Self-Consistent Combination of the Three-Dimensional RISM Theory of Molecular Solvation with Analytical Gradients and the Amsterdam Density Functional Package

Sergey Gusarov,[†] Tom Ziegler,[‡] and Andriy Kovalenko*[,†,§]

National Institute for Nanotechnology, National Research Council of Canada, 9107-116 St., W6-010, ECERF Bldg., Edmonton, Alberta T6G 2V4, Canada, Department of Chemistry, University of Calgary, Calgary, Alberta T2N 1N4, Canada, and Department of Mechanical Engineering, University of Alberta, Edmonton, Alberta T6G 2V4, Canada

Received: August 4, 2005; In Final Form: February 14, 2006

The three-dimensional reference interaction site model with the closure relation by Kovalenko and Hirata (3D-RISM-KH) in combination with the density functional theory (DFT) method has been implemented in the Amsterdam density functional (ADF) software package. The analytical first derivatives of the free energy with respect to displacements of the solute nuclear coordinates have also been developed. This enables study of chemical reactions, including reaction coordinates and transition state search, with the molecular solvation described from the first principles. The method yields all of the features available by using other solvation approaches, for instance infrared spectra of solvated molecules. To evaluate the accuracy of the present method, test calculations have been carried out for a number of small molecules, including four glycine conformers, a set of small organic compounds, and carbon nanotubes of various lengths in aqueous solution. Our predictions for the solvation free energy agree well with other approaches as well as experiment. This new development makes it possible to calculate at modest computational cost the electronic properties and molecular solvation structure of a solute molecule in a given molecular liquid or mixture from the first principles.

1. Introduction

The importance of solvation in chemical and biological processes can hardly be overestimated. Solvent effects on the structure and functionality of solutes are not only important for understanding the mechanisms of many organic and biochemical reactions but also crucial for molecular drug design. Progress in computer hardware and electronic structure methods has stimulated the development of solvent models. Advancements in this direction has been especially impressive over the past few years. Many efforts have been devoted to incorporate solvent effects into quantum chemical approaches in order to accurately model chemical reactions in liquids. The Kohn−Sham density functional theory (KS-DFT)[1,2] has become a widely used model for electronic structure calculations due to its simplicity and low computational cost. Recent improvements of the exchange-correlation functionals based on the generalized gradient approximation (GGA) makes it possible to treat the solute electrons sufficiently accurately. In combination with solvent models, DFT allows one to calculate the potential energy surface (PES) for molecular systems and thus to study chemical reactions.

Methods for including solvent effects are typically subdivided into two types: explicit and continuum solvent models. Explicit (or discrete) methods[3,4] treat a large number of solvent molecules together with the solute, generally at the same or comparable level of accuracy. This approach is in principle the most realistic and accurate model for studying the physical chemistry of solvation, whether the solutes are simple organic molecules or complex biomolecules. However, the main disadvantage of discrete models is that adequate description of solution requires a significant number of solvent molecules. This results in high computational cost and slow convergence. An alternative approach is based on continuum self-consistent reaction field models.[5−11] Here the solvent is represented not by a collection of molecules but rather by a structureless continuous medium with the average properties of the real solvent. A central quantity is the "reaction field", i.e., the electrostatic influence of the solvent on the electronic energy, structure, and properties of the solute molecule. These methods are significantly less expensive computationally than explicit solvent treatments, but such simplified models of solvent are often insufficient to properly represent the specific interactions between the solute and solvent molecules.

Because of the above difficulties, there is significant interest in developing more rapid and accurate hybrid methods. One solution is to make use of the supermolecular approach[12,13] in which the solvated system is modeled by a cluster of the solute and a representative small number of explicit solvent molecules. This "supermolecule" is immersed in a continuous dielectric medium characterized by the macroscopic dielectric constant of the solvent.

Another, theoretically consistent way to introduce a microscopic description of solvation effects involves combination of quantum chemical methods with statistical mechanical theory of molecular solvation based on density distribution functions. One promising approach in this direction is the reference interaction site model (RISM) which yields the radial pair correlation functions between atomic sites constituting the molecules of liquid. It had been pioneered by Chandler and Andersen[14] and extended by Hirata et al. to polar and quadrupolar liquids[15] and to ions in a molecular polar solvent.[16]

* To whom correspondence should be addressed.
[†] National Research Council of Canada.
[‡] University of Calgary.
[§] University of Alberta.

10.1021/jp054344t CCC: $33.50
Published on Web 04/04/2006

The first hybrid of the RISM integral equation theory with electron structure theory is due to Tenno, Hirata, and Kato.[17] They combined the ab initio Hartree−Fock scheme with RISM in a self-consistent field approach referred to as RISM-SCF. This method makes it possible to determine in a self-consistent manner the effect of organic as well as aqueous solvents on the electronic structure of a variety of solute molecules. In addition, Sato et al[18] formulated an analytical energy gradient method for hybrid MCSCF−RISM calculations and applied it to the study of the cis−trans conformational equilibrium of 1,2-difluoroethylene in aqueous solution.

Because RISM is based on the orientational reduction to site−site solute−solvent radially symmetric correlation functions, there is a loss of information about the three-dimensional (3D) spatial organization (structure) of the solvent density around a macromolecular solute of complex shape.[19] Besides the convergence problems, this makes the conventional site−site RISM difficult to use in combination with the KS-DFT scheme which requires spatial density profiles rather than radial site−site distribution functions. To circumvent this limitation, a generalization of RISM theory to three-dimensional space (3D-RISM) has been proposed which yields the 3D correlations of molecular sites for solvent molecules around a solute of arbitrary shape.[20−24] Thus, the 3D-RISM-KH approach is best suited and has been coupled with the KS-DFT in a self-consistent field description of the electronic structure of solvated macromolecules and interfaces.[22,24]

In this paper, we present the implementation of the 3D-RISM method in the Amsterdam density functional (ADF) program.[25] This self-consistent KS-DFT/3D-RISM-KH method includes also calculation of analytical energy gradients. The accuracy of the method is illustrated on a set of examples.

2. Theory

2.1. Kohn−Sham DFT in the Presence of Solvent. The electronic structure of the solute is calculated from the self-consistent KS-DFT equations modified to include the presence of solvent. The total system of the solute and solvent has the Helmholtz free energy defined as

$$A[n_e(\mathbf{r}),\{\rho_\gamma(\mathbf{r})\}] = E_{solute}[n_e(\mathbf{r})] + \Delta\mu_{solv}[n_e(\mathbf{r}),\{\rho_\gamma(\mathbf{r})\}] \quad (1)$$

where E_{solute} is the electronic energy of the solute consisting of the standard components,[9] $\Delta\mu_{solv}$ is the excess chemical potential of solvation coming from the solute−solvent interaction and solvent reorganization due to the presence of the solute, $n_e(\mathbf{r})$ is the electron density distribution, and $\rho_\gamma(\mathbf{r})$ is the classical density distributions of interaction sites $\gamma = 1, ..., s$ of the solvent molecule. The solute energy is determined by the standard KS-DFT expression written in atomic units as

$$E_{solute}[n_e(\mathbf{r})] = T_s[n_e(\mathbf{r})] + \int d\mathbf{r}\, n_e(\mathbf{r})v_i(\mathbf{r}) + \frac{r}{d\mathbf{r}'}\frac{n_e(\mathbf{r})n_e(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{xc}[n_e(\mathbf{r})] \quad (2)$$

where $T_s[n_e(\mathbf{r})]$ is the kinetic energy of a noninteracting electron gas in its ground state with density distribution $n_e(\mathbf{r})$, $E_{xc}[n_e(\mathbf{r})]$ is the exchange-correlation energy, and $v_i(\mathbf{r})$ comprises the external potential and the nuclear attractive potential.

From the minimization of the free-energy functional (1)

$$\frac{\delta A[n_e(\mathbf{r}),\{\rho_\gamma(\mathbf{r})\}]}{\delta n_e(\mathbf{r})} = 0 \quad (3)$$

subject to the normalization condition for N_e valence electrons of solute

$$\int d\mathbf{r}\, n_e(\mathbf{r}) = N_e \quad (4)$$

we obtain the self-consistent KS equation modified due to the presence of solvent[22,24]

$$\left[-\frac{1}{2}\Delta + v_i(\mathbf{r}) + v_h(\mathbf{r}) + v_{xc}(\mathbf{r}) + v_{solv}(\mathbf{r})\right]\psi_j(\mathbf{r}) = \epsilon_j\psi_j(\mathbf{r}) \quad (5)$$

where the Hartree potential is

$$v_h(\mathbf{r}) = \int d\mathbf{r}'\, \frac{n_e(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \quad (6)$$

the electron density distribution is determined by summation over the N_e lowest occupied eigenstates with allowance for their double occupancy by electrons with opposed spins

$$n_e(\mathbf{r}) = \sum_{j=1}^{N_e} |\psi_j(\mathbf{r})|^2 \quad (7)$$

the exchange-correlation potential is the functional derivative

$$v_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[n_e(\mathbf{r})]}{\delta n_e(\mathbf{r})} \quad (8)$$

and the solvent potential is defined as

$$v_{solv}(\mathbf{r}) = \frac{\delta\Delta\mu_{solv}[n_e(\mathbf{r}),\{\rho_\gamma(\mathbf{r})\}]}{\delta n_e(\mathbf{r})} \quad (9)$$

To simplify the calculation of $v_H(\mathbf{r})$ in ADF, we use the fitted density

$$n_e(\mathbf{r}) \approx \sum_{j=1}^{N_e} c_a f_a(\mathbf{r}) \quad (10)$$

which after substitution in eq 6 yields the fitted potential. Here f_a is the set of the single-center Slater functions and the coefficients c_a are determined by least-squares fitting.[26,27] Together with using the locality properties,[28] this allows one to dramatically reduce the amount of calculations necessary for evaluation of the potentials and matrix elements.

The total free energy is calculated as

$$A_{tot} = \sum_{j=1}^{N_e}\epsilon_j - \frac{1}{2}\int d\mathbf{r}\, d\mathbf{r}'\, \frac{n_e(\mathbf{r})n_e(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{xc}[n_e(\mathbf{r})] - \int d\mathbf{r}\, v_{xc}(\mathbf{r})\, n_e(\mathbf{r}) + \Delta\mu_{solv}[n_e(\mathbf{r}),\{\rho_\gamma(\mathbf{r})\}] - \int d\mathbf{r}\, v_{solv}(\mathbf{r})\, n_e(\mathbf{r}) \quad (11)$$

2.2. Three-Dimensional RISM Theory. The classical density distributions of molecular solvent around a solute of arbitrary shape are obtained by using the 3D-RISM integral equations. These equations can be derived from the six-dimensional, molecular Ornstein−Zernike (OZ) equation 28 for the solute−solvent correlation functions by averaging out their dependence on the orientation of solvent molecules but holding the solute orientation intact.[22,24] It has the form

$$h_\gamma(\mathbf{r}) = c_\alpha(\mathbf{r})*(\omega_{\alpha\gamma}(\mathbf{r}) + \rho_\alpha\, h_{\alpha\gamma}(\mathbf{r})) \quad (12)$$

where $h_\alpha(\mathbf{r})$ and $c_\alpha(\mathbf{r})$ are respectively the 3D site total and direct

correlation functions of solvent site α around the solute, $\omega_{\alpha\gamma}(\mathbf{r}) = \delta(\mathbf{r} - l_{\alpha\gamma})/(4\pi l_{\alpha\gamma}^2)$ is the intramolecular matrix of a solvent molecule with site separations $l_{\alpha\gamma}$, ρ_α is the number density of solvent site α, and "*" corresponds to convolution in the direct space and summation over repeating site indices. The site−site radial correlation functions of bulk water, $h_{\alpha\gamma}(\mathbf{r})$, are obtained from the dielectrically consistent RISM theory (DRISM) developed by Perkyns and Pettitt[30] which provides a consistent description of the dielectric properties for ions in polar solvent. Used as input to the 3D-RISM, eq 12 yields the dielectric properties of polar solvent around a solute of general shape at least at the level of a macroscopic dielectric constant.[23,24] The 3D site correlation functions are specified on a 3D linear grid in a rectangular supercell, and the convolution in eq 12 is handled by using the 3D fast Fourier transform (3D-FFT) technique. For a (macro)molecular solute, the box size has to be large enough to ensure decay of the short-range part of the 3D site correlation functions at the supercell boundaries, whereas the long range electrostatic part of the correlations is separated out and handled analytically.[23,24]

The solute−solvent 3D-RISM eq 12 must be complemented with a closure relating the site−site total and direct correlation functions. The approximations which appeared to be successful and adequate for the description of the structure and thermodynamics of various solutes in polar liquids[32] are built in analogy with the closures in theory of simple liquids.[29] The 3D analogue of the so-called hypernetted chain (HNC) closure to the 3D-RISM eq 12 is constructed by assuming

$$g_\gamma(\mathbf{r}) = \exp(-\beta u_\gamma(\mathbf{r}) + h_\gamma(\mathbf{r}) - c_\gamma(\mathbf{r})) \tag{13}$$

where $u_\gamma(\mathbf{r})$ is the 3D interaction potential between solvent site α and the whole solute, $\beta = 1/k_B T$ is the inverse temperature with the Boltzmann constant k_β, and $g_\gamma(\mathbf{r}) = h_\gamma(\mathbf{r}) + 1$ is the 3D site distribution function of solvent site α around the solute. Beglov and Roux[20] have obtained the 3D-RISM/HNC equations, eqs 12 and 13, within the density functional method by reduction of the generalized closure of Chandler, McCoy, and Singer for nonuniform polyatomic systems.[33] In the case of a very deep well of the attractive potential between the solute and individual solvent sites, the 3D-HNC closure (13) can become divergent.[22,34] This artifact is absent in the partial linearization of the HNC closure proposed by Kovalenko and Hirata[22,24] (KH approximation)

$$g_\gamma(\mathbf{r}) = \begin{cases} \exp(\chi_\gamma(\mathbf{r})) & \text{for } \chi_\gamma(\mathbf{r}) \le 0 \\ 1 + \text{for } \chi_\gamma(\mathbf{r}) & \text{for } \chi_\gamma(\mathbf{r}) > 0 \end{cases} \tag{14}$$

$$\chi_\gamma(\mathbf{r}) = -\beta u_\gamma(\mathbf{r}) + h_\gamma(\mathbf{r}) - c_\gamma(\mathbf{r})$$

Equation 14 combines the exponential HNC approximation for the regions of depletion of the distribution function, $g_\gamma(\mathbf{r}) < 1$, the mean spherical approximation (MSA) for the regions of enrichment, $g_\gamma(\mathbf{r}) > 1$, with the function and its first derivative continuous at the joint point $X(\mathbf{r}) = 0$. The 3D-KH approximation (14) enforces proper long-range asymptotics of the direct correlation function $c_\gamma(\mathbf{r})$ in the same way as in the original HNC and MSA closures. The MSA-type linearization prevents the artifact of the distribution function diverging in the regions with a large potential. This partial linearization somewhat reduces and widens high peaks of the distribution functions, whereas it much less affects the coordination numbers of the solvation shells.

The 3D-HNC approximation (13) leads to the excess chemical potential of solvation in the closed analytical form[22,24]

$$\Delta\mu^{\mathrm{HNC}} = \rho kT \sum_\gamma \int \mathrm{d}\mathbf{r} \left[\frac{1}{2}(h_\gamma(\mathbf{r}))^2 - c_\gamma(\mathbf{r}) - \frac{1}{2}(h_\gamma(\mathbf{r})\, c_\gamma(\mathbf{r})\right] \tag{15}$$

equivalent to that derived by Singer and Chandler for the site−site RISM/HNC equations.[35] For the 3D-KH closure (14), the analytical expression is somewhat different[22,24]

$$\Delta\mu^{\mathrm{KH}} = \rho kT \sum_\gamma \int \mathrm{d}\mathbf{r} \left[\frac{1}{2}(h_\gamma(\mathbf{r}))^2 \Theta(-h_\gamma(\mathbf{r})) - c_\gamma(\mathbf{r}) - \frac{1}{2}(h_\gamma(\mathbf{r})\, c_\gamma(\mathbf{r})\right] \tag{16}$$

where Θ in the Heaviside step function which puts the term h^2 in effect in the regions of density depletion only.

Special considerations are required in order to apply the 3D-RISM approach to the case of nonzero net charge of the solute. Within the supercell technique, the Coulomb part of the interaction potential $u(\mathbf{r})$ is synthesized on the super-cell grid by the Ewald summation method.[36] This distorts the long-range Coulomb asymptotics of the direct correlation function $c(\mathbf{r})$. Furthermore, the background compensating charge of the supercell brings about a constant shift in the total correlation function $h(\mathbf{r})$ which departs from unity at a distance from the solute. The resulting error in the solvation chemical potential calculated from expressions (15) or (16) can amount to several tens of kcal/mol. To eliminate these artifacts of the supercell technique, Kovalenko and Hirata[23,24] developed analytical corrections which restore the proper nonperiodic asymptotics of the 3D site total and direct correlation functions following from the 3D-RISM integral equations and cancel out the error in the solvation chemical potential for ions and ion pairs with an accuracy of about 0.1 kcal/mol.

2.3. Effective Potentials and Analytical Gradients. The classical effective potential energy of the solute acting on solvent site γ_α is broken up into the short-range interaction $u_\gamma^{(\mathrm{sr})}(\mathbf{r})$ between the solvent site and the whole solute and the electrostatic energy of the solvent site effective charge q in the fields of the solute nuclei $\phi^{(\mathrm{n})}(\mathbf{r}) = \sum_i Z_i/|\mathbf{r} - \mathbf{R}_i|$ and electrons $\phi^{(\mathrm{e})}(\mathbf{r})$

$$u_\gamma(\mathbf{r}) = u_\gamma^{(\mathrm{sr})}(\mathbf{r}) + q_\gamma(\phi^{(\mathrm{n})}(\mathbf{r}) + \phi^{(\mathrm{e})}(\mathbf{r})) \tag{17}$$

This short-range part is represented by the sum of the 12−6 Lennard-Jones potentials over the solute sites

$$u_\gamma^{(\mathrm{sr})}(\mathbf{r}) = \sum_i 4\epsilon_{i\gamma}\left[\left(\frac{\sigma_{i\gamma}}{r_i}\right)^{12} - \left(\frac{\sigma_{i\gamma}}{r_i}\right)^{6}\right] \tag{18}$$

where $r_i = |\mathbf{r} - \mathbf{R}_i|$ is the separation between the solute nucleus i and solvent site γ and $\sigma_{i\gamma}$ and $\epsilon_{i\gamma}$ are the LJ diameter and energy parameters. The potential of valence electrons acting on a single solvent site, $\phi_\gamma^{(\mathrm{e})}(\mathbf{r})$, is calculated in the density fitting procedure.[25]

The effective potential of solvent acting on the solute electrons, $v_{\mathrm{solv}}(\mathbf{r})$, is the functional derivative of the excess chemical potential of solvation with respect to the electron density distribution of the solute (9). In both the 3D-HNC and 3D-KH forms of the excess chemical potential, eqs 15 and 16, this leads to the expression[22,24]

$$v_{\mathrm{solv}}(\mathbf{r}) = \frac{\delta\Delta\mu_{\mathrm{solv}}}{\delta n_e(\mathbf{r})} = \rho \sum_\gamma \int \mathrm{d}\mathbf{r}'\, h_\gamma(\mathbf{r})\, v_\gamma^{\mathrm{ps}}(|\mathbf{r} - \mathbf{r}'|) \tag{19}$$

TABLE 1: Free Energies (kcal/mol) Predicted by the KS-DFT/3D-RISM-KH Theory for a Set of Small Molecules in Aqueous Solution

	ΔG			ΔG			ΔG	
	LDA	GGA		LDA	GGA		LDA	GGA
				Diatomic				
N_2	2.29	2.46	Cl_2	1.17	1.22	CO	2.34	2.40
HCl	−0.63	−0.49	$O_2(^1\Sigma)$	1.07	1.14	F_2	1.94	2.02
HF	−3.98	−3.91	FCl	1.83	1.91	$O_2(^1\Pi)$	1.01	1.23
OH^-	−116.65	−117.32	SH^-	−79.11	−77.86	CS	0.77	0.84
CN^-	−78.22	−76.15	NO^+	−53.67	−51.02	OCl^-	−83.90	−82.11
				Triatomic				
HCN	−2.14	−2.03	CS_2	2.98	3.11	H_2O	−6.14	−6.08
FNO	−0.98	−0.89	CO_2	1.51	1.68	N_2O	2.04	2.15
O_3	3.32	3.41	ClNO	0.65	0.77	SO_2	0.68	0.77
FCN	−0.01	0.14	ClCN	−1.55	−1.46	ClNS	−1.15	−1.02
NO_2^-	−74.99	−74.15	COS	2.95	3.07	H_2S	−1.10	−1.12
HCC^-	−78.76	−78.52	PH_2^-	−82.58	−82.03	NO_2^+	−51.88	−51.23
N_3^-	−72.44	−71.12	NH_2^-	−112.31	−110.45	HOO^-	−98.75	−97.06
				Tetratomic				
HCCH	−0.56	−0.40	CH_3^-	−89.38	−88.12	NCl_3	3.54	3.67
$FOOF^t$	2.41	2.56	HCNO	−1.20	−1.10	NH_3	−3.99	−3.84
PCl_3	2.19	2.45	$HOOH^g$	−7.72	−7.34	HCCF	2.36	2.45
PF_3	1.38	1.53	HCHO	−2.12	−2.02	$FOOF^g$	1.24	1.41
HCCCl	0.39	0.49	NF_3	3.16	3.24	SOF_2	0.82	0.97
HOCN	−11.84	−11.09	$COCl_2$	2.08	2.31	PH_3	−0.14	0.12
COF_2	2.28	2.41	HNCO	−1.80	−1.56	$HOOH^t$	−8.48	−8.21
CO_3^{2-}	−210.61	−211.04	$HCOO^-$	−79.53	−78.25	CH_3^+	−56.10	−56.03
NO_3^-	−69.65	−68.42	H_3S^+	−59.53	−59.44	H_3O^+	−85.25	−85.01
				Pentatomic				
CH_4	1.14	1.28	POF_3	1.75	1.91	CH_2FCl	−5.56	−5.21
H_2CNH	−3.83	−3.64	CHF_3	1.03	1.25	PSF_3	3.57	3.63
CH_2N_2	0.96	1.02	$HCOOH^s$	−3.84	−3.69	CF_4	3.19	3.25
NOF_3	3.28	3.39	CH_2CO	1.16	1.27	$HCOOH^a$	−9.71	−9.45
CHF_2Cl	−2.19	−1.67	CH_3Cl	1.96	2.17	CH_2F_2	2.06	2.48
$PSCl_3$	2.06	2.34	CH_3Br	8.49	8.55	CF_3Cl	3.49	3.88
SO_2F_2	3.04	3.31	CH_2Br_2	8.35	8.66	CCl_4	2.83	2.99
$CHCl_3$	2.65	2.81	CH_2Cl_2	−2.72	−2.34	CH_3F	−0.07	0.64
SO_2Cl_2	2.84	3.05	$CFCl_3$	1.87	2.02	CF_2Cl_2	3.52	3.77
CH_3O^-	−86.44	−85.14	HPO_3^-s	−9.86	−9.54	$POCl_3$	−0.95	−0.62
HPO_3^-e	−9.95	−9.03	CH_2CN^-	−76.91	−75.74	NH_4^+	−68.32	−68.04
HCO_3^-s	−79.76	−77.95	HCO_3^-e	−77.0	−75.12	PH_4^+	−57.96	−57.11
CH_3S^-	−81.42	−82.12	ClO_4^-	−63.97	−63.11	H_2COH^+	−58.47	−57.10

where $v_\gamma^{ps}(|\mathbf{r} - \mathbf{r}'|)$ is the contribution of site γ into the pseudopotential of a solvent molecule acting on an external electron which is given by the variational derivative of the classical site potential with respect to the valence electron density[22]

$$v_\gamma^{ps}(|\mathbf{r} - \mathbf{r}'|) = \frac{\delta u_\gamma(\mathbf{r})}{\delta n^{(e)}(\mathbf{r}')} \tag{20}$$

Notice that the potential (19) signifies the mean field approximation which follows essentially from the use of the solvation free energy in the form (15) or (16).

The analytical first derivative with respect to the nuclear coordinates R_i is obtained by differentiation of the free energy expression (1)

$$\frac{dA[n_e(\mathbf{r}),\{\rho_\gamma(\mathbf{r})\}]}{dR_i} = \frac{dE_{solute}[n_e(\mathbf{r})]}{dR_i} + \frac{d\,(\Delta\mu_{solv}[n_e(\mathbf{r}),\{\rho_\gamma(\mathbf{r})\}])}{dR_i} \tag{21}$$

where the former term has the same structure as in the case gas phase. The latter term is derived from the excess chemical potential in the form (15) or (16). Its variation can be written as[22,24]

$$\delta\Delta\mu_{solv} = kT\rho^V\sum_\gamma\int d\mathbf{r}\left[h_\gamma(\mathbf{r})\delta h_\gamma(\mathbf{r})\Theta(-h_\gamma(\mathbf{r})) - \delta c_\gamma(\mathbf{r}) - \frac{1}{2}\delta(h_\gamma(\mathbf{r})c_\gamma(\mathbf{r}))\right] = \rho^V\sum_\gamma\int d\mathbf{r}\, g_\gamma(\mathbf{r})\,\delta u_\gamma(\mathbf{r}) \tag{22}$$

Similar to Kovalenko and Hirata,[22,24] we have the following variation of the classical potential between the solute and solvent site γ:

$$\delta u_\gamma(\mathbf{r}) = \int d\mathbf{r}'\left[\left(\frac{\delta u_\gamma(\mathbf{r})}{\delta n^{(e)}(\mathbf{r}')}\right)\delta n^{(e)}(\mathbf{r}') + \left(\frac{\delta u_\gamma(\mathbf{r})}{\delta n_N(\mathbf{r}')}\right)\delta n_N(\mathbf{r}')\right] = \int d\mathbf{r}'\, v_\gamma^{ps}(|\mathbf{r} - \mathbf{r}'|)[\delta n_N(\mathbf{r}') + \delta n^{(e)}(\mathbf{r}')] \tag{23}$$

Substituting eq 23 into eq 22 and using $n_N(\mathbf{r}) = \sum_j q_j\delta(\mathbf{r} - \mathbf{R}_j)$ gives the solvation contribution to the free energy gradients

$$\frac{d(\Delta\mu_{solv})}{dR_i} = \int d\mathbf{r}\left[\rho^V g_\gamma(\mathbf{r})\frac{\partial u_\gamma^{(sr)}(\mathbf{r})}{\partial\mathbf{R}_i} + v_{solv}(\mathbf{r})\frac{\partial n^{(e)}}{\partial R_i}\right] + q_i\frac{\partial v_{solv}(\mathbf{r})}{\partial\mathbf{r}}\bigg|_{\mathbf{r}=R_i} \tag{24}$$

Calculation of the expression (24) does not require much computational effort. The second term in the square brackets is calculated together with the gradients of the exchange-correlation potential,[37] and the rest is calculated in the 3D-RISM procedure. Notice that the first term in the square brackets contributes little to the gradients because large values of the derivatives of the short-range potential $u_\gamma^{(sr)}(\mathbf{r})$ around the core are suppressed by the distribution function $g_\gamma(\mathbf{r})$ exponentially decaying in that region.

We emphasize that the formula (24) differs from the analytical free energy derivative following from the 1D-RISM scheme[18] by the absence of the term representing the change in the solute−solvent correlations with the intramolecular distribution functions (intermolecular matrix).

Notice that the 3D-RISM-KH solvation method can be similarly combined in a self-consistent field approach with any multireference electronic structure theory. Sato et al.[18] pioneered such a combination of 3D-RISM-KH and ab initio CASSCF method.

3. Applications

To illustrate the capability of our combination of the 3D-RISM-KH theory with KS-DFT, we present here preliminary results of the benchmark calculations for a set of simple molecules and organic compounds, taken primarily from ref 39. For the results presented below, the DZP basis set is used and the van der Waals parameters are taken from the CHARMM force fields. All calculations are carried out at the LDA level of theory[40] (unless otherwise specified), and the supercell in 3D-FFT is taken to have 64 × 64 × 64 nodes (or 64 × 64 × 128 for glycine).

3.1. Free Energies for Small Molecules. The calculation of the free energies of solvation ΔG_{solv} were carried out for the representative set of small molecules in aqueous solution (Table 1) and compared to the free energies calculated by other methods. The results are summarized in Table 2. We can see from these data that the predictions of the KS-DFT/3D-RISM-KH combination are in good agreement with the results of the other methods. A noticeable overestimation of the hydration free energy is observed only for strongly charged ions. It can be attributed to the necessity to self-consistently adjust the van der Waals parameters in the case of charged molecules and also to the imperfectness of the closure relation. The latter has considerable space for improvement, for example, by introducing the modification to the KH closure similarly to Omelyan et al.[41]

3.2. Benzene Oxide Valence Tautomerism. The next example is the study of the transition state for the interconversion (valence tautomerization) of benzene oxide and oxepin in (i) isooctane and (ii) 15% aqueous methanol solvents. We chose these solvents to further emphasize the ability of the KS-DFT/3D-RISM-KH method to readily treat arbitrary molecular solvent mixtures. For these systems, it is difficult even to define solvent parameters in COSMO which in this case have no physical meaning because of the complex geometry.[42] The geometry structure of the solute species can be found in ref 39. Table 3 gives presents predictions for these systems which are in good agreement with the results of the DFT/COSMO calculations.[39]

3.3. Glycine Structure in Water Solution. A classical example of geometry optimization in solvent is the study of the hydration effects on the structure and relative stability of glycine. It is the simplest amino acid with the smallest hydrocarbon backbone and is a very instructive biochemical model compound. The importance of this example is that the geometry optimization in solution is required for the zwitterionic form of glycine because this structure does not exist in the gas phase. There are several theoretical investigations on environment effects on the molecular structure of glycine.[39,43,44]

In this study, the calculations were performed for the three neutral conformers 1, 2, and 3 as well as a zwitterionic (ZW) one (Figure 1). The geometrical parameters and its comparison with experiment are presented in Tables 4 and 5.

The neutral conformers have been experimentally characterized in the gas phase through the microwave or IR spectroscopy. They differ in the orientation of the C−N and O−H bonds with respect to the C═O bond and correspondingly in the number and type of hydrogen-bonded interactions. Conformer 1 possesses an internal hydrogen bond between the hydroxyl hydrogen and the nitrogen lone pair and has the largest dipole moment. This form is stable only in gas phase and relaxes into zwitterionic form when the structure is optimized in solution. Zwitterionic (ZW) glycine, the most stable tautomer in water,

TABLE 2: Solvation Free Energy (kcal/mol) Predicted by the KS-DFT-LDA/3D-RISM-KH Theory vs Experiment

obj.	ΔG^{expt}	ΔG
	Diatomic	
N_2	2.3[a]	2.29
CO	2.2[a]	2.34
HF	−5.6[b]	−3.98
HCl	−1.2[b]	−0.63
OH^-	−110[c]	−116.65
CN^-	−75[c]	−78.22
SH^-	−76[c]	−79.11
	Triatomic	
H_2O	−6.3,[d] −4.4[b]	−6.14
HCN	−3.2[c]	−2.14
HCC^-	−73[c]	−78.76
HO_2^-	−101[c]	−98.75
NH_2^-	−95[c]	−112.31
NO_2^-	−73[c]	−74.99
N_3^-	−74[c]	−72.44
PH_2^-	−67[c]	−82.58
	Tetratomic	
NH_3	−4,[b] −4.3[e]	−3.99
HOOH[g]	−8.7[a]	−7.72
HCCH	0.0[a]	−0.56
HCHO	−1.7[a]	−2.12
H_3O^+	−104,[f] −102,[b] −105[c]	−85.25
NO_3^-	−66[c]	−69.65
	Pentatomic	
CH_4	2.0[a]	1.14
NH_4^+	−80.5,[f] −77,[b] −81[c]	−68.32
CH_3O^-	−98[c]	−86.44
CH_2CN^-	−75[c]	−76.91
CH_3S^-	−76[c]	−81.42
PH_4^+	−73[c]	−57.96

[a] Reference 47. [b] Reference 48. [c] Reference 49. [d] Reference 50. [e] Reference 51. [f] Reference 52.

TABLE 3: Free Energies (kcal/mol) for Benzene Oxyde−Oxepine Transformations in the Gas Phase and Two Solvents: Isooctane and 15% Water−Methanol Mixture[a]

		isooctane		H_2O/MeOH	
	gas phase	RISM	COSMO	RISM	COSMO
benzene oxyde	−3.150	−2.911	−2.905	−2.908	−2.902
transition state	−3.143	−2.903	−2.898	−2.899	−2.896
oxepine	−3.147	−2.912	−2.905	−2.910	−2.904
planar	−3.138	−2.903	−2.898	−2.900	−2.896

[a] Results of the KS-DFT-LDA/3D-RISM-KH Theory versus DFT/COSMO.[39]

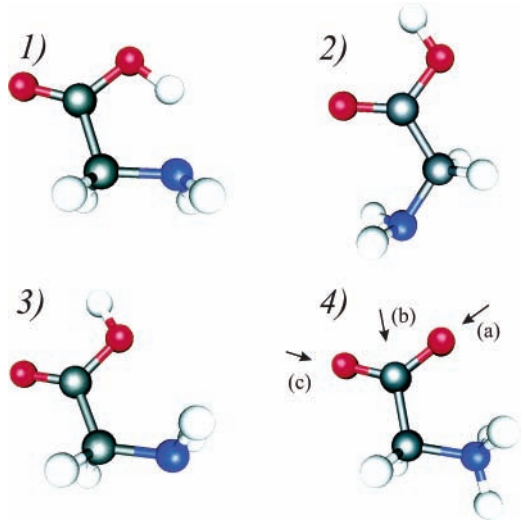

Figure 1. Glycine conformers.

TABLE 4: Geometrical Parameters (Distances in Å and Angles in Degrees) for the Neutral (1) Form of Glycine in Gas and in Aqueous Solution

				water	
coord.	LDA	gas GGA	expt.[a]	LDA	GGA
C_1-O_2	1.207	1.214	1.205	1.214	1.223
C_1-O_3	1.342	1.365	1.355	1.329	1.355
C_4-C_1	1.495	1.518	1.526	1.492	1.519
N_5-C_4	1.428	1.451	1.467	1.430	1.455
O_3-H_6	0.992	0.990	0.966	0.996	0.993
C_4-H_7	1.107	1.102	1.081	1.108	1.103
N_5-H_9	1.029	1.027	1.001	1.030	1.028
$O_3-C_1-O_2$	123.44	123.35		123.67	123.56
$C_4-C_1-O_2$	124.81	125.49	125.1	124.52	125.40
$N_5-C_4-C_1$	115.39	115.85	112.1	115.74	116.07
$H_6-O_3-C_1$	105.47	105.31		108.20	107.77
$H_7-C_4-C_1$	107.64	107.68		107.14	107.26
$H_{10}-N_5-C_4$	109.70	109.16		109.89	109.37
DCH	123.90	123.38		124.02	123.42
DH	57.60	57.09		57.84	57.46

[a] Reference 52.

TABLE 5: Geometrical Parameters (Distances in Å and Angles in Degrees) for ZW Conformer in Aqueous Solution

coord.	LDA	GGA	expt.[a]
C_1-O_2	1.243	1.258	1.250
C_1-O_3	1.262	1.268	1.251
C_4-C_1	1.523	1.547	1.526
N_5-C_4	1.462	1.494	1.476
H_6-N_5	1.076	1.055	
H_7-C_4	1.102	1.097	
H_9-N_5	1.033	1.030	
$O_3-C_1-O_2$	128.49	128.67	
$C_4-C_1-O_2$	117.02	115.52	117.1
$N_5-C_4-C_1$	108.04	109.61	111.9
$H_6-N_5-C_4$	100.06	103.08	
$H_7-C_4-C_1$	110.05	110.23	
$H_9-N_5-C_4$	114.03	113.65	
D1	59.68	60.16	
D2	117.60	117.89	

[a] Reference 53.

has a large gas-phase dipole moment and charge separation along the molecular framework that leads to strong interactions with water.

Table 6 lists the free energies of solvation for the four conformers of glycine and their total dipole moments in the gas phase and aqueous solution. Geometry optimization of the neutral and zwitterionic forms of glycine in water shows that

TABLE 6: Solvation Free Energy ΔG_{Solv} (kcal/mol) and Dipole Moments μ (Debye) of Glycine, Following from the KS-DFT/3D-RISM-KH Theory

	ΔG_{solv}		μ gas		μ water	
	LDA	GGA	LDA	GGA	LDA	GGA
I			5.98	5.71		
II	−10.86	−11.44	1.21	1.12	1.96	1.82
III	−14.54	−16.89	1.85	1.78	2.99	2.89
ZW	−41.54	−43.03			13.07	12.42

solvent has little effect on the structure of the neutral forms but significantly affects the zwitterionic form. The results for ΔG_{solv} obtained by KS-DFT/3D-RISM-KH agree very well with the previous calculations.[39]

The 3D distribution functions calculated from the 3D-RISM-KH theory yields the hydration structure of glycine in aqueous solution in three-dimensional detail. Figure 3 shows the 3D distributions of water oxygen and hydrogen in the first hydration shell around the carboxylate group of glycine. Clearly seen is the hydrogen bonding of water molecules to the carboxylate oxygens of glycine (large blue segments surrounding the two oxygens). The corresponding positions of water oxygen can be determined by the analysis of 3D distribution functions (Figures 2 and 3). Our estimation (outer part of the red shell) agree well with the existing data for a water−glycine cluster obtained in molecular simulations (with force fields adjusted for aqueous solution).[45] As distinct from carboxylate, water molecules around the nitrogen group of glycine are oriented with their hydrogens outward the nitrogen group of glycine (blue segment in the lower part of Figure 3 and blue bands behind the red cloud).

3.4. Timings. To illustrate the efficiency of our implementation, we present the timings comparison for single-point ADF calculations of (6,6) carbon nanotubes of different sizes with the KS-DFT-LDA/3D-RISM-KH and KS-DFT-LDA/COSMO solvation models. To properly assess the convergence rate for different tube lengths, we use a CPU time per one iteration of the self-consistent potential loop. This includes the calculation and diagonalization of the Fock matrix and the evaluation of solvent effect (either the by the 3D-RISM-KH or COSMO). This presentation per self-consistent potential loop is convenient to compare performance of calculation for gas and liquid phases which converge in different numbers of iteration. These numbers for the 3D-RISM-KH and COSMO solvation schemes appear to be rather similar but strongly dependent from the size, electronic configuration, starting guess, etc.

The rest of the CPU time which is spent for evaluation of the solvent effects can be easily estimated using of gas-phase timings also presented in Figure 4. It can be subdivided into the time for solving the equations of the of solvation model and that for calculating the matrix elements of the solvation potential. The latter is comparably small and similar for the 3D-RISM-KH and COSMO cases. The time for solving the 3D-RISM integral equations scals mainly as the 3D discrete Fourier transform in $O(N \log_2 N)$ operations with the number of 3D grid points N. Therefore, for a system big enough (about 180−200 atoms) we expect that the computational efforts for solving the linear system $Aq = B$ with large arrays A and B in the COSMO method[39] become higher than for solving the 3D-RISM-KH equations. This crossover is clearly seen in Figure 4. Overall, the timing efficiency of the KS-DFT/3D-RISM-KH strongly depends on the 3D grid size to accommodate the given solute. It is slower than KS-DFT/COSMO for small molecules, but quickly becomes much more efficient for large macromolecules.

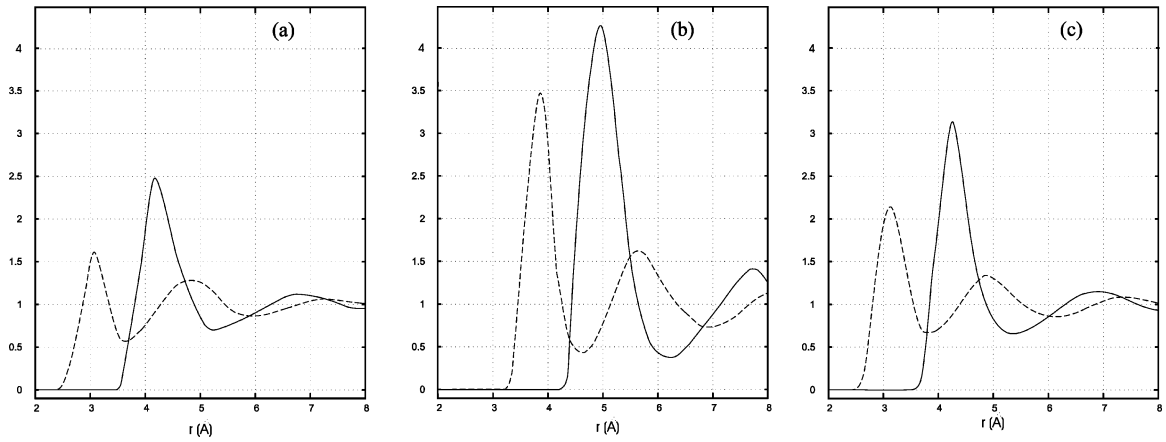

Figure 2. Sections of the 3D distribution functions between glycine and the oxygen (solid lines) and hydrogen (dashed lines) sites of water solvent molecule (a) along C−O axis; (b) along C−C axis; (c) along C−O axis (see Figure 1).

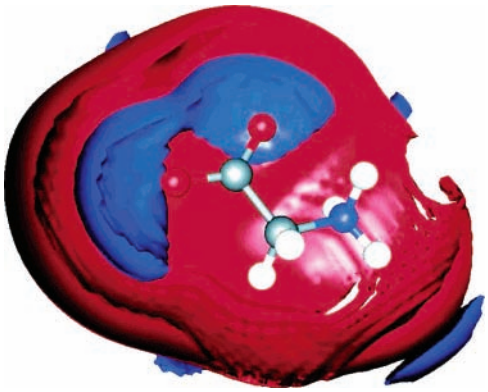

Figure 3. First hydration shell around the carboxylate group of ZW glycine in water: oxygen (red) and hydrogen (blue).

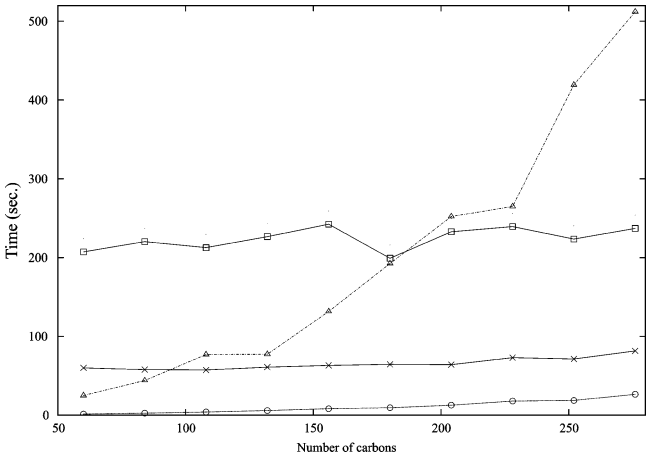

Figure 4. Timings, in seconds, per one self-consistent potential iteration for (6,6) carbon nanotubes of various sizes in the KS-DFT-LDA/3D-RISM-KH (solid line with crosses for 64 × 64 × 128 and squares for 128 × 128 × 128 cell) and KS-DFT-LDA/COSMO[38] (dot−dash line with triangles) solvation models and gas phase (short-dash line with circles).

4. Conclusions

The implementation of the self-consistent KS-DFT/3D-RISM-KH method and the related technique of analytical energy gradients provides a powerful tool to study reaction dynamics of large systems in molecular solvents and mixtures of an arbitrary composition, including electrolyte solutions. (A similar hybrid approach can be applied to any MCSCF/3D-RISM-KH combination.) Calculation of analytical energy gradients requires small additional computational load and is realized using the analytical expression for the solvation free energy and the 3D solvent distribution functions which are most naturally and effectively obtained from the 3D-RISM-KH theory.

Test results for a number of simple molecules and organic compounds are presented in order to show the correctness and accuracy of the KS-DFT/3D-RISM-KH method. It enables efficient treatment of transition state structures and mechanisms of chemical reactions for large nanosystems in solution, with the solvation effects accounted from the first principles. Among other features, the approach also yields IR spectra of nanosystems with molecular solvation effects included (to be considered in detail in subsequent studies). The preliminary benchmarks we obtained show that for nanosystems larger than 150−200 atoms the presented method becomes more efficient than continuum solvation approaches in both time and memory resources.

We emphasize that as distinct from any continuum solvation scheme, the KS-DFT/3D-RISM-KH method yields the solvation shell structure and thermodynamics from the first principles. This is achieved with the accuracy comparable to KS-DFT/MD type approaches but with much smaller computational efforts.

Acknowledgment. This work is supported by the National Research Council of Canada. T.Z. thanks the Canadian government for a Canada Research Chair.

Note Added after ASAP Publication. This article was published on Articles ASAP on April 4, 2006. The author added a new ref 24 and changed reference citation numbers in the text. A paragraph was added at the end of section 2.3 and ref 38 was changed. The corrected version was reposted on Articles ASAP on April 17, 2006.

References and Notes

(1) Hohenberg, P. W. Kohn. *Phys. Rev. B* **1964**, *136*, 864.
(2) Kohn, W.; Sham, L. J. *Phys. Rev. A* **1965**, *140*, 1133.
(3) van Gunsteren, W.; Berendsten, H. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 992.
(4) Vaidehi, N.; Wesolowski, T.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 4264.
(5) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117.
(6) Miertus, S.; Tomasi, J. *Chem. Phys.* **1982**, *65*, 239.
(7) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans.* **1993**, *2*, 799.
(8) Andzelm, J.; Kölmel, C.; Klamt, A. *J. Chem. Phys.* **1995**, *103*, 9312.
(9) Cramer, C.; Truhlar, D. *Chem. Rev.* **1999**, *99*, 2161.
(10) Marten, B.; Kim, K.; Cortis, C.; Friesner, R.; Murphy, R.; Ringnalda, M.; Sitkoff, D.; Honig, B. *J. Phys. Chem.* **1996**, *100*, 11775.
(11) Bashford, D.; Case, D. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129.
(12) Pullman, A.; Pullman, B. *Q. Rev. Biophys.* **1975**, *7*, 505.

(13) Morokuma, K. *J. Am. Chem. Soc.* **1982**, *104*, 3732.
(14) Chandler, D.; Andersen, H. C. *J. Chem. Phys.* **1972**, *57*, 1930.
(15) Hirata, F.; Rossky, P. *Chem. Phys. Lett.* **1981**, *83*, 329. Hirata, F.; Petitt, B.; Rossky, P. *J. Chem. Phys.* **1982**, *77*, 509.
(16) Hirata, F.; Rossky, P.; Pettitt, B. *J. Chem. Phys.* **1983**, *78*, 4133.
(17) Ten-no, S.; Hirata, F.; Kato, S. *Chem. Phys. Lett.* **1993**, *214*, 391.
(18) Sato, H.; Hirata, F.; Kato, S. *J. Chem. Phys.* **1996**, *105*, 1546.
(19) Akiyama, R.; Hirata, F. *J. Chem. Phys.* **1998**, *108*, 4904.
(20) Beglov, D.; Roux, B. *J. Phys. Chem. B* **1997**, *101*, 7821.
(21) Kovalenko, A.; Hirata, F. *Chem. Phys. Lett.* **1998**, *290*, 237.
(22) Kovalenko, A.; Hirata, F. *J. Chem. Phys.* **1999**, *110*, 10095.
(23) Kovalenko, A.; Hirata, F. *J. Chem. Phys.* **2000**, *112*, 10391.
(24) Kovalenko, A. Three-dimensional RISM theory for molecular liquids and solid-liquid interfaces. In *Molecular Theory of Solvation*; Hirata, Fumio, Ed.; Understanding Chemical Reactivity (series); Mezey, Paul G., Series Ed.; Kluwer Acadamic Publishers: Dordrecht, The Netherlands, 2003; Vol. 24, pp 169−275.
(25) Velde, G.; Bickelhaupt, F.; van Gisbergen, S.; Guerra, C.; Baerends, E.; Snijders, J.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931. Guerra, C.; Snijders, J.; te Velde, G.; Baerends, E. *Theor. Chem. Acc.* **1998**, *99*, 391. ADF2004.01, SCM, Theoretical Chemistry; Vrije Universiteit: Amsterdam, The Netherlands, http://www.scm.com.
(26) Baerends, E. J.; Ros, P.; Ellis, D. E. *Chem. Phys.* **1973**, *2*, 41.
(27) Te Velde, G.; et al. *J. Comput. Chem.* **2001**, *22*, 931.
(28) Guerra, C.; Snijders, J.; te Velde, G.; Baerends, E. *Theor. Chem. Acc.* **1998**, *99*, 391.
(29) Hansen, J.; McDonald, I. *Theory of Simple Liquids*, 2nd ed.; Academic: London, 1986.
(30) Perkyns, J. S.; Pettitt, B. M. *Chem. Phys. Lett.* **1992**, *190*, 626; *J. Chem. Phys.* **1992**, *97*, 7656.
(31) Kovalenko, A.; Ten-no, S.; Hirata, F. *J. Comput. Chem.* **1999**, *20*, 928.
(32) Hirata, F. *Bull. Chem. Soc. Jpn.* **1998**, *71*, 1483.
(33) Chandler, D.; McCoy, J.; Singer, S. *J. Chem. Phys.* **1997**, *85*, 5971.
(34) Kovalenko, A.; Hirata, F. *J. Phys. Chem. B* **1999**, *103*, 7942.
(35) Singer, S. J.; Chandler, D. *Mol. Phys.* **1985**, *55*, 621.
(36) Allen, M.; Tildesley, D. *Computer simulation of liquids*; Oxford University Press: New York, 1987.
(37) Verslus, L.; Ziegler, T. *J. Chem. Phys.* **1988**, *88*, 322.
(38) Sato, H.; Kovalenko, A.; Hirata, F. *J. Chem. Phys.* **2000**, *112*, 9463.
(39) Pye, C.; Ziegler, T. *Theor. Chem. Acc.* **1999**, *101*, 396.
(40) Ceperley, D. M.; Alder, B. J. *Phys. Rev. Lett.* **1980**, *45*, 566.
(41) Omelyan, I.; Kovalenko, A.; Hirata, F. *Chem. Phys. Lett.* **2004**, *397*, 368.
(42) Pollack, G.; Kennan, R.; Himm, J.; Carr, P. *J. Chem. Phys.* **1989**, *90*, 6569.
(43) Hall, N.; Smith, B. *J. Phys. Chem. A* **1998**, *102*, 3985.
(44) Romano, S.; Clementi, E. *Int. J. Quantum Chem.* **1978**, *14*, 839.
(45) Watanabe, T.; Hashimoto, K.; Takase, H.; Kikuchi, O. *J. Mol. Struct.* **1997**, *397*, 113.
(46) Giesen, D.; Cramer, C.; Truhlar, D. *J. Phys. Chem. B* **1995**, *99*, 7137.
(47) Pearson, R. *J. Am. Chem. Soc.* **1986**, *108*, 6109.
(48) Dolney, D.; Hawkins, G.; Winget, P.; Liotard, D.; Cramer, C.; Truhlar, D. *J. Comput. Chem.* **2000**, *21*, 340.
(49) Marcus, Y. *Ion Solvation*; Wiley: Chichester, U.K., 1985.
(50) Fox, B.; Beyer, M.; Achatz, U.; Joos, S.; Niedner-Schatteburg, G.; Bondybey, V. *J. Phys. Chem. A* **2000**, *104*, 1147.
(51) Chambers, C.; Hawkins, G.; Cramer, C.; Truhlar, D. *J. Phys. Chem.* **1996**, *100*, 16385.
(52) Iijima, K.; Tanaka, K.; Onuma, S. *J. Mol. Struct.* **1925**, *246*, 791.
(53) Jonsson, P.; Kvick, Å. *Acta Crystallog. B* **1972**, *28*, 1822.