JOURNAL OF CHEMICAL INFORMATION AND MODELING

Review

Cite This: J. Chem. Inf. Model. 2017, 57, 2646-2656

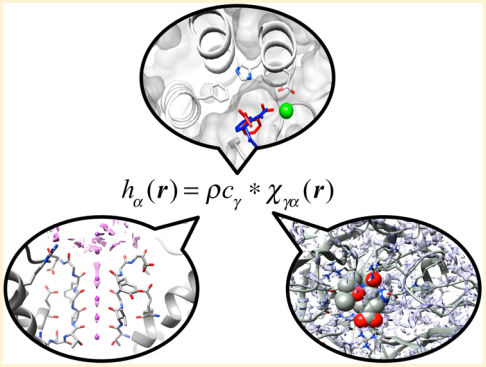
pubs.acs.org/jcim

Role of Solvation in Drug Design as Revealed by the Statistical Mechanics Integral Equation Theory of Liquids

Norio Yoshida*

Department of Chemistry, Graduate School of Science, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395 Japan

ABSTRACT: Recent developments and applications in theoretical methods focusing on drug design and particularly on the solvent effect in molecular recognition based on the three-dimensional reference interaction site model (3D-RISM) theory are reviewed. Molecular recognition, a fundamental molecular process in living systems, is known to be the functional mechanism of most drugs. Solvents play an essential role in molecular recognition processes as well as in ligand−protein interactions. The 3D-RISM theory is derived from the fundamental statistical mechanics theory, which reproduces all solvation thermodynamics naturally and has some advantages over conventional solvation methods, such as molecular simulation and the continuum model. Here, we review the basics of the 3D-RISM theory and methods of molecular recognition in its applications toward drug design.



KEYWORDS: *Drug design, Molecular recognition, Solvation, 3D-RISM*

INTRODUCTION

Molecular recognition is one of the most important fundamental processes in biological systems and is known to be the functional mechanism of most drugs.[1] This is defined as a molecular process in which a ligand molecule is bound at a particular position in host molecules, such as proteins, with high selectivity. Many drugs are selectively bound to target host proteins where they inhibit or activate their intrinsic function. For example, oseltamivir, a well-known antiviral drug, is selectively bound to the neuraminidase enzyme of influenza virus and blocks its function.[2,3] Amantadine is also an antiviral drug, which works as a channel blocker for the M2 proton channel of influenza A virus.[4,5] This drug is bound to the M2 proton channel and blocks proton conduction through the membrane. The molecular recognition of these drugs consequently causes the inhibition of viral replication.

Molecular recognition processes are governed by the difference in free energy between the ligand-binding and -unbinding states.[1] Here, let us consider a molecular recognition process characterized by the free energy change between the binding and unbinding states, ΔG

$$\Delta G = G_{\mathrm{complex}} - (G_{\mathrm{ligand}} + G_{\mathrm{protein}}) \quad (1)$$

which is referred to as binding free energy, where G_{complex}, G_{ligand}, and G_{protein} denote the free energy of the ligand−protein complex, ligand, and host protein in the unbinding states, respectively. A negative value of the binding free energy indicates that the molecular recognition process occurs spontaneously. The binding free energy originates from the changes in the conformation of the molecule, interaction between ligand and protein, and solvation structure; hence eq 1 can be rewritten as

$$\Delta G = \Delta G^{\mathrm{conf}} + \Delta G^{\mathrm{int}} + \Delta G^{\mathrm{solv}} \quad (2)$$

where ΔG^{conf}, ΔG^{int}, and ΔG^{solv} denote the changes in the conformational energy, the interaction energy between ligand and protein, and the solvation free energy, respectively. The solvation free energy change, the last term on the right-hand side of eq 2, is attributed to the solvation structure change due to the ligand binding by the host protein. When there is no ligand molecule at the binding pocket or recognition site of the host protein, these regions are usually occupied by water molecules, and the ligand molecules are simply solvated in solution. For a ligand molecule to bind to the recognition site, the solvation structure of both the ligand and host protein must change substantially. For example, the hydrogen bond between a ligand and a solvent water is cleaved, resulting in an alternative interaction between the ligand and the host protein. The breaking of the hydrogen bond causes reorganization of the hydrogen bond network in the solvent water. Such a structural change in solvation also causes the entropy of the solvent to change because the degrees of freedom allowed in the solvent water are changed substantially by the enlargement of the space for translational motion and changes in the restriction of rotational motion. Therefore, to understand the molecular recognition process, which is a thermodynamic process, an appropriate treatment of the solvation is highly anticipated.

Molecular simulation is the most popular way to handle the solvation.[6] In molecular simulation, all of the degrees of freedom of the solvent molecules, namely coordinates and orientations, are treated explicitly; therefore, they are known as

Received: June 27, 2017
Published: October 9, 2017

DOI: 10.1021/acs.jcim.7b00389
J. Chem. Inf. Model. 2017, 57, 2646−2656

the "explicit solvent model." To compute thermodynamic properties using molecular simulation, the configuration integrals for the coordinates and orientations of all solvent molecules need to be performed under a specific ensemble. The coordinates and orientations of solvent molecules are generated by time propagation of the Newtonian equation of motion in a molecular dynamic (MD) simulation or by random sampling in a Monte Carlo (MC) simulation. The molecular simulation provides a detailed molecular view and accurate thermodynamic quantities if a sufficiently good ensemble average is taken. However, the computational cost of taking "a sufficiently good ensemble average" to reproduce solvation free energy including entropy is very high. It is unpractical to apply these methods in virtual screening that targets large amounts of compounds.

A continuum solvation model, such as the generalized Born (GB)[7] and Poisson−Boltzmann (PB)[8] methods, is another candidate for determining the solvation thermodynamics of biomolecules.[9−11] In this model, the solvent environment is described as a spherical or molecular-shaped cavity surrounded by a continuum medium characterized by a dielectric constant. Due to this treatment of the solvent environment, the model only describes the electrostatic interaction between solute and solvent medium, but not nonelectrostatic, cavitation, and hydrophobic interactions. To incorporate the missing energy term for the hydrophobic contribution in solvation free energy models, PB and GB employ a simple model dependent on the solvent accessible surface area (SA) of the solute molecule, known as the PBSA or GBSA methods.[12−14] These methods are combined with molecular mechanics (MM/PBSA or MM/GBSA methods) and are widely used in docking simulations that are implemented in popular molecular simulation packages. The information on the solvation obtained by the models is rather limited compared to the explicit solvent model due to the simple expression of the solvent environment. However, the computational cost to obtain the solvation free energy is much lower than that of the explicit solvent model; therefore, these methods are a mainstay of the solvation model for virtual screening in modern drug design.

The three-dimensional reference interaction site model (3D-RISM) theory has been developed as an alternative approach for the solvation of biomolecules. It is based on the statistical mechanics integral equation theory of molecular liquids.[15−19] Although there are many theories to obtain the solvation structure of biomolecules based on the integral equation theory of liquids,[20−24] the 3D-RISM theory is most widely used. The 3D-RISM equation is derived from the molecular Ornstein−Zernike (MOZ) equation for a solute−solvent system by taking a statistical average over the orientation of the solvent species. In the 3D-RISM theory, the host protein and ligand molecules are regarded as "solutes," and water, ions, and all surrounding molecules are regarded as "solvents." By solving the 3D-RISM equation coupled with an appropriate closure relation under the solute−solvent interaction potential, one can obtain the three-dimensional distribution function (3D-DF) of solvent species around solute molecules. The 3D-DF represents the probability of finding the solvent species as a function of the Cartesian coordinate of the solvent. Such a function is especially suited to complex molecular systems. Furthermore, the 3D-DFs are obtained through a complete ensemble average over the entire configuration space of solvent molecules in the thermodynamic limits.[18,25] This means that the 3D-DFs obtained by 3D-RISM theory can reproduce the solvation structure in the binding pocket, channel, and cavity inside the protein where the dielectric constant of the solvent differs from the bulk.

Several review articles related to 3D-RISM theory have been published to date. Yoshida et al. reviewed some earlier works on the application of this approach to molecular recognition processes; these studies were discussed mainly on the basis of 3D-DF, which is a main output of the 3D-RISM theory.[19] Other review papers introduced multiscale methodological developments of the 3D-RISM theory combined with other theoretical methods, such as quantum mechanics methods and MD simulation methods, and their applications.[26−29] Further reviews concentrated on more specific applications, such as channel proteins,[30] ligand mapping,[31] and electrolyte solvation.[32] Ratkova, Palmer, and Fedorov reviewed general aspects of applying integral equation theories to model thermodynamics of solvation.[33]

In the present article, we complement the previous reviews discussed above by examining some recent papers from a different viewpoint; that is, we wish to bring together methods that can be used for drug design based on 3D-RISM and related theories. A brief overview of the 3D-RISM theory is described in the next section, followed by the introduction of two different types of analysis for investigating the molecular recognition processes, namely, the DF-based and binding free energy analyses. The development of the solvent and ligand placement algorithm based on the 3D-RISM theory is also introduced.

BASICS OF THE 3D-RISM THEORY

We briefly introduce the basics of the 3D-RISM theory. The details of the theory were presented in the literature.[15−18] The 3D-RISM theory is a DF theory describing the solvation structure of macromolecules, which is derived from the MOZ integral equation. To derive the 3D-RISM equation, let us start by defining the pair density DF of polyatomic molecules. The pair density DF of polyatomic molecules is a function of both the position $\mathbf{r}$ and the orientations $\mathbf{\Omega}$ of the molecules. The pair density DF is defined as

$$\rho(\mathbf{r}_1, \mathbf{r}_2, \mathbf{\Omega}_1, \mathbf{\Omega}_2) = \left\langle \sum_i \sum_{j\neq i} \delta(\mathbf{r}_1 - \mathbf{r}_i)\delta(\mathbf{r}_2 - \mathbf{r}_j) \right.$$
$$\left. \delta(\mathbf{\Omega}_1 - \mathbf{\Omega}_i)\delta(\mathbf{\Omega}_2 - \mathbf{\Omega}_j) \right\rangle \quad (3)$$

Under the assumption of the translational invariance on the pair density distribution, the pair DF is derived as

$$g(\mathbf{r}_{12}, \mathbf{\Omega}_1, \mathbf{\Omega}_2) = \left(\frac{\Omega}{\rho}\right)^2 \rho(\mathbf{r}_{12}, \mathbf{\Omega}_1, \mathbf{\Omega}_2) \quad (4)$$

where $\mathbf{r}_{12} = \mathbf{r}_2 - \mathbf{r}_1$ and $\Omega \equiv \int d\mathbf{\Omega}$. Using this function, the MOZ equation is given as

$$h(\mathbf{r}_{12}, \mathbf{\Omega}_2, \mathbf{\Omega}_2) = c(\mathbf{r}_{12}, \mathbf{\Omega}_1, \mathbf{\Omega}_2) + \frac{\rho}{\Omega} \int c(\mathbf{r}_{13}, \mathbf{\Omega}_1, \mathbf{\Omega}_3) h(\mathbf{r}_{32}, \mathbf{\Omega}_3, \mathbf{\Omega}_2) d\mathbf{r}_3 d\mathbf{\Omega}_3 \quad (5)$$

where $h = g + 1$ is the total correlation function and c is the direct correlation function which is defined through the MOZ equation. Since the correlation functions in the MOZ formalism have positional and orientational variables, it is difficult to handle them even numerically. This problem can be solved using a method to reduce the functions by introducing

DOI: 10.1021/acs.jcim.7b00389
J. Chem. Inf. Model. 2017, 57, 2646−2656

the interaction-site model, which was originally proposed by Chandler and Andersen.[34] The 3D total correlation function is derived by the integration of the molecular pair correlation function over the orientation of solvent molecules:

$$h_\gamma(\mathbf{r}) = \frac{1}{\Omega}\int h(\mathbf{r}_{12}, \mathbf{\Omega}_1, \mathbf{\Omega}_2)\delta(\mathbf{r}_{12} + l_{2\gamma}(\mathbf{\Omega}_2) - \mathbf{r})d2 \quad (6)$$

where $l_{2\gamma}$ denotes the vector connecting the molecular center and the interaction site γ on molecule 2. In addition, they applied a superposition approximation to the direct correlation function:

$$c(\mathbf{r}_{12}, \mathbf{\Omega}_1, \mathbf{\Omega}_2) = \sum_\gamma c_\gamma(\mathbf{r}) \quad (7)$$

By applying eqs 6 and 7, the MOZ equation is reduced to the molecular-site OZ or 3D-RISM equations:

$$h_\gamma(\mathbf{r}) = \sum_{\gamma'} [c_{\gamma'} * (\omega_{\gamma'\gamma} + \rho_{\gamma'} h_{\gamma'\gamma})] \quad (8)$$

where * denotes the convolution integral and $\omega_{\gamma\gamma'}$ is the intramolecular correlation function:

$$\omega_{\gamma\gamma'}(r) = \delta_{\gamma\gamma'}\delta(r) + \frac{1}{4\pi l_{\gamma\gamma'}^2}(1 - \delta_{\gamma\gamma'})\delta(r - l_{\gamma\gamma'}) \quad (9)$$

with $l_{\gamma\gamma'}$ the distance between the interaction sites γ and γ'. $h_{\gamma\gamma'}$ is a total correlation function of neat solvent, which is obtained beforehand by solving the RISM calculation for the solvent system. Because the intramolecular correlation function is determined from the molecular structure, the 3D-RISM equation has two unknown functions, h and c. To determine h and c uniquely, we need one more relation between them, which is called the closure equation. The closure equation must be a function of variable r to be used in the RISM: hence, it is given as an equation for the site−site correlation function. For example, the site−site form of the hyper-netted chain (HNC) closure is approximately given by

$$g_\gamma(\mathbf{r}) = \exp\left[-\frac{1}{k_B T}\mu_\gamma(\mathbf{r}) + h_\gamma(\mathbf{r}) - c_\gamma(\mathbf{r})\right] \quad (10)$$

where k_B, T, and u are the Boltzmann constant, thermodynamic temperature, and a pair interaction potential, respectively. A more useful closure relation was proposed by Kovalenko and Hirata, and is known as the Kovalenko−Hirata (KH) closure:[35,36]

$$g_\gamma(\mathbf{r}) = \begin{cases} \exp[d_\gamma(\mathbf{r})] & d_\gamma(\mathbf{r}) < 0 \\ 1 + d_\gamma(\mathbf{r}) & d_\gamma(\mathbf{r}) \geq 0 \end{cases}$$

$$d_\gamma(\mathbf{r}) = -\frac{1}{k_B T}\mu_\gamma(\mathbf{r}) + h_\gamma(\mathbf{r}) - c_\gamma(\mathbf{r}) \quad (11)$$

The KH closure has improved the convergence in iterative solutions of the 3D-RISM equation substantially and has also yielded reasonable physical properties.

By solving 3D-RISM-HNC or 3D-RISM-KH equations, one can obtain the 3D-DFs and the correlation functions of solvent around the solute macromolecules. Physical quantities related to solvation are also obtained by correlation functions, such as solute−solvent interaction energy, partial molar volume, and solvation free energy. The solvation free energy is given by

$$\Delta\mu = \sum_\gamma \rho_\gamma \int_0^1 d\lambda \int d\mathbf{r}\, u_\gamma(\mathbf{r})\, g_\gamma^\lambda(\mathbf{r}) \quad (12)$$

where λ is a solute−solvent coupling parameter and g^λ is a DF under the solute−solvent coupling state λ. The coupling parameter takes 0 when there is no interaction between the solute and solvent and takes 1 when the solute molecules immersed in the solvent and system are in a state of equilibrium. Unlike molecular simulation, the integration on λ can be performed analytically when the HNC or KH closure is adopted. The resulting analytical expression of $\Delta\mu$ is given as

$$\Delta\mu = \frac{\rho}{\beta}\sum_\gamma \int\left\{-c_\gamma(\mathbf{r}) + \frac{1}{2}h_\gamma(\mathbf{r})^2 - \frac{1}{2}h_\gamma(\mathbf{r})c_\gamma(\mathbf{r})\right\}d\mathbf{r} \quad (13)$$

for the HNC closure and

$$\Delta\mu = \frac{\rho}{\beta}\sum_\gamma \int\left\{-c_\gamma(\mathbf{r}) + \frac{1}{2}h_\gamma(\mathbf{r})^2\Theta(-h_\gamma(\mathbf{r})) - \frac{1}{2}h_\gamma(\mathbf{r})c_\gamma(\mathbf{r})\right\}d\mathbf{r} \quad (14)$$

for the KH closure, where Θ denotes the Heaviside step function.[35−37] Using these expressions, one can obtain the solvation free energy with a single 3D-RISM calculation, which is one of the great advantages over the molecular simulation methods.

DISTRIBUTION FUNCTION-BASED ANALYSIS

In this section, we review the molecular recognition study based on the 3D-DFs evaluated by the 3D-RISM theory. The molecular recognition process is defined as a molecular process in which a ligand molecule is bound to the host molecules at a particular position and orientation with high selectivity. This means that the ligand has a high affinity to the binding pocket of host molecules and the probability of finding the ligand molecule at the binding pocket should be high. In other words, the DF of the ligand molecule shows a conspicuous peak at the binding pocket. Therefore, the position of the binding site, ligand-binding mode, and affinity of the binding from the DF can be determined simultaneously. In such applications, the ligand molecule is treated as one of the "solvent" components in the 3D-RISM framework. Because the 3D-DFs obtained by 3D-RISM theory are obtained through a complete ensemble average over the entire configuration space of solvent molecules including ligands in the thermodynamic limits, the 3D-DFs automatically include the configuration of solvents in the binding pocket, channel, and cavity inside the protein.

Pioneering work leading to a series of studies for molecular recognition based on the 3D-DFs obtained by 3D-RISM theory was conducted by Imai et al.[38] They found that the water distribution inside the cavity of hen egg-white lysozyme evaluated by the theory completely reproduced the experimental observation. Following their work, Yoshida et al. applied the theory to selective ion binding by human lysozyme.[39,40] In this study, the 3D-DFs of ions were evaluated by the theory for wild-type human lysozyme and its mutants with different ion selectivity. The experimental and theoretical (3D-RISM) positions of the Ca^{2+} ion at the ion binding site of mutated human lysozyme are compared in Figure 1. The position of the highest Ca^{2+} peak of 3D-DF almost completely coincides with that of the experiments. The distance between

DOI: 10.1021/acs.jcim.7b00389
J. Chem. Inf. Model. 2017, 57, 2646−2656

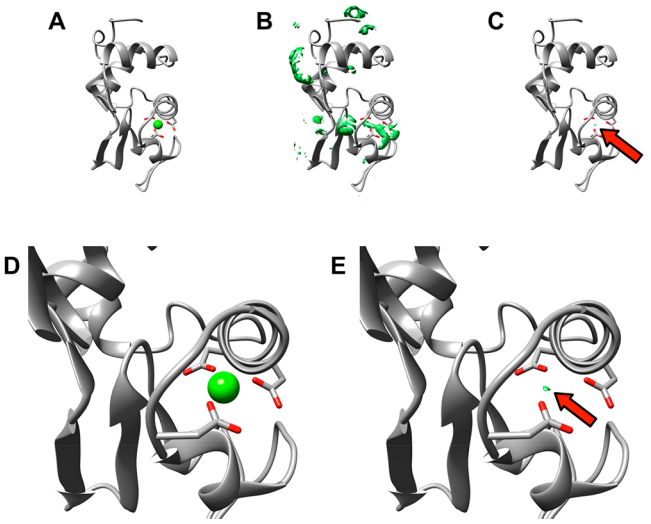


Figure 1. Comparison of the Ca^{2+} binding position of mutated human lysozyme. (A) X-ray crystal structure of mutated human lysozyme with the Ca^{2+} ion colored green. (B, C) Isosurface plots of the 3D-DFs of the Ca^{2+} ion evaluated by 3D-RISM-KH theory with $g > 10$ and $g > 50$, respectively. (D) Expansion of panel A around the Ca^{2+} binding site. (E) Expansion of panel C around the Ca^{2+} binding site. Red arrows in panels C and E indicate the positions of the highest peak in the 3D-DFs for the Ca^{2+} ion.

the 3D-RISM and experimental Ca^{2+} positions is only 0.6 Å, which is much smaller than the resolution of the X-ray crystallographic spectra. The peak positions based on 3D-DF analysis of water, ions, and xenon were compared with the experimental observations for several systems, such as the KcsA potassium channel,[41] Na-dependent F-type ATP synthase,[42] and myoglobin,[43] and all show good agreement with the experimental data. Comparisons with the MD simulation are also reported for hen−egg lysozyme[38] and GroEL/ES,[44] and the papers concluded that the 3D-RISM approach can yield 3D-DF values that are very close to those obtained through MD simulations.

These results imply that channel proteins are suitable targets to demonstrate the power of 3D-RISM theory because the molecular transportation through the channel is usually slow and a rare event on the molecular simulation time-scale. Most channel proteins play a variety of vital roles in living systems by transporting water, ions, and ligands, selectively.[45] Therefore, clarifications of the molecular mechanism are essential for diagnosing related diseases or designing drugs.

As an example, we introduce our 3D-RISM study on the influenza A M2 channel.[46] The M2 channel found on the viral membrane of the influenza A virus is a target of anti-influenza drugs because of its important role in proton transport and viral replication. For example, an amantadine drug family is known as a channel blocker to inhibit proton transportation by the M2 channel.[4,5] The M2 channel is highly selective for protons through gating controlled by pH. To investigate the mechanism of pH-dependent proton transport in the M2 channel, Phongphanphanee et al. considered the distributions of water and hydronium ions inside the channel by the concerted use of 3D-RISM theory and MD simulation.[46] It was suggested that the pH-controlled gating is accomplished by a change in the protonation state of the four histidines (His37) in the channel. The experiments also suggested that tryptophan 41 (Trp41) acts as a gate to move in conjunction with the protonation state of the histidines. The protonated states of histidines are regulated by pH from the nonprotonated state (0H) to the quadruple protonated state (4H) in the decreasing order of pH. They performed to obtain the M2 channel structure depending on the protonated state of histidines. Based on those structures, the 3D-RISM calculations were conducted and the 3D-DFs of water and hydronium ion, which is a proton analogue, were obtained.

The pH-dependent structure of the Trp41 gate, 3D-DFs of water, and hydronium ions are shown in Figure 2. The 3D-DFs

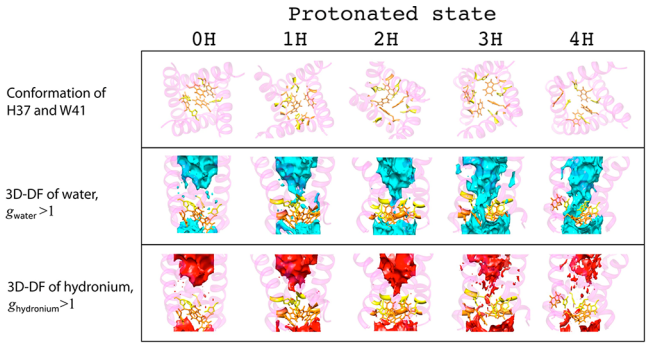


Figure 2. Structure of the Trp41 gate with histidines in a different protonated state (upper panel) and 3D-DFs of water (cyan in middle panel) and hydronium ions (red in lower panel) with $g > 1$.

of water clearly show that the accessibility of water through the channel increases with the protonated state of the channel in the order 0H < 1H < 2H < 3H < 4H. This result can be explained readily in terms of the pore diameter, which is widened by the electrostatic repulsion within protonated His37. Then the Trp41 turns to open the gate. The results suggest that the 0H, 1H, and 2H states are considered to be closed forms, whereas the 3H and 4H states are open forms.

The potential mean forces (PMFs) through the channel are shown in Figure 3. The PMFs can readily be derived from 3D-DFs. The PMFs of water show high barriers only in the 0H, 1H, and 2H states. However, the PMF of water in 3H and 4H is negative along the entire pore channel, which indicates that water is permeating through the channel.

In the closed states, 0H, 1H, and 2H, hydronium ions exhibit a behavior similar to water. This indicates that a hydronium ion or a proton is also prevented from being transported across the channel. The PMF of the hydronium ions in 3H and 4H exhibit small barrier. The barrier heights in the 3H and 4H forms are just 2−3 and 5−7 kJ/mol, respectively, which are comparative to the thermal energy. The barrier height for protons is higher in 4H than in 3H, which contrasts with the heuristic argument based on the pore size around the gating region. These observations can be explained by two competing factors. One of these competing factors is the electrostatic repulsion between the protonated His37 residues, which makes the pore channel larger and enhances the distribution of protons and water. The other factor is the electrostatic repulsion between protons and the protonated His37 residues, which will increase the PMF barrier. The two effects are balanced at the 3H state to maximize the proton distribution in the channel. These results are consistent with those by Voth and colleagues.[47]

SOLVENT/LIGAND PLACEMENT TECHNIQUE

A 3D-DF provides the probability of finding molecules at a position obtained through an ensemble average over the entire

DOI: 10.1021/acs.jcim.7b00389
J. Chem. Inf. Model. 2017, 57, 2646−2656

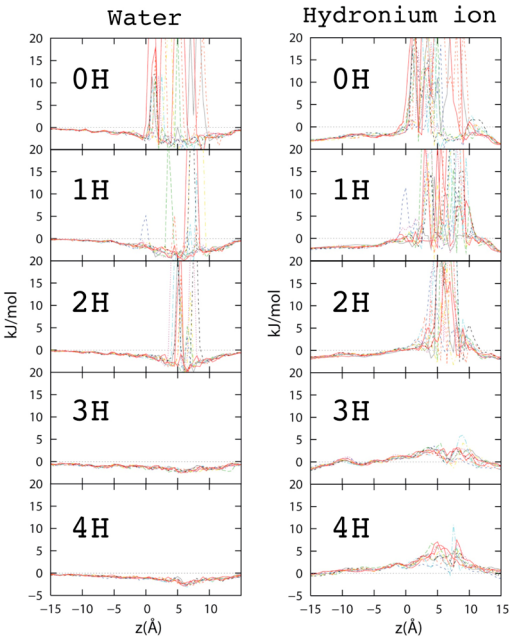


Figure 3. Potential mean force (PMF) of water (left) and hydronium ions (right) in different states of protonation. Each line represents the PMF calculated with a different M2 channel conformation in the protonation state.

configuration space of solvent molecules. Although 3D-DF itself is a useful physical quantity, the most probable position of the solvent or ligand molecules may also be required, such as complementation experimental data and preparation of the initial configuration for a molecular simulation. Many computational methods that reproduce or predict the explicit solvent configuration or ligand binding modes have been proposed.[48−52] Here, we review several methods based on the 3D-RISM theory.[42,53−56]

The simplest way to specify the most probable position of a solvent site or atom is to find the highest peak of the 3D-DF.[54] With this view, Sindhikara et al. proposed a method to place the solvent at the highest peak position of 3D-DF, i.e., the so-called placevent algorithm.[42] The scheme to place the solvent is as follows: (1) the highest peak position of the 3D-DF is identified; (2) the explicit solvent atom is placed at the position; and (3) the 3D-DF is updated by evacuating one population unit of the solvent atom to conserve the total population of the solvent atoms. These steps are continued until there are a satisfactory number of explicit solvent atoms. In Figure 4, the coordinates of the explicit water oxygen and sodium ion in Na-dependent F-type ATP synthase (PDB: 2WGM[57]) evaluated by the placevent algorithm were compared with those of the experimental results. As shown in Figure 4, the placevent algorithm reproduces the coordination structure well. The average error compared to experimental data was 0.39 ± 0.11 Å for sodium ion and 0.65 ± 0.24 Å for water oxygen, these values are much lower than the experimental resolution, 2.35 Å.

The placevent algorithm places simple ions and solvent atoms based on a single 3D-DF. Because most drugs are polyatomic molecules, an extension of the method for polyatomic ligands is desired. Unlike with simple ions, identifying the binding mode or position and orientation of polyatomic ligand molecules has some difficulties. The information on the orientation and molecular structure of the

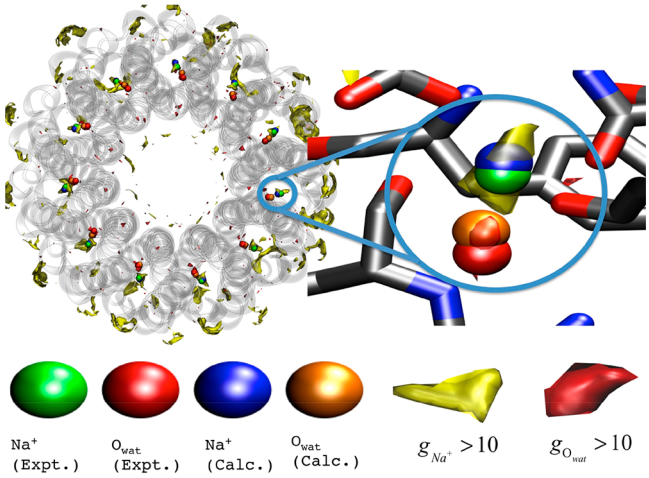


Figure 4. Comparison of the position of sodium ion and water oxygen between computational (calc) and experimental (expt) results. The green and red colored surfaces denote the iso-surface plot of the sodium ion and water oxygen atom, respectively.

ligand is averaged out by applying the interaction site model. Therefore, it is difficult to extract the most probable binding mode of a ligand from the 3D-DFs directory. Several methods have been proposed to realize the ligand placement based on 3D-DFs.[53,55,56,58] Pioneering work was conducted by Imai et al.[53] They applied the ligand placement algorithm to map small ligand molecules on protein surfaces to further develop fragment-based drug design. Kiyota et al. proposed an efficient scheme to obtain the 3D-DFs of larger ligand molecules and used the ligand placement algorithm to obtain the 3D-DFs of complex ligand molecules.[55] Following this study, Sindhikara and Hirata also proposed a similar method with improved numerical stability and realized a more detailed analysis of the binding affinity of the ligand.[56] Nikolić et al. expanded these methods to enable their use with flexible ligand molecules by including the AUTODOCK protocol.[58,59]

Although these methods employed different scoring functions and grid generation algorithms, they adopted a similar strategy for the placement of the ligand. Here, we introduce the basic method used for the ligand placement from 3D-DFs obtained by 3D-RISM theory; we describe the method reported by Kiyota et al. as an example.[55] The method defines two scoring function types: one is for the position of ligands and the other is for their orientation, in terms of 3D-DFs and a trial geometry of a ligand molecule. A scheme of the method is shown in Figure 5 where the binding mode of an aspirin molecule to phospholipase A2 (PDB: 1OXR[60]) was examined as an example. First, the 3D-DFs of the ligand, aspirin, around the host protein, phospholipase A2, were evaluated by 3D-RISM theory (see Figure 5a). Second, to specify the position of ligand binding, a function for the distribution center or net distribution of the ligand is introduced. The iso-surface plot of the net distribution of ligand aspirin is shown in Figure 5b. A conspicuous distribution can be found at the binding site suggested by experiments. Finally, one can investigate the most probable orientation of ligands at the position identified in the previous step by using the scoring function for the orientation. In Figure 5c, the result of the orientational search of aspirin at the binding site determined in the previous step is compared with the ligand coordinate taken from the PDB. It is worth noting that both the orientation and the position of the aspirin

DOI: 10.1021/acs.jcim.7b00389
J. Chem. Inf. Model. 2017, 57, 2646−2656

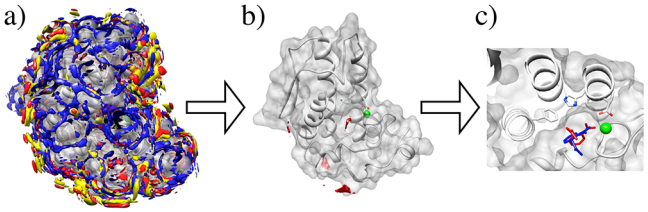


Figure 5. Scheme for placing a ligand molecule based on 3D-DFs. (a) Iso-surface plot of 3D-DFs of the aspirin molecule around phospholipase A2 with the threshold $g(\mathbf{r}) > 2$; (red) COOH; (gray) aromatic ring; (blue) $OCOCH_3$ group, respectively. (b) Results of the box search algorithm. Red surfaces denote the net distribution of aspirin molecules. (c) Predicted binding mode of aspirin (red-colored sticks) as a result of the orientational-search algorithm compared with those taken from the X-ray structure (blue-colored sticks).

ligand are sufficiently consistent. This application clearly demonstrates that the method based on the 3D-RISM theory is a prospective tool for searching the binding site and determining the binding mode.

BINDING FREE ENERGY ANALYSIS

In the field of structure-based drug design or drug molecular modeling, docking simulation is a mainstay method that predicts the binding mode and affinity of drug candidate ligands to the host protein. Usually, the analyses are performed when the location of the binding pocket is known. MM/GBSA or MM/PBSA are popular methods to evaluate the binding free energy of ligands; they are based on the MM of biomolecules combined with the GB or PB and SA methods.[4,5,7,8] These methods are widely used in docking simulations for virtual screening in modern drug design because of their low computational cost and reasonable accuracy. Because these methods employ the implicit solvent model to consider the solvent effects, the description of the solvation has some flaws; i.e., the interaction originates to molecular shape, such as hydrogen bonding, hydrophobic interaction, and solvent entropic term.[9,11] To circumvent these difficulties, a method using the 3D-RISM theory instead of GBSA or PBSA has been proposed, which is called the MM/3D-RISM method.[61−63] In addition, by employing the 3D-RISM theory, not only the solvation free energy but also detailed solvation structure can be obtained.

The accuracy of the MM/3D-RISM method has been assessed for the biotin−avidin system and for DNA.[61,64] Genheden et al. examined the binding free energy of seven biotin analogues to avidin protein and compared the results obtained by MM/3D-RISM to those calculated using MM/PBSA and MM/GBSA methods.[61] The mean absolute deviations (MADs) and the correlation coefficient of binding free energies from the experimental values are summarized in Table 1. The experimental values for the correlation coefficient of the binding free energy (0.80−0.93) are comparable to those obtained by MM/PBSA (0.91−0.93) and MM/GBSA (0.59−0.89). On the other hand, MADs of the calculated binding free energies from the experimental values are poor in all the methods, although the PB^{Delphi} shows best performance. The MAD of 3D-RISM-KH is relatively small. Yesudas et al. calculated the binding energy of short DNA (9−20-mers) from single-strand oligonucleotides by MM/3D-RISM and compared the results to those obtained using MM/PBSA and MM/GBSA.[64] In their study, both single-trajectory and three-trajectory methods were examined. Here, the single-trajectory method only uses the MD trajectory for protein−ligand complex systems, whereas the three-trajectory method uses three individual trajectories for the ligands, the protein, and their complex. The three-trajectory method gives better results than those obtained with the single-trajectory method. The correlation coefficient of the three-trajectory MM/3D-RISM (0.98) is similar to those of MM/PBSA (0.99) and MM/GBSA (0.99). The results noted above indicate that the MM/3D-RISM method is comparable to other methods and yields reasonable values for practical use. The MM/3D-RISM method and its extensions have been successfully applied to several systems.[65−67]

Table 1. Comparison of Computed Binding Free Energy with Experimental Values[a]

solvation method[b]	MAD	correlation coefficient
3D-RISM-KH-GF	70.7 ± 1.0	0.80 ± 0.02
3D-RISM-KH	36.8 ± 10	0.90 ± 0.02
PB^{AMBER}	48.2 ± 1.0	0.91 ± 0.01
PB^{Delphi}	16.3 ± 1.0	0.93 ± 0.02
GB^{HCT}	42.1 ± 0.8	0.59 ± 0.03
GB^{OBC1}	40.5 ± 0.9	0.69 ± 0.03
GB^{OBC2}	86.8 ± 1.0	0.85 ± 0.01
GBn	93.3 ± 1.1	0.89 ± 0.01

[a]Units of MAD are kilojoules per mole.[18] [b]3D-RISM-KH-GF denotes the values obtained using the Gaussian fluctuation solvation free energy formula with the correlation function calculated by 3D-RISM-KH theory. PB^{AMBER} and PB^{Delphi} are the PB methods implemented in Amber[73] and Delphi II,[74] respectively. GB^{HCT}, GB^{OBC1}, GB^{OBC2}, and GBn are the GB methods available in Amber.[73] See the work of Genheden et al.[61] for details of the methods.

Here, one of the applications of the MM/3D-RISM method is briefly introduced to demonstrate its use in drug design. Phanich et al. have applied the MM/3D-RISM method to elucidate the molecular mechanism for the oseltamivir resistance of mutated influenza B neuraminidase.[67] For influenza B virus neuraminidase, oseltamivir-resistant variants containing the single mutations E119G, R152 K, and D198N have been reported.[68−72] These residues are placed at the oseltamivir-binding pocket (see Figure 6).

The results of the MM/3D-RISM calculation are summarized in Table 1, which shows that the binding free energy for wild-type is negative while those of the mutants are positive. These results indicate that oseltamivir binding by mutated

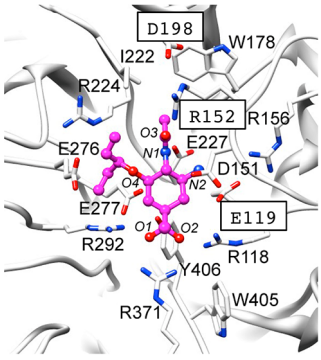


Figure 6. Oseltamivir and the bound residues of influenza B neuraminidase. The labeled residues are mutated for the oseltamivir-resistant mutants.

DOI: 10.1021/acs.jcim.7b00389
J. Chem. Inf. Model. 2017, 57, 2646−2656

Table 2. Binding Free Energy and its Components for Oseltamivir Binding by Influenza B Neuraminidase Compared with Those of Mutants[a]

energetics	wild-type	E119G	R152K	D198N
ΔG	−0.3 ± 1.0	2.9 ± 2.5	0.9 ± 2.0	0.1 ± 2.1
ΔG^{int}	−181.2 ± 5.6	−148.8 ± 2.7	−177.6 ± 9.3	−178.2 ± 1.9
ΔG^{solv}	180.9 ± 4.6	151.8 ± 2.5	178.6 ± 7.4	178.3 ± 2.5
$\delta\Delta G^{b}$		3.2	1.3	0.4
$\delta\Delta G^{IC50,b}$		2.0[c]	3.4[c]/2.7[d]	1.3[d]

[a]Units are kilocalories per mole. [b]δ is the value relative to the wild-type. The ΔG^{IC50} values are converted from experimental IC50 values. [c]Jackson et al.[72] [d]Mishin et al.[71]

neuraminidase is unfavorable. The relative binding free energy from the wild-type, $\delta\Delta G$, is compared with those converted from the experimental IC50 values. They showed a qualitatively good agreement, namely, E119G and R152K show greater resistance and D198N shows lesser resistance. As seen in eq 2, the binding free energy can be split into three components, which are shown in Table 2. The conformational energy changes caused by the ligand binding, ΔG^{conf}, are not considered because single trajectory analysis is employed. For all cases, ΔG^{int} takes a negative value while ΔG^{solv} is positive. Creation of hydrogen bonding between the oseltamivir ligand and the surrounding residues is a major factor for the negative ΔG^{int} value, whereas the positive contribution to ΔG^{solv} originates from the partial dehydration of both oseltamivir ligand and host neuraminidase. These results indicate that the loss of hydrogen bonds with solvent water and solute molecules is almost compensated for by the gain of hydrogen bonds between the ligand and the residues around the binding pocket. In the case of the wild-type, the stabilization by the gain of the hydrogen bonds is slightly greater than the dehydration penalty, consequently, the wild-type shows oseltamivir affinity. Contrary to the wild-type, the compensation for the dehydration by ligand−host interaction is not sufficient for all the mutants. In the case of the E119G mutant, because the hydrophilic glutamate residue is replaced by the nonpolar and short side chain of glycine, the hydrogen bond between residue 119 and the oseltamivir ligand is not observed (Figure 7). At the same time, this means the hydrogen bond between the residue 119 and solvent water also disappears. The change in the hydrogen

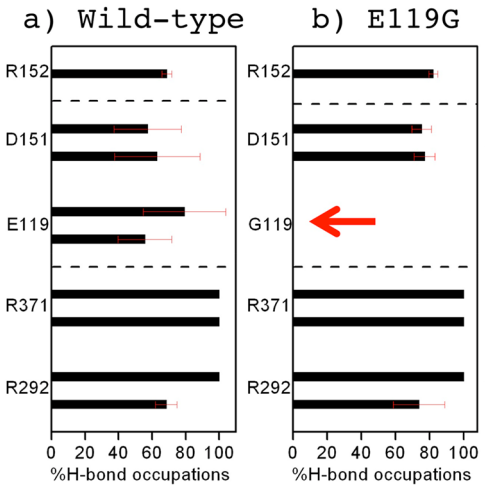


Figure 7. Occupation of the hydrogen bond formation between the specific residues and oseltamivir over the molecular dynamic trajectories. Red arrow indicates the mutated residue.

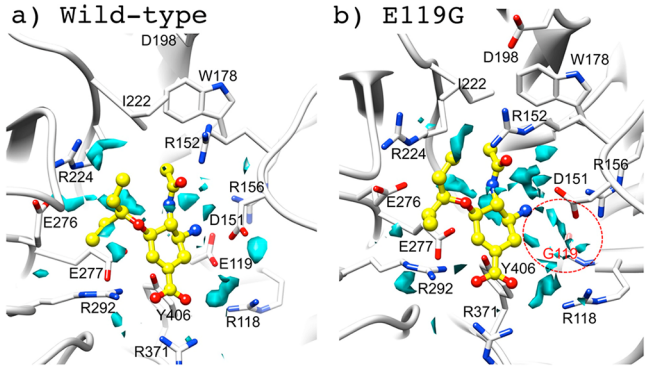


Figure 8. Isosurface plot of 3D-DFs of water oxygen colored blue inside the binding pocket of (a) wild-type and (b) E119G mutant with threshold $g(\mathbf{r}) \geq 5$. The red circle indicates the distribution caused by mutation.

bond formation shifts the ΔG^{int} positive and ΔG^{solv} negative, respectively. In addition, the substitution of the large side chain of glutamate to the small side chain of glycine makes space for the solvent water after ligand binding. As seen in Figure 8, solvent water distribution can be observed around G119; therefore, the oseltamivir ligand can keep the solvation partially and the dehydration penalty of the ligand is reduced. Similar analyses have been performed for R152K and D198N mutants, and detailed pictures of the binding affinity reduction of the oseltamivir-resistant mutants were clarified.

COMPUTATIONAL COST

Given that there are usually thousands to millions of drug candidate molecules, drug screening requires a huge computational cost. This computational cost is thus a crucial element to be considered in developing theoretical methods for drug design. In terms of the computational cost of evaluating solvation free energy, it is evident that implicit solvent models, such as GBSA and PBSA, are vastly superior to explicit solvent models. For example, the solvation free energy calculation for proteins that have several hundreds of amino acid residues can be easily performed by the implicit solvation model within 1 s with a commercially available single-node workstation and software package, such as Amber.

On the other hand, the computational cost of 3D-RISM is relatively high, because it requires nonlinear simultaneous equations to be solved in an iterative manner. Therefore, efforts to accelerate the program have been made. By using OpenMP parallelization, the 3D-RISM takes approximately 30 min to solve a single protein solvation structure using commercially available multicore central processing units (CPUs).

One way to overcome the large computational cost in solving the 3D-RISM equation is to use a supercomputer system, which

DOI: 10.1021/acs.jcim.7b00389
J. Chem. Inf. Model. 2017, 57, 2646−2656

have numerous nodes with many-core CPUs. For example, the RIKEN K supercomputer has over 700 000 CPU cores and each CPU has eight cores. To use the 3D-RISM program efficiently in such a supercomputer system, a massively parallel implementation of the program is essential. A three-dimensional, fast Fourier transform (3D-FFT) routine is required to solve the convolution integral in the 3D-RISM equation (eq 8), and the greatest challenge concerning the use of massively parallel implementation of 3D-RISM program is a parallelization of this 3D-FFT. The usual parallel 3D-FFT employs slab-type parallelization, in which the data array is distributed across nodes along only one axis. Therefore, the maximum number of nodes is limited to the number of grids on one axis. To overcome this difficulty, Maruyama et al. employed the volumetric 3D-FFT algorithm.[75] The 3D-RISM calculations on 16,384 nodes were tested, and the 3D-RISM program with volumetric 3D-FFT achieved excellent performance running on a RIKEN K supercomputer. Indeed, the calculations took only 4 and 11 s for 512^3 and 2048^3 grids, respectively. Another approach to realize massively parallel computing has been proposed by Yokogawa et al.[76] They proposed a theory free from 3D-FFT, called multicenter-molecular OZ (MC-MOZ). Because the MC-MOZ theory does not require 3D-FFT computation, the massively parallel computation is easy to perform with supercomputer systems.

An alternative way to accelerate the 3D-RISM program is to use the general-purpose computing facilities on graphics processing units (GPUs). Advantages of GPUs are high memory bandwidth and large numbers of cores. However, since GPUs typically have relatively small amounts of working memory, it is difficult to implement the 3D-RISM program. This is because the normal 3D-RISM program employs an algorithm called the modified method of direct inversion in iterative subspace (MDIIS)[77] to accelerate convergence, and this algorithm requires a large amount of memory. To overcome this problem, Maruyama and Hirata proposed a modified Anderson method to implement the 3D-RISM program on GPUs.[78] Given that the modified Anderson method requires only about one-third to one-fourth of the working memory compared with the MDIIS method, it made it possible to implement the 3D-RISM program on GPUs. By running the 3D-RISM program with a Tesla-K40 GPU, a single protein solvation structure can be solved in only a few minutes.

These implementations and the development of the methods will allow the practical use of 3D-RISM and related theories for drug screening.

SUMMARY AND PERSPECTIVE

In this paper, we reviewed the development and application of the theoretical method focusing on solvent effects in molecular recognition based on the 3D-RISM and related theories. The 3D-RISM theory realized an analysis that was difficult with the conventional method.

The 3D-DFs evaluated by the 3D-RISM theory are obtained through a complete ensemble average over the entire configuration space of solvent molecules in the thermodynamic limits. It is extremely difficult to achieve a similar level of ensemble average with an MD simulation if the solute molecule has a complex structure like a protein. Because the 3D-DFs by the 3D-RISM theory can reproduce the solvation structure even in the binding pocket, channel, and cavity inside the protein, it is easy to deduce the binding affinity of ligands, including solvent, ions, and cofactors, at these regions. By exploiting this feature of 3D-DFs, methods to predict the location of binding sites and the most probable binding modes of ligands have been proposed. The method is effective not only for drug design, but also for preparing the initial configuration of an MD simulation.

The 3D-RISM theory naturally reproduces all of the solvation thermodynamic properties, including solute−solvent interaction energy, solvation entropy, and solvation free energy, and their derivatives, such as the partial molar volume. The changes in these thermodynamic properties are useful in understanding the mechanism of molecular recognition at a molecular level. Although the 3D-RISM theory reproduces the change in solvation free energy due to molecular recognition, the accuracy of the absolute value of solvation free energy shows only qualitative agreements with experiments. To improve the accuracy of the solvation free energy, a pioneering work was conducted by Palmer et al.,[79,80] who found that the partial molar volume of the solute molecule had a linear correlation with the difference of the solvation free energy between the experimental and computational values. They proposed an empirical correction method of solvation free energy using linear correlation. Following their work, studies to improve the accuracy continue in many theoretical groups.[81−88]

Another serious concern in molecular recognition or drug design is an accurate treatment of the ligand−host protein interaction. To address this, quantum chemical methods are employed, such as quantum mechanics and molecular mechanics (QM/MM) or fragment molecular orbital (FMO) methods.[89−94] To consider the solvent effects, a hybrid of these methods with the 3D-RISM theory has been proposed by Yoshida et al.,[95,96] which is an extension of a series of studies on the development of the hybrid approach of the electronic structure theory and integral equation theory of liquids.[35,97−99]

In this review, we have focused on the advantages of the 3D-RISM theory compared to molecular simulation and continuum models. However, there are many areas where these methods are superior to the 3D-RISM theory. Theoretical development and applications by the concerted use of these methods are needed to tackle the unresolved problems of solvation and improve the accuracy further.

AUTHOR INFORMATION

Corresponding Author

*E-mail: noriwo@chem.kyushu-univ.jp.

ORCID

Norio Yoshida: 0000-0002-2023-7254

Funding

This work was supported by Grants-in-Aid (JP16H00842, JP16K05519) from MEXT, Japan.

Notes

The author declares no competing financial interest.

ACKNOWLEDGMENTS

The author is grateful to Prof. Fumio Hirata (Toyota RIKEN) and Prof. Haruyuki Nakano (Kyushu University) for invaluable discussions.

REFERENCES

(1) Michaelis, L.; Menten, M. L. The Kinetics of the Inversion Effect. *Biochem Z.* **1913**, *49*, 333−369.

(2) Govorkova, E. A.; Ilyushina, N. A.; Boltz, D. A.; Douglas, A.; Yilmaz, N.; Webster, R. G. Efficacy of Oseltamivir Therapy in Ferrets

DOI: 10.1021/acs.jcim.7b00389
J. Chem. Inf. Model. 2017, 57, 2646−2656

Inoculated with Different Clades of H5n1 Influenza Virus. *Antimicrob. Agents Chemother.* **2007**, *51*, 1414−1424.
(3) Sugaya, N.; Mitamura, K.; Yamazaki, M.; Tamura, D.; Ichikawa, M.; Kimura, K.; Kawakami, C.; Kiso, M.; Ito, M.; Hatakeyama, S.; Kawaoka, Y. Lower Clinical Effectiveness of Oseltamivir against Influenza B Contrasted with Influenza a Infection in Children. *Clin. Infect. Dis.* **2007**, *44*, 197−202.
(4) Helenius, A. Unpacking the Incoming Influenza-Virus. *Cell* **1992**, *69*, 577−578.
(5) Pinto, L. H.; Holsinger, L. J.; Lamb, R. A. Influenza-Virus M2 Protein Has Ion Channel Activity. *Cell* **1992**, *69*, 517−528.
(6) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Academic Press: San Diego, CA, 2001.
(7) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(8) Gilson, M. K.; Honig, B. Calculation of the Total Electrostatic Energy of a Macromolecular System - Solvation Energies, Binding-Energies, and Conformational-Analysis. *Proteins: Struct., Funct., Genet.* **1988**, *4*, 7−18.
(9) Tomasi, J.; Persico, M. Molecular-Interactions in Solution - an Overview of Methods Based on Continuous Distributions of the Solvent. *Chem. Rev.* **1994**, *94*, 2027−2094.
(10) Cramer, C.; Truhlar, D. G. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.* **1999**, *99*, 2161−2200.
(11) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, *105*, 2999−3093.
(12) Cheatham, T. E.; Srinivasan, J.; Case, D. A.; Kollman, P. A. Molecular Dynamics and Continuum Solvent Studies of the Stability of Polyg-Polyc and Polya-Polyt DNA Duplexes in Solution. *J. Biomol. Struct. Dyn.* **1998**, *16*, 265−280.
(13) Swanson, J. M. J.; Henchman, R. H.; McCammon, J. A. Revisiting Free Energy Calculations: A Theoretical Connection to Mm/Pbsa and Direct Calculation of the Association Free Energy. *Biophys. J.* **2004**, *86*, 67−74.
(14) Foloppe, N.; Hubbard, R. Towards Predictive Ligand Design with Free-Energy Based Computational Methods? *Curr. Med. Chem.* **2006**, *13*, 3583−3608.
(15) Beglov, D.; Roux, B. An Integral Equation to Describe the Solvation of Polar Molecules in Liquid Water. *J. Phys. Chem. B* **1997**, *101*, 7821−7826.
(16) Beglov, D.; Roux, B. Solvation of Complex Molecules in a Polar Liquid: An Integral Equation Theory. *J. Chem. Phys.* **1996**, *104*, 8678−8689.
(17) Kovalenko, A.; Hirata, F. Three-Dimensional Density Profiles of Water in Contact with a Solute of Arbitrary Shape: A RISM Approach. *Chem. Phys. Lett.* **1998**, *290*, 237−244.
(18) Hirata, F. *Molecular Theory of Solvation*; Kluwer: Dordrecht, 2003.
(19) Yoshida, N.; Imai, T.; Phongphanphanee, S.; Kovalenko, A.; Hirata, F. Molecular Recognition in Biomolecules Studied by Statistical-Mechanical Integral-Equation Theory of Liquids. *J. Phys. Chem. B* **2009**, *113*, 873−886.
(20) Yokogawa, D.; Sato, H.; Sakaki, S. An Integral Equation Theory for 3d Solvation Structure: A New Procedure Free from 3d Fourier Transform. *Chem. Phys. Lett.* **2006**, *432*, 595−599.
(21) Yokogawa, D.; Sato, H.; Sakaki, S. The Position of Water Molecules in Bacteriorhodopsin: A Three-Dimensional Distribution Function Study. *J. Mol. Liq.* **2009**, *147*, 112−116.
(22) Zhao, S.; Ramirez, R.; Vuilleumier, R.; Borgis, D. Molecular Density Functional Theory of Solvation: From Polar Solvents to Water. *J. Chem. Phys.* **2011**, *134*, 194102.
(23) Ishizuka, R.; Yoshida, N. Extended Molecular Ornstein-Zernike Integral Equation for Fully Anisotropic Solute Molecules: Formulation in a Rectangular Coordinate System. *J. Chem. Phys.* **2013**, *139*, 084119.
(24) Jeanmairet, G.; Levesque, M.; Sergiievskyi, V.; Borgis, D. Molecular Density Functional Theory for Water with Liquid-Gas Coexistence and Correct Pressure. *J. Chem. Phys.* **2015**, *142*, 154112.
(25) Hansen, J. P.; McDonald, I. R. *Theory of Simple Liquids*, 3rd ed.; Academic Press: Amsterdam, 2006.
(26) Kovalenko, A.; Blinov, N. Multiscale Methods for Nano-chemistry and Biophysics in Solution. *J. Mol. Liq.* **2011**, *164*, 101−112.
(27) Kovalenko, A.; Kobryn, A. E.; Gusarov, S.; Lyubimova, O.; Liu, X. J.; Blinov, N.; Yoshida, M. Molecular Theory of Solvation for Supramolecules and Soft Matter Structures: Application to Ligand Binding, Ion Channels, and Oligomeric Polyelectrolyte Gelators. *Soft Matter* **2012**, *8*, 1508−1520.
(28) Kovalenko, A. Multiscale Modeling of Solvation in Chemical and Biological Nanosystems and in Nanoporous Materials. *Pure Appl. Chem.* **2013**, *85*, 159−199.
(29) Sato, H. A Modern Solvation Theory: Quantum Chemistry and Statistical Chemistry. *Phys. Chem. Chem. Phys.* **2013**, *15*, 7450−65.
(30) Phongphanphanee, S.; Yoshida, N.; Hirata, F. Molecular Recognition Explored by a Statistical-Mechanics Theory of Liquids. *Curr. Pharm. Des.* **2011**, *17*, 1740−1757.
(31) Imai, T. A Novel Ligand-Mapping Method Based on Molecular Liquid Theory. *Curr. Pharm. Des.* **2011**, *17*, 1685−94.
(32) Maruyama, Y.; Yoshida, N.; Hirata, F. Electrolytes in Biomolecular Systems Studied with the 3D-RISM/RISM Theory. *Interdiscip. Sci.: Comput. Life Sci.* **2011**, *3*, 290−307.
(33) Ratkova, E. L.; Palmer, D. S.; Fedorov, M. V. Solvation Thermodynamics of Organic Molecules by the Molecular Integral Equation Theory: Approaching Chemical Accuracy. *Chem. Rev.* **2015**, *115*, 6312−6356.
(34) Chandler, D.; Andersen, H. C. Optimized Cluster Expansions for Classical Fluids. 2. Theory of Molecular Liquids. *J. Chem. Phys.* **1972**, *57*, 1930−1937.
(35) Kovalenko, A.; Hirata, F. Self-Consistent Description of a Metal-Water Interface by the Kohn-Sham Density Functional Theory and the Three-Dimensional Reference Interaction Site Model. *J. Chem. Phys.* **1999**, *110*, 10095−10112.
(36) Kovalenko, A.; Hirata, F. Potentials of Mean Force of Simple Ions in Ambient Aqueous Solution. I. Three-Dimensional Reference Interaction Site Model Approach. *J. Chem. Phys.* **2000**, *112*, 10391−10402.
(37) Singer, S. J.; Chandler, D. Free Energy Functions in the Extended RISM Approximation. *Mol. Phys.* **1985**, *55*, 621−625.
(38) Imai, T.; Hiraoka, R.; Kovalenko, A.; Hirata, F. Water Molecules in a Protein Cavity Detected by a Statistical-Mechanical Theory. *J. Am. Chem. Soc.* **2005**, *127*, 15334−15335.
(39) Yoshida, N.; Phongphanphanee, S.; Maruyama, Y.; Imai, T.; Hirata, F. Selective Ion-Binding by Protein Probed with the 3D-RISM Theory. *J. Am. Chem. Soc.* **2006**, *128*, 12042−12043.
(40) Yoshida, N.; Phongphanphanee, S.; Hirata, F. Selective Ion Binding by Protein Probed with the Statistical Mechanical Integral Equation Theory. *J. Phys. Chem. B* **2007**, *111*, 4588−4595.
(41) Phongphanphanee, S.; Yoshida, N.; Oiki, S.; Hirata, F. Distinct Configurations of Cations and Water in the Selectivity Filter of the Kcsa Potassium Channel Probed by 3D-RISM Theory. *J. Mol. Liq.* **2014**, *200*, 52−58.
(42) Sindhikara, D. J.; Yoshida, N.; Hirata, F. Placevent: An Algorithm for Prediction of Explicit Solvent Atom Distribution-Application to Hiv-1 Protease and F-Atp Synthase. *J. Comput. Chem.* **2012**, *33*, 1536−1543.
(43) Kiyota, Y.; Hiraoka, R.; Yoshida, N.; Maruyama, Y.; Imai, T.; Hirata, F. Theoretical Study of Co Escaping Pathway in Myoglobin with the 3D-RISM Theory. *J. Am. Chem. Soc.* **2009**, *131*, 3852−3853.
(44) Stumpe, M. C.; Blinov, N.; Wishart, D.; Kovalenko, A.; Pande, V. S. Calculation of Local Water Densities in Biological Systems: A Comparison of Molecular Dynamics Simulations and the 3D-RISM-Kh Molecular Theory of Solvation. *J. Phys. Chem. B* **2011**, *115*, 319−28.
(45) Hill, B. *Ion Channels of Excitable Membranes*, 3rd ed.; Sinauer Associates Inc.: Sunderland, MA, 2001.

DOI: 10.1021/acs.jcim.7b00389
J. Chem. Inf. Model. 2017, 57, 2646−2656

(46) Phongphanphanee, S.; Rungrotmongkol, T.; Yoshida, N.; Hannongbua, S.; Hirata, F. Proton Transport through the Influenza a M2 Channel: Three-Dimensional Reference Interaction Site Model Study. *J. Am. Chem. Soc.* **2010**, *132*, 9782−9788.
(47) Chen, H.; Wu, Y.; Voth, G. A. Origins of Proton Transport Behavior from Selectivity Domain Mutations of the Aquaporin-1 Channel. *Biophys. J.* **2006**, *90*, L73−L75.
(48) Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory. *J. Phys. Chem. B* **1998**, *102*, 3531−3541.
(49) Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 2. Applications to Simple Fluids. *J. Phys. Chem. B* **1998**, *102*, 3542−3550.
(50) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. Motifs for Molecular Recognition Exploiting Hydrophobic Enclosure in Protein-Ligand Binding. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 808−813.
(51) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding. *J. Am. Chem. Soc.* **2008**, *130*, 2817−2831.
(52) Nguyen, C. N.; Young, T. K.; Gilson, M. K. Grid Inhomogeneous Solvation Theory: Hydration Structure and Thermodynamics of the Miniature Receptor Cucurbit[7]Uril. *J. Chem. Phys.* **2012**, *137*, 044101.
(53) Imai, T.; Oda, K.; Kovalenko, A.; Hirata, F.; Kidera, A. Ligand Mapping on Protein Surfaces by the 3D-RISM Theory: Toward Computational Fragment-Based Drug Design. *J. Am. Chem. Soc.* **2009**, *131*, 12430−12440.
(54) Hirano, K.; Yokogawa, D.; Sato, H.; Sakaki, S. An Analysis of 3d Solvation Structure in Biomolecules: Application to Coiled Coil Serine and Bacteriorhodopsin. *J. Phys. Chem. B* **2010**, *114*, 7935−7941.
(55) Kiyota, Y.; Yoshida, N.; Hirata, F. A New Approach for Investigating the Molecular Recognition of Protein: Toward Structure-Based Drug Design Based on the 3D-RISM Theory. *J. Chem. Theory Comput.* **2011**, *7*, 3803−3815.
(56) Sindhikara, D. J.; Hirata, F. Analysis of Biomolecular Solvation Sites by 3D-RISM Theory. *J. Phys. Chem. B* **2013**, *117*, 6718−6723.
(57) Meier, T.; Krah, A.; Bond, P. J.; Pogoryelov, D.; Diederichs, K.; Faraldo-Gomez, J. D. Complete Ion-Coordination Structure in the Rotor Ring of Na+-Dependent F-Atp Synthases. *J. Mol. Biol.* **2009**, *391*, 498−507.
(58) Nikolić, D.; Blinov, N.; Wishart, D.; Kovalenko, A. 3D-RISM-D Ock: A New Fragment-Based Drug Design Protocol. *J. Chem. Theory Comput.* **2012**, *8*, 3356−3372.
(59) Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. A Semiempirical Free Energy Force Field with Charge-Based Desolvation. *J. Comput. Chem.* **2007**, *28*, 1145−1152.
(60) Singh, R. K.; Ethayathulla, A. S.; Jabeen, T.; Sharma, S.; Kaur, P.; Singh, T. P. Aspirin Induces Its Anti-Inflammatory Effects through Its Specific Binding to Phospholipase a(2): Crystal Structure of the Complex Formed between Phospholipase a(2) and Aspirin at 1.9 Angstrom Resolution. *J. Drug Targeting* **2005**, *13*, 113−119.
(61) Genheden, S.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Ryde, U. An Mm/3D-RISM Approach for Ligand Binding Affinities. *J. Phys. Chem. B* **2010**, *114*, 8505−8516.
(62) Luchko, T.; Gusarov, S.; Roe, D. R.; Simmerling, C.; Case, D. A.; Tuszynski, J.; Kovalenko, A. Three-Dimensional Molecular Theory of Solvation Coupled with Molecular Dynamics in Amber. *J. Chem. Theory Comput.* **2010**, *6*, 607−624.
(63) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Götz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *Amber12*; University of California: San Francisco, CA, 2012.
(64) Yesudas, J. P.; Blinov, N.; Dew, S. K.; Kovalenko, A. Calculation of Binding Free Energy of Short Double Stranded Oligonucleotides Using Mm/3D-RISM-Kh Approach. *J. Mol. Liq.* **2015**, *201*, 68−76.
(65) Sugita, M.; Hirata, F. Predicting the Binding Free Energy of the Inclusion Process of 2-Hydroxypropyl-Beta-Cyclodextrin and Small Molecules by Means of the Mm/3D-RISM Method. *J. Phys.: Condens. Matter* **2016**, *28*, 384002.
(66) Tanimoto, S.; Higashi, M.; Yoshida, N.; Nakano, H. The Ion Dependence of Carbohydrate Binding of Cbm36: An Md and 3D-RISM Study. *J. Phys.: Condens. Matter* **2016**, *28*, 344005.
(67) Phanich, J.; Rungrotmongkol, T.; Sindhikara, D.; Phongphanphanee, S.; Yoshida, N.; Hirata, F.; Kungwan, N.; Hannongbua, S. A 3D-RISM/RISM Study of the Oseltamivir Binding Efficiency with the Wild-Type and Resistance-Associated Mutant Forms of the Viral Influenza B Neuraminidase. *Protein Sci.* **2016**, *25*, 147−158.
(68) Gubareva, L. V.; Kaiser, L.; Matrosovich, M. N.; Soo-Hoo, Y.; Hayden, F. G. Selection of Influenza Virus Mutants in Experimentally Infected Volunteers Treated with Oseltamivir. *J. Infect. Dis.* **2001**, *183*, 523−531.
(69) Gubareva, L. V.; Matrosovich, M. N.; Brenner, M. K.; Bethell, R. C.; Webster, R. G. Evidence for Zanamivir Resistance in an Immunocompromised Child Infected with Influenza B Virus. *J. Infect. Dis.* **1998**, *178*, 1257−1262.
(70) Barnett, J. M.; Cadman, A.; Burrell, F. M.; Madar, S. H.; Lewis, A. P.; Tisdale, M.; Bethell, R. In Vitro Selection and Characterisation of Influenza B/Beijing/1/87 Isolates with Altered Susceptibility to Zanamivir. *Virology* **1999**, *265*, 286−295.
(71) Mishin, V. P.; Hayden, F. G.; Gubareva, L. V. Susceptibilities of Antiviral-Resistant Influenza Viruses to Novel Neuraminidase Inhibitors. *Antimicrob. Agents Chemother.* **2005**, *49*, 4515−4520.
(72) Jackson, D.; Barclay, W.; Zurcher, T. Characterization of Recombinant Influenza B Viruses with Key Neuraminidase Inhibitor Resistance Mutations. *J. Antimicrob. Chemother.* **2005**, *55*, 162−169.
(73) Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668−88.
(74) Rocchia, W.; Alexov, E.; Honig, B. Extending the Applicability of the Nonlinear Poisson-Boltzmann Equation: Multiple Dielectric Constants and Multivalent Ions. *J. Phys. Chem. B* **2001**, *105*, 6507−6514.
(75) Maruyama, Y.; Yoshida, N.; Tadano, H.; Takahashi, D.; Sato, M.; Hirata, F. Massively Parallel Implementation of 3D-RISM Calculation with Volumetric 3d-Fft. *J. Comput. Chem.* **2014**, *35*, 1347−1355.
(76) Yokogawa, D.; Sato, H.; Imai, T.; Sakaki, S. A Highly Parallelizable Integral Equation Theory for Three Dimensional Solvent Distribution Function: Application to Biomolecules. *J. Chem. Phys.* **2009**, *130*, 064111.
(77) Kovalenko, A.; Ten-No, S.; Hirata, F. Solution of Three-Dimensional Reference Interaction Site Model and Hypernetted Chain Equations for Simple Point Charge Water by Modified Method of Direct Inversion in Iterative Subspace. *J. Comput. Chem.* **1999**, *20*, 928−936.
(78) Maruyama, Y.; Hirata, F. Modified Anderson Method for Accelerating 3D-RISM Calculations Using Graphics Processing Unit. *J. Chem. Theory Comput.* **2012**, *8*, 3015−3021.
(79) Palmer, D. S.; Frolov, A. I.; Ratkova, E. L.; Fedorov, M. V. Towards a Universal Method for Calculating Hydration Free Energies: A 3d Reference Interaction Site Model with Partial Molar Volume Correction. *J. Phys.: Condens. Matter* **2010**, *22*, 492101.
(80) Palmer, D.; Chuev, G.; Ratkova, E.; Fedorov, M. In Silico Screening of Bioactive and Biomimetic Solutes Using Molecular Integral Equation Theory. *Curr. Pharm. Des.* **2011**, *17*, 1695−1708.
(81) Truchon, J.-F.; Pettitt, B. M.; Labute, P. A Cavity Corrected 3D-RISM Functional for Accurate Solvation Free Energies. *J. Chem. Theory Comput.* **2014**, *10*, 934−941.

DOI: 10.1021/acs.jcim.7b00389
J. Chem. Inf. Model. 2017, 57, 2646−2656

(82) Miyata, T.; Thapa, J. Accuracy of Solvation Free Energy Calculated by Hypernetted Chain and Kovalenko-Hirata Approximations for Two-Component System of Lennard-Jones Liquid. *Chem. Phys. Lett.* **2014**, *604*, 122−126.
(83) Sergiievskyi, V.; Jeanmairet, G.; Levesque, M.; Borgis, D. Solvation Free-Energy Pressure Corrections in the Three Dimensional Reference Interaction Site Model. *J. Chem. Phys.* **2015**, *143*, 184116.
(84) Chong, S. H.; Ham, S. Thermodynamic-Ensemble Independence of Solvation Free Energy. *J. Chem. Theory Comput.* **2015**, *11*, 378−380.
(85) Sumi, T.; Mitsutake, A.; Maruyama, Y. A Solvation-Free-Energy Functional: A Reference-Modified Density Functional Formulation. *J. Comput. Chem.* **2015**, *36*, 1359−1369.
(86) Miyata, T.; Miyazaki, S. Accuracy of Temperature-Derivative of Radial Distribution Function Calculated under Approximations in Ornstein-Zernike Theory for One-Component Lennard-Jones Fluid. *Chem. Phys. Lett.* **2016**, *658*, 224−229.
(87) Miyata, T.; Ebato, Y. Thermodynamic Significance to Correct the Location of First Rising Region in Radial Distribution Function Approximately Estimated from Ornstein-Zernike Integral Equation Theory for Lennard-Jones Fluids. *J. Mol. Liq.* **2016**, *217*, 75−82.
(88) Ebato, Y.; Miyata, T. A Pressure Consistent Bridge Correction of Kovalenko-Hirata Closure in Ornstein-Zernike Theory for Lennard-Jones Fluids by Apparently Adjusting Sigma Parameter. *AIP Adv.* **2016**, *6*, 055111.
(89) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment Molecular Orbital Method: An Approximate Computational Method for Large Molecules. *Chem. Phys. Lett.* **1999**, *313*, 701−706.
(90) Fukuzawa, K.; Mochizuki, Y.; Tanaka, S.; Kitaura, K.; Nakano, T. Molecular Interactions between Estrogen Receptor and Its Ligand Studied by the Ab Initio Fragment Molecular Orbital Method. *J. Phys. Chem. B* **2006**, *110*, 16102−16110.
(91) Sawada, T.; Fedorov, D. G.; Kitaura, K. Role of the Key Mutation in the Selective Binding of Avian and Human Influenza Hemagglutinin to Sialosides Revealed by Quantum-Mechanical Calculations. *J. Am. Chem. Soc.* **2010**, *132*, 16862−16872.
(92) Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum-Mechanical and Molecular Mechanical Potential for Molecular-Dynamics Simulations. *J. Comput. Chem.* **1990**, *11*, 700−733.
(93) Gao, J. L. Hybrid Quantum and Molecular Mechanical Simulations: An Alternative Avenue to Solvent Effects in Organic Chemistry. *Acc. Chem. Res.* **1996**, *29*, 298−305.
(94) Aqvist, J.; Warshel, A. Simulation of Enzyme-Reactions Using Valence-Bond Force-Fields and Other Hybrid Quantum-Classical Approaches. *Chem. Rev.* **1993**, *93*, 2523−2544.
(95) Yoshida, N.; Kiyota, Y.; Hirata, F. The Electronic-Structure Theory of a Large-Molecular System in Solution: Application to the Intercalation of Proflavine with Solvated Dna. *J. Mol. Liq.* **2011**, *159*, 83−92.
(96) Yoshida, N. Efficient Implementation of the Three-Dimensional Reference Interaction Site Model Method in the Fragment Molecular Orbital Method. *J. Chem. Phys.* **2014**, *140*, 214118.
(97) Sato, H.; Kovalenko, A.; Hirata, F. Self-Consistent Field, Ab Initio Molecular Orbital and Three-Dimensional Reference Interaction Site Model Study for Solvation Effect on Carbon Monoxide in Aqueous Solution. *J. Chem. Phys.* **2000**, *112*, 9463−9468.
(98) Ten-No, S.; Hirata, F.; Kato, S. A Hybrid Approach for the Solvent Effect on the Electronic-Structure of a Solute Based on the RISM and Hartree-Fock Equations. *Chem. Phys. Lett.* **1993**, *214*, 391−396.
(99) Sato, H.; Hirata, F.; Kato, S. Analytical Energy Gradient for the Reference Interaction Site Model Multiconfigurational Self-Consistent-Field Method: Application to 1,2-Difluoroethylene in Aqueous Solution. *J. Chem. Phys.* **1996**, *105*, 1546−1551.

DOI: 10.1021/acs.jcim.7b00389
J. Chem. Inf. Model. 2017, 57, 2646−2656