

A New Approach for Investigating the Molecular Recognition of Protein: Toward Structure-Based Drug Design Based on the 3D-RISM Theory

Yasuomi Kiyota,[†] Norio Yoshida,^{†,‡} and Fumio Hirata^{*,†,‡}

[†]Department of Theoretical Molecular Science, Institute for Molecular Science, Okazaki 444-8585, Japan

[‡]Department of Functional Molecular Science, The Graduate University for Advanced Studies, Okazaki 444-8585, Japan

ABSTRACT: A new approach to investigate a molecular recognition process of protein is presented based on the three-dimensional reference interaction site model (3D-RISM) theory, a statistical mechanics theory of molecular liquids. Numerical procedure for solving the conventional 3D-RISM equation consists of two steps. In step 1, we solve ordinary RISM (or 1D-RISM) equations for a solvent mixture including target ligands in order to obtain the density pair correlation functions (PCF) among molecules in the solution. Then, we solve the 3D-RISM equation for a solute–solvent system to find three-dimensional density distribution functions (3D-DDF) of solvent species around a protein, using PCF obtained in the first step. A key to the success of the method was to regard a target ligand as one of “solvent” species. However, the success is limited due to a difficulty of solving the 1D-RISM equation for a solvent mixture, including large ligand molecules. In the present paper, we propose a method which eases the limitation concerning solute size in the conventional method. In this approach, we solve a solute–solute 3D-RISM equations for a protein–ligand system in which both proteins and ligands are regarded as “solutes” at infinite dilution. The 3D- and 1D-RISM equations are solved for protein–solvent and ligand–solvent systems, respectively, in order to obtain the 3D- and 1D-DDF of solvent around the solutes, which are required for solving the solute–solute 3D-RISM equation. The method is applied to two practical and noteworthy examples concerning pharmaceutical design. One is an odorant binding protein in the *Drosophila melanogaster*, which binds an ethanol molecule. The other is phospholipase A2, which is known as a receptor of acetylsalicylic acid or aspirin. The result indicates that the method successfully reproduces the binding mode of the ligand molecules in the binding sites measured by the experiments.

1. INTRODUCTION

The molecular recognition (MR) in living systems is a crucial elementary process for biomolecules to perform their functions as, for example, enzymes or ion channels. The MR process can be defined as a molecular process in which one or few guest molecules are bound in high probability at a particular site, a cleft or a cavity, of a host molecule in a particular orientation. The process is governed essentially by the two physicochemical properties: (1) difference in the thermodynamic stability (or free energy) between the bound and unbound states of host and guest molecules, and (2) structural fluctuation of molecules. In this article, we propose a new approach to describe the molecular recognition process based on the statistical mechanics of molecular liquids.

In the last three decades, many computational methodologies for investigating a MR process have been proposed.^{1–18} As is mentioned above, the focus of the MR in silico is concerned with the prediction of ligands or drugs that would be strongly bound to key regions of a receptor or an enzyme. The popular “docking simulation” for drug design uses essentially a trial and error scheme to find a “best-fit complex” of host and guest molecules based on geometrical and/or energetic criteria.^{3,4} However, the best-fit complex in a geometrical sense is not necessarily the most stable one in terms of the thermodynamics because it cannot account for the solvent; so neither the dehydration penalty nor the entropy barrier is taken into account. By the “dehydration

penalty,” we mean a free energy penalty concerning a molecular process in which a water molecule detaches from a binding site.

The so-called implicit solvent models, the generalized Born⁵ and the Poisson–Boltzmann equations,⁶ which have been used most popularly for evaluating the solvation thermodynamics of biomolecules, are not accurate and insightful for this problem under concern, because by definition they do not have a molecular view for solvent. It is impossible to define a dielectric constant of solvent inside a host cavity, and therefore, it cannot account for the dehydration penalty, especially that from the host cavity. At best, those quantities can be calculated by fitting the empirical parameters, such as the boundary conditions and the dielectric constants, with experimental data, but then it loses credibility as a theory to predict the phenomena.

The molecular simulation, on the other hand, can provide the most detailed molecular view for the process. The simulation methods, molecular dynamics (MD) and Monte Carlo (MC), sample the configuration space of a ligand–receptor system in solvent using the numerical integration of the Newtonian equation (MD) or the probabilistic search along the Markov path (MC) in order to evaluate the free energy difference between the bound and unbound states of the host–guest system. However, this type of simulation does not work for the

Received: May 29, 2011

Published: September 15, 2011

problem well, because a MR process is usually slow as well as rare events. A common strategy adopted by the simulation community to overcome the difficulty is a non-Boltzmann-type sampling which defines a “reaction coordinate” or an “order parameter” onto which all other degrees of freedoms are projected. The best example is the “umbrella” sampling to realize the potential of mean force or the free energy along a conduction path of an ion in an ion channel.⁷ The method is quite powerful for sampling the configuration space around an order parameter if the parameter is unique and if the configuration space to be projected on the parameter is sufficiently small. Unfortunately, the problems in the biochemical processes are not so simple as can be described by a unique order parameter. So, it is often the case that the results of the simulation depend on choices of order parameters and on “scheduling” of the sampling. The other methodology employed to accelerate the sampling is to apply an artificial external force on the system. That kind of simulation should verify that the configuration of water satisfies the Boltzmann distribution.

A different approach to the MR process has been developed based on the three-dimensional reference interaction site model (3D-RISM) theory during the last five years.^{8–18} The method integrates *analytically* the configuration space of a ligand–receptor system in solvent by means of the statistical mechanics. The analytical integration which extends over the infinitely large configuration space is the advantage which distinguishes the method from the molecular simulation. Due to this advantage, the 3D-RISM theory is free from the difficulties which the simulation methods are facing.

The 3D-RISM equation was derived from the molecular Ornstein–Zernike (MOZ) equation, the most fundamental equation to describe the density pair correlation of liquids, for a solute–solvent system in the infinite dilution by taking a statistical average over the orientation of solvent molecules.^{19–21,35–37} By solving the combined 3D-RISM with RISM equations, the latter providing the solvent structure in terms of the site–site density pair correlation functions, one can get the “solvation structure” or the solvent distributions around a solute. The high peak of the solvent distributions indicates that the solvent affinity of target protein or receptor at that point is high. Therefore, the MR process can be probed by the solvent or ligand distribution. The method produces naturally all the solvation thermodynamics as well, including energy, entropy, free energy, and their derivatives, such as the partial molar volume and compressibility. Unlike the molecular simulation, there is no necessity for concern about size of the system and “sampling” of the configuration space, because the method treats essentially the infinite number of molecules and integrates over the entire configuration space of a system.²²

By the way, in all previous studies of MR by 3D-RISM theory, the receptor protein and ligand molecule were regarded as solute and solvent, respectively. In those cases, the MR process is analyzed in terms of solvent distribution around a solute molecule, which is called a sol’u’té–sol’v’ent density distribution function (uv-DDF). The RISM equation for solvent system should be solved before a 3D-RISM calculation. There, “solvent” consists of ligand molecules, water, and other components of solution. Because the RISM and 3D-RISM equations coupled with closure relation (i.e., hypernetted chain or Kovalenko–Hirata (KH) closure) are nonlinear integral equations, those are solved in an iterative manner.^{23,24}

Although many theoretical and methodological efforts have been devoted to solve the RISM and 3D-RISM equations, the

system has been largely limited to that including relatively small solvent molecules, such as water, ions, carbon monoxide, and the largest being glycerol.^{14,25,26} A reason why is because numerical solution of the RISM equation for solution including large ligands becomes increasingly unstable as the size of ligands increases. However, many ligands of biological interests, including ordinary drug molecules, are not so small. Therefore, we propose a new approach to tackle MR of large ligand molecules by protein based on the 3D-RISM and RISM theories. The strategy of the method is to regard a ligand molecule as a solute in addition to a receptor protein, which are immersed in solvent in the infinite dilution. The distribution of ligand molecules around a receptor protein is described by the sol’u’té–sol’u’té density distribution function (uu-DDF), instead of uv-DDF. In this sense, the new method is named as “uu-3D-RISM.” Under the treatment of this method, interactions between ligand molecules can be ignored completely from the consideration, because the density of ligand molecule is vanishingly small at the limit. Therefore, it is not necessary to solve the ligand–ligand RISM equation, the most unstable equation, anymore. This assumption stabilizes the numerical solutions of a set of the 3D-RISM and RISM equations dramatically.

An approach to uu-DDF has already been proposed by Kovalenko and Hirata to investigate the potential of mean force between two molecular ions in a polar molecular solvent.²¹ In their method, uu-DDF is a function of position and orientation of two solute molecules. Therefore, in order to obtain the uu-DDF for all possible positions and orientations of a ligand molecule, the method requires sampling in the entire coordinate space. On the contrary, the uu-DDF can be obtained by a single calculation in the present method, because one of the solute molecules, usually a ligand molecule, is treated in terms of an interaction-site model.

This paper is organized as follows. In Section 2, we briefly review the RISM and 3D-RISM theories in order to identify each member of the RISM family and derive the uu-3D-RISM equation. We also clarify how the molecular recognition process of biological system is treated with the uu-3D-RISM method. Section 3 is devoted to applications of the uu-3D-RISM method to two practical and noteworthy examples, odorant binding protein²⁷ and phospholipase A2.^{28–31} Section 4 concludes the paper.

2. METHOD

2.1. Outline of 3D-RISM Theory. The MOZ integral equation for a multicomponent system is written as

$$h^{ij}(1,2) = c^{ij}(1,2) + \sum_l \int c^{ij}(1,3)\rho^l h^{lj}(3,2)d(3) \quad (1)$$

where ρ^l is the number density of species l , $h^{ij}(1,2)$ and $c^{ij}(1,2)$ denote the total and direct correlation functions between a pair of molecular species i and j in a solution, respectively.³² The numbers in the parentheses represent the coordinates of molecules in the liquid system, including both the position \mathbf{r} and the orientation Ω . The total correlation function $h^{ij}(1,2)$ is related to the density pair correlation functions $g^{ij}(1,2)$ by $h^{ij}(1,2) = g^{ij}(1,2) - 1$. The summation in the right-hand side runs over species in a mixture.

The eq 1 depends essentially on six coordinates in the Cartesian space, and they include a six-fold integral. This integral is the one which had prevented the theory from applying to

Polyatomic molecules. It is the interaction-site model and the RISM approximation proposed by Chandler and Andersen³³ that enabled one to solve the equations. The idea behind the model is to project the functions onto the one-dimensional space along the distance between a pair of interaction sites, usually placed on the center of atoms, by taking the statistical average over the angular coordinates of molecules, fixing the separation between two interaction sites. The projection can be accomplished by the following equation:

$$h_{\alpha\gamma}(r) = \frac{1}{\Omega^2} \int \delta(|\mathbf{r}_1 + \mathbf{l}_1^\alpha|) \delta(|\mathbf{r}_2 + \mathbf{l}_2^\gamma| - r) h(1, 2) d(1) d(2) \quad (2)$$

where \mathbf{r}_1 and \mathbf{l}_1^α indicate the position of molecule 1 in laboratory frame and the position of site α of molecule 1 in molecular frame, respectively.

Now, we classify molecular species in the system into two categories, "solute" and "solvent," respectively, as previous works.^{19–21} After this, the superscripts "u" and "v" denote solute and solvent, respectively. For example, $h^{vv'}$ is the total correlation function between different molecular species v and v' in solvent, and ρ^v and ρ^u also are elements of diagonal matrices which denote density of each species in solvent and solute. The summations concerning v and v' in the equations run over solvent species, while u and u' run over solute species. The most interesting case to investigate "solvation" of a biomolecule can be realized by taking the infinite dilution limit for all the solute species, namely, $\rho^u \rightarrow 0$. Then, eqs 1 and 2 are constructed from solvent–solvent, solute–solvent, and solute–solute systems, and you note that these equations can be solved sequentially, because the former equation is independent from the later.

The RISM equation can be derived from eqs 1 and 2 with a super position approximation for the direct correlation function, which reads

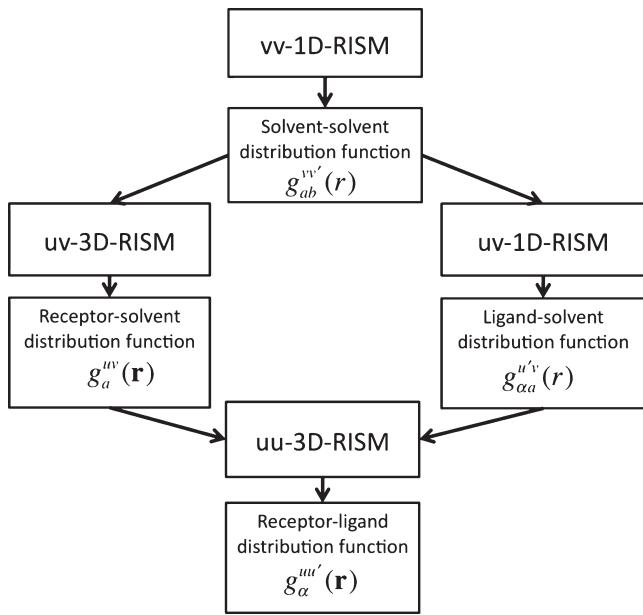
$$h_{\alpha\lambda}^{vv'}(r) = \sum_{\substack{\eta \in v' \\ \gamma \in v}} \omega_{\alpha\gamma}^v * c_{\gamma\eta}^{vv'} * \omega_{\eta\lambda}^{v'}(r) + \sum_{\substack{\nu'' \\ \eta \in v'' \\ \gamma \in v}} \rho^{\nu''} \sum_{\substack{\eta \in v'' \\ \gamma \in v}} \omega_{\alpha\gamma}^v * c_{\gamma\eta}^{vv''} * h_{\eta\lambda}^{v''v'}(r) \quad (3)$$

where ω is an intramolecular correlation function, the asterisk denotes the convolution integrals, and ρ^v denotes the number density of solvent species v. For clarity, we refer to eq 3 as vv-1D-RISM equation hereafter. A similar equation can be derived from eqs 1 and 2 for a solute–solvent system in the infinite dilution limit as follows:

$$h_{\alpha\lambda}^{uv}(r) = \sum_{\substack{\eta \in v \\ \gamma \in u}} \omega_{\alpha\gamma}^u * c_{\gamma\eta}^{uv} * \omega_{\eta\lambda}^v(r) + \sum_{\substack{\nu' \\ \eta \in v' \\ \gamma \in v}} \rho^{\nu'} \sum_{\substack{\eta \in v' \\ \gamma \in v}} \omega_{\alpha\gamma}^u * c_{\gamma\eta}^{uv'} * h_{\eta\lambda}^{v'v}(r) \quad (4)$$

The theory has been proven to be so successful for describing structure and thermodynamics of liquid and liquid mixtures, including a variety of aqueous solutions.²⁰ However, the theory has exhibited serious breakdown, especially when it was applied to solutions including macromolecules, such as protein as a solute.³⁴ Then, in order to avoid the problem, alternative

Scheme 1. Scheme of uu-RISM Method



approaches have been developed during two decades.^{19,35} We have derived from eq 1 by taking an average over angular coordinates only for solvent coordinates, not for solute coordinates.^{36,37}

$$h_a^{uv}(r) = \frac{1}{\Omega} \int \delta(\mathbf{r}_2 + \mathbf{l}_2^\alpha - \mathbf{r}) h^{uv}(1, 2) d(2) = \sum_{v' \in \text{solvent}} \sum_{\gamma \in v'} c_a^{uv'} * [\omega_{\gamma\alpha}^{v'} + \rho^{v'} h_{\gamma\alpha}^{v'v}](\mathbf{r}) \quad (5)$$

where v' runs over the all solvent species. Where we also employed a super position approximation for the solute–solvent direct correlation function:

$$c^{uv}(1, 2) = \sum_{\alpha \in v} c_\alpha^{uv}(\mathbf{r}) \quad (6)$$

This is the basic assumption of the 3D-RISM theory. The solvent–solvent (vv) total correlation function appeared in the right-hand side of eq 5 is evaluated from the vv-1D-RISM equation, or eq 4, in advance. The $\rho h(\mathbf{r})$ is essentially the "second moment of the density fluctuation" of two spatial points, which can be identified as a "mean excess density" due to the method devised by Percus.³⁸

The MR phenomena can be described by solving the vv-1D-RISM and 3D-RISM equations sequentially, considering that a receptor protein is immersed in solvent–ligand mixture in the infinite dilution limit of the receptor. In other words, ligand molecules are treated as one of components of a solvent mixture. MR is realized in terms of the $\rho h(\mathbf{r})$ of ligand atoms at a binding site, relative to bulk solutions; if $\rho h(\mathbf{r})$ is greater than zero, then we conclude that the ligand is "recognized" by the site. So, the procedure of realizing MR by 3D-RISM is quite straightforward. There is no necessity to define order parameters, such as "reaction coordinates" and "umbrella", for exploring the configuration space of ligands, which is the case in the molecular simulations.

Table 1. Summary of Performed Calculations^a

species	odorant binding protein		phospholipase A2		
	uu-RISM	uv-RISM	uu-RISM	uv-RISM	
protein	1OOF	3D representation (solute)	3D representation (solute)	1OXR	3D representation (solute)
ligand	ethanol	site representation (solute)	site representation (solvent)	aspirin	site representation (solute)
solvent	water	site representation (solvent)	site representation (solvent)	water	site representation (solvent)

^a Note that the type of ligand is different between uu- and uv-3D-RISM.

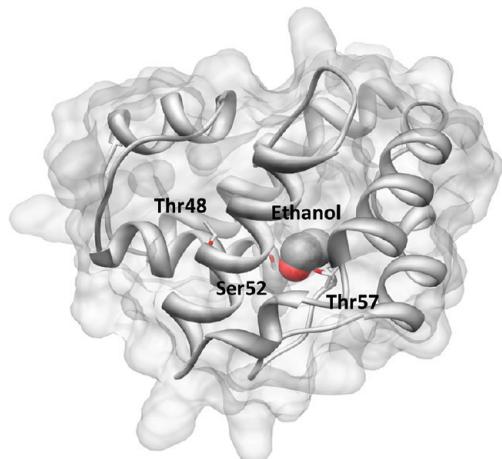


Figure 1. X-ray crystal structure of the odorant binding protein LUSH from *Drosophila melanogaster*. It has a specific alcohol binding site which can bind a series of short-chain n-alcohols. The structure taken from PDB (PDB ID: 1OOF) includes one ethanol molecule which is represented by the VDW surface. The protein surfaces are represented as a gray transparent surface. The binding site is constructed by a group of amino acids, Thr57, Ser52, and Thr48. These amino acids form a network of concerted hydrogen bonds between the protein and the alcohol.

It is the inclusion of ligand species in solvent that gave 3D-RISM/vv-1D-RISM a great advantage. In fact, many applications of 3D-RISM to MR processes so far have been so successful as long as small ligands including water, CO, NH₃, and metal ions, etc. are concerned. However, the advantage turns into disadvantage when large molecules including most of the drug compounds are involved. The problem originates in the vv-1D-RISM equation, not in the 3D-RISM equation. To describe a MR process with 3D-RISM, we have to solve the vv-1D-RISM equation for solvent mixture, including ligand species in advance. However, according to our experience, numerical solutions of the vv-1D-RISM equation for a mixture including a large compound are quite unstable due to inherent nonlinearity of the integral equation, which increases with increasing size and complexity of molecules.

In the following subsection, we develop a new approach which is derived from the solute–solute MOZ equation in which both receptor and ligand molecules are dissolved in a solvent mixture at the infinite dilution.

2.2. uu-3D-RISM Equation. The strategy of the new approach is to regard a ligand molecule as a solute molecule, in addition to a receptor protein, which is immersed in solvent in the infinite dilution limit. By this assumption, eq 1 can be regarded as a protein–ligand uu-MOZ equation. In the present approach,

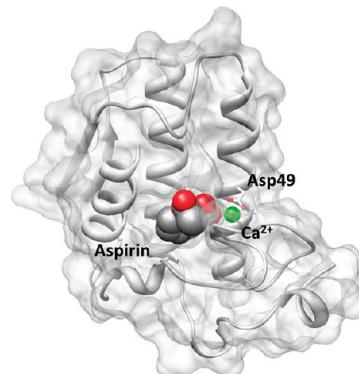


Figure 2. X-ray crystal structure of the complex formed between phospholipase A2 (PLA2) and 2-acetoxybenzoic acid, aspirin. It can be taken from PDB (PDB ID: 1OXR). Phospholipase A2 can bind aspirin, which is represented by VDW surface, for anti-inflammatory effects in its specific binding site. The protein surfaces are represented as a gray transparent surface. The aromatic ring of aspirin is embedded in the hydrophobic environment, and other substituted groups form several important attractive interactions with calcium ion, His48, and Asp49. Calcium ion is shown as a green sphere.



Figure 3. Structure aspirin depicted with two different presentations.

since a ligand molecule can be assumed to be reasonably small, we apply the interaction site model to a ligand molecule in a manner similar to solvent. The uu-3D-RISM can be derived from eq 1 by taking an average over angular coordinates for ligand(solute) coordinates, not for protein(solute) coordinate. The solute–solute total correlation functions can be written as

$$\tilde{h}_\alpha^{\text{uu}'}(\mathbf{r}) = \frac{1}{\Omega} \int \delta(\mathbf{r}_2 + \mathbf{l}_2^\alpha - \mathbf{r}) h^{\text{uu}'}(1, 2) d(2) \quad (7)$$

Accordingly, the Fourier transform of total correlation functions is obtained by orientational reduction of Fourier transform of the molecular total correlation functions:

$$\tilde{h}_\alpha^{\text{uu}'}(\mathbf{k}) = \frac{1}{\Omega} \int d\Omega_2 e^{i\mathbf{k}\cdot\mathbf{r}_2} \tilde{h}^{\text{uu}'}(k, \Omega_1, \Omega_2) \quad (8)$$

We employ a super position approximation for the solute–solute direct correlation function:

$$c^{\text{uu}'}(1, 2) \equiv \sum_{\alpha \in v} c_\alpha^{\text{uu}'}(\mathbf{r}) \quad (9)$$

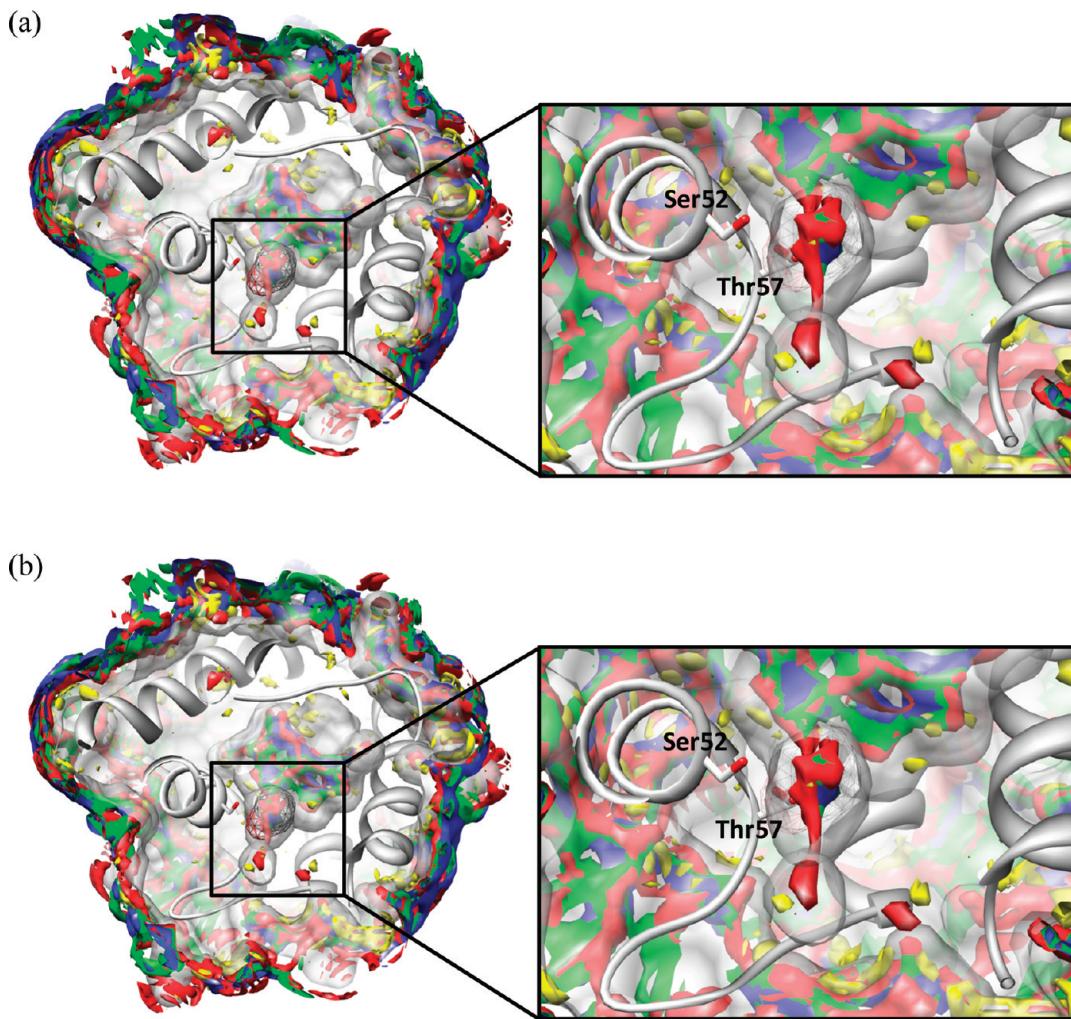


Figure 4. The 3D-DDF of ethanol around and inside odorant binding protein, LUSH, obtained by (a) uu-3D- and (b) uv-3D-RISM with the threshold $g_\gamma(\mathbf{r}) > 2$: blue, CH_3 ; green, CH_2 ; red, oxygen atom of hydroxyl group; yellow, hydrogen atom of hydroxyl group. The protein surfaces are represented as a gray transparent surface. The location of ethanol in X-ray structure is depicted with a wire frame.

The Fourier transform of solute–solute direct correlation functions can be defined as

$$\begin{aligned} \tilde{c}^{\text{uu}'}(k, \Omega_1, \Omega_2) &= \int d\mathbf{r}_{12} e^{ik\mathbf{r}_{12}} c^{\text{uu}'}(r_{12}, \Omega_1, \Omega_2) \\ &= \sum_{\alpha} e^{-ik\mathbf{r}_{\alpha 2}} c_{\alpha}^{\text{uu}'}(\mathbf{k}) \end{aligned} \quad (10)$$

The solute–solute MOZ eq 1 in Fourier space can be written as

$$\begin{aligned} \tilde{h}^{\text{uu}'}(k, \Omega_1, \Omega_2) &= \int d\mathbf{r}_{12} e^{ik\mathbf{r}_{12}} h^{\text{uu}'}(r_{12}, \Omega_1, \Omega_2) \\ &= \int d\mathbf{r}_{12} e^{ik\mathbf{r}_{12}} c^{\text{uu}'}(r_{12}, \Omega_1, \Omega_2) \\ &+ \frac{1}{\Omega} \sum_{\nu} \int d\mathbf{r}_{12} \int d(3) e^{ik\mathbf{r}_{12}} c^{\text{uv}}(r_{12}, \Omega_1, \Omega_2) \rho^{\nu} h^{\text{vu}'}(r_{12}, \Omega_1, \Omega_2) \end{aligned} \quad (11)$$

From eqs 2 and 8–11, the uu-3D-RISM equation is obtained as

$$h_{\alpha}^{\text{uu}'}(\mathbf{r}) = \sum_{\gamma} c_{\gamma}^{\text{uu}'} * \omega_{\gamma\alpha}^{\text{u}'}(\mathbf{r}) + \sum_{\nu} \sum_{\gamma} c_{\gamma}^{\text{uv}} * \rho^{\nu} h_{\gamma\alpha}^{\text{vu}'}(\mathbf{r}) \quad (12)$$

In order to solve those 1D- and 3D-RISM equations obtained above, we need another equation which complements or “closes” the equations. Here, we employ the KH closure which reads³⁹

$$\begin{aligned} g_{\alpha}^{\text{uu}'}(\mathbf{r}) &= \begin{cases} \exp(d_{\alpha}^{\text{uu}'}(\mathbf{r})) & \text{for } d_{\alpha}^{\text{uu}'}(\mathbf{r}) \leq 0 \\ 1 + d_{\alpha}^{\text{uu}'}(\mathbf{r}) & \text{for } d_{\alpha}^{\text{uu}'}(\mathbf{r}) > 0 \end{cases} \\ d_{\alpha}^{\text{uu}'}(\mathbf{r}) &= -\beta u_{\alpha}^{\text{uu}'}(\mathbf{r}) + h_{\alpha}^{\text{uu}'}(\mathbf{r}) - c_{\alpha}^{\text{uu}'}(\mathbf{r}) \end{aligned} \quad (13)$$

The procedure to obtain the receptor–ligand distribution function is shown in Scheme 1. First, vv-DDF is evaluated by vv-1D-RISM, where solvent includes water, electrolyte, organic solvent, and so on. The vv-DDF is used in both uv-3D- and in uv-1D-RISM calculations. The uv-3D- and uv-1D-RISM calculations are carried out to obtain receptor–solvent and ligand–solvent DDF, respectively. By inserting these two DDFs, uu-3D-RISM can be solved to get receptor–ligand DDF.

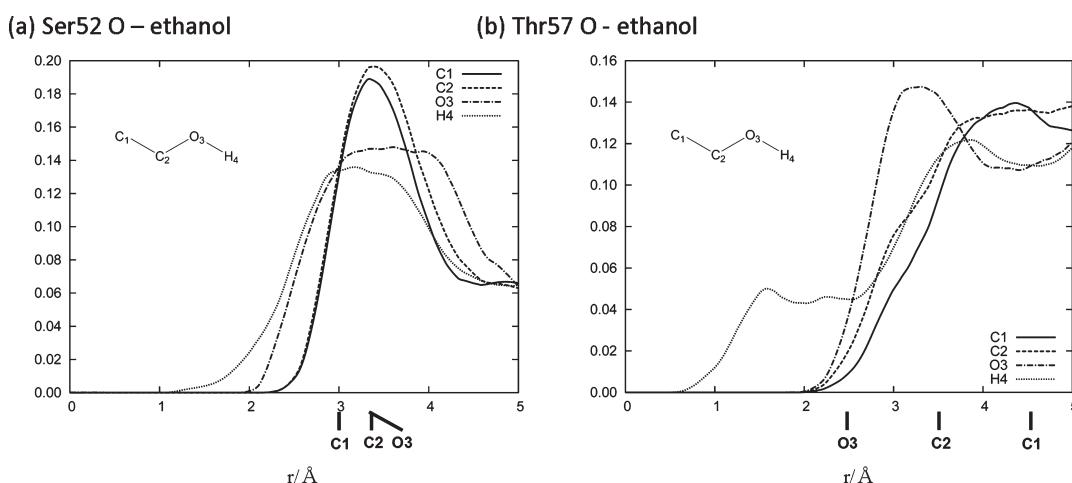


Figure 5. RDFs of ethanol around hydroxyl groups of (a) Ser52 and (b) Thr57, respectively. Oxygen atoms in each hydroxyl groups were chosen as the averaging center.

Table 2. Distance (Å) Matrix between the Specific Sites of LUSH and Ethanol

sites	Ser52-O	Thr57-O
C1	3.0	4.6
C2	3.3	3.5
O3	3.3	2.3

From receptor–ligand DDF, a MR analysis, such as a binding pocket search, can be performed.

2.3. Extracting Binding Mode of Ligand Inside Protein from 3D-DDF. At this point, we would like to make a comment on general difficulty to extract “binding mode,” or position and orientation, of a ligand inside protein from 3D-DDF. The problem is common to the experimental methodologies, such as X-ray and neutron diffractions, since both the theory and the experiments are observing essentially the same property, namely, the density distribution of ligand atoms, which are atomic positions statistically averaged over thermal motion. For the purpose of comparing the results from 3D-DDF with that from the experiment, it will be best if we can do it directly by defining a measure for “distance” between the two distributions. However, such a methodology has not been well developed yet. Moreover, general measure for 3D-DDF is available neither in literature nor in the Protein Data Bank. Instead, the most probable binding mode of a ligand is presented, which of course depends on a way of analyzing the 3D-DDF data.

Considering such a situation, we are developing two methods which provide us information concerning binding mode of ligand inside a protein. One of those methods is based on the radial distribution functions (RDF) of ligand atoms from atoms in amino acid residues of protein. A position of the first peak in RDF corresponds to an average distance between a pair of atoms of a ligand and a protein. By making the analyses of distances among the several atoms in ligand and protein, we can extract the binding mode of a ligand inside of protein. When the size of ligand is small enough, such as carbon monoxide and ethanol, the method works quite well to determine the binding mode.

The other method we are developing to abstract the binding mode of a ligand inside a protein from 3D-DDF is similar

essentially to that used in the analysis of 3D-DDF from the diffraction measurement. The method defines two “score” functions, one corresponding to the position of ligand and the other to its orientation, in terms of 3D-DDF and of trial geometry of a ligand. The level of agreement between a trial geometry and 3D-DDF is ranked according to the score functions.

2.4. Computational Detail. The approach proposed in this article overcomes the difficulty associated with the uv-3D-RISM approach, and it will provide a new tool for the rational drug design. Here, we demonstrate robustness and capability of the uu-3D-RISM theory by applying the approach to two systems which are of great interest in biochemistry and pharmacology. One is an odorant binding protein known as LUSH, and the other is phospholipase A2. Table 1 shows the outline of these works.

2.4.1. Odorant Binding Protein (PDB ID: 1OOF). In order to examine robustness of the new approach, we consider binding of an ethanol molecule to an odorant binding protein (LUSH), see Figure 1. The ligand is small enough to be treated with the uv-3D-RISM, so that one can compare the results from the methods with that from the uu-3D-RISM theory. In the case of uv-3D-RISM, the ligand is regarded as a component of solvent, while in the case of uu-3D-RISM, it is considered as a solute. In both cases, the odorant binding protein is treated as a solute receptor protein.

The Amber-99 parameter set⁴⁰ was employed for the protein, and the general amber force field (GAFF)⁴¹ was employed for the ligand ethanol and for the acetic acids, which is part of receptor protein. TIP3P water⁴² was chosen as solvent at 298 K and 0.9979 g/cm³. The uu-3D- and uv-3D-RISM equations were solved on a grid of 160³ points in a cubic supercell of 80 Å³. The density of ethanol was so chosen that the volume ratio of water to ethanol becomes 99:1%.

2.4.2. Phospholipase A2 (PDB ID: 1OXR). In order to demonstrate the capability of the new approach, we examine MR of aspirin by PLA2 (Figure 2). Since aspirin is a rather large ligand, having 14 specific interaction sites (Figure 3) in the neutral state, it may not be treated readily with the ordinary uv-3D-RISM due to the difficulty stated above. So, this is a good example to demonstrate capability of the new method. The Amber-99 parameter set was employed for the protein, while GAFF was

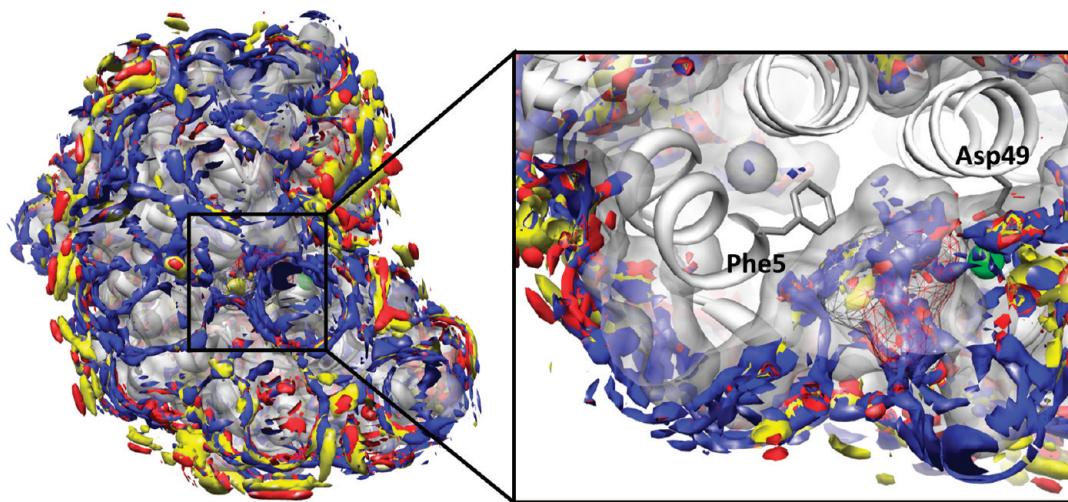


Figure 6. The 3D-DDF of protonated (neutral) aspirin around and inside phospholipase A2, obtained by uu-3D-RISM with the threshold $g_{\gamma}(r) > 2$: red, COOH; yellow, aromatic ring; and blue, OCOCH₃. The protein surfaces are represented as a gray transparent surface. The location of aspirin in X-ray structure is depicted with a wire frame.

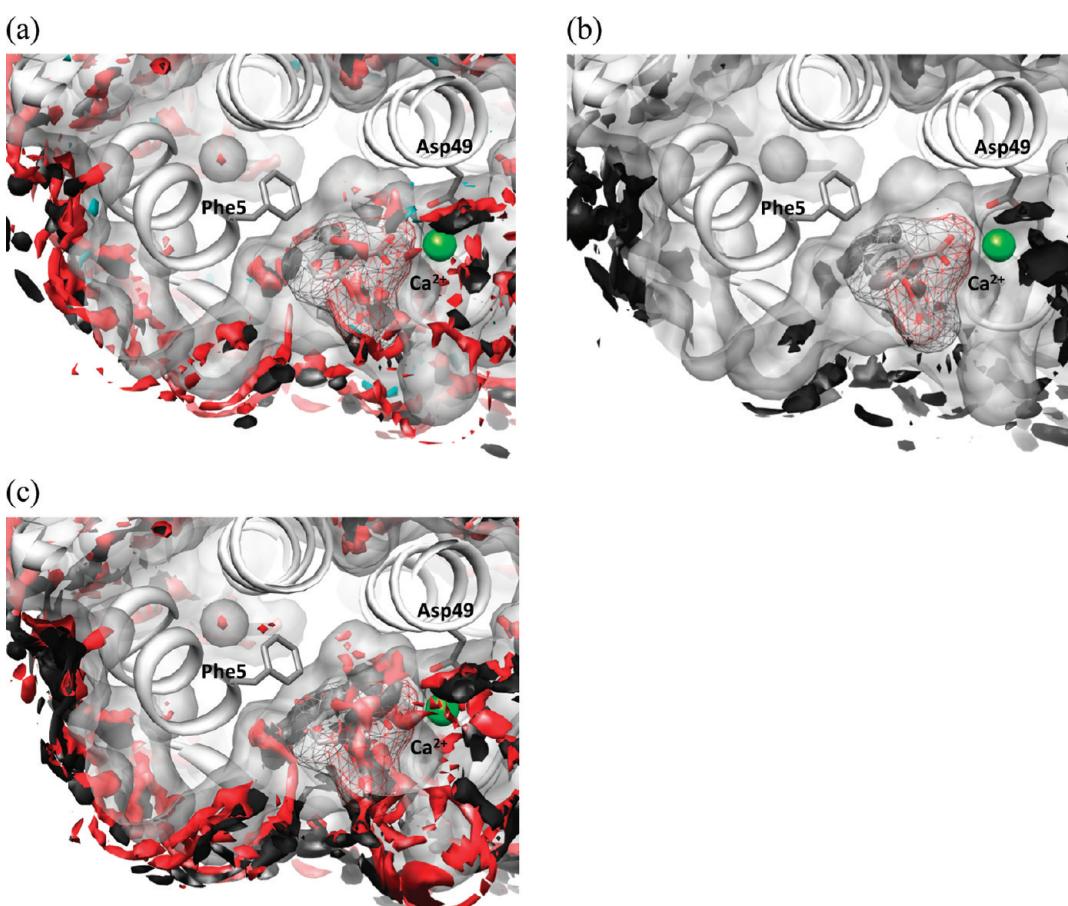


Figure 7. The 3D-DDF of neutral aspirin around and inside phospholipase A2, obtained by uu-3D-RISM with the threshold $g_{\gamma}(r) > 2$. (a) Carboxyl group, COOH; (b) aromatic ring; and (c) acetoxymethyl group, OCOCH₃. The color code is assigned to oxygen (red), carbon (black), and hydrogen (cyan). The protein surfaces are represented as a gray transparent surface.

employed for aspirin with a united-atom modification concerning hydrogen atoms. TIP3P water was chosen as solvent at 298 K

and 0.9979 g/cm³. The 3D-RISM equation was solved on a grid of 200³ points in a cubic supercell of 100 Å³.

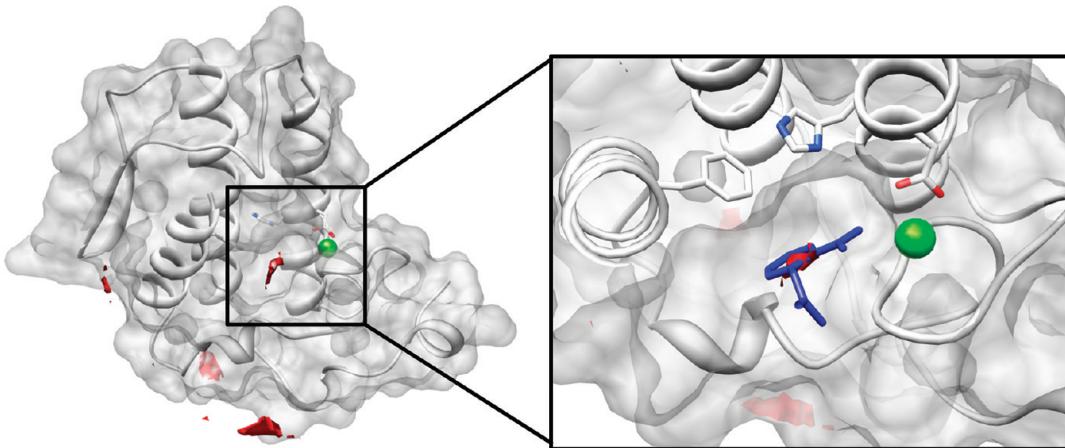


Figure 8. Affinity of an aspirin molecule to binding site in the phospholipase A2 estimated by the function of DC based on local PMF with the threshold $f_{DC}(x) > 1.45$. The protein surfaces are represented as a gray transparent surface. In the top view, the location of aspirin in X-ray structure is depicted with blue sticks.

3. RESULTS AND DISCUSSION

3.1. Odorant binding protein, LUSH. *3.1.1. 3D-Distribution Functions of Ethanol around and inside LUSH.* In Figure 4, the 3D-DDFs of ethanol obtained by uu-3D- and uv-3D-RISM are compared. The 3D-DDFs are depicted by isosurface representation with the threshold $g_\gamma(r) > 2$. This threshold implies that the probability of finding site γ at the position r is twice as large as that in the bulk. The gray surface represents the protein, whereas the blue, green, red, and yellow surfaces depict the distribution of CH_3 , CH_2 , O, and H sites of alcohol, respectively. At a glance, DDF from uu-3D-RISM, which is depicted in Figure 4a, shows good agreement with that from uv-3D-RISM shown in Figure 4b, especially nearby the binding site. Both 3D-DDFs are also in accord with the results from X-ray crystallography.

It is clear from the formulation described in the previous section that the uu-3D-DDF is equivalent to uv-3D-DDF in the low density limit of ligand concentration. The difference of these two 3D-DDFs are measured by the root-mean-square deviation (rmsd) $d_{\text{rmsd}} = (\sum_{i=1}^n (g_\gamma^{\text{uu}}(x_i) - g_\gamma^{\text{uu}}(x_i))^2 / n)^{1/2}$, x_i denotes each grid point); CH_3 site, $d_{\text{max}} = 0.0764$ and $d_{\text{rmsd}} = 0.0055$; CH_2 site, $d_{\text{max}} = 0.0433$ and $d_{\text{rmsd}} = 0.0066$; O site, $d_{\text{max}} = 0.1015$ and $d_{\text{rmsd}} = 0.0099$; and H site, $d_{\text{max}} = 0.1362$ and $d_{\text{rmsd}} = 0.0079$. These differences are small enough for an actual application to systems in which the ligand concentration is low. In reality, the difference is well within a thermal fluctuation of the atomic position of ligands at a binding site in protein. Therefore, the uu-3D-DDF evaluated by uu-3D-RISM can be employed to molecular recognition problem instead of uv-3D-DDF for ligands in a low-concentration region.

3.1.2. Radial Distribution Functions of Ethanol from Hydroxyl Groups in a Binding Site. The RDFs between atoms in a ligand and those belonging to amino acids in protein allow us to investigate binding modes, position and orientation, of ligands inside protein. The RDFs can be obtained by averaging the 3D-DDF over the direction around a specified center:

$$g_a^{1D}(r, \mathbf{r}_0) = \frac{1}{4\pi} \int g_a(\mathbf{r}_0 + \mathbf{r}) d\hat{\mathbf{r}} \quad (14)$$

where $\hat{\mathbf{r}}$ is the direction of \mathbf{r} , and \mathbf{r}_0 indicates a center for averaging. The averaging centers were selected near the binding

Table 3. Top Three Peaks List of Affinity Based on the Overlap between the Ligand and Its Distribution Function

	order	α	β	γ	overlap
(a)	1	10	330	50	2.33×10^5
(b)	2	40	260	40	2.02×10^5
(c)	3	60	220	50	1.83×10^5

sites. In the present case, we choose the oxygen atoms of hydroxyl groups in Ser52 and Thr57 as an averaging center. The RDFs around these sites are shown in Figure 5. For comparing the peak positions of RDFs with those from the X-ray structure, we refer to the distance between these specific sites and each site of an ethanol molecule obtained from the experiment (Table 2). The distances experimentally determined are also marked by bars in the x -axis of the Figure 5a and 5b. Although each peak of RDF is not very sharp reflecting thermal fluctuation of the ligand inside the binding site, the positions of peaks in RDFs are consistent with those deduced from the X-ray crystallography.

The RDFs of hydrogen atoms, which are not treated by the X-ray diffraction measurement, are also depicted in Figure 5. The results may provide an orientation of hydrogen atom of the ligand molecule. In the case of Ser52, any peak indicating such orientation does not appear between two oxygen atoms, while a discernible peak appears at $r \sim 1.6 \text{ \AA}$ in the case of Thr57. It is not clear at this moment whether the peak indicates the existence of a hydrogen bond or not. However, it is clear that the hydroxyl group is oriented toward Thr57, which is also consistent with the result from the X-ray crystallography.

3.2. Phospholipase A2, PLA2. It is not a straightforward task to apply the uv-3D-RISM method to the problem due to the reason described in detail in the previous section: The ligand molecule or aspirin is too large to get a convergent result of the vv-RISM equation for the solvent mixture, including the ligand molecules. Here, we only apply the uu-3D-RISM to the molecular recognition of aspirin to PLA2.

Aspirin, acetylsalicylic acid, is a weak acid in aqueous solution. We employed a neutral state, which was shown in Figure 3, because the affinity of the neutral state to binding site is much higher than a charged state.

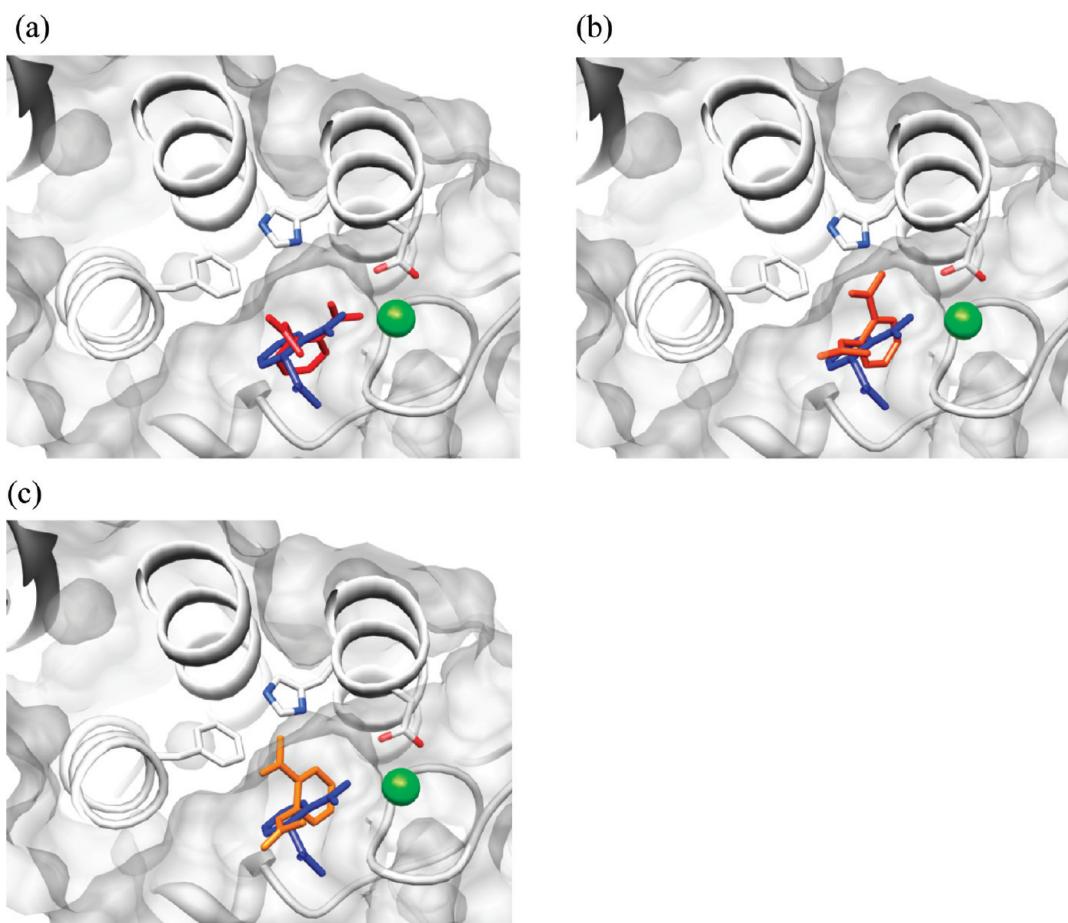


Figure 9. Predicted structures of ligand from the top three peaks of affinity based on the overlap between the ligand and its distribution function (Table 3). The protein surfaces are represented as a gray transparent surface. The location of aspirin in X-ray structure is depicted with blue sticks.

3.2.1. 3D-DDF of Aspirin Around and Inside Phospholipase A2. The 3D-DDF of aspirin at the nonprotonated state is shown in Figure 6 with the threshold $g_\gamma(\mathbf{r}) > 2$. The red, yellow, and blue surfaces are the distributions of carboxyl group, COOH (number of sites is 4), the aromatic ring (number of sites is 6), and the acetoxy group, OCOCH_3 (number of sites is 4), respectively. We can easily observe these distributions not only inside the binding site but also around the protein. Since the distributions shown in Figure 6 are jumbled inside the biding site, the contribution from each site group, COOH, aromatic ring, and OCOCH_3 , is separately depicted in Figure 7. The color code assigned to each atom in the figures is as follows: red, oxygen; black, carbon; and blue, hydrogen.

As you can see in Figure 7a and c, the carboxyl and acetoxy groups are widely distributed inside and around the binding site region. On the other hand, the distribution of the aromatic ring is seen only inside the binding site and is apparently accommodated well within the pocket. Indeed, the binding pocket is formed by hydrophobic amino acid residues, like leucine, phenylalanine, and so on. In that sense, this result represents the case in which the hydrophobic effect makes essential contributions to the molecular recognition for the aromatic ligand.

3.2.2. Binding Mode of an Aspirin Molecule to Binding Site in the Phospholipase A2. Since the final goal of our study is to establish a method to probe a large ligand molecule recognized by protein using 3D-RISM, we must determine the location of

binding site which has the highest affinity of ligand in target protein. The 3D-DDFs or the potential of mean force (PMF) are a good indicator to evaluate the affinity. Those have been successfully applied to measure the affinity or selectivity of solvent in protein.^{12–14} However, since 3D-DDF is the distribution function of an individual site consisting a ligand molecule, it is difficult to evaluate the affinity of whole ligand molecule directly.

In this section, we introduce a function for distribution center (DC) of ligand to measure the affinity of ligand molecule. The function of DC is defined as

$$f_{\text{DC}}(\mathbf{x}) = \begin{cases} \left(\prod_{\gamma}^N \frac{1}{V_{\text{box}} - V_{\text{protein}}(\mathbf{x})} \int_{V_{\text{box}}(\mathbf{x})} g_\gamma(\mathbf{r}) d\mathbf{r} \right)^{1/N} & \text{for } V_{\text{box}} - V_{\text{protein}}(\mathbf{x}) \geq V_{\text{ligand}} \\ 0 & \text{for } V_{\text{box}} - V_{\text{protein}}(\mathbf{x}) < V_{\text{ligand}} \end{cases} \quad (15)$$

where \mathbf{x} denotes the center of box, N is the total number of sites of ligand molecule for normalization, V_{box} is the volume of the box, and $V_{\text{protein}}(\mathbf{x})$ is the excluding volume of the solute protein in the box. Therefore, $V_{\text{box}} - V_{\text{protein}}(\mathbf{x})$ denotes the space where ligand can be distributed. Note that the integrations in right-hand side of eq 15 are only performed inside V_{box} centered at \mathbf{x} . The size of box is adjusted to the length of a ligand molecule. The uu-3D-DDF is integrated in the box, and the result is projected to

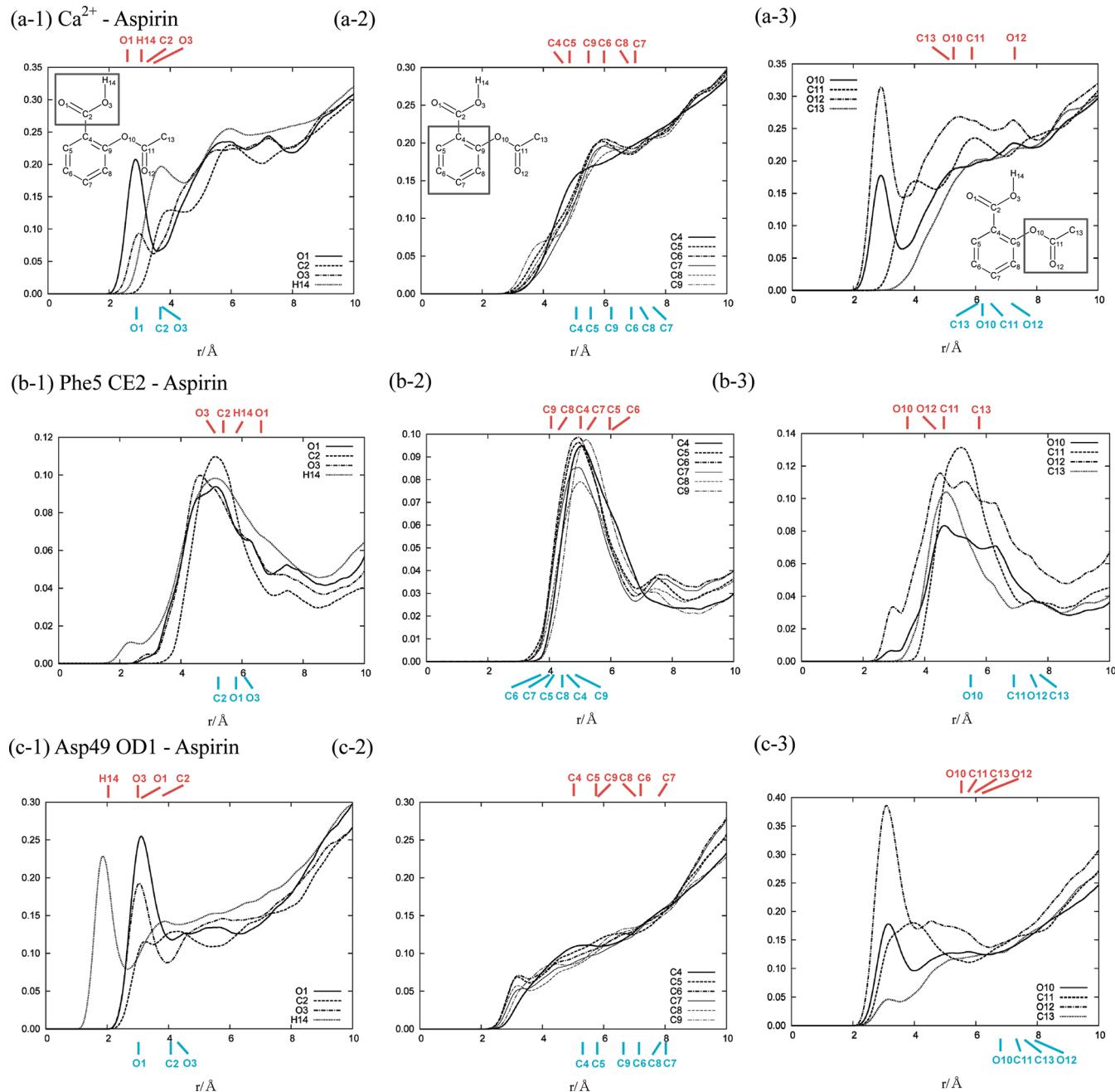


Figure 10. RDFs of aspirin around (a) calcium ion, (b) Phe5, and (c) Asp49, respectively. In part b, CE2 atom was chosen as the averaging center. In part c, OD1 atom was chosen as the averaging center. The distances between each specific site and the atom of predicted aspirin as top peak are marked by upper bars in the x -axis (red indices). The distances between the sites and the atom of aspirin in X-ray structure are also marked by lower bars (blue indices).

the center of box. If the value is larger than one, the probability of finding “an aspirin molecule” at the position is higher than bulk. Although the DC only gives us the rough estimate of the location of binding site, it is helpful to guide a further analysis concerning the binding mode in more detail using information of RDFs, which will be discussed later.

We preformed the calculation of DC based on eq 15 in order to estimate the affinity of ligand to the binding site. The result of DC is shown in Figure 8 with the threshold $f > 1.45$. The maximum value is 1.56. The DC function does not take quite a high value because it is averaged over the sites and volume.

Note that the result of DC is projected onto the center of the calculated box. In Figure 8, we observe the highest peak at the center of the binding site, which is determined by the X-ray crystallography. The results demonstrate that the new method is capable of locating the binding site in the protein and the affinity of a ligand to the site properly.

In this paper, although we focused the highest peak on the binding site, there is another important finding in Figure 8. One may identify a long distribution stretched from the binding site to the bulk. This peak may imply the pathway through which the ligand is entering and escaping. Other peaks observed around the

surface of protein are of interest, because those peaks may be related to the escaping and entering pathways of ligand in diffusive motion. Further analysis of ligand distributions around protein surface will be presented in future.

In order to understand a mechanism of molecular recognition process, it is important to determine an explicit structure of the ligand inside the specific binding site, because we need the distinct structure to calculate some physiological properties, like free energy. Actually, we can also investigate the orientation of the ligand by calculating the overlap between the structure of the ligand and the 3D distribution functions which are obtained by uu-3D-RISM. We define the target function for the orientation by the following equation:

$$f_{\text{ori}}(\mathbf{x}, \Omega) = \prod_{\gamma}^N g_{\gamma}(\mathbf{x} + \mathbf{l}_{\gamma} \cdot \hat{\mathbf{R}}(\Omega)) \quad (16)$$

where \mathbf{x} denotes center of box which is obtained by eq 15, \mathbf{l}_{γ} denotes the internal coordinate of site γ , and $\hat{\mathbf{R}}(\Omega)$ denotes

Table 4. Distance (\AA) between the Specific Sites of PLA2 and the Atoms of Aspirin

sites	Ca^{2+}		Phe5–CE2		Asp49–OD1	
	theory	exptl.	theory	exptl.	theory	exptl.
O1	2.4	2.8	6.7	5.8	3.1	3.0
C2	3.3	3.7	5.5	5.3	3.5	4.0
O3	3.4	3.7	5.2	6.0	3.0	4.2
C4	4.4	5.0	5.1	4.5	5.0	5.4
C5	4.9	5.5	6.0	4.1	5.9	5.8
C6	6.2	6.9	6.0	3.9	7.2	7.1
C7	7.0	7.7	5.3	4.0	7.8	8.0
C8	6.8	7.4	4.2	4.3	7.2	7.9
C9	5.6	6.2	4.0	4.5	5.9	6.6
O10	5.6	6.2	3.5	5.4	5.5	6.7
C11	5.9	6.3	4.4	6.9	5.8	7.3
O12	6.8	7.1	4.3	7.6	6.3	8.0
C13	5.6	6.1	5.9	7.8	6.1	7.6

rotational matrix with the Euler angles to search the entire orientational space. As we mentioned above, the ligand molecule just fits the size of box. It means that the center of the box coincides with the center of the molecule, which is the origin of the rotational matrix at the same spatial point. The advantage of this approach is that we can measure the affinity quantitatively as the degree of overlap. The results of this searching are summarized in Table 3 and Figure 9. In Table 3, the top three orientations, ranked based on the degree of overlap, are listed in terms of the Euler angles, α , β , and γ . Figure 9 shows the explicit ligand structures corresponding to the top three orientations listed in Table 3. The location and orientation of aspirin in the X-ray structure are depicted with blue sticks. It is worthwhile to note that the structure corresponding to one with the greatest overlaps has the same orientation with the X-ray structure, concerning the molecular axis aligning C_2 , C_4 , and C_7 atoms, although the rotational angle around the axis is somewhat different from each other. This orientation seems to be induced by the calcium ion located at the binding site, since the carboxylic group of aspirin faces to the calcium ion. In case of the other two structures with lower score, the carboxylic groups are facing toward His48, which is positively polarized as well.

3.2.3. RDFs of Aspirin around Specific Sites Inside or around Binding Sites of PLA2. In order to find the orientation of aspirin in the binding pocket, we examine the RDFs of each site of aspirin using eq 14. Three specific sites of the residues around the binding pocket are chosen as the averaging centers in order to calculate RDFs. These are the calcium ion, the CE2 atom in Phe5, and the OD2 atom in Asp49. The reasons why those atoms are chosen as the averaging centers are because the calcium ion and the OD2 atom in Asp49 help aspirin to bind in the pocket through the carboxyl group and because the CE2 atom exists in the side opposite to the calcium ion across the pocket. The RDFs are shown in Figure 10. For the purpose of comparing the peaks of RDFs with the corresponding information from the X-ray structure, the distances between these specific sites of the amino acid residue and each atomic site of the aspirin molecule, determined by the X-ray crystallography, are summarized in Table 4. The distances are marked by bars (blue) in the x -axis of the Figure 10. The distances corresponding to the structure

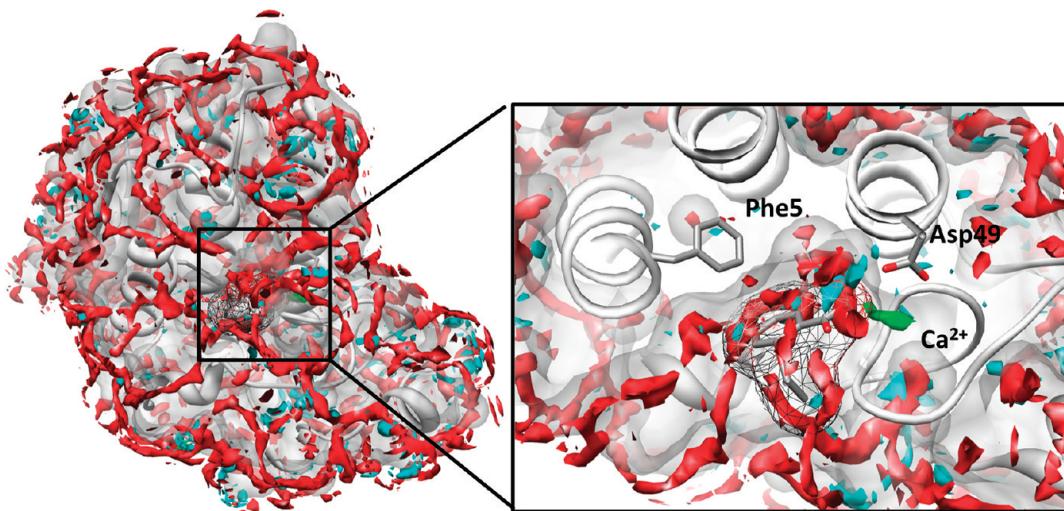


Figure 11. The 3D-DDFs of water around and inside phospholipase A2 are obtained by uv-3D-RISM with the threshold $g_{\gamma}(r) > 3$: red, oxygen atom of water; and cyan, hydrogen atom of water. The 3D-DDF of calcium ion is also obtained with the threshold $g_{\text{Ca}^{2+}}(r) > 40$ as a green spot. The protein surfaces are represented as a gray transparent surface. The location of aspirin in X-ray structure is depicted with a wire frame.

(position and orientation) deduced from the spatial distribution functions due to uu-3D-RISM, with the highest score (Figure 9a), are also marked with red in the same figures. The position of each peak in RDFs is in general consistent with corresponding distance from the X-ray diffraction as well as from the 3D-RISM, except for the distinct peaks in those corresponding to O12 and O10 in Figure 10 (a-3) and to O10 and O12 in (c-3). Those peaks are assigned to distributions of the corresponding atoms existing outside the binding site, and they are irrelevant to the ligand bound in the active site.

Especially interesting among RDFs is the carboxyl group around the calcium ion Figure 10 (a-1) and the aromatic ring around Phe5 (b-2). The peak positions of the RDFs coincide well with those deduced from the orientation of ligands determined both by the experiment and from the analysis of our spatial distribution functions. These suggest strongly the importance of a role played by the calcium ion in recognizing aspirin in the active site. We have confirmed the distinct binding of a calcium ion ($g_{Ca^{2+}}(r) > 40$) at the binding site by means of the 3D-RISM calculation in Figure 11.

The RDFs of carboxyl group around Asp49, shown in Figure 10 (c-1), are worthwhile to draw special attention, since they are suggestive of a mechanism concerning the recognition of aspirin by the protein. According to the results, the ligand molecule is forming a hydrogen bond with one of the carboxylic oxygen atoms of Asp49 through its carboxylic hydrogen atom; note the sharp peak around $r = 1.8 \text{ \AA}$ in the RDF of hydrogen (H14). Apparently, the carboxylic oxygen of Asp49 was supposed to make a hydrogen bond with solvent water, if the position was not invaded by the ligand.

Figure 11 shows the 3D-DDFs of water molecule at the binding site without the ligand. The region of binding site is constructed by hydrophobic residues, like phenylalanine, leucine, isoleucine, and so on. However, water molecules can be bound with main chain or with a charged residue, such as Asp49, through hydrogen bonds. Water molecules are apparently making a hydrogen-bond network or train inside the binding site. It suggests that the dehydration penalty will be extremely high when a ligand replaces those water molecules, and ordinary docking algorithms might not be able to find the binding site.³¹

So, we can draw a hypothetical scheme concerning the recognition mechanism of aspirin to PLA2. The recognition process is largely motivated by the calcium ion, which was already bound at the active site before any aspirin is put in the solution. The recognition process is initiated first by the coulomb interaction between the calcium ion and the carbonyl–oxygen of the carboxyl group, which is followed by formation of the hydrogen bond between the carboxyl–oxygen of Asp49 and the carboxyl–hydrogen of the ligand. In the latter process, a water molecule which was hydrogen bonded to the carboxyl–oxygen, prior to the ligand invasion, is excluded from the binding site. The ligand is further stabilized by the hydrophobic interaction between the phenyl group of the ligand and the hydrophobic residues consisting of the other side of the binding site.

4. CONCLUDING REMARK

We proposed a new approach, the uu-3D-RISM theory, to investigate the molecular recognition in biological system. A motivation to develop the new approach was that the ordinary RISM/3D-RISM approach has difficulty in solving the solvent–solvent RISM equation involving large ligand molecules, which of course have vital importance in the rational drug design. The uu-

3D-RISM is formulated from the general equation of molecular Ornstein–Zernike by considering both a receptor and a ligand as “solutes” immersed in solvent at the infinite dilution limit.

In order to confirm the robustness of the new approach, we calculated the spatial distribution of ethanol at the active site of an odorant binding protein, LUSH, based on the two methods, the ordinary RISM/3D-RISM theory and the uu-3D-RISM, since an ethanol molecule is small enough to be handled with the old method. The new approach reproduced the results from the old method, with subtle difference expected from the discrepancy in the concentration of ligand: one in a finite concentration and the other in the infinite dilution. The analysis based on the radial distribution function (RDF) indicates that the position and the orientation of the ligand inside the binding pocket are consistent with those from the experimental results due to the X-ray crystallography. Robustness of the new approach was thus verified.

We then applied the new approach to an aspirin binding protein, phospholipase A2 (PLA2), with aspirin as a ligand. The process may not be tractable by the old method due to the reason stated above. Since the size of aspirin is much larger and more complex than previous application, or ethanol, analyzing the spatial distribution (uu-DDF) of the ligand inside the binding site, obtained from uu-3D-RISM, is not a trivial problem anymore. So, we developed a new approach to analyze uu-DF, defining a new function referred to as “distribution center (DC),” which locates the center of the most probable distribution of ligand. The position and orientation of aspirin inside the binding site of PLA2 were determined from DC and RDFs of atomic sites of the ligand around particular residues consisting the binding pocket. The binding configuration of the ligand inside the pocket was in fair agreement with that determined from the X-ray crystallography. We will report details about analyses for the origin of binding affinity in a following paper.

The second application of the uu-3D-RISM method clearly demonstrates that the theory is a prospective tool for discovering or designing a new drug, because aspirin itself is already one of the most popular drugs in the market.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: hirata@ims.ac.jp. Telephone: +81 (Japan)-564-55-7314.

■ ACKNOWLEDGMENT

Authors are also grateful to Prof. Sato in Kyoto U. for his helpful discussion. This work is supported by Grants-in-Aid for Scientific Research on Innovate Areas of Molecular Science of Fluctuations toward Biological Functions from the MEXT in Japan. Authors are also supported by the Next Generation Super Computing Project, Nanoscience Program, a Grant-in-Aid for Scientific Research. N.Y. is grateful to a Grant-in-Aid for Young Scientists. Y.K. is grateful to the support by the grant from the Japan Society for the Promotion of Science (JSPS) and the fellow. Molecular graphics images were produced using the UCSF Chimera⁴³ package and gOpenMol.^{44,45} A part of this work was carried out by super computer in the Research Center for Computational Science (RCCS), Okazaki Research Facilities, National Institutes of Natural Sciences (NINS). The uu-3D-RISM program has been implemented to RISM/3D-RISM solver developed by our group in IMS.^{46,47}

■ REFERENCES

- (1) Jorgensen, W. L. *Science* **2004**, *303*, 1813–1818.
- (2) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nature Reviews* **2004**, *3*, 935–949.
- (3) Karplus, M.; McCammon, A. *Nat. Struct. Biol.* **2002**, *9*, 646–652.
- (4) Zwier, M. C.; Chong, L. T. *Curr. Opin Pharmacol.* **2010**, *10*, 1–8.
- (5) Essex, J. W.; Severance, D. L.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **1997**, *101*, 9663–9666.
- (6) Chang, M. *Monte Carlo simulation for the pharmaceutical industry: concepts, algorithms, and case studies*; CRC Press: Boca Raton, FL, 2010.
- (7) Raha, K.; Peters, M. B.; Wang, B.; Yu, N.; Wollacott, A. M.; Westerhoff, L. M.; Merz, K. M., Jr. *Drug Discovery Today*. **2007**, *12*, 725–731.
- (8) Imai, T.; Hiraoka, R.; Kovalenko, A.; Hirata, F. *J. Am. Chem. Soc. Commun.* **2005**, *127*, 15334–15335.
- (9) Imai, T.; Hiraoka, R.; Seto, T.; Kovalenko, A.; Hirata, F. *J. Phys. Chem. B* **2007**, *111*, 11585–11591.
- (10) Yoshida, N.; Phongphanphanee, S.; Maruyama, Y.; Imai, T.; Hirata, F. *J. Am. Chem. Soc. Commun.* **2006**, *273*, 12042–12043.
- (11) Yoshida, N.; Phongphanphanee, S.; Hirata, F. *J. Phys. Chem. B* **2007**, *111*, 4588–4595.
- (12) Phongphanphanee, S.; Yoshida, N.; Hirata, F. *Chem. Phys. Lett.* **2007**, *449*, 196–201.
- (13) Phongphanphanee, S.; Yoshida, N.; Hirata, F. *J. Am. Chem. Soc. Commun.* **2008**, *130*, 1540–1541.
- (14) Phongphanphanee, S.; Yoshida, N.; Hirata, F. *J. Phys. Chem. B* **2010**, *114*, 7967–7973.
- (15) Phongphanphanee, S.; Rungrotmongkol, T.; Yoshida, N.; Hannongbua, S.; Hirata, F. *J. Am. Chem. Soc.* **2010**, *132*, 9782–9788.
- (16) Kiyota, Y.; Hiraoka, R.; Yoshida, N.; Maruyama, Y.; Imai, T.; Hirata, F. *J. Am. Chem. Soc. Commun.* **2009**, *131*, 3852–3853.
- (17) Kiyota, Y.; Yoshida, N.; Hirata, F. *J. Mol. Liq.* **2011**, *159*, 93–98.
- (18) Imai, T.; Oda, K.; Kovalenko, A.; Hirata, F.; Kidera, A. *J. Am. Chem. Soc.* **2009**, *131*, 12430–12440.
- (19) Roux, B.; Yu, H. A.; Karplus, M. *J. Phys. Chem.* **1990**, *94*, 4683–4688.
- (20) Hirata, F. *Molecular Theory of Solvation*; Kluwer: Dordrecht, The Netherlands, 2003.
- (21) Kovalenko, A.; Hirata, F. *J. Phys. Chem. B* **1999**, *103*, 7942–7957.
- (22) Yoshida, N.; Imai, T.; Phongphanphanee, S.; Kovalenko, A.; Hirata, F. *J. Phys. Chem. B* **2009**, *113*, 873–886.
- (23) Kovalenko, A.; Hirata, F. *J. Chem. Phys.* **2000**, *112*, 10391–10402.
- (24) Minezawa, N.; Kato, S. *J. Chem. Phys.* **2007**, *126*, 054511 (15 pages).
- (25) Kovalenko, A.; Ten-No, S.; Hirata, F. *J. Comput. Chem.* **1999**, *20*, 928–936.
- (26) Kinoshita, M.; Okamoto, Y.; Hirata, F. *J. Chem. Phys.* **1997**, *107*, 1586–1599.
- (27) Kruse, S. W.; Zhao, R.; Smith, D. P.; Jones, D. N. M. *Nat. Struct. Biol.* **2003**, *10*, 694–700.
- (28) Dennis, E. A. *J. Biol. Chem.* **1994**, *269*, 13057–13060.
- (29) Nicolas, J. P.; Lin, Y.; Lambeau, G.; Ghomashchi, F.; Lazdunski, M.; Gelb, M. H. *J. Biol. Chem.* **1997**, *272*, 7173–7181.
- (30) Argiolas, A.; Pisano, J. J. *J. Biol. Chem.* **1983**, *258*, 13697–13702.
- (31) Chang, D. T.; Oyang, Y.; Lin. *J. Nuc. Acids Res.* **2005**, *33*, W233–W238.
- (32) Hansen, J. P.; McDonald, I. R. *Theory of Simple Liquids*, 3rd ed.; Academic: London, 2006.
- (33) Chandler, D.; Andersen, H. C. *J. Chem. Phys.* **1972**, *57*, 1930–1937.
- (34) Harano, Y.; Imai, T.; Kovalenko, A.; Kinoshita, M.; Hirata, F. *J. Chem. Phys.* **2001**, *114*, 9506–9511.
- (35) Chandler, D.; McCoy, D. J.; Singer, S. J. *J. Chem. Phys.* **1986**, *85*, 5971–5977.
- (36) Kovalenko, A.; Hirata, F. *Chem. Phys. Lett.* **1998**, *290*, 237–244.
- (37) Kovalenko, A.; Hirata, F. *J. Chem. Phys.* **1999**, *110*, 10095–10112.
- (38) Percus, J. K. *Phys. Rev. Lett.* **1962**, *8*, 462–463.
- (39) Kovalenko, A.; Hirata, F. *Chem. Phys. Lett.* **2001**, *346*, 496–502.
- (40) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.
- (41) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (42) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (43) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605–1612.
- (44) Laaksonen, L. *J. Mol. Graphics* **1992**, *10*, 33–34.
- (45) Bergman, D. L.; Laaksonen, L.; Laaksonen, A. *J. Mol. Graphics Modell.* **1997**, *15*, 301.
- (46) Yoshida, N.; Hirata, F. *J. Comput. Chem.* **2006**, *27*, 453–462.
- (47) Yoshida, N.; Kiyota, Y.; Hirata, F. *J. Mol. Liq.* **2011**, *159*, 83–92.