JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Integral Equation Theory of Molecular Solvation Coupled with Quantum Mechanical/Molecular Mechanics Method in NWChem Package

Gennady N. Chuev,*[,†,‡] Marat Valiev,*[,§] and Marina V. Fedotova[∥]

[†]Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, Leipzig, 04103, Germany

[‡]Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences, Pushchino, Moscow Region, 142290, Russia

[§]William R. Wiley Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, P. O. Box 999, Richland, Washington 99352, United States

[∥]Institute of Solution Chemistry, Russian Academy of Sciences, Akademicheskaya Street, 1, 153045 Ivanovo, Russia

ABSTRACT: We have developed a hybrid approach based on a combination of integral equation theory of molecular liquids and quantum mechanical/molecular mechanics (QM/MM) methodology in NorthWest computational Chemistry (NWChem) software package. We have split the evaluations into consequent QM/MM and statistical mechanics calculations based on the one-dimensional reference interaction site model, which allows us to reduce significantly the time of computation. The method complements QM/MM capabilities existing in the NWChem package. The accuracy of the presented method was tested through computation of the water structure around several organic solutes and their hydration free energies. We have also evaluated the solvent effect on the conformational equilibria. The applicability and limitations of the developed approach are discussed.

I. INTRODUCTION

Chemical reactions in solutions play an important role in various chemical and biological processes. Considerable effort has been devoted to developing efficient computational methods, which are able to reveal mechanisms of the reactions in solutions. Although significant progress has been made in the field of computational chemistry in recent decades, highly accurate quantum mechanical (QM) calculations are still limited. For reactions in solutions, it is often possible to simplify the problem and treat only part of the system by density functional or high-level ab initio methods. In this case, QM description is used only for the regions where electronic structure transformations are occurring, while the rest of the system, whose chemical identity remains essentially the same, is treated approximately using classical description. The latter can be developed in a number of different ways and can be broadly classified into two types—continuum or explicit (discrete) solvent models. Continuum self-consistent reaction field methods are computationally efficient[1−3] but often inadequate to properly represent the specific molecular interactions between the solute and solvent molecules. The discrete molecular mechanics (MM) models are, in principle, more realistic,[4,5] since they treat explicitly intermolecular interactions. However, their use dramatically increases the computational time necessary to carry out sufficient statistical sampling.

Given the above difficulties, there is an ongoing effort to develop new accurate and efficient solvation models that can be combined with QM calculations. Among the recent developments in this area is a combination of QM methods with integral equation theories of molecular liquids.[6,7] Similar to explicit solvation models, this approach provides detailed information about intermolecular interactions in terms of solute−solvent distribution functions. There are various modifications of the method depending on the level of the QM treatment and the type of the applied integral equations. The original model was based on the combination of the one-dimensional reference interaction site model (1D RISM) proposed by Chandler and Andersen[8] with ab initio Hartree−Fock scheme. Since then, several variations have been developed including 1D RISM self-consistent field (SCF),[6] and the combination of extended RISM and quantum mechanical solvation model (XSOL),[9] treating partial solute charges in a different manner. The analytical energy gradient formula has been proposed[10] within the 1D RISM SCF approach and successfully applied to a large number of systems (see review in ref 11).

In the 1D RISM model, the solvent description is built around radially symmetric site−site distribution functions, and the detailed information about the three-dimensional (3D) structure of the solvent density around a solute of complex shape is not available. The latter leads to problems in calculations of 3D electrostatic potential induced by solvent, which is required for accurate treatment of solute electronic structure. Recently two methods[12,13] have been proposed to circumvent this limitation. The first one is based on the solution of molecular Ornstein−Zernike (MOZ) integral equation, referred to as the MOZ-SCF method. The MOZ-SCF theory incorporates the exchange repulsion/charge-transfer interactions for calculating the solute−solvent interactions and explores the three-dimensional (3D) solvent density profile through the rotational invariant expansions of the correlation functions. Another generalization of the method is the 3D RISM SCF[14] with the closure proposed by Kovalenko

Received: December 25, 2011
Published: March 14, 2012

and Hirata (3D-RISM-KH),[15] which has been implemented in a self-consistent manner in the Amsterdam density functional (ADF) package.[13] Although the 3D RISM-KH and the MOZ-SCF models provide more sophisticated calculations, they are also more time-consuming by 1−2 orders than the 1D RISM-SCF method. Today there is a large family of one-dimensional (1D) and 3D RISM SCF models depending on the treatment of electrostatic charges (see, for example, refs 16 and 17) or the level of QM calculations.[18−20] Recently, Yoshida et al. proposed an RISM/QM/MM approach in which solute regions are treated by the ordinary QM/MM method, while the solvent region is handled with the RISM theory.[21]

In this paper we present a new hybrid methodology that combines the 1D RISM approach with combined quantum mechanical/molecular mechanics (QM/MM) methodology. The approach utilizes a dual solvent model which includes discrete solvent representation in terms of MM force field and site−site radial distribution description through the RISM model. Both solvent descriptions are coupled to the same QM representation of the solute leading to QM/MM and QM/RISM descriptions, respectively. The QM/MM representation allows one to provide a highly accurate description of the solute structure and its electrostatic field in terms of point electrostatic potential (ESP) charges. The latter are then utilized in the QM/RISM model, providing efficient description of the solvent structure and energetic. The entire program is implemented as part of open source NorthWest computational Chemistry (NWChem) software package developed in the W. R. Wiley Environmental Molecular Sciences Laboratory (EMSL) at the Pacific Northwest National Laboratory.[22] It takes advantage of the QM/MM module in NWChem[23,25,26] and provides an efficient alternative to otherwise time-consuming QM/MM free energy calculations. While initial implementation is built around the 1D RISM model, the provided software infrastructure can accommodate 3D RISM implementation as well. The implementation is tested by the analysis of hydrated structure and thermodynamic properties of several organic polar solutes such as acetic acid, methylamine, and their ionized forms: acetate and methylammonium ions. Solvent effects on the structure and functionality of the solutes are important for understanding the mechanisms of biochemical reactions, because the hydration of these solutes represents a typical example of biomimetic behavior. The applicability of the method is also tested for the calculation of conformational equilibria in aqueous solution using syn- and anti-isomers of acetic acid.

II. METHODOLOGY

Within the framework of the 1D RISM approach,[8] the solute and solvent molecules are modeled as a set of sites interacting via pairwise distance-dependent potentials $u_{i\alpha}(r)$, which are usually represented by the long-range electrostatic $u_{i\alpha}^{el}(r)$ and the short-range Lennard-Jones (LJ) $u_{i\alpha}^{LJ}(r)$ as

$$u_{i\alpha}(r) = u_{i\alpha}^{el}(r) + u_{i\alpha}^{LJ}(r) \quad u_{i\alpha}^{el}(r) = \frac{q_i q_\alpha}{r}$$
$$u_{i\alpha}^{LJ}(r) = 4\epsilon_{i\alpha}\left[\left(\frac{\sigma_{i\alpha}}{r}\right)^{12} - \left(\frac{\sigma_{i\alpha}}{r}\right)^{6}\right] \quad (1)$$

where we use Greek letters to indicate solvent sites, q_i, q_α are the partial electrostatic charges of the corresponding solute and solvent sites, and $\epsilon_{i\alpha}$, $\sigma_{i\alpha}$ are the LJ solute−solvent interaction parameters.

The solute and the solvent intramolecular structure of molecules are described by the relevant intramolecular correlation functions $\omega_{ij}(r)$ and $\omega_{\alpha\beta}(r)$; the former is written as

$$\omega_{ij}(r) = \delta_{ij} + [1 - \delta_{ij}]\frac{\delta(r - r_{ij})}{4\pi r_{ij}^2} \quad (2)$$

where r_{ij} is the distance between the solute sites and $\delta(r-r_{ij})$ is the Dirac δ-function. The solvent intramolecular function $\omega_{\alpha\beta}(r)$ has a similar form and is fixed, since the solvent molecules are considered to be rigid. We note also that coordinates of solute sites and distances between them as well as solute partial charges q_i are input data for 1D RISM calculations. They can be be obtained by the QM optimization of solute geometry and electronic structure or with the use of the conventional force field data.

The intermolecular solute−solvent correlations are described within the 1D RISM method by the pairwise total correlation functions $h_{i\alpha}(r)$ related with the relevant radial distribution functions (RDF) $g_{i\alpha}(r) = h_{i\alpha}(r) + 1$. These functions are coupled with the direct correlation functions $c_{i\alpha}(r)$ via the set of RISM integral equations:[8]

$$h_{i\alpha}(r) = \sum_{j=1}^{N_u}\sum_{\beta=1}^{N_v}\int_{R^3}\int_{R^3}\omega_{ij}(|\mathbf{r} - \mathbf{r}'|)\, c_{j\beta}(|\mathbf{r}' - \mathbf{r}''|)\, \chi_{\beta\alpha}(\mathbf{r}'')\, d\mathbf{r}'\, d\mathbf{r}'' \quad (3)$$

where $\chi_{\beta\alpha}(r) = \omega_{\beta\alpha}(r) + \rho h_{\beta\alpha}(r)$ are the partial structure factors of the bulk solvent and ρ is the bulk density, while N_u and N_v are the number of sites of solute and solvent, correspondingly. To make eq 3 complete, we should apply closure relations. The most popular is the hypernetted closure (HNC) written as[27]

$$h_{i\alpha}(r) = \exp[-\beta u_{i\alpha}(r) + h_{i\alpha}(r) - c_{i\alpha}] - 1 \quad (4)$$

where $\beta = 1/k_B T$, k_B is the Boltzmann constant, and T is the temperature. However, in many cases the application of HNC leads to a slow convergence rate, and even divergence of the numerical solution due uncontrolled growth of the exponent. Kovalenko and Hirata[15] have proposed the KH closure by linearization of the exponent:

$$h_{s\alpha}(r) = \begin{cases} \exp[\Xi_{i\alpha}(r)] - 1 & \text{when } \Xi_{i\alpha}(r) < 0 \\ \Xi_{i\alpha}(r) & \text{when } \Xi_{i\alpha}(r) > 0 \end{cases}$$
$$\Xi_{i\alpha}(r) = -\beta u_{i\alpha}(r) + h_{i\alpha}(r) - c_{i\alpha}(r) \quad (5)$$

The evaluations of thermodynamic properties of solute is straightforward within the framework of the HNC and the KH approximations, if the data on $h_{i\alpha}(r)$ and $c_{i\alpha}(r)$ are known. The change ΔG in the free energy of the solute is given by[6]

$$\Delta G = E_{reor} + \Delta\mu \quad (6)$$

where the first term E_{reor} is the reorganization energy caused by changes in charge distribution and geometry of the solute, while the second term $\Delta\mu$ is the excess chemical potential produced by changes in solvent structure around the solute. Within the framework of the RISM theory there are several expressions

which allow one to obtain values of the $\Delta\mu$. We have applied the most popular expressions, namely, the Gaussian fluctuations (GF),[28] the HNC,[27] and the KH expressions,[15] which are given by

$$\Delta\mu^{GF} = \frac{2\pi\rho}{\beta}\sum_{i=1}^{N_u}\sum_{\alpha=1}^{N_v}\int_0^\infty [-2c_{i\alpha}(r) - c_{i\alpha}(r)\, h_{i\alpha}(r)]\, r^2\, dr$$

$$\Delta\mu^{HNC} = \Delta\mu^{GF} + \frac{2\pi\rho}{\beta}\sum_{i=1}^{N_u}\sum_{\alpha=1}^{N_v}\int_0^\infty h_{i\alpha}{}^2(r) r^2\, dr$$

$$\Delta\mu^{KH} = \Delta\mu^{GF} + \frac{2\pi\rho}{\beta}\sum_{i=1}^{N_u}\sum_{\alpha=1}^{N_v}\int_0^\infty \Theta(-h_{i\alpha}(r))\, h_{i\alpha}{}^2(r) r^2\, dr \quad (7)$$

where $\Theta(x)$ is the Θ function.

The set of the RISM eqs 3 and 4 or 5 can be solved by the standard numerical iterative scheme using Bessel−Fourier transforms for the calculation of the convolution integrals. The input data for such calculations consist of solute charges q_i, the LJ parameters of solute−solvent interactions, the solute intramolecular functions $\omega_{ij}(r)$, the partial structure factors of the solvent, temperature, and bulk solvent density. The details of the calculations are described in section III. The output of these calculations are the solute−solvent total and direct correlation functions. We note also that in the original RISM-SCF approach iterations are continued until both the electronic and solvent structures become self-consistent within a given convergence criteria,[6] while the self-consistent calculations require $N_{QM}N_{RISM}$ cycles, where N_{QM} is the number of loops for QM evaluations, while N_{RISM} is the same for conventional 1D RISM calculations. We use another procedure in which we split the QM and RISM cycles. Initially the partial charges and geometry of the solute are optimized with the conventional QM/MM approach with short-time molecular simulations, which requires $N_{QM/MM}$ cycles, and then we run the RISM cycle calculating solute−solvent correlation functions and thermodynamics of solvation. Thus, the total number of cycles is $N_{QM/MM} + N_{RISM}$ which is sufficiently less than $N_{QM}N_{RISM}$. Such an algorithm allows us to use the conventional QM/MM routine used in the NWChem package[23] and significantly reduce the time of calculations.

III. COMPUTATIONAL DETAILS

A. QM/MM Calculations. The solute molecules have been treated by the conventional QM/MM routine implemented in the NWChem package.[23,25] The QM treatment was based on the DFT/B3LYP level of theory with the 6-31+G* basis set. In the case of hydration in pure water the organic molecules were solvated in a 30 Å cubic box of 1304 classical SPC/E water molecules. First we performed an initial optimization of the entire system (including solvent) with the use of multilevel QM/MM optimization. This method performs a sequence of alternating optimization cycles of the QM and MM regions. Then the solvent part of the system was equilibrated at 100 ps constant temperature (T = 298.15 K) by classical molecular dynamics simulations. During this simulation the QM region was represented by fixed point charges. These charges have been derived by the conventional procedure fitting ESP[29,30] based on the uniform rectangular grid with 0.5 Å separation. The grid was further pruned by removing points that were outside 30 Å from the closest atomic center and points that were inside the atomic radius. The atomic radius was set to the van der Waals radius plus probe radius of 0.7 Å. The effective charges were recalculated in each optimization cycle by fitting the electrostatic field outside the QM region to that produced by the full electron density representation. After the equilibration the system was optimized once more using multilevel QM/MM optimization. The resulting structure of the solute along with computed electrostatic potential charges was used as an input for 1D RISM calculations. At the same time, constant temperature (T = 298.15 K) classical molecular dynamics simulations of the solvent (again with QM region represented by fixed point charges) were continued for ∼1 ns to provide relevant radical distribution functions for comparison with RISM calculations.

The reorganization free energies were calculated by performing separate DFT/B3lYP optimizations of the solute molecules in the gas phase and comparing them with corresponding QM/MM results. In our calculations, we apply

$$E_{reor} = \Delta G_r = \Delta E_r + \Delta\Delta G_r \quad (8)$$

where ΔE_r is the change in the zero-point energies (difference between the intramolecular energy of solute and the reference energy in the gas state), while the term $\Delta\Delta G_r$ takes entropic contributions into account. In each case, zero-point energies, enthalpies, and entropy contributions were calculated at the same DFT/B3lYP level with a frequency scaling factor of 0.9857.[34] Those were also combined with CCSD(T)[35] QM/MM and gas-phase energy calculations with aug-cc-pvdz[64] basis set to get reorganization free energies at the CCSD(T) level.

B. One-Dimensional RISM Calculations. The optimized structure and partial charges of the solutes are taken from the QM/MM loop in the NWChem package or from the PubChem database[31] and the GAFF parameters.[24] The LJ parameters for the quantum solute and solvent were taken from standard Amber force field definitions.[24] Similarly to ref 32, we have used a modified LJ diameter σ_H = 0.8 Å for lthe hydroxyl hydrogen of water to provide a better agreement for formation of hydrogen bonding in pure water. The Lorentz−Berthelot combination rules for the Lennard-Jones parameters are used for potentials between different species. Two routines have been implemented for calculations of solvent and solute−solvent structure. Both routines are based on the fast Fourier transform applied for solution of the RISM integral equations. The conventional method is applied to decompose the long- and the short-range parts of direct correlation functions. The modified direct inversion in the iterative subspace (MDIIS) method[33] has been used to accelerate convergence of the numerical solution. The iteration process was stopped when the accuracy of the nth iteration had reached the threshold $\epsilon = 10^{-10}$. Initially we calculate solvent structure on the basis of the standard RISM theory, and then the partial solvent structure factors stored in k-space are used to calculate solute−solvent correlation functions.

IV. RESULTS AND DISCUSSION

A. Solute Structure. First, to reveal the effect of self-consistency on the evaluations of charges, we compare partial charges and bonds of acetic acid and acetate ion obtained by QM/MM NWChem module and by RISM-SCF calculations.[36,37] Table 4A yields the comparison for methyl carbon

Figure 1. QM/MM optimized structures of acetate (left) and acetic acid (right) with distances shown in angstroms.

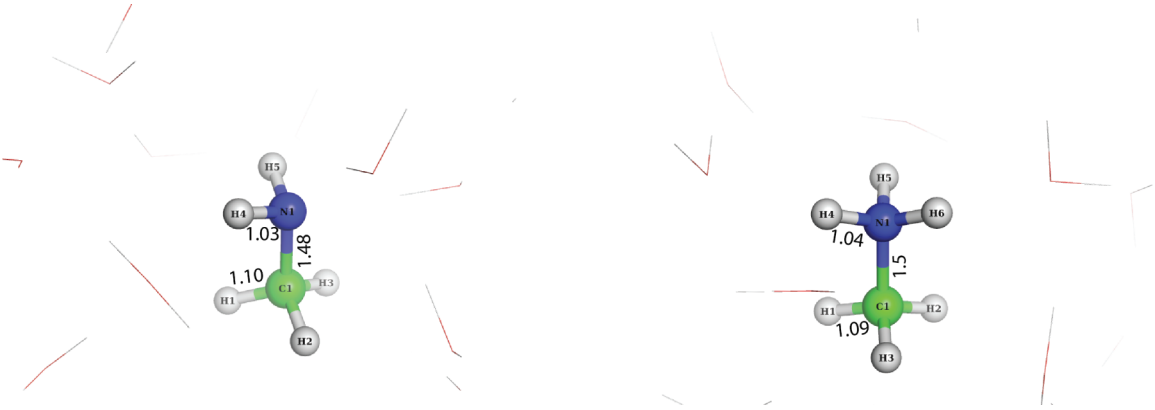

Figure 2. QM/MM optimized structures of methylamine (left) and methylammonium (right) with distances shown in angstroms.

Table 1. Partial Charges and Dipole Moments, D (in Debye), of Hydrated Acetic Acid and Acetate Ion Calculated by the Present Method and RISM-SCF[36] (in Parentheses)

	charges						
	C_{me}	H_{me}	C	O	O_H	H	**D** (D)
CH_3COO^-	−0.47 (−0.42)	0.11 (0.09)	1.28 (1.63)	−1.04 (−1.25)	−1.10 (−1.25)		7.00 (7.23)
CH_3COOH	−0.44 (−0.68)	0.14 (0.20)	1.00 (1.26)	−0.77 (−0.90)	−0.73 (−0.82)	0.52 (0.55)	2.78 (3.00)

C_{me}, methyl hydrogens H_{me}, carbonyl carbon C, carbonyl oxygen O, protonated carbonyl oxygen O_H, and the corresponding proton H. As it is seen, there is an agreement between our results and those obtained by RISM-SCF.[36] Although RISM-SCF provides the charges whose absolute values exceed slightly that obtained by the present method, the difference between the charges calculated by these methods is less than 30%. The same trend takes place for partial charges of methylamine and methylammonium ion. Our calculations yield −1.0 and −0.18 for nitrogen atoms of these solutes, while the RISM-SCF provides −1.23 and −0.12, respectively.[38] Although the RISM-SCF calculations provide solute bonds which are less than those obtained by our method, nevertheless the difference between the structures of the hydrated acetic acid optimized by the methods are rather small (Figures 1 and 2). The latter leads also to a good agreement between dipole moments of acetic acid and acetate ion, calculated by the present method and the RISM-SCF methods; the deviation between the results does not exceed 7% (Table 1) that is sufficiently less than the spread obtained by self-consistent XSOL calculations[9] in which a distinct method has been applied to evaluate partial charges. At the same time, we should indicate also that the obtained partial charges are rather different from those derived from the GAFF parameters.[31] For example, the difference between the partial charges obtained by NWChem and those used in GAFF[24] exceed 50% for carbonyl atoms of acetate ion. Therefore, we have indicated that a split of the quantum and the classical loops, which calculate separately the optimized solute structure and solute−solvent correlations, does not effect sufficiently the quality of the optimization of solute geometry and charges.

B. Solvent Structure. *Solvation of Methyl Group.* As a first step, we investigated solvation of the methyl group for acetic acid, methylamine, and their ionized states. For this purpose we have calculated the methyl carbon−water oxygens ($C-O_w$) radial distribution functions (RDFs) with GAFF/RISM, QM/RISM, and QM/MM approaches (see Figure 3).

The overall shape of methylammonium and methylamine RDFs looks fairly similar to each other. We observe a broad peak corresponding to the first solvation shell around the methyl group. All three methods show that charged methylammonium is characterized by slightly higher peak intensity and shorter distances compared to its neutral counterpart methylamine. At the same time, the more extended solvation shell around neutral methylamine has slightly higher

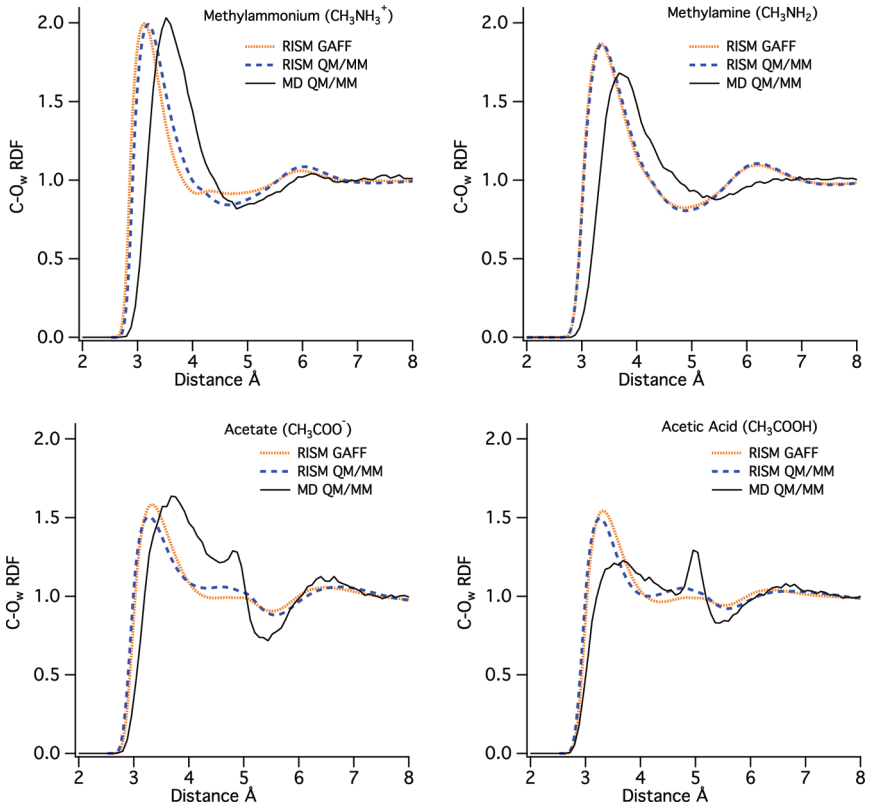

Figure 3. Methyl carbon−water oxygen (C−O_w) radial distribution function for methylammonium (top left), methylamine (top right), acetate (bottom left), and acetic acid (bottom right),.

coordination numbers (see Table 2). Aside from the small differences for charged methylammonium molecule, both GAFF/RISM and QM/RISM calculations produce nearly the same RDFs. Compared to the QM/MM MD, we observe slight underestimation of the position of the first peak (0.3 Å for methylamine and 0.4 Å for methylammonium). Regarding the intensity of the first peak, both RISM methods show good agreement with QM/MM simulation for methylammonium but overestimate it for methylamine by about 10%. Both RISM schemes do fairly well with regard to coordination numbers (see Table 2), with GAFF/RISM showing agreement somewhat with MD data.

Table 2. Intensity of First Peak of the RDF, Its Position (in Parentheses), and Coordination Numbers (in Square Brackets) for Methylamine, Methylammonium, Acetic Acid, and Acetate

	RDF	GAFF/RISM	QM/RISM	QM/MM
$CH_3NH_3^+$	C_mO_w	2.0 (3.1 Å) [13.5]	2.0 (3.2 Å) [13.0]	2.0 (3.5 Å) [13.8]
	H_aO_w	1.5 (1.8 Å) [1.1]	1.1 (1.9 Å) [1.0]	1.8 (1.9 Å) [1.4]
	N_aO_w	2.0 (3.0 Å) [6.2]	2.3 (2.9 Å) [5.7]	3.5 (2.9 Å) [5.4]
CH_3NH_2	C_mO_w	1.9 (3.4 Å) [14.9]	1.9 (3.4 Å) [15.1]	1.7 (3.7 Å) [14.5]
	H_aO_w	0.7 (1.8 Å) [0.5]	0.7 (1.8 Å) [0.4]	0.6 (2.1 Å) [0.6]
	N_aH_w	1.5 (1.8 Å) [1.4]	1.4 (1.7 Å) [1.2]	1.1 (1.8 Å) [1.2]
$CH_3CO_2^-$	C_mO_w	1.6 (3.3 Å) [10.7]	1.5 (3.3 Å) [9.6]	1.6 (3.7 Å) [11.3]
	O_cH_w	3.4 (1.6 Å) [3.0]	4.8 (1.6 Å) [3.5]	5.6 (1.7 Å) [3.7]
CH_3CO_2H	C_mO_w	1.5 (3.3 Å) [9.7]	1.5 (3.3 Å) [7.8]	1.3 (3.7 Å) [9.0]
	$H_{oh}O_w$	2.1 (1.7 Å) [0.9]	2.5 (1.7 Å) [1.0]	3.0 (1.8 Å) [1.0]
	O_cH_w	1.4 (1.7 Å) [1.6]	2.5 (1.6 Å) [2.0]	2.4 (1.8 Å) [2.3]

As with methylammonium and methylamine, both RISM schemes predict similar solvation structure for acetic acid and acetate ion. There are however some differences compared to QM/MM MD results. An interesting feature of QM/MM MD RDF in this case is the appearance of the distinct secondary peak in the region between 4 and 5 Å. Our analysis shows that this peak is associated with the solvent structure around the carboxylic group on the other end of the molecule. Neither RISM calculation manages to fully capture this secondary peak, but it is more pronounced in the QM/RISM calculations. This can be explained by the fact that QM/RISM calculation provides better description of the solvation shell around the carboxylic group than GAFF/RISM (see Figures 6 and 7 and the discussion that follows later). The change in the charge state from −1 in the case of acetate ion to neutral for acetic acid has little effect on the location of the first peak, but it does lower its intensity. While rather small within the RISM treatment, this change in intensity is quite pronounced according to dynamical QM/MM simulation. Given the differences in RDF noted above, we observe more variation in coordination numbers between RISM and MD calculations. Again GAFF/RISM shows the best agreement with MD data in this case, underestimating coordination numbers by 0.6 for acetate ion and overestimating by 0.7 for acetic acid.

dx.doi.org/10.1021/ct2009297 | J. Chem. Theory Comput. 2012, 8, 1246−1254

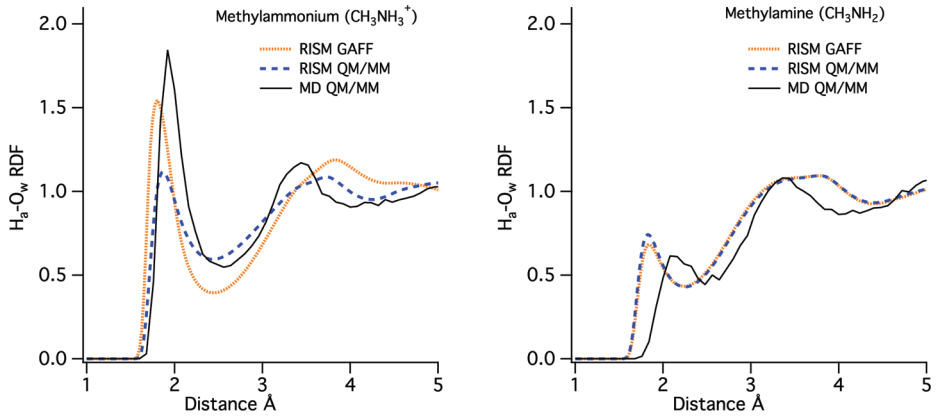

Figure 4. Amine hydrogen−-water oxygen radial distribution function for $CH_3NH_3^+$ (left) and CH_3NH_2 methylamine (right).

Solvation of Amine Group. To investigate solvation of amine groups of methylamine and methylammonium ion, we have computed amine hydrogen−water oxygen (H_a−O_w) and amine nitrogen−water hydrogen (N_a−H_w) RDFs.

The amine hydrogen−water oxygen RDF (see Figure 4) provides a good probe of the hydrogen-bonding environment around the amine group in methylamine and methylammonium. All three methods show a well-defined peak around 1.8−2.0 Å that corresponds to hydrogen acceptor bonds formed between amide hydrogens and water molecules. The latter interaction is much stronger in methylammonium due to the excess positive charge. The two RISM schemes show similar lengths of hydrogen acceptor bonds (as inferred from the position of the first peak), which agree well with MD results for $CH_3NH_3^+$ but slightly underestimated for CH_3NH_2. As far as the strength of these interactions, the largest disagreement among all three methods is observed for $CH_3NH_3^+$ where RISM substantially underestimates the intensity of the peak. The agreement is somewhat better for CH_3NH_2, with RISM treatment now overestimating the intensity. Taking the area under the first peak as an estimator of the number of acceptor hydrogens bonds, we find that RISM predicts 0.5 hydrogen bonds per solute hydrogen for CH_3NH_2, which compares well with our QM/MM MD value of 0.6 and previous simulations.[41] For $CH_3NH_3^+$, RISM calculations show 1 hydrogen bond, which correlates well with the 1.4 value from our QM/MM MD simulations and previous estimates lying in the range of 1−1.2.[41,43]

In CH_3NH_2 the lone on pair on amide nitrogen can also be from hydrogen donor bonds with the solvent. This interaction can be analyzed through N_a−H_w RDF (see Figure 5). We observe that RISM treatment overestimates the strength of this interaction but does a reasonable job regarding peak location or the length of the corresponding hydrogen bond. MD simulations predict that on average there will be 1 water molecule that is hydrogen bonded to amide nitrogen. QM/RISM calculation leads to the same result, and GAFF/RISM slightly overestimates it.

Solvation of Carboxylate Group. To analyze solvation of the carboxylic group in acetic acid and acetate ion, we have compared RDFs of carbonyl oxygen and water hydrogen (O−H_w) and also hydroxyl hydrogen and water oxygen (H−H_w) in the case of acetic acid (see Figures 7 and 6).

All three simulations produce nearly the same values for the positions of the first RDF peak, with RISM predicting slightly lower values (0.1 Å) compared to MD. There is less agreement

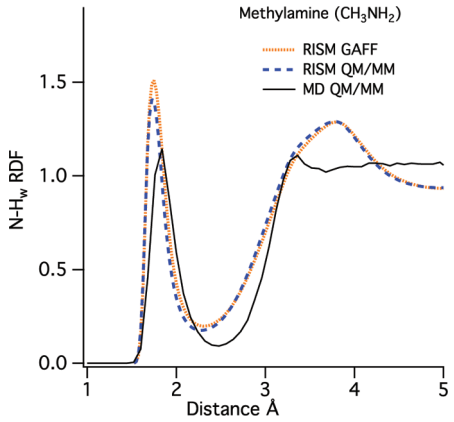

Figure 5. Amine nitrogen−water hydrogen RDF for CH_3NH_2.

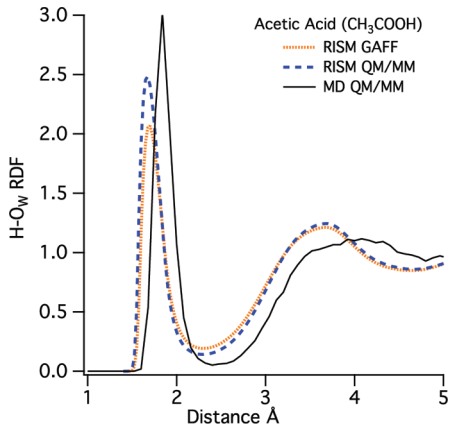

Figure 6. Hydroxyl hydrogen−water oxygen RDF for acetic acid.

regarding the peak intensities. GAFF/RISM yields consistently lower values than QM/RISM, with the latter showing the best agreement with the QM/MM MD simulations. For the acetic acid, the analysis of H−H_w RDF (see Figure 6) shows that both RISM calculations predict 1.0 hydrogen bonds for the hydroxyl group. This agrees well with the 1.0 value obtained in our QM/MM MD calculations and prior Monte Carlo simulations.[44] For the carbonyl oxygen (see O−H_w RDF on Figure 7) we obtain 1.6 hydrogen bonds with GAFF/RISM and 2.0 with QM/RISM. QM/RISM predictions in this case are in good agreement with the 1.9 value obtained in prior semiempirical

dx.doi.org/10.1021/ct2009297 | J. Chem. Theory Comput. 2012, 8, 1246−1254

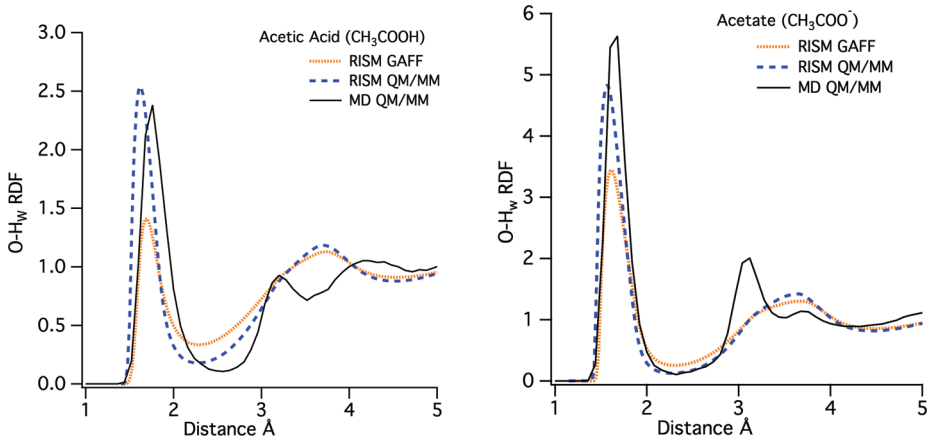

Figure 7. Carbonyl oxygen−water hydrogen RDF for acetic acid (left) and acetate ion (right).

Table 3. Comparison of Excess Chemical Potential Calculated with the HNC and the KH Closures with Various Expressions for Chemical Potential in the Present Work (QM/RISM) and Prior RISM-SCF[36,37] and XSOL-AM1[9] Simulations[a]

property		CH_3NH_2	$CH_3NH_3^+$	$CH_3CO_2H^s$	$CH_3CO_2H^a$	CH_3COO^-
$\Delta\mu^{HNC}$	HNC	5.5	−44.5	−1.4 (−1.4,[36] −2.8 [37])	−12.2 (−12.5 [37])	−114.2 (−117.4 [36])
$\Delta\mu^{GF}$	HNC	−5.2 (−5.4[9])	−59.2	−18.3 (−8.7 [9])	−29.8	−140.4
$\Delta\mu^{KH}$	KH	−2.0	−54.9	−12.5	−23.3	−116.2
$\Delta\mu^{GF}$	KH	−3.3	−56.5	−14.8	−25.9	−123.4
ΔE_r	B3LYP	5.0	2.9	4.4	9.5	18.2
	CCSD(T)	6.7	3.4	7.5	14.3	28.9
	AM1[9]	1.0		2		
	SCF[36,37]	1.0		4.8,[36] 5.5 [37]	10.2 [37]	20.4 [36]
ΔG_r	B3LYP	7.0	4.3	7.3	12.2	24.3
	CCSD(T)	8.7	4.8	10.5	17.0	35.0

[a]Also shown are internal energies and free energies of solute reorganization. Subscripts s and a for CH_3CO_2H denote syn- and anti-conformers, respectively. The units are kilocalories per mole.

Table 4. Hydration Energies ΔG_{hyd} of Organic Solutes from Experiment[55,56] and Calculated by the RISM-QM and the GAFF/RISM Methods with the Use of the KH Closure and the KH and the GF Expressions (in Parentheses) for Chemical Potential[a]

solute	ΔG(GAFF)	ΔG(B3LYP)	ΔG(CCSD(T))	ΔG(expt)
CH_3NH_2	−2.6 (−3.9)	5.0 (3.7)	6.7 (5.4)	−4.6 [55]
$CH_3NH_3^+$	−67.6 (−69.1)	−62.8 (−64.4)	−62.3 (−63.9)	−73 ± 5 [56]
syn−CH_3CO_2H	−4.2 (−5.8)	−5.2 (−7.5)	−2.2 (−4.3)	−6.7 [55]
CH_3COO^-	−82.1 (−86.5)	−91.9 (−99.1)	−81.2 (−88.4)	−82 ± 5 [56]

[a]The units are kilocalories per mole.

QM/MM MC simulations[44] and are slightly below the 2.3 value provided by QM/MM MD simulations.

Removal of proton, leading to acetate ion, enhances the solvation of the carboxylic group. This trend is captured well by both RISM calculations. The number of the hydrogen bonds made by the carbonyl oxygen is 3.0 with GAFF/RISM and 3.5 with QM/RISM. QM/RISM in this case agrees well with the 3.7 value predicted by QM/MM MD simulations. These values correlate well with NMR spectroscopy,[45] where the hydration number of 5.5−6.5 was found for the COO^- group, and prior simulations[40,42] yield an average number of 6−7 solvent molecules around the COO^- group with 3−3.5 water molecules per each carbonyl oxygen.

C. Energetics. The results of our QM/RISM calculations of excess chemical potential $\Delta\mu$ with the HNC and the KH closures and various expressions (HNC, KH, and GF) for chemical potential are presented in Table 3. For comparison, we also included similar data obtained by prior RISM-SCF[36,37] and XSOL[9] simulations. Both energy and free energy reorganization values at the B3lYP and the CCSD(T) levels of theory are also provided. Although the values of $\Delta\mu$ vary significantly, depending on the type of closure and the expression used, the difference between excess chemical potentials calculated by different routines but with use of the same expression does not exceed 0.1 kcal/mol for acetic acid and its conformer and less than 3 kcal/mol for acetate ion.

We have tested the KH as well as the HNC closures and found that the former is to be more superior for prediction of the hydration free energy, although this closure underestimates strongly peaks of RDFs.[46] The results for hydration free energy obtained with the use of KH and the GF expressions for chemical potential are compared with the experimental data in Table 4. In this table we include also the results obtained by the conventional 1D RISM with the use of GAFF parameters and the structure obtained from a database.[31] Numerous attempts have indicated that 1D RISM cannot predict accurately absolute values of hydration energy, especially in the case of molecular ions (see, for example, refs 47−51). In evaluations of hydration

free energy we take into account the electrostatic potential drop at the liquid/vacuum boundary which is not included in RISM calculations. Therefore following ref 52, we have to modify eq 6 and calculate the hydration energy as

$$\Delta G_{\mathrm{hyd}} = \Delta\mu + \Delta G_{\mathrm{r}} + ze\psi_{\mathrm{s}} \quad (9)$$

where ΔG_{r} is the reorganization free energy value obtained by our calculations, z is the charge of the system, and ψ_{s} is the surface potential at the water/vacuum boundary, which is equal to 0.527 V following ref 53. The KH expression for chemical potential provides somewhat more accurate estimates for hydration energy.

Conformational Equilibria. We have evaluated the conformational equilibria between syn- and anti-isomers of acetic acid. Using the NWChem package, we have obtained that in the gas phase the syn-conformer is more stable than the anti-conformer by 6.1 kcal/mol at DFT/B3LYP and 5.1 kcal/mol at CCSD(T) levels of theory. This agrees well with previous estimates of 6.0 [44] and 6.9 kcal/mol.[37] Upon hydration the internal energy difference between the two conformers increase further to 11.2 kcal/mol at DFT/B3LYP and 11.9 kcal/mol at CCSD(T) levels of theory (with the syn-conformer being more stable). This is balanced by the solvation effects, which, opposite to the internal energy trend, favor the anti-conformer by 10.8 kcal/mol (using HNC closure and energy expression). Similar solvation stabilization of 9.7 kcal/mol was observed in prior investigation.[37] Overall, the syn-conformer still remains to be a more stable species than the anti-conformer, and the overall free energy difference (including solute thermal contributions) between the two conformers at the CCSD(T) level of theory and HNC approximation is found to be 0.5 kcal/mol. This is in agreement with the previous QM/MM and RISM-SCF calculations yielding 1.1 ± 0.3 [54] and 1.7,[37] respectively.

V. CONCLUSIONS AND PERSPECTIVE

We have presented a hybrid approach based on a combination of integral equation theory of molecular liquids and QM/MM capabilities of the NWChem package. We have split the evaluations into consequent QM/MM and 1D RISM calculations. In general, the complete procedure requires 100−200 RISM iterations after QM calculations for the tested examples. The test of the chosen examples has indicated that the obtained results are not affected by the splitting, whereas the choice of a way for estimating partial solute charges affects more drastically the results of the calculations. Several options are now available for the 1D RISM calculations in the NWChem package. The GAFF parameters or the solute structure and partial charges optimized by QM/MM may be used as 1D RISM input. The HNC and the KH closures have been applied in NWChem with various expressions for the excess chemical potential. Although we have used MDIIS algorithm to provide stability of convergence, more sophisticated algorithms based on multiresolution analysis such as the wavelet[59,60] or the multigrid[61] methods can be applied to accelerate the convergence of solution.

The test of solvent structure around the organic solutes has indicated advantages and limitations of 1D RISM calculations. Although 1D RISM calculations provide more or less reasonable values of positions of RDF peaks, the RISM predications of their amplitude are quite poor. The obtained coordination numbers are also overestimated due to that fact. At the same time, 1D RISM provides reasonable estimates for the number of hydrogen bonds in all of the tested examples. The overestimation of RDFs peaks as well as the coordination numbers is not a new fact, it has been obtained before in numerous 1D RISM evaluations. In general, the quality of the prediction may be improved by a choice of an appropriate bridge like it is obtained in ref 57. Apart from it, we may realize an alternative strategy in NWChem, since we may provide short-time molecular dynamics and evaluate the first peak of RDFs. In this case, the 1D RISM integral equations can be used to extrapolate the information from the MD to a long-range structure. Such an approach has been applied in ref 58, where the MD data on RDFs are used to constrain bridges. Our calculations confirmed that, in general, the 1D RISM calculations are able to predict a trend in changes of solvation energy due to substitution of some chemical groups,[36,38] or due to conformational equilibria.[37] Nevertheless, our results show that the absolute values of the hydration free energies are far from the experimental data. This gap may be up to several kilocalories per mole for organic polar solutes. Moreover, the application of the appropriate bridge improved RDFs may not result in further improvements for computing free energies (see, for example, ref 57). A way to overcome this bottleneck is to use semiempirical corrections for the free energy calculations. The simplest ways have been proposed in ref 62, which accounts only for the excluded volume and hydrogen bonding corrections. The concept of using empirical corrections to improve the accuracy of hydration free energies calculated by 1D RISM has recently been extended in the structural descriptor correction (SDC) model.[63] The method is based on a combination of 1D RISM with a small number of chemical descriptors associated with the main features of the chemical structure of solutes. The SDC model substantially increases the accuracy of 1D RISM calculations, while correction parameters are transferable between different chemical classes, which allows one to cover a wide range of organic solutes. Therefore, there is a tremendous opportunity to further improve the accuracy of the present 1D RISM calculations implemented in NWChem and incorporate these new theoretical developments.

AUTHOR INFORMATION

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

G.N.C. acknowledges the financial support from the FP7-PEOPLE-IIF-2008 program (Grant No. 235064, Programme: People). M.V.F. and G.N.C. thank the Russian Foundation for Basic Research (Grant No. 12-03-97508-r-centre-a) for partial support of this work. M.V. acknowledges support from the U.S. Department of Energy's (DOE) Office of Basic Energy Sciences, Chemical Sciences program. Calculations were performed using the Molecular Science Computing Facility (MSCF) in the William R. Wiley Environmental Molecular Sciences Laboratory, a DOE national scientific user facility located at the Pacific Northwest National Laboratory (PNNL). We are thankful to Dr. Maxim Fedorov and professor Michail V. Basilevsky for useful discussions.

REFERENCES

(1) Cramer, C.; Truhlar, D. *Chem. Rev.* **1999**, *99*, 2161.
(2) Bashford, D.; Case, D. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129.

(3) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999.
(4) van Gunsteren, W.; Berendsten, H. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 992.
(5) Vaidehi, N.; Wesolowski, T.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 4264.
(6) Ten-no, S.; Hirata, F.; Kato, S. *Chem. Phys. Lett.* **1993**, *214*, 391.
(7) Ten-no, S.; Hirata, F.; Kato, S. *J. Chem. Phys.* **1994**, *100*, 7443.
(8) Chandler, D.; Andersen, H. C. *J. Chem. Phys.* **1972**, *57*, 1930.
(9) Shao, L.; Yu, H.; Gao, J. L. *J. Phys. Chem. A* **1998**, *102*, 10366.
(10) Sato, H.; Hirata, F.; Kato, S. *J. Chem. Phys.* **1996**, *105*, 1546.
(11) Sato, H. In *Continuum solvation models in chemical physics: From theory to applications*; Mennucci, B., Cammi, R., Eds.; Wiley: Chichester, U.K., 2007.
(12) Yoshida, N.; Kato, S. *J. Chem. Phys.* **2000**, *113*, 4974.
(13) Gusarov, S.; Ziegler, T.; Kovalenko, A. *J. Phys. Chem. A* **2006**, *110*, 6083.
(14) Sato, H.; Kovalenko, A.; Hirata, F. *J. Chem. Phys.* **2000**, *112*, 9463.
(15) Kovalenko, A.; Hirata, F. *J. Chem. Phys.* **1999**, *110*, 10095.
(16) Yoshida, N.; Hirata, F. *J. Comput. Chem.* **2006**, *27*, 453.
(17) Yokogawa, D.; Yokogawa, Sato, H.; Sakaki, S. *J. Chem. Phys.* **2007**, *126*, 244504.
(18) Luchko, T.; Gusarov, S.; Roe, D. R.; Simmerling, C.; Case, D. A.; Tuszynski, J.; Kovalenko, A. *J. Chem. Theory Comput.* **2010**, *6*, 607.
(19) Genheden, S.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Ryde, U. *J. Phys. Chem. B* **2010**, *114*, 8505.
(20) Kloss, T.; Heil, J.; Kast, S. M. *J. Phys. Chem. B* **2008**, *112*, 4337.
(21) Yoshida, N.; Kiyota, Y.; Hirata, F. *J. Mol. Liq.* **2011**, *159*, 83.
(22) Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Van Dam, H. J. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L.; de Jong, W. A. *Comput. Phys. Commun.* **2010**, *181*, 1477.
(23) Valiev, M.; Garrett, B. C.; Tsai, M. K.; Kowalski, K.; Kathmann, S. M.; Schenter, G. K.; Dupuis, M. *J. Chem. Phys.* **2007**, *127*, 051102.
(24) Case, D. A.; Cheatham, T. E. III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M. Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. *J. Comput. Chem.* **2005**, *26*, 1668. (http://www.ambermd.org/)
(25) Valiev, M.; Yang, J.; Adams, J. A.; Taylor, S. S.; Weare, J. H. *J. Phys. Chem. B* **2007**, *111*, 13455.
(26) Valiev, M.; Bylaska, E. J.; Dupuis, M.; Tratnyek, P. G. *J. Phys. Chem. A* **2008**, *112*, 2713.
(27) Singer, S. J.; Chandler, D. *Mol. Phys.* **1985**, *55*, 621.
(28) Chandler, D.; Singh, Y.; Richardson, D. M. *J. Chem. Phys.* **1984**, *81*, 1975.
(29) Cox, S. R.; Williams, D. E. *J. Comput. Chem.* **1981**, *2*, 304.
(30) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129.
(31) Bolton, E. E.; Wang, Y.; Thiessen, P. A.; Bryant, S. H. PubChem: Integrated platform of small molecules and biological activities. *Ann. Rep. Comp. Chem.* **2008**, *4*, 217−241.
(32) Ishida, T.; Rossky, P. J.; Castner, E. W. Jr. *J. Phys. Chem. B* **2004**, *108*, 17583.
(33) Kovalenko, A.; Ten-no, S.; Hirata, F. *J. Comput. Chem.* **1999**, *20*, 928.
(34) Merrick, J. P.; Moran, D.; Radom, L. *J. Phys. Chem. A* **2007**, *111*, 11683.
(35) Bartlett, R. J.; Musial, M. *Rev. Mod. Phys.* **2007**, *79*, 291.
(36) Kawata, M.; Ten-no, S.; Kato, S.; Hirata, F. *J. Phys. Chem.* **1996**, *100*, 1111.
(37) Sato, H.; Hirata, F. *J. Mol. Struct.* **1999**, *113*, 461.
(38) M. Kawata, M.; Ten-no, S.; Kato, S.; Hirata, F. *Chem. Phys.* **1996**, *203*, 53.
(39) Jorgensen, W. L.; Gao, J. *J. Phys. Chem.* **1986**, *90*, 2174.
(40) Meng, E. C.; Kollman, P. A. *J. Phys. Chem.* **1996**, *100*, 11460.
(41) Meng, E. C.; Caldwell, J. W.; Kollman, P. A. *J. Phys. Chem.* **1996**, *100*, 2367.
(42) Alagona, G.; Ghio, C.; Kollman, P. *J. Am. Chem. Soc.* **1986**, *108*, 185.
(43) Rizzo, R. C.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1999**, *121*, 4827.
(44) Gao, J. L. *J. Phys. Chem.* **1992**, *96*, 6432.
(45) Kuntz, I. D. *J. Am. Chem. Soc.* **1971**, *93*, 514.
(46) Perkyns, J. S.; Lynch, G. C.; Howard, J. J.; Pettitt, B. M. *J. Chem. Phys.* **2010**, *132*, 064106.
(47) Lim, C.; Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 5610.
(48) Lee, P. H.; Maggiora, G. M. *J. Phys. Chem.* **1993**, *97*, 10175.
(49) Chuev, G. N.; Fedorov, M. V.; Chiodo, S.; Russo, N.; Sicilia, E. *J. Comput. Chem.* **2008**, *29*, 2406.
(50) Chuev, G. N.; Chiodo, S.; Erofeeva, S. E.; Fedorov, M. V.; Russo, N.; Sicilia, E. *Chem. Phys. Lett.* **2006**, *418*, 485.
(51) Fedotova, M. V.; Kruchinin, S. E. *J. Mol. Liq.* **2011**, *164*, 201.
(52) Horinek, D.; Mamatkulov, S. I.; Netz, R. R. *J. Chem. Phys.* **2009**, *130*, 124507.
(53) Warren, G. L.; Patel, S. *J. Chem. Phys.* **2007**, *127*, 064509.
(54) Gao, J. L.; Pavelites, J. J. *J. Am. Chem. Soc.* **1992**, *114*, 1912.
(55) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, l. *J. Solution Chem.* **1981**, *10*, 563.
(56) Florian, J.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 5583.
(57) Marucho, M.; Kelley, C. T.; Pettitt, B. M. *J. Chem. Theory Comput.* **2008**, *4*, 385.
(58) Du, Q.; D. Beglov, D.; Roux, B. *J. Phys. Chem. B* **2000**, *104*, 796.
(59) Chuev, G. N.; Fedorov, M. V. *J. Comput. Chem.* **2004**, *25*, 1369.
(60) Chuev, G. N.; Fedorov, M. V. *J. Chem. Phys.* **2004**, *120*, 1191.
(61) Sergiievskyi, V. P.; Hackbusch, W.; Fedorov, M. V. *J. Comput. Chem.* **2011**, *32*, 1982.
(62) Chuev, G. N.; Fedorov, M. V.; Crain, J. *Chem. Phys. Lett.* **2007**, *448*, 198.
(63) Ratkova, E. L.; Chuev, G. N.; Sergiievskyi, V. P.; Fedorov, M. V. *J. Phys. Chem. B* **2010**, *114*, 12068.
(64) Kendall, R. A.; Dunning, T. H. Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.