Analytical energy gradient formula was derived for reference interaction site model self-consistent field explicitly including spatial electron density distribution (RISM-SCF-SEDD). RISM-SCF-SEDD is a combination method of ab initio electronic structure theory and statistical mechanics for molecular liquids [D. Yokogawa et al., J. Chem. Phys. 126, 244504 (2007)]. As shown previously, RISM-SCF-SEDD is numerically stable and has expanded the applicability of the solvation theory. The energy gradient is an indispensable tool to compute molecular geometry and its implementation further extends the capability of RISM-SCF-SEDD. The present method was applied to chemical systems in aqueous solution; hydration structure and geometry of phosphate anion PO$_4^{3-}$ and tautomerization between 2-pyridone and 2-hydroxypyridine. Compared to available experimental data, the present method correctly reproduced the geometries and relative energies of solvated molecules with microscopic solvent distribution. It is clearly shown that highly sophisticated quantum chemical calculation such as coupled cluster with single and double and perturbative triple excitations coupled with solvation effect is a powerful tool to accurately evaluate molecular properties. © 2009 American Institute of Physics.

[doi:10.1063/1.3265856]
with density functional theory (DFT) and coupled cluster expansion using single and double excitations with perturbation treatment of triples [CCSD(T)].

This article is organized as follows. In Sec. II, the equations employed in this work were derived. The detail of the calculation was described in Sec. III. The present method was applied to two demonstrative examples; hydration structure and geometry of phosphate anion PO₄³⁻ in aqueous phase and the tautomerization between 2-pyridone (PY) and 2-hydroxypyridine (HP) in aqueous phase. The results were compared with available experimental data in Sec. IV.

II. THEORY

Before deriving the equation of energy gradient, we briefly summarize key equations employed in this work. Solvation structure around a solute molecule is calculated by RISM equation with hypernetted-chain (HNC) closure,

\[ h_{as}^{UV} = \sum_{\gamma} \omega_{\alpha\gamma}^{s} * c_{\gamma\gamma}^{UV} * \chi_{\alpha}^{UV}, \]

where \( \omega_{\alpha\gamma}^{s} \) is the number density of the solvent site \( \alpha \) and \( \gamma \) is the solvent site. \( h_{as}^{UV} \) is the total correlation function between solute and solvent molecules, and \( \{ \phi_{as}^{UV} \} \) is the solute-solvent interaction potential. \( \{ \omega_{\alpha\gamma}^{s} \} \) and \( \{ \phi_{as}^{UV} \} \) are intramolecular correlation functions, which define solute and solvent molecular geometries, respectively. \( \rho_{s}^{t} \) is the number density of the solvent site \( t \).

The solute-solvent interaction potential, \( \{ \phi_{as}^{UV} \} \), is the sum of Coulombic and Lennard-Jones (LJ) terms,

\[ \phi_{as}^{UV}(r) = \phi_{as}^{CL}(r) + 4 \epsilon_{as} \left( \frac{a_{as}}{r} \right)^{12} - \left( \frac{a_{as}}{r} \right)^{6}. \]

The Coulombic term \( \phi_{as}^{CL}(r) \) is usually written as classical expression, namely, \( q_{s}d_{f}(r) \). However, in RISM-SCF-SEDD, it is calculated directly from electron density \( \rho_{s}(r) \) of the solute molecule. \( \rho_{s}(r) \) is expanded with the spherical auxiliary basis sets (ABSs), \( \{ f_{i} \} \), on each solute atomic site,

\[ \rho_{s}(r) = \sum_{i} d_{f}(r), \]

with the coefficient \( d \) and ABS, \( \phi_{as}^{CL}(r) \) is thus expressed as follows:

\[ \phi_{as}^{CL}(r) = -q_{s} \sum_{i} d_{i} \int f_{i}(r') \frac{d_{f}(r')}{|r-r'|} dr' + \frac{q_{s} Z_{\alpha}}{r}, \]

where \( q_{s} \) is a partial charge on a solvent site \( s \), \( Z_{\alpha} \) is the nuclear charge of a solute atomic site \( \alpha \), and \( R_{\alpha} \) is the position vector of atomic site \( \alpha \). The expansion coefficient \( \{ d_{i} \} \) is evaluated by the following equation:

\[ d = X^{-1} \text{tr}(PY) - \frac{Z^{T} X^{-1} \text{tr}(PY) - N_{e} X^{-1} Z}{Z^{T} X^{-1} Z}, \]

where \( P \) is usual “P-matrix” related to the electron density and \( N_{e} \) is the number of electron in solute molecule. \( X, Y \), and \( Z \) are matrices which are defined with the given ABSs, atomic orbitals, and geometry of solute molecule, as follows:

\[ X_{ij} = \int \chi_{i}(r) \chi_{j}(r) dr_{1} dr_{2}, \]

\[ Y_{\mu
u,ij} = \int \chi_{\mu}(r) \chi_{\nu}(r) \chi_{i}(r) \chi_{j}(r) dr_{1} dr_{2}, \]

\[ Z_{i} = \int f_{i}(r) dr, \]

where \( \chi \) is the basis function employed in molecular orbital calculation.

In the RISM-HNC framework, excess chemical potential \( (\Delta \mu) \) given by

\[ \Delta \mu = -\frac{1}{\beta} \sum_{as} \rho_{s} \int dr \left[ c_{\alpha}^{UV}(r) - \frac{1}{2} (h_{as}^{UV}(r))^{2} + \frac{1}{2} h_{as}^{UV}(r) c_{\alpha}^{UV}(r) \right] + \frac{1}{2} \sum_{as} \rho_{s} \int dr \left[ c_{\alpha}^{UV}(r) c_{\alpha}^{UV}(r) - 1 - t_{as}^{UV}(r) \right] \]

\[ + \frac{1}{2 (2\pi)^{5}} \int dk \left[ \sum_{as} \rho_{s}^{UV}(k) c_{\alpha}^{UV}(k) - \frac{1}{2} \sum_{as} \rho_{s}^{UV}(k) c_{\alpha}^{UV}(k) \right]. \]

where \( t_{as}^{UV} = h_{as}^{UV} - c_{\alpha}^{UV} \) and hat (\( \hat{\cdot} \)) means the function in reciprocal space. RISM equation and HNC closure are derived from Eq. (11) by the well-known variational stationarity. In analogy with the RISM-HNC as well as with the original RISM-SCF method, a corresponding to the Helmholtz free energy of the system is defined as follows:

\[ A = E_{\text{solute}} + \Delta \mu, \quad \text{where } E_{\text{solute}} = \langle \Psi | H | \Psi \rangle. \]

\( H \) is the Hamiltonian for a solute molecule in the standard quantum chemical method and \( \Psi \) is the wave function of the solute molecule in solvent. Thanks to the variational stationarity of RISM-HNC theory, the first derivative of \( A \) with respect to the nuclear coordinate of the solute molecule \( R_{s} \) is simply given by (see Appendix for the detail)
\[
\frac{\partial A}{\partial R_a} = \frac{\partial E_{\text{solvent}}}{\partial R_a} + \frac{\partial \Delta \mu}{\partial R_a}
- \frac{\partial E_{\text{solvent}}}{\partial R_a} - V_i \frac{\partial d}{\partial R_a}
= \frac{1}{2(2\pi)^3} \sum_{\alpha \beta} \frac{\partial \epsilon_{\alpha \beta}^{UV}(k)}{\partial \epsilon_{\alpha \beta}^{UV}(k)}
+ \frac{1}{2(2\pi)^3} \sum_{\alpha \beta} \frac{\partial \epsilon_{\alpha \beta}^{VV}(k)}{\partial \epsilon_{\alpha \beta}^{VV}(k)}.
\]

The first and third terms in the last equation are the same as those obtained by previous work. Column vector \( V \) in the second term represents the electrostatic potential on ABSs induced by averaged solvent distribution, whose \( i \)th element \( (V_i) \) is given by

\[
V_i = \sum_{s} \sum_{\alpha} q_s \int \frac{F_i(r')}{|r - r'|} \chi_{\alpha s}(|r - r'|) dr dr'.
\]

The derivative of \( d \) with respect to \( R_a \) in the second term is

\[
\frac{\partial d}{\partial R_a} = - X^{-1} \frac{\partial X}{\partial R_a} - \left[X^{-1} \frac{\partial X}{\partial R_a} \right] \left[X^{-1} \frac{\partial X}{\partial R_a} \right]
+ X^{-1} \frac{\partial X}{\partial R_a} \frac{\partial Y}{\partial R_a},
\]

where \( S \) is the usual overlap matrix. In our program code, the matrices, \( X, Y, \) and \( Z \), and the first derivative of \( X \) and \( Y \) are calculated by the Obara–Saika recursion relation.

III. COMPUTATIONAL DETAILS

All calculations were performed using the GAMESS program package, where RISM-SCF-SEDD and the energy gradient technique were implemented by us. Spherical convolution was computed with Ohura’s fast Fourier transform program. For the acceleration of the RISM-SCF convergence, M-DIIS method was employed. Dielectric correction for RISM was not used and the standard geometry optimization algorithm implemented in GAMESS was employed.

The present calculation was performed with DFT using the B3LYP functional (VWN 5) and CCSD(T). Hydration effect was evaluated by RISM-SCF-SEDD and for phosphate molecule, C-PCM (Ref. 19) calculation was also performed for comparison. The basis set for PO₄³⁻ was cc-pV(T+d)Z for P atom and cc-pVTZ for O atom. For all atoms in PY and HP, aug-cc-pVDZ basis set was used. ABS was constructed from the s-type primitive atomic orbital basis of these basis sets [please see Ref. 8 in our previous work (Ref. 11) for the details]. The diffuse functions were not included in the ABS construction for PY/HP.

The LJ parameters for each atom are summarized in Table I and potentials between atoms were evaluated with the Lorentz–Berthelot mixing rule \( \sigma_{ij} = (\sigma_i + \sigma_j)/2 \) and \( \epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j} \). For solvent water molecule, the SPC model was employed with the LJ parameters of the hydrogen sites \( (\sigma=1.0 \text{ Å} \) and \( \epsilon=0.056 \text{ kcal/mol} \). All calculations were performed at the temperature of 298.15 K and at the number of density of 0.033 426 molecule/Å³ (=1 g cm⁻³).

IV. RESULTS AND DISCUSSION

In this work, two chemical systems were studied; hydration structure and geometry of phosphate anion PO₄³⁻ in aqueous phase and tautomerization between PY and HP in aqueous phase. The former system was recently studied with neutron scattering experiments, and hydration structure as well as intramolecular bond lengths were precisely reported. The latter system has been widely studied with theoretical and experimental methods.

A. Phosphate anion PO₄³⁻

The geometry optimizations of PO₄³⁻ were performed with DFT (B3LYP) calculation under the \( T_d \) symmetry. In aqueous solution, both of RISM-SCF-SEDD and C-PCM were employed. The geometrical parameters optimized with gas phase (geom I), RISM-SCF-SEDD (geom II), and with C-PCM (geom III) are shown in Table II together with experimental data. In gas phase, P–O and O–O distances were 1.582 and 2.583 Å, respectively. These distances became significantly shorter in aqueous solution; the optimized distances of P–O and O–O with RISM-SCF-SEDD were 1.539 and 2.514 Å, respectively, which were very close to the experimental data (1.53 Å for P–O and 2.49 Å for O–O). Computation with CCSD(T) method gives essentially the same results. Although the optimized distances by C-PCM also became shorter than those of geom I, they were somewhat longer than experimental data.

The computed radial distribution functions (RDFs) \( [g(r)] \) of water oxygen (\( O_w \)) and water hydrogen (\( H_w \)) around solute P and O atoms are shown in Fig. 1, together with the each coordination numbers \( [N_i(r)] \).

### Table II. Geometrical parameters of PO₄³⁻ (Å) in gas phase and in aqueous solution.

<table>
<thead>
<tr>
<th></th>
<th>Gas phase</th>
<th>Aqueous solution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>geom I</td>
<td>RISM-SCF-SEDD</td>
</tr>
<tr>
<td>P–O</td>
<td>1.582</td>
<td>1.539</td>
</tr>
<tr>
<td>O–O</td>
<td>2.583</td>
<td>2.514</td>
</tr>
</tbody>
</table>

1Reference 21.
show a good agreement, especially the P–Ow distance is because Ow does not interact directly with PO4 scattering experiments. but reasonably accord with those obtained by the neutron numbers corresponding to the first shell are somewhat larger, almost identical to the experiment data. The coordinationates around Hw. The peak position of RDF and coordination

\[
g(r) = 4\pi\rho(r)\int_0^r x^2 g_X(x)dx \quad (X = P \text{ or } O),
\]

where \(\rho(r)\) is the number density of water. A high and narrow peak was seen at 1.69 Å in \(g_{OHw}(r)\), indicating a strong hydrogen bonding between O sites of PO4\(^{3-}\) and Hw. The P–Ow peak at 3.67 Å was broad compared to O–Hw peak. This is because Ow does not interact directly with PO4\(^{3-}\) and fluctuates around Hw. The peak position of RDF and coordination number obtained by RISM-SCF-SEDD is compared to the experimental data\(^{21}\) in Table III. The computed distances show a good agreement, especially the P–Ow distance is almost identical to the experiment data. The coordination numbers corresponding to the first shell are somewhat larger, but reasonably accord with those obtained by the neutron scattering experiments.

**B. PY/HP**

In this molecule, tautomeric isomerization occurs between amide form (PY) and imidic acid forms (HP). Three conformers, PY, HP\(_{anti}\), and HP\(_{syn}\) are considered in terms of the position of the transferring H atom. Geometries of these conformers were optimized under \(C_s\) symmetry and shown in Fig. 2. The obtained geometry reasonably agrees with the experimental data determined by microwave spectroscopy.\(^{23}\) Root mean square deviation in the bond length between the DFT (B3LYP) calculations and experimentally obtained data in gas phase was 0.015 Å for PY and 0.010 Å for HP\(_{syn}\). The deviation is satisfactorily small, suggesting that calculation is reliable for the geometry optimization. On the other hand, much highly sophisticated quantum chemical method is necessary to adequately evaluate the relative energy. The energy obtained by DFT (B3LYP) shows that PY is slightly stable, contradicting the experimental data. By performing CCSD(T) calculation the relative stability between PY and HP was properly reproduced (Table IV).

The geometries in aqueous phase calculated by RISM-SCF-SEDD are also shown in Fig. 2. The solvation does not affect the geometry of HP\(_{syn}\) and HP\(_{anti}\) so much. The maximum change in distances from gas to aqueous phase was only 0.004 Å for HP\(_{syn}\) and 0.007 Å for HP\(_{anti}\). On the other hand, the geometry of PY in aqueous phase is quite different from that in gas phase. In particular, the bond lengths of C2–N1 and C3–C2 decrease by 0.029 and 0.022 Å, respectively, and that of O7–C2 increases by 0.033 Å. This large change may be understood with resonance structures in PY. As shown in Scheme 1, PY has zwitterionic resonance structures in addition to amide form. In gas phase, amide form is stable compared to zwitterionic form, while the zwitterionic resonance structures are much stabilized in aqueous phase by solvent water. In aqueous phase, the double bond character becomes greater in C2–N1 and C3–C2 bonds while single bond character becomes greater in O7–C2 bond.

The relative free energy in aqueous phase calculated by the present method is shown in Table IV, where those computed by the original RISM-SCF (HF level)\(^{24}\) and by QM/MM (Ref. 25) are also shown. In the case of CCSD(T) calculation in aqueous phase, the energies were also evaluated with a different type of LJ parameter, general amber force field (GAFF).\(^{26}\) A qualitative trend, namely, PY became more stable compared to HP in aqueous solution, is reproduced by all methods, but there are quantitative differences among the methods. DFT level calculation with RISM-SCF-SEDD and the original RISM-SCF overestimate the stability

<table>
<thead>
<tr>
<th>TABLE III. The RDF peak position and the coordination number.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>This work</strong></td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>P–Ow distance (Å)</td>
</tr>
<tr>
<td>O–Hw distance (Å)</td>
</tr>
<tr>
<td>(N) (first shell of P–Ow)</td>
</tr>
<tr>
<td>(N) (first shell of O–Hw)</td>
</tr>
</tbody>
</table>

\(^a\)Reference 21.
of PY. The agreement is certainly improved with the present RISM-SCF-SEDD coupled with CCSD(T) calculation. In the calculation, the sensitivity of LJ parameters was significantly small.

Figure 3 displays the radial distribution functions of hydrogen (Hw) of solvent water around N and O atoms in the three isomers obtained by RISM-SCF-SEDD method. In HPsyn and HPanti, the peak around 2 Å corresponds to the hydrogen bonding. Hw coordinates to both N and O atoms, which are exposed to the solvent. In the case of PY, N atom is already occupied by the transferring H atom and Hw cannot interact, leading to the disappearance of the peak around 2 Å seen in HP. At the same time, O atom in PY shows alkoxide character and its charge becomes more negative compared to that in HP. Consequently, O strongly attracts Hw, which makes RDF of O–Hw higher. This is an important factor that PY is strongly stabilized in aqueous phase.

V. CONCLUSIONS

We derived an analytical energy gradient formula based on RISM-SCF-SEDD. The present method was applied to two demonstrative systems in aqueous phase, phosphate anion PO4−, and tautomerization between PY and HP. The computational procedure of RISM-SCF-SEDD is numerically stable compared to the original RISM-SCF. The new method has expanded the applicability of ab initio electronic structure calculation coupled with solvation effect and the development of energy gradient technique reported here further extends the capability of the method. It should be emphasized that both of the original RISM-SCF and RISM-SCF-SEDD methods are equivalent to ab initio QM/MM method, in which the electronic structure of the solute molecule is determined under the influence of solvent molecules. Thanks to the analytical treatment based on statistical mechanics, the computational time to evaluate the statistical ensemble is negligibly short. This feature enables us to combine highly sophisticated electronic structure theory, e.g., CCSD(T), DFT. We believe that the present method will become the powerful tools to analyzing interesting chemical systems, such as chemical reactions in solution, with reasonable computational cost.

ACKNOWLEDGMENTS

This work was financially supported in part by the Grant-in Aid for Scientific Research on Priority Areas “Water and Biomolecules” (Grant No. 430-18031019), “Molecular Theory for Real Systems” (Grant No. 461), “Ionic Liquids” (Grant No. 452-20031014), “Molecular Science of Fluctuations” (Grant No. 21107511), and by the Grant-in-Aid for General Research (Grant No. 19350010). D.Y. thanks the Grant-in-Aid for JSPS Fellows. All of them were supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) Japan.

APPENDIX: DERIVATION OF EQUATION (13)

After some mathematical manipulation, the derivative of excess chemical potential Δμ with respect to nuclear coordinate Rα is obtained as follows:

\[
\Delta \mu \left( R_\alpha \right) = \sum_{\beta} \frac{\partial \mu_{\text{excess}}}{\partial R_\beta} \frac{\partial R_\beta}{\partial R_\alpha}
\]

This is a general expression for the derivative of the excess chemical potential with respect to a nuclear coordinate. It involves the partial derivatives of the electronic and solvation contributions to the chemical potential.

Table IV. Computed and experimentally obtained relative energies (kcal/mol). The energies in aqueous phase calculated by CCSD(T) with GAFF parameter are shown in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>Gas phase</th>
<th>Aqueous solution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PY</td>
<td>HPanti</td>
</tr>
<tr>
<td>DFT</td>
<td>0.0</td>
<td>5.3</td>
</tr>
<tr>
<td>CCSD(T)</td>
<td>0.0</td>
<td>3.7</td>
</tr>
<tr>
<td>HF</td>
<td>0.0</td>
<td>⋯</td>
</tr>
<tr>
<td>QM/MM (AM1)</td>
<td>0.0</td>
<td>⋯</td>
</tr>
<tr>
<td>Expt</td>
<td>0.0</td>
<td>⋯</td>
</tr>
</tbody>
</table>

Notes:
1. This work.
3. Reference 25.
4. References 23, 31, and 32.
\[
\frac{\partial \Delta \mu}{\partial R_\alpha} = - \sum_{\alpha \neq i} \rho_\alpha^V \int dr \left[ (e^{-\beta \phi_{\alpha\ar}^U(r)} + h_{\alpha\ar}^U(r) - 1 - h_{\alpha\ar}^U(r)) \right]
\]

\[
\times \left( \frac{\partial^2 \phi_{\alpha\ar}^U(r)}{\partial r_\alpha} - (h_{\alpha\ar}^U(r) + h_{\ar\alpha}^U(r) - c_{\alpha\ar}^U(r)) \frac{\partial h_{\alpha\ar}^U(r)}{\partial R_\alpha} \right) + \left( - h_{\alpha\ar}^U(r) + \sum_{\beta \neq \alpha} \omega_{\alpha}^U \frac{c_{\beta\alpha}^U(r)}{\partial r_\alpha} \frac{\partial c_{\beta\alpha}^U(r)}{\partial R_\alpha} \right) + \left( - \beta e^{-\beta \phi_{\alpha\ar}^U(r)} + h_{\alpha\ar}^U(r) - c_{\alpha\ar}^U(r) \right) \frac{\partial \phi_{\alpha\ar}^U(r)}{\partial R_\alpha} \right]
\]

(A1)

The terms with the derivatives of \(t_{\alpha\ar}^U\), \(h_{\alpha\ar}^U\), and \(c_{\alpha\ar}^U\) are vanished using RISM and HNC equations. Because the distance \(r\) between solute and solvent atoms in the solute-solvent interaction potential \(\{\phi_{\alpha\ar}^U\}\) does not depend on the nuclear coordinates \(R_\alpha\) of solute atoms, the derivative of \(\phi_{\alpha\ar}^U\) in Eq. (A1) is given by

\[
\frac{\partial \phi_{\alpha\ar}^U}{\partial R_\alpha} = - d_i \sum_{\alpha \neq \alpha} \frac{\partial d_i}{\partial R_\alpha} \int f(r') \left( \frac{1}{|r - r'|} \right) dr'.
\]

(A2)

By substituting Eq. (A2) into Eq. (A1), Eq. (13) is obtained.

---

16. Useful libraries for FFT calculation can be download from the following web page: http://www.kurims.kyoto-u.ac.jp/~ourou/
18. CCSD(T) calculation in this work was performed using the subroutine implemented in GAMESS (Ref. 27).
22. To check the accuracy of the present calculation level, the optimized distances were also computed by RISM-SCF-SEDD with CCSD calculation. Because analytical gradient of CCSD(T) was not available in the current version of GAMESS package, the obtained geometry was determined numerically by evaluating free energy surface. The obtained distances were 1.536 Å for P-O and 2.507 Å for O-O, which are very close to those obtained by DFT (B3LYP) calculation.