculations, the OPLS parameter set requires some small modifications. Namely, whenever the OPLS parameter set had assigned values of 0 to the Lennard-Jones parameters of certain hydrogen atoms, these were reassigned nonzero values of $\epsilon = 0.046$ and $\sigma = 0.40$.¹⁰ This change was made in order to preclude the appearance of high artificial peaks in the calculated radial distribution functions. To compare with the stand-alone RISM/HNC calculations, a few coupled RISM/MC calculations were also repeated using this modified OPLS parameter set. The linear extrapolation procedure mentioned below was performed without using a bridge correction or low- k integration cutoff, but instead continuously assigning $\hat{c}^{\text{RISMsr}}(k)$ a constant value for all $k = 10\Delta k = 1.46$.

IV. RESULTS

Table II gives solvation free-energy results obtained by running our coupled RISM/MC program for several test molecules. The second-to-last column displays the results of our bridge-function corrected calculations found by using Eq. (25), for which we have applied the low- k integration cutoff described in the preceding section. Most of these results show good agreement with experiment. The main exceptions are methane and methylamine. However, part of the problem with these molecules may be attributable to the force field itself. For example, calculated solvation free energies for al-

TABLE III. Free energies of solvation (kcal/mol).

Solute	28×10^6 configurations	32×10^6 configurations	36×10^6 configurations	Average ^a
Acetic acid	-6.27	-6.04	-6.50	-6.27 ± 0.23
Ammonia	-3.16	-2.11	-2.09	-2.45 ± 0.61
Ethane	1.30	0.39	1.47	1.05 ± 0.58
Methylamine	-2.65	-2.07	-2.75	-2.49 ± 0.37
Methane	-1.08	-1.40	-1.46	-1.31 ± 0.20
Methanol	-4.36	-3.95	-4.71	-4.34 ± 0.38
Methanethiol	-2.24	-2.12	-2.54	-2.30 ± 0.22

^aAverage was computed from values corresponding to 28 million, 32 million, and 36 million configurations, and the error represents the non-biased standard deviation from the average.

kanes such as methane are quite potential dependent, being contingent upon small differences in shapes and sizes of water pockets.

We also include in Table II results where different options are turned off in order to see their relative importance in the calculated total free energy of solvation. In particular, columns, respectively, display the total free-energy integral without including the bridge correction or any low- k integration cutoff, the same integral including the bridge correction, and this integral starting at the low- k cutoff point but without any bridge correction. The values given in the second column, corresponding to unmodified RISM/MC, in almost all cases do not reflect the correct ordering of the solvation free energies of the eight molecules. Furthermore, these values are all but one too positive when compared to the experimental values, with an average absolute error of 5.75 kcal/mol. The function $\Sigma_{\alpha,\gamma} \hat{h}_{\alpha\gamma}(k) \hat{c}_{\alpha\gamma}^{\text{RISMsr}}(k)$ shown in Fig. 1 exhibits a large maximum at $k=0$. However the area under the curve corresponding to k between 0 and about $10\Delta k$ should not contribute to the solvation free energy nearly as substantially as is calculated using unmodified RISM/MC, because this area is partially cancelled by the area under the curve $\Sigma_{\alpha,\gamma} 2\hat{h}_{\alpha\gamma}(k) \hat{\beta}_{\alpha\gamma}(k)$ also seen in Fig. 1. The former area contributes a positive quantity to the energies in the second column, accounting for their excessively positive values.

Because the bridge-function corrections are subtracted from the sum in Eq. (25), the values in the third column of Table II are a great deal more negative than those in the second column, and the bridge correction leads to values generally much closer to experimental ones. Nevertheless, the ordering of most of the solvation free energies remains incorrect.

Right around $k=0$, the accuracy of both $\hat{h}(k)$ and $\hat{\beta}(k)$ is questionable, and so the effect of truncation starting at $k=5\Delta k$ is shown in the fourth column. We have neglected the bridge function, and so the values are again for the most part too positive. The orders of water and acetic acid are reversed, as are the orders of the pairs ammonia/methylamine and methane/ethane.

When the bridge function is included along with this truncation, most results move closer to experimental values, the notable exception being methane, for which our calculated result becomes even more erroneously negative. Acetic acid is now correctly predicted to be least solvated, although

it shares this property with water. The average absolute error for the molecules is now only 1.11 kcal/mol.

Concern about the need to check the reasonableness of the convergence and quality of simulations prompted us to estimate the statistical error associated with our free-energy results. To analyze the sensitivity of our results to the simulation length, we recalculated solvation free energies using radial distribution functions averaged over eight or nine batches of 4×10^6 configurations each. Solvation free energies obtained using these 32×10^6 or 36×10^6 configurations are shown in Table III, along with those obtained using our original 28×10^6 configurations. Average values from these three simulation lengths are given in Table III together with the statistical error, representing one standard deviation from the average. These errors range in value from 0.20 to 0.61 kcal/mol. Smaller errors in calculated results can be expected if simulations are run over longer time intervals.

Table IV compares the results of the HNC, GF, and PW methods for the molecules methanol and water. It can be seen that the partial wave approach produces values closest to the experimental ones, lying somewhere in between the HNC values, which are too positive, and the GF values, which are excessively negative.

We have observed that if, instead of solving the RISM equation at low k , we simply linearly extrapolate \hat{c}^{RISMsr} as a constant value at low k , reasonable results can also be obtained. In fact this was the procedure originally used to carry out the calculations for our previously published results.¹⁴ The similarity between our original and present results can be explained as follows: A large part of the low- k contribution of \hat{c}^{RISMsr} to the solvation free energy is canceled by that made by $\hat{\beta}$. By extrapolating this contribution at low k , we are merely cutting off the steep rise in this function as k tends to 0. If this extrapolation is carried over to stand-alone RISM/HNC calculations, then the results of these calculations are likewise improved. This can be seen in Table V. In the second column, we show the results of stand-alone

TABLE IV. Free energies of solvation (kcal/mol).

Solute	Expt.	$\Delta\mu^{\text{HNC}}$	$\Delta\mu^{\text{GF}}$	$\Delta\mu^{\text{PW}}$
TIP3P water	-6.3	-5.10	-9.76	-6.27
Methanol	-5.1	5.73	-8.00	-4.36

TABLE V. Free energies of solvation (kcal/mol).

Solute	RISM(PW)/HNC		Expt.
	RISM(PW)/HNC	(extrapolated)	
Acetic acid	-1.63	-8.95	-6.7
Ammonia	-2.97	-4.82	-4.3
Ethane	7.45	0.55	1.83
Methylamine	-1.67	-7.04	-4.6
Methane	5.37	1.87	2.0
Methanol	-2.58	-6.19	-4.67
Methanethiol	5.10	0.77	-1.2
TIP3P water	-7.17	-8.31	-5.90

RISM/HNC calculations without using the extrapolation procedure. In the same way as we observed excessively positive values in Table II when using unmodified RISM/MC, so also the results seen here, with the exception of water, are all more positive than the experimental values. They are, moreover, in the wrong sequence. In the third column, we show stand-alone RISM/HNC results improved by using the extrapolation procedure. It can be noted that this extrapolation improves both the ordering and the values of calculated solvation free energies, remarkably. A few results of RISM/MC calculations using the modified OPLS parameter set and the extrapolation procedure are given for comparison in the fourth column, showing some improvement over the extrapolated RISM/HNC values.

Some representative radial distribution functions for ammonia in water are shown in Fig. 3.

V. CONCLUSIONS

We have derived an expression for solvation free energy that is invariant with respect to small fluctuations in the calculated direct correlation function c . This allows the amplification of small errors in \hat{h} to large errors in \hat{c}^{RISMsr} , by the high condition numbers of the matrices occurring in the RISM equation at low k , to contribute minimally to the error in the solvation free energy.

Condition numbers converging to ∞ as $k \rightarrow 0$ magnify discrepancies between the RISM and HNC equations to a

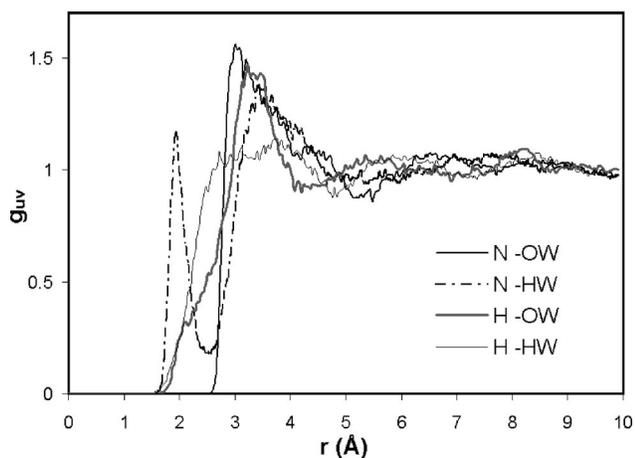


FIG. 3. MC-calculated solute/solvent radial distribution functions for ammonia in TIP3P water.

large difference $\hat{B} = \hat{c}^{\text{RISM}} - \hat{c}^{\text{HNC}}$ in the direct correlation function, near $k=0$. The solvation free-energy expression must be improved to account for this difference, or bridge function. We used the expression for solvation free energy with a bridge correction developed by Haymet and co-workers. Our results demonstrate that much of the contribution to free energy made by the large values taken on by the integrand in the solvation free-energy expression in the range of k close to 0 is canceled by the bridge-function contribution of opposite sign. As our results demonstrate, careful handling of this integrand at low k can also be expected to lead to improved absolute solvation free-energy values when using the stand-alone RISM/HNC method. By inverting the RISM expression for \hat{h} in terms of \hat{c} regularly worked with, to one for \hat{c} in terms of \hat{h} , we have obtained insight into the source of some of the difficulties associated with the stand-alone RISM/HNC method.

Reasonable results were obtained by applying RISM/MC to a small set of neutral test cases. Many questions remain concerning how the accuracy of the RISM/simulation methodology compares to the stand-alone RISM/HNC approach, or to free-energy simulations. Although our results look good, yet some of these have significant errors and there is certainly room for improvement.

We expect that studies focused on searching for better-quality bridge functions will lead to improvement of both the RISM/MC and the stand-alone RISM/HNC methods for calculating solvation free energy. In particular, the correct determination of the quantity $\sum_{\alpha,\gamma} 2\hat{h}_{\alpha\gamma}(k)\hat{\beta}_{\alpha\gamma}(k)$ near $k=0$ is a large factor contributing to the accuracy of each of these two methods. Especially in the case of RISM/MC, further work directed towards the accurate evaluation of $\sum_{\alpha,\gamma} \hat{h}_{\alpha\gamma}(k)\hat{c}_{\alpha\gamma}^{\text{RISMsr}}(k)$ near $k=0$ is also needed.

One appeal of the RISM/simulation methodology is its simplicity, requiring only one MD or MC simulation, followed by algebraic evaluation of the direct correlation function, from which a RISM solvation free energy can be calculated, without requiring any iterative solution of the RISM and closure equations. We are in the process of performing an application of the RISM/simulation methodology to the alanine dipeptide. Because this system has been studied so widely, we will be able to turn for comparison not only to experimental findings, but also to a wide range of theoretical results. By examining in detail how best to accurately solve for the direct correlation functions satisfying the RISM equation, we have provided a basis for extending applications of the coupled RISM/simulation method to larger systems.

ACKNOWLEDGMENTS

We are grateful to Professor Shao-Wen Zhang for assistance in the implementation of the RISM method and to Professor Monte Pettitt, Professor John Perkyns, and Professor Frank Stenger for helpful discussion. We thank the Center for High Performance Computing at the University of Utah for computer time and support.

APPENDIX: REDUCTION OF MATRIX DIMENSIONALITY IN THE RISM EQUATION AT LOW k

The matrix reduction procedure used to solve the equation

$$[\hat{h}_{uv} - \hat{\gamma}_{uv}] = [\hat{\omega}_u][\hat{c}_{uv}^{\text{RISMsr}}][\hat{\omega}_v + \rho_v \hat{h}_{vv}] \tag{A1}$$

for \hat{c}^{RISMsr} [see Eq. (10)] at low k is shown below. Note that as $k \rightarrow 0$, the matrix $\hat{\gamma}_{uv}$ converges to the 0 matrix. In the treatment shown here, pairs of proximal solute sites $2u+1$ and $2u$ are merged into a single site in the approximation $\hat{\omega}_{2u+1,u'}(k) \approx \hat{\omega}_{2u,u'}(k)$.

$$\begin{bmatrix} (\hat{h} - \hat{\gamma})_{00} & (\hat{h} - \hat{\gamma})_{01} & (\hat{h} - \hat{\gamma})_{02} \\ (\hat{h} - \hat{\gamma})_{10} & (\hat{h} - \hat{\gamma})_{11} & (\hat{h} - \hat{\gamma})_{12} \\ (\hat{h} - \hat{\gamma})_{20} & (\hat{h} - \hat{\gamma})_{21} & (\hat{h} - \hat{\gamma})_{22} \\ (\hat{h} - \hat{\gamma})_{30} & (\hat{h} - \hat{\gamma})_{31} & (\hat{h} - \hat{\gamma})_{32} \\ (\hat{h} - \hat{\gamma})_{40} & (\hat{h} - \hat{\gamma})_{41} & (\hat{h} - \hat{\gamma})_{42} \\ \vdots & \vdots & \vdots \\ (\hat{h} - \hat{\gamma})_{s0} & (\hat{h} - \hat{\gamma})_{s1} & (\hat{h} - \hat{\gamma})_{s2} \end{bmatrix} = \begin{bmatrix} 1 & \hat{\omega}_{01} & \hat{\omega}_{02} & \hat{\omega}_{03} & \hat{\omega}_{04} & \cdots & \hat{\omega}_{0s} \\ \hat{\omega}_{10} & 1 & \hat{\omega}_{12} & \hat{\omega}_{13} & \hat{\omega}_{14} & \cdots & \hat{\omega}_{1s} \\ \hat{\omega}_{20} & \hat{\omega}_{21} & 1 & \hat{\omega}_{23} & \hat{\omega}_{24} & \cdots & \hat{\omega}_{2s} \\ \hat{\omega}_{30} & \hat{\omega}_{31} & \hat{\omega}_{32} & 1 & \hat{\omega}_{34} & \cdots & \hat{\omega}_{3s} \\ \hat{\omega}_{40} & \hat{\omega}_{41} & \hat{\omega}_{42} & \hat{\omega}_{43} & 1 & \cdots & \hat{\omega}_{4s} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\omega}_{s0} & \hat{\omega}_{s1} & \hat{\omega}_{s2} & \hat{\omega}_{s3} & \hat{\omega}_{s4} & \vdots & 1 \end{bmatrix} \begin{bmatrix} \hat{c}_{00}^{\text{RISMsr}} & \hat{c}_{01}^{\text{RISMsr}} & \hat{c}_{02}^{\text{RISMsr}} \\ \hat{c}_{10}^{\text{RISMsr}} & \hat{c}_{11}^{\text{RISMsr}} & \hat{c}_{12}^{\text{RISMsr}} \\ \hat{c}_{20}^{\text{RISMsr}} & \hat{c}_{21}^{\text{RISMsr}} & \hat{c}_{22}^{\text{RISMsr}} \\ \hat{c}_{30}^{\text{RISMsr}} & \hat{c}_{31}^{\text{RISMsr}} & \hat{c}_{32}^{\text{RISMsr}} \\ \hat{c}_{40}^{\text{RISMsr}} & \hat{c}_{41}^{\text{RISMsr}} & \hat{c}_{42}^{\text{RISMsr}} \\ \vdots & \vdots & \vdots \\ \hat{c}_{s0}^{\text{RISMsr}} & \hat{c}_{s1}^{\text{RISMsr}} & \hat{c}_{s2}^{\text{RISMsr}} \end{bmatrix} \tag{A2}$$

$$\times \begin{bmatrix} \hat{\chi}_{00} & \hat{\chi}_{01} & \hat{\chi}_{02} \\ \hat{\chi}_{10} & \hat{\chi}_{11} & \hat{\chi}_{12} \\ \hat{\chi}_{20} & \hat{\chi}_{21} & \hat{\chi}_{22} \end{bmatrix}$$

↓

$$\begin{bmatrix} (\hat{h} - \hat{\gamma})_{00} & (\hat{h} - \hat{\gamma})_{01} & (\hat{h} - \hat{\gamma})_{02} \\ (\hat{h} - \hat{\gamma})_{20} & (\hat{h} - \hat{\gamma})_{21} & (\hat{h} - \hat{\gamma})_{22} \\ (\hat{h} - \hat{\gamma})_{40} & (\hat{h} - \hat{\gamma})_{41} & (\hat{h} - \hat{\gamma})_{42} \\ \vdots & \vdots & \vdots \\ (\hat{h} - \hat{\gamma})_{s0} & (\hat{h} - \hat{\gamma})_{s1} & (\hat{h} - \hat{\gamma})_{s2} \end{bmatrix} = \begin{bmatrix} 1 & \hat{\omega}_{02} & \hat{\omega}_{04} & \cdots & \hat{\omega}_{0s} \\ \hat{\omega}_{20} & 1 & \hat{\omega}_{24} & \vdots & \hat{\omega}_{2s} \\ \hat{\omega}_{40} & \hat{\omega}_{42} & 1 & \cdots & \hat{\omega}_{4s} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\omega}_{s0} & \hat{\omega}_{s2} & \hat{\omega}_{s4} & \cdots & 1 \end{bmatrix}$$

$$\times \begin{bmatrix} \hat{c}_{00}^{\text{RISMsr}} + \hat{c}_{10}^{\text{RISMsr}} & \hat{c}_{01}^{\text{RISMsr}} + \hat{c}_{11}^{\text{RISMsr}} & \hat{c}_{02}^{\text{RISMsr}} + \hat{c}_{12}^{\text{RISMsr}} \\ \hat{c}_{20}^{\text{RISMsr}} + \hat{c}_{30}^{\text{RISMsr}} & \hat{c}_{21}^{\text{RISMsr}} + \hat{c}_{31}^{\text{RISMsr}} & \hat{c}_{22}^{\text{RISMsr}} + \hat{c}_{32}^{\text{RISMsr}} \\ \hat{c}_{40}^{\text{RISMsr}} + \hat{c}_{50}^{\text{RISMsr}} & \hat{c}_{41}^{\text{RISMsr}} + \hat{c}_{51}^{\text{RISMsr}} & \hat{c}_{42}^{\text{RISMsr}} + \hat{c}_{52}^{\text{RISMsr}} \\ \vdots & \vdots & \vdots \\ \hat{c}_{s0}^{\text{RISMsr}} & \hat{c}_{s1}^{\text{RISMsr}} & \hat{c}_{s2}^{\text{RISMsr}} \end{bmatrix} \begin{bmatrix} \hat{\chi}_{00} & \hat{\chi}_{01} & \hat{\chi}_{02} \\ \hat{\chi}_{10} & \hat{\chi}_{11} & \hat{\chi}_{12} \\ \hat{\chi}_{20} & \hat{\chi}_{21} & \hat{\chi}_{22} \end{bmatrix}. \tag{A3}$$

To better facilitate our solvation free-energy determination using the solution \hat{c}^{RISMsr} , we implement this as

$$\begin{bmatrix} (\hat{h} - \hat{\gamma})_{00} & (\hat{h} - \hat{\gamma})_{01} & (\hat{h} - \hat{\gamma})_{02} \\ 0 & 0 & 0 \\ (\hat{h} - \hat{\gamma})_{20} & (\hat{h} - \hat{\gamma})_{21} & (\hat{h} - \hat{\gamma})_{22} \\ 0 & 0 & 0 \\ (\hat{h} - \hat{\gamma})_{40} & (\hat{h} - \hat{\gamma})_{41} & (\hat{h} - \hat{\gamma})_{42} \\ \vdots & \vdots & \vdots \\ (\hat{h} - \hat{\gamma})_{s0} & (\hat{h} - \hat{\gamma})_{s1} & (\hat{h} - \hat{\gamma})_{s2} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \hat{\omega}_{02} & 0 & \hat{\omega}_{04} & \cdots & \hat{\omega}_{0s} \\ 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ \hat{\omega}_{20} & 0 & 1 & 0 & \hat{\omega}_{24} & \cdots & \hat{\omega}_{2s} \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 \\ \hat{\omega}_{40} & 0 & \hat{\omega}_{42} & 0 & 1 & \cdots & \hat{\omega}_{4s} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\omega}_{s0} & 0 & \hat{\omega}_{s2} & 0 & \hat{\omega}_{s4} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \hat{c}_{00}^{\text{RISMsr}} & \hat{c}_{01}^{\text{RISMsr}} & \hat{c}_{02}^{\text{RISMsr}} \\ \hat{c}_{10}^{\text{RISMsr}} & \hat{c}_{11}^{\text{RISMsr}} & \hat{c}_{12}^{\text{RISMsr}} \\ \hat{c}_{20}^{\text{RISMsr}} & \hat{c}_{21}^{\text{RISMsr}} & \hat{c}_{22}^{\text{RISMsr}} \\ \hat{c}_{30}^{\text{RISMsr}} & \hat{c}_{31}^{\text{RISMsr}} & \hat{c}_{32}^{\text{RISMsr}} \\ \hat{c}_{40}^{\text{RISMsr}} & \hat{c}_{41}^{\text{RISMsr}} & \hat{c}_{42}^{\text{RISMsr}} \\ \vdots & \vdots & \vdots \\ \hat{c}_{s0}^{\text{RISMsr}} & \hat{c}_{s1}^{\text{RISMsr}} & \hat{c}_{s2}^{\text{RISMsr}} \end{bmatrix}$$

$$\times \begin{bmatrix} \hat{\chi}_{00} & \hat{\chi}_{01} & \hat{\chi}_{02} \\ \hat{\chi}_{10} & \hat{\chi}_{11} & \hat{\chi}_{12} \\ \hat{\chi}_{20} & \hat{\chi}_{21} & \hat{\chi}_{22} \end{bmatrix}, \tag{A4}$$

where all nondiagonal elements of even rows and columns of the matrix $\hat{\omega}$ have been replaced with zeros, as well as all elements of even rows of the matrix $\hat{h} - \hat{\gamma}$. Because the solution \hat{c}^{RISMsr} will now have all zeros in its even rows, our function

for solvation free energy applies exactly the same as for the values of k for which this matrix reduction is unnecessary. A loop can be written into the program being used to find \hat{c}^{RISMsr} , so that the dimensionality of the matrix $\hat{\omega}$ is gradually reduced by powers of 2 as the severity of the ill conditioning increases in the limit as $k \rightarrow 0$.

- ¹F. Hirata and P. J. Rossky, Chem. Phys. Lett. **83**, 329 (1981); F. Hirata, B. M. Pettitt, and P. J. Rossky, J. Chem. Phys. **77**, 509 (1982).
- ²P. J. Rossky and H. L. Friedman, J. Chem. Phys. **72**, 5694 (1980).
- ³H. C. Andersen and D. Chandler, J. Chem. Phys. **57**, 1918 (1972); D. Chandler and H. C. Andersen, *ibid.* **57**, 1930 (1972).
- ⁴J. P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, 2nd ed. (Academic, London, 1986).
- ⁵T. Morita and K. Hiroike, Prog. Theor. Phys. **23**, 1003 (1960).
- ⁶S. J. Singer and D. Chandler, Mol. Phys. **55**, 621 (1985).
- ⁷D. A. Zichi and P. A. Rossky, J. Chem. Phys. **84**, 1712 (1985).
- ⁸M. Orozco and F. J. Luque, Chem. Rev. (Washington, D.C.) **100**, 4187 (2000).
- ⁹H. A. Yu and M. Karplus, J. Chem. Phys. **89**, 2366 (1988).
- ¹⁰H. A. Yu, B. Roux, and M. Karplus, J. Chem. Phys. **92**, 5020 (1990).
- ¹¹A. Agarwal, F. Brown, and M. R. Reddy, in *Free Energy Calculations in Rational Drug Design*, edited by M. R. Reddy and M. D. Erion (Kluwer Academic/Plenum, New York, 2001).
- ¹²A. Kitao, F. Hirata, and N. Go, J. Phys. Chem. **97**, 10231 (1993).
- ¹³H.-A. Yu, B. M. Pettitt, and M. Karplus, J. Am. Chem. Soc. **113**, 2425 (1991); T. Lazaridis and M. Karplus, in *Thermodynamics in Biology*, edited by E. Di Cera (Oxford University Press, New York, 2000), p. 3; T. Ichiye and D. Chandler, J. Phys. Chem. **92**, 5257 (1988).
- ¹⁴H. Freedman and T. N. Truong, Chem. Phys. Lett. **381**, 362 (2003).
- ¹⁵B. M. Pettitt and M. Karplus, Chem. Phys. Lett. **121**, 194 (1985); B. M. Pettitt, M. Karplus, and P. J. Rossky, J. Phys. Chem. **90**, 6335 (1986); N. V. Prabhu, J. S. Perkyns, H. D. Blatt, P. E. Smith, and B. M. Pettitt, Biophys. Chem. **78**, 113 (1999); P. E. Smith, B. M. Pettitt, and M. Karplus, J. Phys. Chem. **97**, 6907 (1993).
- ¹⁶M. Kinoshita, Y. Okamoto, and F. Hirata, J. Am. Chem. Soc. **120**, 1855 (1998); J. Chem. Phys. **110**, 4090 (1999).
- ¹⁷S. Kast, Phys. Chem. Chem. Phys. **3**, 5087 (2001).
- ¹⁸F. Hirata, P. J. Rossky, and B. M. Pettitt, J. Chem. Phys. **78**, 4133 (1983).
- ¹⁹A. Kovalenko and F. Hirata, J. Chem. Phys. **112**, 10391 (2000); M. Kinoshita, Y. Okamoto, and F. Hirata, J. Comput. Chem. **18**, 1320 (1997).
- ²⁰D. Chandler, Y. Singh, and D. Richardson, J. Chem. Phys. **81**, 1975 (1984).
- ²¹S. Ten-no, J. Chem. Phys. **115**, 3724 (2001).
- ²²S. Ten-no and S. Iwata, J. Chem. Phys. **111**, 4865 (1999).
- ²³D. Chandler, R. Silbey, and B. M. Ladanyi, Mol. Phys. **46**, 1335 (1982); P. J. Rossky and R. A. Chiles, *ibid.* **51**, 661 (1984).
- ²⁴M. T. Heath, *Scientific Computing: An Introductory Survey*, 2nd ed. (McGraw-Hill, New York, 2002).
- ²⁵D. M. Duh and A. D. J. Haymet, J. Chem. Phys. **97**, 7716 (1992).
- ²⁶D.-M. Duh and A. D. J. Haymet, J. Chem. Phys. **103**, 2625 (1995).
- ²⁷D.-M. Duh and D. Henderson, J. Chem. Phys. **104**, 6742 (1996).
- ²⁸W. L. Jorgensen, boss, Yale University, New Haven, CT, 2000.
- ²⁹*CRC Handbook of Chemistry and Physics*, 83rd ed. (CRC, Boca Raton, FL, 2004).
- ³⁰A. Kovalenko, S. Ten-No, and F. Hirata, J. Comput. Chem. **20**, 928 (1999).