PROTEINS: Structure, Function, and Bioinformatics 66:804–813 (2007)

Locating Missing Water Molecules in Protein Cavities by the Three-Dimensional Reference Interaction Site Model Theory of Molecular Solvation

Takashi Imai,[1]* Ryusuke Hiraoka,[1] Andriy Kovalenko,[2,3] and Fumio Hirata[4]
[1]*Department of Bioscience and Bioinformatics, Ritsumeikan University, Kusatsu, Shiga 525–8577, Japan*
[2]*National Institute for Nanotechnology, National Research Council of Canada, Edmonton, Alberta T6G 2V4, Canada*
[3]*Department of Mechanical Engineering, University of Alberta, Edmonton, Alberta T6G 2V4, Canada*
[4]*Department of Theoretical Studies, Institute for Molecular Science, Okazaki, Aichi 444–8585, Japan*

ABSTRACT **Water molecules confined in protein cavities are of great importance in understanding the protein structure and functions. However, it is a nontrivial task to locate such water molecules in protein by the ordinary molecular simulation and modeling techniques as well as experimental methods. The present study proves that the three-dimensional reference interaction site model (3D-RISM) theory, a recently developed statistical–mechanical theory of molecular solvation, has an outstanding advantage in locating such water molecules. In this paper, we demonstrate that the 3D-RISM theory is able to reproduce the structure and the number of water molecules in cavities of hen egg-white lysozyme observed commonly in the X-ray structures of different resolutions and conditions. Furthermore, we show that the theory successfully identified a water molecule in a cavity, the existence of which has been ambiguous even from the X-ray results. In contrast, we confirmed that molecular dynamics simulation is helpless at present to find such water molecules because the results substantially depend on the initial coordinates of water molecules. Possible applications of the theory to problems in the fields of biochemistry and biophysics are also discussed. Proteins 2007; 66:804–813.**

Key words: hen egg-white lysozyme; protein hydration; molecular recognition; 3D-RISM theory; molecular dynamics simulation

INTRODUCTION

Water molecules are commonly observed in internal cavities of proteins.[1–5] Such internally buried water molecules are highly conserved in homologous protein families,[1,6] as the key residues are. This fact implies that the internal water molecules play a vital role in stabilizing the native structure of proteins and in their function. In fact, by reducing the water content in the crystallization[7] or by using mutant proteins,[8] it has been demonstrated that the internal water molecules contribute to the structure and stability of proteins. The roles of internal water molecules in the protein function have also been revealed by other kinds of biomolecules, including enzyme,[9] ion channel,[10] and proton pump[11] proteins. Water molecules bridging the cavity between a receptor protein and its ligand can be a clue to the structure-based rational drug design. For instance, HIV protease inhibitors were designed by displacing and mimicking the hydrogen-bonding between the receptor and a bridging water molecule.[12] Thus, finding such water molecules in proteins is essential in understanding and predicting the structure and function of proteins.

Considerable efforts have therefore been made to observe these water molecules inside and around proteins by several experimental methods such as X-ray[13] and neutron diffraction,[14] and NMR.[15,16] However, it is still a nontrivial task. In reality, the hydration models in protein crystals determined by the X-ray analysis depend on the experimental conditions and the resolution.[3] The water molecules detected by X-ray and neutron diffraction are inconsistent with each other in some cases.[3,6] Besides, the difference between the water sites in crystal and the hydration structure in aqueous solution might have given substantial problems. Although the NMR techniques are useful for obtaining information about the hydration properties of proteins, especially their dynamics,[15,16] in aqueous solution, they cannot yield three-dimensional (3D) spatial information of the hydration structure.

Molecular simulation has been prevalently used to analyze protein hydration and dynamics. Various simulation methods have been developed and successfully applied to solve the various problems in analyzing protein structure

Grant sponsor: Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan; Grant number: 15076212; Grant sponsors: NAREGI Nanoscience Project, MEXT; "High-Tech Research Center" Project for Private Universities, MEXT; and National Research Council of Canada.

*Correspondence to: Takashi Imai, Department of Bioscience and Bioinformatics, Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan. E-mail: t-imai@is.ritsumei.ac.jp

Received 8 April 2006; Revised 14 September 2006; Accepted 23 October 2006

Published online 21 December 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21311

and function. However, it is one of the most difficult tasks for the molecular simulation to predict the presence of internal water molecules ab initio. That is because they are most likely trapped in the protein through a process of large conformational fluctuation. If water molecules were not initially put at a cavity isolated from the bulk aqueous phase, they could not reach the cavity within the ordinary simulation time of the nanosecond order. Indeed, water molecules could not penetrate into or escape from some internal sites hidden deep in protein in molecular dynamics (MD) simulation of the nanosecond order, though they could come and go between cavities or clefts near to the surface and the bulk phase.[17–19] The consequence is also expected from NMR studies, which have shown that the residence time of the water molecules buried in internal cavities or trapped in narrow clefts is in the range from about 10^{-9} to 10^{-2} second (1 ns to 10 ms), whereas that of the water molecules at the protein surface is in the subnanosecond range.[15,16] It is obvious that the simulation of the millisecond order, which is comparable to the ab initio protein folding simulation, is by far impossible to be carried out.

Integral equation theories of solutions[20] are prospective candidates for the theoretical approach that can describe the hydration structure of protein. Given the molecular interactions and the thermodynamic conditions, they yield the intermolecular correlation functions, such as the radial distribution function (RDF), through statistical–mechanical relations. The obtained functions are free from statistical errors and satisfy complete ergodicity, which ensures the proper penetration–escape equilibrium of water in protein, for example.

The reference interaction site model (RISM) theory[20–22] is one of the integral equation theories of molecular solutions. It adopts the site–site description for the intermolecular correlation functions. The theory has succeeded in describing various kinds of solution structure as well as in providing the thermodynamic quantities of solution.[22–25] However, it has been found that the RISM theory breaks down when applied to larger molecular systems such as proteins. For instance, it systematically underestimates one of the thermodynamic quantities, the partial molar volume (PMV), of amino acids,[26] and sometimes gives negative PMV values for polypeptides and proteins,[27] which is apparently unphysical.

Recently, we have demonstrated that the 3D generalization[28–31] of the RISM theory provides a drastic improvement in the quantitative estimation of the PMV of biomolecules.[27,32,33] The 3D-RISM theory has great advantages of not only providing the thermodynamic properties quantitatively, but also yielding the 3D distribution functions of solvent sites (water-oxygen and hydrogen). This gives a more detailed description of the hydration structure around the solute (protein), compared to the RDFs between solute sites and solvent sites obtained by the ordinary RISM theory. However, it was an open question whether the 3D-RISM theory can describe water molecules confined in small cavities of protein. We finally answered it in the latest study applying the 3D-RISM theory to hen egg-white lysozyme in aqueous solution.[34] The preliminary result demonstrated that, in comparison with the available X-ray structure, the theory gives an excellent description of water molecules within a cavity. So, we are convinced now that the 3D-RISM theory is the most promising approach to locate the places where water molecules are present in a protein.

In this study, we apply the 3D-RISM theory to locate entrapped water in all cavities of hen egg-white lysozyme in aqueous solution. Lysozyme has three major cavities of sizes enough to hold at least one water molecule, which are illustrated in Figure 1. Table I compiles the data of water molecules detected in the cavities, based on some available X-ray structures of lysozyme. In Cavity 1, which is the largest, there are four water molecules (W1 to W4, numbered in the order from the bottom of the cavity), apart from a few exceptions. Cavity 2 is quite small and holds only one water molecule (W5). Cavity 3 is larger than Cavity 2, but most of the X-ray analyses detected no water molecules in that cavity. Some other analyses located one water molecule (W6) in the cavity, but it has a large B-factor. It has therefore been unclear whether or not a water molecule is actually present there. Here, we clarify where and how many water molecules exist in those cavities from the distribution functions obtained by the 3D-RISM theory. The distribution functions and the hydration numbers are compared with the results of the MD simulation. As mentioned earlier, simulation results concerning the hydration number in the cavities may not be free from ambiguity, depending on the initial condition. Therefore, we examined two types of simulation, changing the initial configuration of water molecules: (i) all the water molecules are initially set at their X-ray positions in the cavities, and (ii) the water molecules are initially removed from the cavities.

METHODS

3D-RISM Theory

The three-dimensional (3D) distribution functions of water oxygen and hydrogen sites around a lysozyme molecule are obtained by the 3D-RISM theory.[28–31] Here, we briefly outline the theory.

For a solute–solvent system at infinite dilution, the 3D-RISM integral equation is written as[28–31]

$$h_\gamma(\mathbf{r}) = \sum_{\gamma'} c_{\gamma'}(\mathbf{r}) \ * \ \left(w^{vv}_{\gamma'\gamma}(r) + \rho h^{vv}_{\gamma'\gamma}(r) \right) \qquad (1)$$

where $h_\gamma(\mathbf{r})$ and $c_\gamma(\mathbf{r})$ are, respectively, the 3D total and direct correlation functions of solvent site γ around the solute, the asterisk denotes a convolution integral in direct space, $w^{vv}_{\gamma'\gamma}(r)$ is the site–site intramolecular correlation function of solvent molecules, ρ is the number density of solvent, and $h^{vv}_{\gamma'\gamma}(r)$ is the total correlation functions of pure solvent, obtained independently from a single-component RISM theory. We adopted the dielectrically consistent RISM theory[35,36] to obtain the correla-

TABLE I. B-Factors of Water Molecules in Three Cavities of Lysozyme Detected by X-ray Analyses

		Cavity 1[a]				Cavity 2	Cavity 3
PDB ID	Resolution	W1	W2	W3	W4	W5	W6
4LZT	0.95	7.99	9.47	15.24	14.15	29.09	—
193L	1.33	11.03	23.33	22.62	14.52	11.46	—
1HEL[b]	1.70	9.12	21.39	28.04	15.91	11.69	—
1LSE	1.70	14.28	28.77	17.84	—	17.25	46.26
5LTM	1.80	15.30	17.15	26.02	22.62	30.01	—
5LYT	1.90	13.25	25.70	8.81	25.61	5.15	49.80
2LYM	2.00	16.81	22.55	18.26	20.49	14.36	67.12
2LZT	2.00	4.00	12.33	13.82	9.57	6.66	—
2LYZ	2.00	(1.59)[c]	(1.70)[c]	(1.80)[c]	—	(1.42)[c]	—
Average[d]		11.47	20.09	18.83	17.55	15.71	54.39

[a]Water sites are numbered in the order from the bottom of the cavity.
[b]The structure used in the 3D-RISM calculation.
[c]Atomic radius.
[d]Excluding the data of 2LYZ and undetected ones.

tion functions of pure water in this study. The 3D distribution function $g_\gamma(\mathbf{r})$ is defined by $g_\gamma(\mathbf{r}) = h_\gamma(\mathbf{r}) + 1$.

The 3D-RISM equation is complemented by the 3D-HNC closure, including analytical corrections to the 3D correlation functions for the supercell periodicity artifact,[28,31]

$$h_\gamma(\mathbf{r}) = \exp\left(-\beta u_\gamma(\mathbf{r}) + h_\gamma(\mathbf{r}) - c_\gamma(\mathbf{r}) - \Delta Q_\gamma\right) + \Delta Q_\gamma - 1 \quad (2)$$

where $u_\gamma(\mathbf{r})$ is the interaction potential between solvent site γ and the whole solute, which is calculated on the supercell grid using the minimum image convention and the Ewald summation method. The correction term ΔQ_γ is calculated in the Fourier space by

$$\Delta Q_\gamma = \frac{4\pi\beta}{V_{\text{cell}}} q \lim_{k\to 0} \sum_{\gamma'} \frac{q_{\gamma'}}{k^2} \left(w^{vv}_{\gamma'\gamma}(k) + \rho h^{vv}_{\gamma'\gamma}(k)\right) \quad (3)$$

where V_{cell} is the supercell volume, q is the net charge of the solute, and q_γ is the partial site charge of solvent site γ.

3D-RISM Calculation

The 3D atomic coordinates of hen egg-white lysozyme were taken from Protein Databank (PDB) code 1HEL,[37] and then the hydrogen atoms were added at the proper positions. In the 3D-RISM calculation, all the crystallographic water molecules were removed from the coordinate data.

The 3D potential was constructed from the site–site pair potential, which consists of the Lennard-Jones 12-6 and Coulomb potential terms. We used the potential parameters for lysozyme sites from Amber (parm99)[38] and SPC/E[39] for water with a repulsive term added to the hydrogen site[40] (σ_H = 0.4 Å and ε_H = 0.05 kcal/mol). The total charge on lysozyme was neutralized by uniformly shifting the charges of the protein atoms. The neutralization is not required by the 3D-RISM calculation but by the MD simulation. The uniform charge shifts were implemented instead of including counter ions in order to present a fair comparison between the 3D-RISM and MD calculations, because it is a hard task for MD simulation to reach the complete equilibrium in an electrolyte solution. The result of the 3D-RISM calculation is not sensitive to this small charge shifts.

The 3D-RISM/HNC equations were solved on a grid of 256^3 points in a cubic supercell of size 128 Å, with ambient water at temperature 298.15 K and number density 0.033329 $Å^{-3}$, by using the modified direct inversion in the iterative subspace method.[28,41] The RDFs $g_{\alpha\gamma}(r)$ between solute site α and solvent site γ were obtained by numerical averaging of the 3D distribution function $g_\gamma(\boldsymbol{r})$ over the solute orientation. The averaging employed the 700-point repulsion scheme based on the spherical harmonic reduction or elimination by a weighted distribution quadrature method.[42] The isosurfaces of the 3D distribution functions were visualized by the VMD software.[43]

Molecular Dynamics Simulation

We completed two MD simulations using the public program package GROMACS.[44,45] For Simulation (i), all the crystallographic water molecules were put in their crystallographic positions, and then other water molecules were automatically added while avoiding overlap so that the density was kept usual. Additionally, the water molecule in Cavity 3 was manually set at the site because it was not included in the PDB data, 1HEL. For Simulation (ii), all water molecules were added automatically. Then, the water molecules accidentally set in the cavities were removed.

Each simulation was carried out in the NVT ensemble using Nosé-Hoover algorithm, at temperature 298.15 K in a cubic box of size 64 Å, with the periodic boundary conditions. The long-range electrostatic interactions were

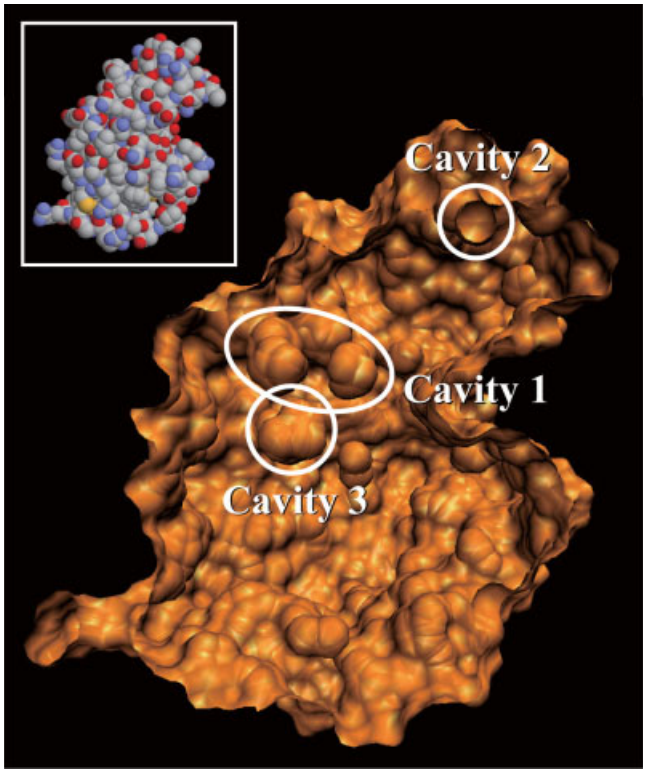

Fig. 1. Molecular surface representation of lysozyme. The front half of the surface is clipped off to indicate three cavities inside the protein. The upper left inset is the van der Waals model of lysozyme. The atomic coordinates are taken from 1HEL.

treated by the particle mesh Ewald method. The box contained one lysozyme molecule and 8038 water molecules in Simulation (i) or 8030 in Simulation (ii). The atomic coordinates of lysozyme were fixed at the positions of the X-ray structure to perform the direct comparison with the 3D-RISM theory. The potential parameters were also identical with those of the 3D-RISM calculation. For each simulation, the data were taken from the 10-ns simulation after the 0.1-ns equilibration.

RESULTS

Figure 2(a) shows the 3D distribution functions $g(\mathbf{r})$ of water oxygen and hydrogen in Cavity 1, calculated by the 3D-RISM theory, which are represented by isosurfaces of $g(\mathbf{r}) > 8$ each for oxygen (red) and hydrogen (white). Four and eight major peaks, respectively, for water oxygen and hydrogen are found in the cavity. The hydration structure is in excellent agreement with the coordinates of four water molecules obtained from the X-ray diffraction. The hydrogen-bonding networks between the water molecules and the polar sites in the cavity are also reasonable. (More details are described in our previous communication.[34])

We compare this result with that of MD simulation in terms of the radial distribution function (RDF) and the running coordination number viewed from W2 site of the

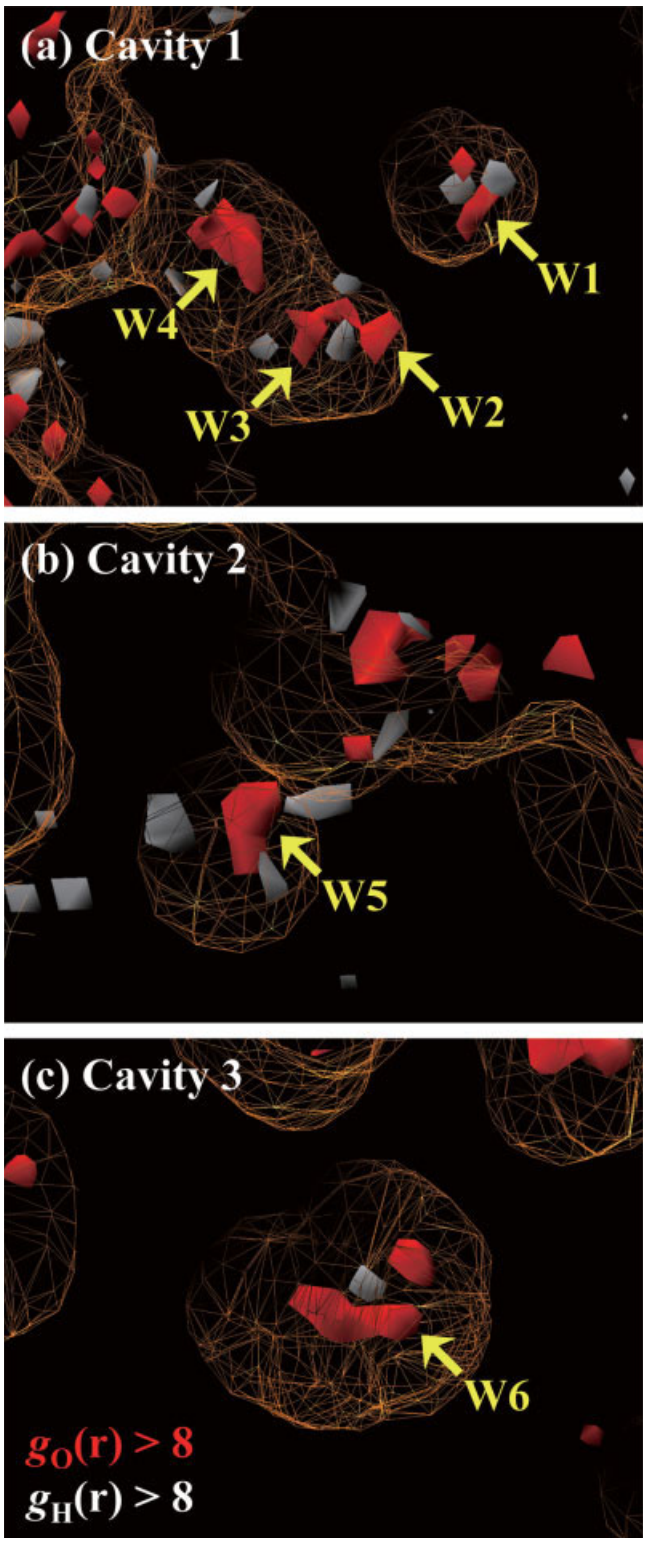

Fig. 2. Isosurface representation of the 3D distribution functions $g(\mathbf{r})$ of water oxygen and hydrogen in the three cavities of lysozyme, calculated by the 3D-RISM theory. The surfaces where $g(\mathbf{r}) > 8$ are drawn in red and white for water oxygen and hydrogen, respectively. The lysozyme surfaces are represented by mesh. (**a**) Cavity 1, (**b**) Cavity 2, and (**c**) Cavity 3.

X-ray structure, which is almost at the center of the cavity. The running coordination number is defined as follows:

$$N(r) = 4\pi\rho \int_0^r g(r) r^2 dr. \quad (4)$$

It signifies the hydration number within a sphere of radius r from the origin. Figure 3 shows these numbers as a function of the radial distance r, obtained by the 3D-RISM theory and two sets of the MD simulation: (i) the four water molecules are initially set at their X-ray positions in the cavity and (ii) the four water molecules are initially removed from the cavity.

First, we compare the results of the 3D-RISM theory and Simulation (i). Concerning the general feature of RDF, the 3D-RISM prediction is in fair agreement with the simulation result. However, the details are somewhat different: the peak height in RDF from 3D-RISM is generally lower, and the width is broader. Nevertheless, the running coordination numbers obtained from the 3D-RISM theory and Simulation (i) are in good coincidence. The hydration numbers within the cavity ($r < 5.2$) following from the 3D-RISM theory and from Simulation (i) are 3.6 and 3.5, respectively. (In the previous paper,[34] the hydration number from the 3D-RISM theory was reported to be 3.8. That value was calculated based on the RDF from the side chain oxygen of S91 and was found to include a little contribution, about 0.1, from the other sites outside the cavity.)

However, the hydration number 2.7 obtained from Simulation (ii) agrees neither with the 3D-RISM theory nor with Simulation (i). The discrepancy of about 1 in the hydration number is due to a very simple reason. In fact, any water molecule could not reach W1 site in the simulation time in the present study. That is because W1 site is completely isolated from the bulk aqueous phase, while the other sites have access to the bulk phase. Consequently, the hydration number obtained from the simulation crucially depends on the initial coordinates of water molecules.

The hydration numbers in the cavity obtained from the 3D-RISM theory, 3.6, and MD simulation, 3.5, were less than the number of water sites, 4, detected by the 3D-RISM theory as well as X-ray diffraction. To explain the reason, we analyze the time course of the positions of water molecules in the cavities using the trajectory of Simulation (i). Figure 4(a) shows the time course of the distances between the four water molecules and W2 site of the X-ray structure. Water 4 sometimes leaves and enters the site in a timescale of the nanosecond order, which is consistent with the timescale found in previous NMR[15,16] and simulation[17–19] studies. Water 3 stays in its position except for a moment, though it largely fluctuates when Water 4 leaves its site. Water 2 also keeps the position and shows a behavior similar to that of Water 3. The fluctuation of Water 1 is highly limited because W1 site is small and isolated from the others. As a result, the hydration number had a value between 3 and 4. The time course from Simulation (ii), shown in Figure 4(b), exhibits a similar behavior for Waters 2, 3, and 4. However, no water molecule ever reached W1 site during the simulation time. We emphasize that although the 3D-RISM theory takes no explicit account of the dynamics of molecules, it can provide a reasonable hydration number through the statistical–mechanical relations.

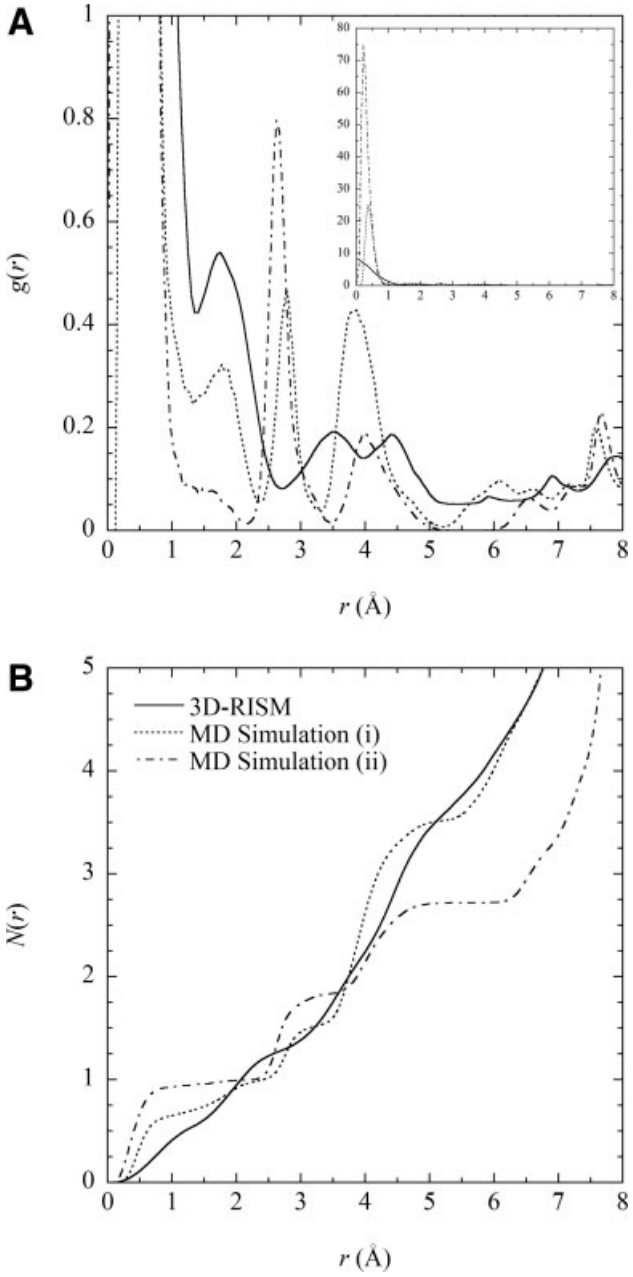

Fig. 3. Radial distribution functions and running coordination numbers of water oxygen viewed from site W2 in Cavity 1. The cavity is within $r = 5.2$ Å. Results of the 3D-RISM theory (solid lines), the MD simulations (i) initially including the water molecules in the cavities (dotted lines), and (ii) initially excluding these water molecules (dot-dashed lines). (**a**) Radial distribution functions, and (**b**) running coordination numbers.

Figure 2(b) shows the isosurface plots of the 3D distribution functions of water in Cavity 2. One peak of water

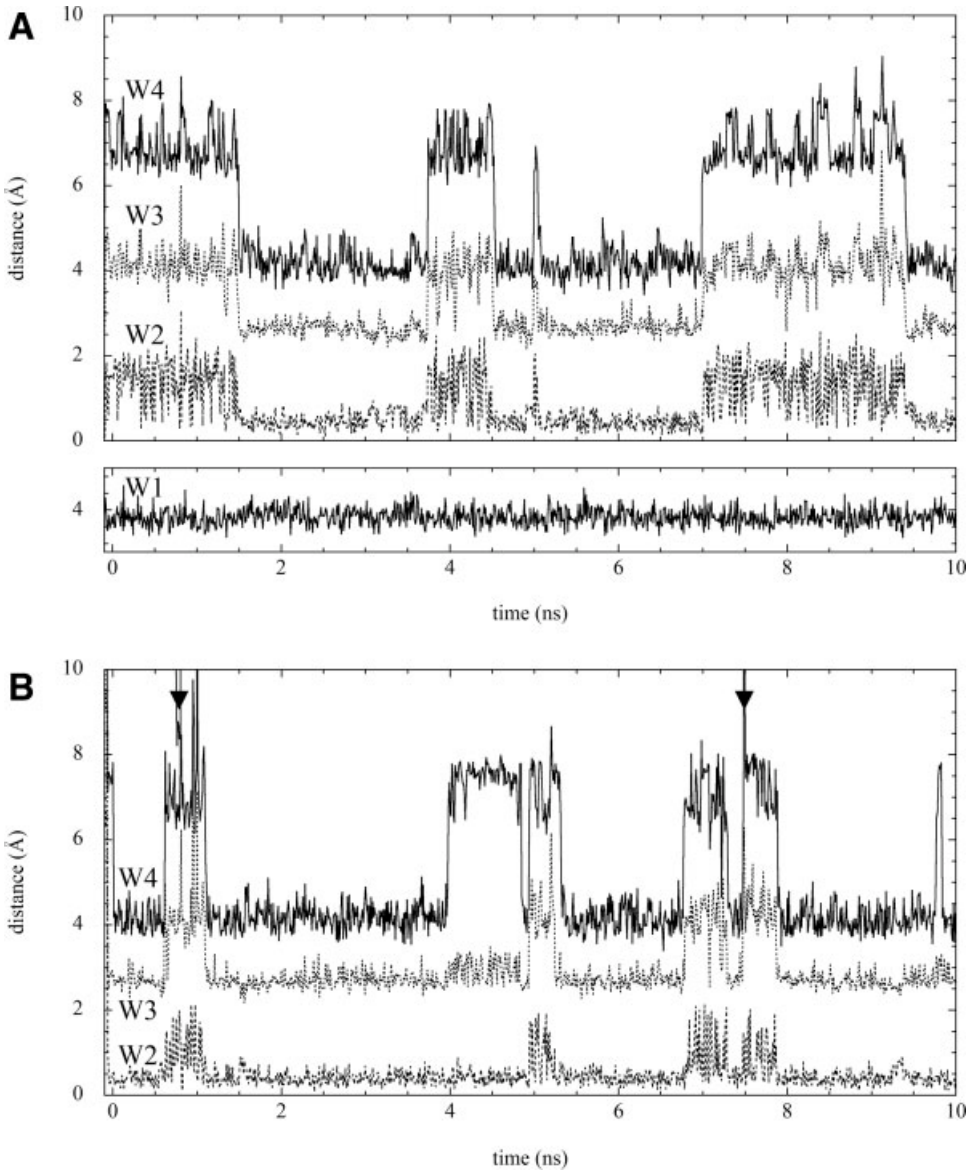

Fig. 4. Time course of the distance between site W2 and each of the four water molecules (W1 to W4) in Cavity 1. (**a**) Simulation (i) in which the water molecules were initially set in the X-ray positions. (**b**) Simulation (ii) in which they were initially removed from the sites. In Simulation (ii), water W4 is exchanged for the bulk phase water twice at the moments indicated by the arrows. Negative time corresponds to the equilibration process.

oxygen and two peaks of water hydrogen are there, implying there is one water site. Figure 5 shows the RDF and the running coordination number of water from the W5 site of the X-ray structure. The hydration number within the cavity ($r < 1.7$) is 0.8, from the 3D-RISM result. This result implies that one water molecule is in the cavity, much as the X-ray analysis indicates, but the water molecule does not always stay there. This point will be further discussed in the next section.

In contrast, Simulation (i) gives a trivial result: the water molecule initially set at the site has never got out of the cavity, and the hydration number is 1. On the other hand, Simulation (ii) indicates that the hydration number is 0.4. From the trajectory, it was found that no water molecule could enter the cavity until about 6 ns, and then a water molecule entered there by chance and had been trapped for the rest of the simulation time of 10 ns. That is because the route from the bulk phase to Cavity 2 is very narrow and a water molecule cannot easily pass through there. Thus, the molecular simulation of the nanosecond order cannot conclude how much water prefers staying there.

In the aforementioned examinations for Cavities 1 and 2, we have demonstrated that the 3D-RISM theory can successfully detect water molecules in protein cavities. Then it is of great interest to see how many water molecules are detected in Cavity 3 by the successful theory,

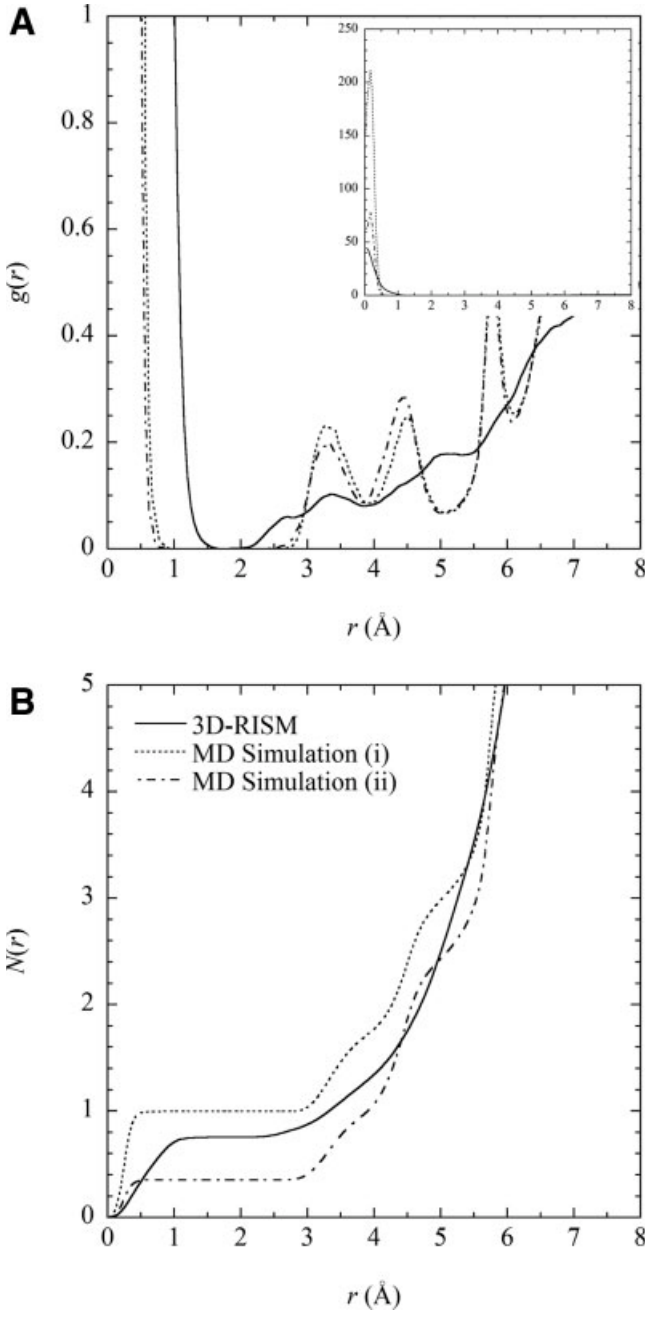

Fig. 5. Radial distribution functions and running coordination numbers of water oxygen viewed from site W5 in Cavity 2. The cavity is within $r = 1.7$ Å. Results of the 3D-RISM theory (solid lines), the MD simulations (i) initially including the water molecules in the cavities (dotted lines), and (ii) initially excluding these water molecules (dot-dashed lines). (**a**) Radial distribution functions, and (**b**) running coordination numbers.

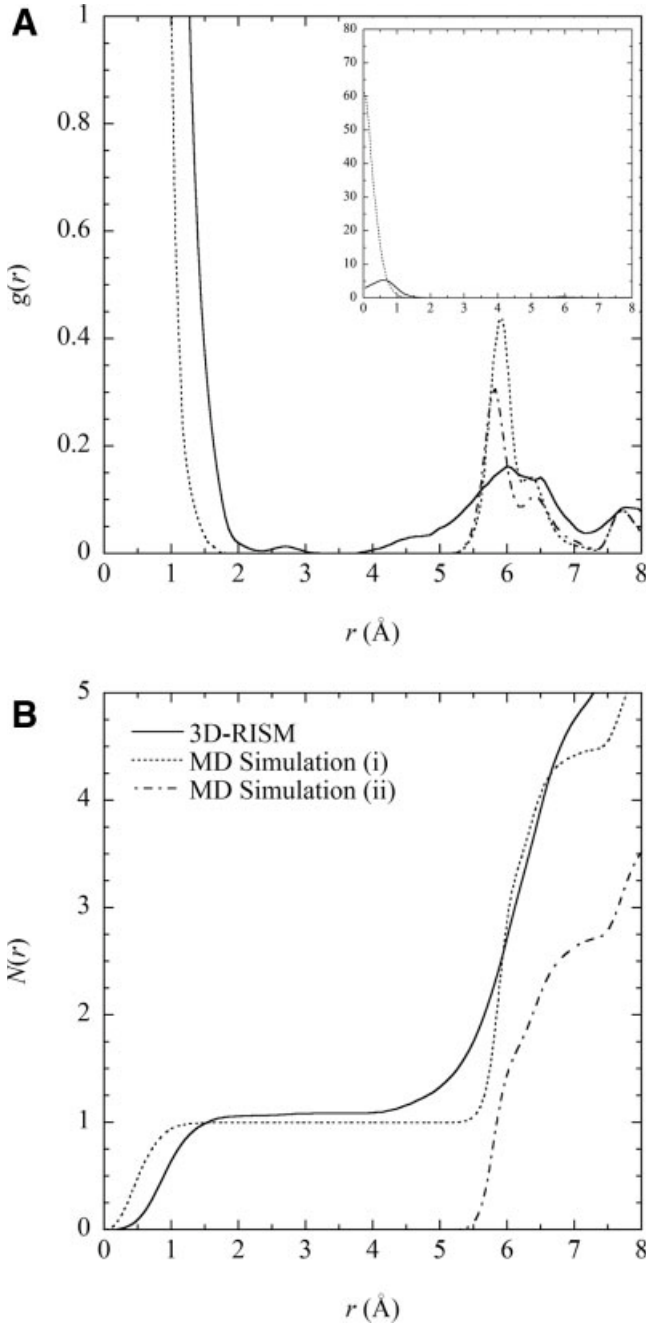

Fig. 6. Radial distribution functions and running coordination numbers of water oxygen viewed from site W6 in Cavity 3. The cavity is within $r = 2.3$ Å. The line nomenclature is the same as in Figure 5.

because there is a contradiction between X-ray analyses, as shown in Table I: some X-ray structures indicate that the cavity holds a water molecule, but the others do not. From the 3D-RISM calculation, it was found that there is some distribution in the cavity, as shown in Figure 2(c). The hydration number within the cavity is 1.1, which means that essentially one water molecule is included in the cavity. The hydration number was obtained from the running coordination number at $r = 2.3$ Å, which is the distance from W6 site. Since the X-ray data 1HEL does not include the water, we used the average coordinate of the water molecule in Simulation (i) as W6 site coordinate to produce the RDF and the running coordination number. The functions are given in Figure 6. It has thus been clarified by the 3D-RISM theory that Cavity 3 actually holds one water molecule.

The MD simulations give trivial results again and are helpless to determine the hydration number in the cavity. As can be seem from Figure 5(b), the hydration number is 1 and 0 from Simulation (i) and (ii), respectively. This is because Cavity 3 is completely isolated from the bulk phase on the present timescale of simulation and a water molecule can neither enter nor leave the cavity.

DISCUSSION

It has been demonstrated in this paper that the 3D-RISM theory can successfully detect water molecules confined in protein cavities. The result is in excellent agreement with the X-ray structure, as well as a particular molecular simulation in which the required number of water molecules were initially included in the cavities in order to provide the results consistent with the X-ray results. Furthermore, it is shown that the theory has an outstanding ability to determine the actual water site in a cavity, which is not completely determined by the X-ray analysis and molecular simulations. The present results strongly suggest that the 3D-RISM theory is a promising method to locate water molecules in protein cavities.

On the other hand, it has been shown that the results of MD simulation depend substantially on the initial coordinates of water molecules. The results cast a serious question on ordinary molecular simulation, concerning its predictive capability for the structure and number of water molecules in a protein cavity. Whether one finds water molecules in a cavity or not depends crucially on the length of the simulation. So, unless one knows in advance that there are water molecules in the cavity, one can never be sure for the prospective as well as present MD simulation techniques if the simulation is converged for the structure and hydration number of water molecules in the cavity. This is a serious drawback of molecular simulations, because the number as well as the position of water molecules and ions in a protein cavity have not always been identified by the experimental methods. A typical example of such a case was Cavity 3, in which no water molecules are detected by most of the X-ray analyses. It should be noted that the 3D-RISM theory does not require such a priori information to predict the location of water molecules.

The 3D-RISM theory explains why some X-ray analyses failed in detecting the water molecule in Cavity 3. The 3D distribution function and the RDF [Figs. 2(c) and 6(a)] show a relatively wide distribution of water in the cavity, compared to those of Cavity 2 [Figs. 2(b) and 5(a)]. This result indicates that the water molecule highly fluctuates within the cavity. Such a water molecule is difficult to be detected by the X-ray diffraction. This is also supported by the fact that some available B-factors of the water in Cavity 3 are much larger than those in the other cavities, as given in Table I.

The 3D-RISM theory can provide information not only about the existence and positions of water molecules but also about their affinity to the sites. As described earlier, the theory can provide the hydration number with decimal fraction. The decimal hydration number is allowed in the following meaning of statistical mechanics. Suppose that the hydration number was 0.8, for example, which is the case of Cavity 2. The hydration number 0.8 means that the water molecule can enter and leave the cavity through large conformational fluctuation beyond the native state in a longer timescale, and it stays there 80% of the time during which the protein takes the native conformation, even if the cavity is completely isolated from the bulk aqueous phase in the conformation. In another view point, 80% of the protein molecules in the ensemble hold one water molecule and the other 20% do not. Thus, the calculated hydration number itself indicates the affinity. In this sense, Cavity 3, for which the hydration number is found to be 1.1, can hold two water molecules with a low probability of 10%.

Although the hydration number obtained by the 3D-RISM theory is excellent, the RDFs are slightly different from the MD results: the peak height is generally lower and the width is broader. This is because the RDFs from the 3D-RISM theory are reconstructed by orientationally averaging the 3D distribution function with a spatial grid resolution of 0.5 Å. Adopting finer grids would improve the fine structure of the RDF as well as the 3D distribution function, though it would little affect the hydration number. However, the grid resolution of 0.5 Å was the best balance between computational cost and accuracy. The present precision is enough to appropriately describe and discuss the hydration structure and hydration numbers, to be associated with the protein structure and functions.

In this study, we performed MD simulation with the protein structure fixed at the X-ray structure, so as to make a direct comparison with the 3D-RISM result. One may expect that taking the conformational fluctuation of the protein into account would improve the ability of molecular simulation in detecting water molecules in protein cavities. That was partially true based on an additional simulation in which the constraints on the atomic coordinates of the protein were released. In fact, it became easier for a water molecule to penetrate Cavity 2 in the simulation, starting from the initial coordinates of Simulation (ii) in which no water is put in the cavities and allowing conformational fluctuation of the protein. However, Cavity 3 still did not receive any water molecule. For Cavity 1, the situation became worse after releasing the constraints. The protein structure was somewhat modified by removing the water molecules in the cavity. As a result, only two water molecules were able to pass into the cavity. This result suggests that these water molecules are essential to maintain the protein structure and their presence is necessary as the initial conditions for molecular simulation of at least nanosecond order.

CONCLUSIONS

The present study has proven that the 3D-RISM theory has an unusual ability to locate water molecules

in protein cavities, even though other methods fail to detect them. On some generalization, the theory is actually useful not only for this particular task but also for a number of issues in biochemistry and biophysics, including recognition of a drug molecule by a receptor protein, accommodation of ions by an ion channel, and enzymatic reaction. We can replace water solvent considered in the present calculation by aqueous solution of ligand molecules, such as drugs, ions, or substrates. Such extensions are straightforward and well-established in this theory, which will produce 3D distributions of cosolvent (ligand) and water molecules in protein cavities as well as surfaces. Analyzing them, we may find peaks of the ligand or water in a protein site, depending on the ratio of their affinities to the site. A consequence is nothing but molecular recognition by protein. If we analyze the dependence of the peak intensity or the binding number on the ligand concentration, we obtain the dissociation constant K_d of the ligand to the protein sites, which is the most fundamental parameter in considering the ligand–protein affinity in an experiment. This approach is much better than the conventional molecular modeling methods involving the "notorious" free energy calculation. A study along this line is in progress in our group.

REFERENCES

1. Teeter MM. Water–protein interactions: theory and experiment. Annu Rev Biophys Biophys Chem 1991;20:577–600.
2. Levitt M, Park BH. Water: now you see it, now you don't. Structure 1993;1:223–226.
3. Karplus PA, Faerman C. Ordered water in macromolecular structure. Curr Opin Struct Biol 1994;4:770–776.
4. Rashin AA, Iofin M, Honig B. Internal cavities and buried waters in globular proteins. Biochemistry 1986;25:3619–3625.
5. Williams MA, Goodfellow JM, Thornton JM. Buried waters and internal cavities in monomeric proteins. Protein Sci 1994;3: 1224–1235.
6. Finer-Moore JS, Kossiakoff AA, Hurley JH, Earnest T, Stroud RM. Solvent structure in crystals of trypsin determined by X-ray and neutron diffraction. Protein Struct Funct Genet 1992;12: 203–222.
7. Nagendra HG, Sukumar N, Vijayan M. Role of water in plasticity, stability, and action of proteins: the crystal structure of lysozyme at very low levels of hydration. Protein Struct Funct Genet 1998;32:229–240.
8. Takano K, Yamagata Y, Yutani K. Buried water molecules contribute to the conformational stability of a protein. Protein Eng 2003;16:5–9.
9. Meyer E. Internal water molecules and H-bonding in biological macromolecules: a review of structural features with functional implications. Protein Sci 1992;1:1543–1562.
10. Zhou Y, Morais-Cabral JH, Kaufman A, MacKinnon R. Chemistry of ion coordination and hydration revealed by a K^+ channel-Fab complex at 2.0 Å resolution. Nature 2001;414:43–48.
11. Tanimoto T, Furutani Y, Kandori H. Structural changes of water in the Schiff base region of bacteriorhodopsin: proposal of a hydration switch model. Biochemistry 2003;42:2300–2306.
12. Lam PYS, Jadhav PK, Eyermann CJ, Hodge CN, Ru Y, Bacheler LT, Meek JL, Otto MJ, Rayner MM, Wong YN, Chang C-H, Weber PC, Jackson DA, Sharpe TR, Erickson-Viitanen S. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. Science 1994;263:380–384.
13. Nakasako M. Water-protein interactions from high-resolution protein crystallography. Phil Trans R Soc Lond B Biol Sci 2004;359:1191–1206.
14. Niimura N, Chatake T, Kurihara K, Maeda M. Hydrogen and hydration in proteins. Cell Biochem Biophys 2004;40:351–370.
15. Otting G, Liepinsh E, Wüthrich K. Protein hydration in aqueous solution. Science 1991;254:974–980.
16. Denisov VP, Halle B. Protein hydration dynamics in aqueous solution. Faraday Discuss 1996;103:227–244.
17. García AE, Hummer G. Water penetration and escape in proteins. Protein Struct Funct Genet 2000;38:261–272.
18. Sterpone F, Ceccarelli M, Marchi M. Dynamics of hydration in hen egg white lysozyme. J Mol Biol 2001;311:409–419.
19. Damjanovic A, García-Moreno B, Lattman EE, García AE. Molecular dynamics study of water penetration in staphylococcal nuclease. Protein Struct Funct Genet 2005;60:433–449.
20. Hansen J-P, McDonald IR. Theory of simple liquids, 2nd ed. London: Academic Press; 1986. 556 pp.
21. Chandler D, Andersen HC. Optimized cluster expansions for classical fluids. II. Theory of molecular liquids J Chem Phys 1972;57:1930–1937.
22. Hirata F. Chemical processes in solution studied by an integral equation theory of molecular liquids. Bull Chem Soc Jpn 1998; 71:1483–1499.
23. Hirata F. Theory of molecular liquids. In: Hirata F, editor. Molecular theory of solvation. Dordrecht, The Netherlands: Kluwer; 2003. pp 1–60.
24. Sato H. Electronic structure and chemical reaction in solution. In: Hirata F, editor. Molecular theory of solvation. Dordrecht, The Netherlands: Kluwer; 2003. pp 61–99.
25. Kinoshita M. Conformational stability of biomolecules in solution. In: Hirata F, editor. Molecular theory of solvation. Dordrecht, The Netherlands: Kluwer; 2003. pp 101–168.
26. Imai T, Kinoshita M, Hirata F. Theoretical study for partial molar volume of amino acids in aqueous solution: implication of ideal fluctuation volume. J Chem Phys 2000;112:9469–9478.
27. Harano Y, Imai T, Kovalenko A, Kinoshita M, Hirata F. Theoretical study for partial molar volume of amino acids and polypeptides by the three-dimensional reference interaction site model. J Chem Phys 2001;114:9506–9511.
28. Kovalenko A. Three-dimensional RISM theory for molecular liquids and solid-liquid interfaces. In: Hirata F, editor. Molecular theory of solvation. Dordrecht, The Netherlands: Kluwer; 2003. pp 169–275.
29. 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.
30. 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.
31. 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.
32. Imai T, Kovalenko A, Hirata F. Solvation thermodynamics of protein studied by the 3D-RISM theory. Chem Phys Lett 2004;395:1–6.
33. Imai T, Kovalenko A, Hirata F. Partial molar volume of proteins studied by the three-dimensional reference interaction site model theory. J Phys Chem B 2005;109:6658–6665.
34. 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.
35. Perkyns JS, Pettitt BM. A dielectrically consistent interaction site theory for solvent–electrolyte mixtures. Chem Phys Lett 1992;190:626–630.
36. Perkyns J, Pettitt BM. A site–site theory for finite concentration saline solutions. J Chem Phys 1992;97:7656–7666.
37. Wilson KP, Malcolm BA, Matthews BM. Structural and thermodynamic analysis of compensating mutations within the core of chicken egg white lysozyme. J Biol Chem 1992;267:10842–10849.
38. Wang J, Cieplak P, Kollman PA. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J Comput Chem 2000;21:1049–1074.

39. Berendsen HJC, Grigera JR, Straatsma TP. The missing term in effective pair potentials. J Phys Chem 1987;91:6269–6271.
40. Pettitt BM, Rossky PJ. Integral equation predictions of liquid state structure for waterlike intermolecular potentials. J Chem Phys 1983;77:1451–1457.
41. 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.
42. Edén M, Levitt MH. Computation of orientational averages in solid-state NMR by Gaussian spherical quadrature. J Magn Reson 1998;132:220–239.
43. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graphics 1996;14:33–38.
44. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: a message-passing parallel molecular dynamics implementation. Comput Phys Commun 1995;91:43–56.
45. Lindahl E, Hess B, van der Spoel D. Gromacs 3.0: a package for molecular simulation and trajectory analysis. J Mol Model 2001;7:306–317.