J|A|C|S ARTICLES

Published on Web 08/05/2009

Ligand Mapping on Protein Surfaces by the 3D-RISM Theory: Toward Computational Fragment-Based Drug Design

Takashi Imai,*,† Koji Oda,‡ Andriy Kovalenko,§ Fumio Hirata,|| and Akinori Kidera†,⊥

Computational Science Research Program, RIKEN, Wako, Saitama 351-0112, Japan, Taisho Pharmaceutical Co., Ltd., Saitama, Saitama 331-9530, Japan, National Institute for Nanotechnology, National Research Council of Canada, and Department of Mechanical Engineering, University of Alberta, Edmonton, Alberta T6G 2M9, Canada, Department of Theoretical and Computational Molecular Science, Institute for Molecular Science, Okazaki, Aichi 444-8585, International Graduate School of Arts and Sciences, Yokohama City University, Tsurumi, Yokohama 230-0045, Japan

Received June 19, 2009; E-mail: takashi.imai@riken.jp

Abstract: In line with the recent development of fragment-based drug design, a new computational method for mapping of small ligand molecules on protein surfaces is proposed. The method uses three-dimensional (3D) spatial distribution functions of the atomic sites of the ligand calculated using the molecular theory of solvation, known as the 3D reference interaction site model (3D-RISM) theory, to identify the most probable binding modes of ligand molecules. The 3D-RISM-based method is applied to the binding of several small organic molecules to thermolysin, in order to show its efficiency and accuracy in detecting binding sites. The results demonstrate that our method can reproduce the major binding modes found by X-ray crystallographic studies with sufficient precision. Moreover, the method can successfully identify some binding modes associated with a known inhibitor, which could not be detected by X-ray analysis. The dependence of ligand-binding modes on the ligand concentration, which essentially cannot be treated with other existing computational methods, is also investigated. The results indicate that some binding modes are readily affected by the ligand concentration, whereas others are not significantly altered. In the former case, it is the subtle balance in the binding affinity between the ligand and water that determines the dominant ligand-binding mode.

Introduction

Fragment-based drug design or discovery (FBDD)[1−3] is currently recognized as one of the most promising alternatives to traditional methods of drug development, such as lead identification and optimization through a high-throughput screening (HTS) campaign using biochemical assay techniques. In FBDD, a series of fragments of drug compounds is first screened against the target protein; and then a few "active" fragments bound to the protein are linked to construct a much more active lead compound (this procedure is called fragment-linking). In more practical cases, two fragments can be merged at the common structure (fragment-merging), or some functional groups can be attached to a selected fragment (fragment-growth). The greatest advantage of FBDD is that a significantly fewer number of the fragment compounds has to be prepared and examined than in an HTS campaign, because the number of substructures that commonly compose drug compounds is relatively limited and it is unnecessary to prepare all possible combinations with FBDD. On the other hand, FBDD also has certain drawbacks. One of these is that the fragmentation generally reduces the binding affinity, so that more sensitive experiments or a higher concentration of fragment molecules are necessary for detection. In addition, detailed information regarding molecular position and orientation in the binding site is required for efficient linking, merging, or growth of the fragments.

One of the most successful experimental approaches in this line is the so-called structure−activity relationships by nuclear magnetic resonance (SAR by NMR)[4,5] In this method, fragment molecules bound to proximal subsites of the target protein are identified by changes in the chemical shift of two-dimensional NMR spectra, and linked together on the basis of the obtained structural data. The technique has been successfully applied to the design of new lead compounds for drugs.[6,7]

Another experimental approach for FBDD is X-ray crystallography, known as multiple solvent crystal structures (MSCS).[8,9]

† RIKEN.
‡ Taisho Pharmaceutical Co., Ltd.
§ National Institute for Nanotechnology and University of Alberta.
|| Institute for Molecular Science.
⊥ Yokohama City University.

(1) Rees, D. C.; Congreve, M.; Murray, C. W.; Carr, R. *Nat. Rev. Drug Discovery* **2004**, *3*, 660.
(2) Hajduk, P. J.; Greer, J. *Nat. Rev. Drug Discovery* **2007**, *6*, 211.
(3) Congreve, M.; Chessari, G.; Tisi, D.; Woodhead, A. J. *J. Med. Chem.* **2008**, *51*, 3661.
(4) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. *Science* **1996**, *274*, 1531.
(5) Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. *Science* **1997**, *278*, 497.
(6) Hajduk, P. J.; et al. *J. Am. Chem. Soc.* **1997**, *119*, 5818.
(7) Oltersdorf, T.; et al. *Nature* **2005**, *435*, 677.
(8) Allen, K. N.; Bellamacina, C. R.; Ding, X.; Jeffery, C. J.; Mattos, C.; Petsko, G. A.; Ringe, D. *J. Phys. Chem.* **1996**, *100*, 2605.
(9) Mottos, C.; Ringe, D. *Nat. Biotechnol.* **1996**, *14*, 595.

 10.1021/ja905029t CCC: $40.75

In this approach, small organic-solvent molecules are regarded as smaller fragments, which generally include a single functional group. The favorable binding positions on the protein surface are determined for each of several organic solvents by the crystal structure of the protein bound by the organic-solvent ligands, which are generated from the crystal soaked in a concentrated mixture of the organic solvent and water. The MSCS approach has proved successful for the prediction of ligand binding sites of the applied proteins.[8−12]

Computational methodologies for the identification of favorable binding sites of functional moieties have also been proposed. They can be roughly categorized into two strategies: one is based on a grid search algorithm, and the other uses energy minimization and/or molecular simulation techniques. The procedure known as GRID[13] is a representative method classified as the former. In the GRID method, a spherical probe imitating a small chemical group is moved through a grid of points around the target protein, and at each point, the interaction energy between the probe and protein is calculated using an empirical energy function. Energetically favorable binding regions for the probe are then identified, based on the energies calculated at the grid points. The GRID method has been practically employed in part of drug optimization.[14] A recently proposed algorithm named FTMAP[15] is also based on a grid search, but the direct grid search calculation is replaced by a fast Fourier transform (FFT) correlation approach to efficiently reduce the computational costs. FTMAP uses molecular probes representing small organic solvents, which are as large as those used in the MSCS approach. An advantage of these approaches is that they allow a global search of the entire protein surface, whereas their disadvantages are that they require empirical parametrization to employ a spherical representation of the probe or expression of the interaction energy function in the form of a convolution integral. Moreover, the most significant drawback is the lack of water molecules in the model. The methods are not able to take the effects of water into account at the atomic level, and therefore cannot handle a situation where water molecules bridge interactions between the ligand and protein.

Multiple copy simultaneous search (MCSS)[16] is a well-known computational method that can be categorized as employing the latter strategies of molecular mechanics and/or simulation techniques. In this method, several thousand copies of a functional group, which is as large as a small organic solvent, are randomly distributed around the binding sites and subjected to simultaneous energy minimization to determine the energetically favorable positions of the functional group as energy minima. In addition to the functionality mapping algorithm, a simple scheme for joining MCSS functional group sites has also been proposed in line with the concept of FBDD.[17,18] An advantage of the MCSS method is that it makes reasonable use of physicochemical potential functions and incorporates the standard molecular simulation framework, which makes the method transferable to various ligands and proteins. However, it is difficult to include the cooperative effects of water, as is for the case with the GRID method. Most recently, an attempt to detect binding sites of a small organic molecule, namely isopropanol, was reported using a molecular dynamics simulation with explicit water.[19] In the study, a concentrated aqueous solution of isopropanol (20% v/v mixture) was used to enhance sampling of the ligand binding. Such a concentrated solution was adopted simply for practical reasons and is reasonable only under the assumption that the interaction between ligand molecules does not affect binding to the protein.

In order to resolve the problems in the current computational approaches, we propose the use of the recently developed integral equation theory of liquids, known as the three-dimensional reference interaction site model (3D-RISM),[20−22] to identify the most probable positions and orientations of fragment molecules on a protein surface. The 3D-RISM theory yields distribution functions of the solvent sites at 3D grid points around the solute, or protein. It requires input of only the solute−solvent interaction potentials, a force field used in standard molecular simulation, and solvent thermodynamic conditions (temperature, density and composition). For the present purpose, we consider a fragment/ligand molecule as a component of the solvent, although the theoretical framework can regard it as part of either the solute or solvent. Applying the 3D-RISM theory to the target protein in a solvent mixture of ligand and water, we obtain the 3D-spatial distributions of atomic sites of the ligand, in addition to those of water around the protein. Each outstanding peak of the ligand-site distributions indicates a very binding position of an atomic site of the ligand. We have already demonstrated successful applications of this approach for some small ligands, such as alkali ions,[23,24] noble gases,[25] and carbon dioxide,[26] in addition to water.[27,28]

Although the reduction from the peaks of the ligand-site distributions to the positions and orientations of the ligand molecule is an obvious operation in the case of atomic and diatomic ligand molecules, it is nontrivial for polyatomic molecules because the site distributions are determined through an ensemble average of different combinations of molecular configurations. Therefore, we also propose an efficient algorithm to determine the most probable positions and orientations, or binding modes, of the ligand molecule from the 3D distribution functions of ligand sites. The algorithm is based on a grid search procedure in which the distribution function acts as a kind of score function. In other words, the ligand molecule is placed so that the atomic sites exhibit simultaneously strong overlap with the corresponding peaks of the distribution functions.

(10) Mottos, C.; Bellamacina, C. R.; Peisarch, E.; Pereira, A.; Vitkup, D.; Petsko, G. A.; Ringe, D. *J. Mol. Biol.* **2006**, *357*, 1471.
(11) English, A. C.; Done, S. H.; Caves, L. S. D.; Groom, C. R.; Hubbard, R. E. *Proteins Struct., Funct., Genet.* **1999**, *37*, 628.
(12) English, A. C.; Groom, C. R.; Hubbard, R. E. *Protein Eng.* **2001**, *14*, 47.
(13) Goodford, P. J. *J. Med. Chem.* **1984**, *28*, 849.
(14) von Itzstein, M.; et al. *Nature* **1993**, *363*, 418.
(15) Brenke, R.; Kozakov, D.; Chuang, G.-Y.; Beglov, D.; Hall, H.; Landon, M. R.; Mattos, C.; Vajda, S. *Bioinformatics* **2009**, *25*, 621.
(16) Miranker, A.; Karplus, M. *Proteins Struct., Funct., Genet.* **1991**, *11*, 29.
(17) Joseph-McCarthy, D.; Hogle, J. M.; Karplus, M. *Proteins Struct., Funct., Genet.* **1997**, *29*, 32.
(18) Joseph-McCarthy, D.; Tsang, S. K.; Filman, D. J.; Hogle, J. M.; Karplus, M. *J. Am. Chem. Soc.* **2001**, *123*, 12758.
(19) Seco, J.; Luque, F. J.; Barril, X. *J. Med. Chem.* **2009**, *52*, 2363.
(20) Kovalenko, A. In *Molecular Theory of Solvation*; Hirata, F., Ed; Kluwer: Dordrecht, The Netherlands, 2003; p 169.
(21) Kovalenko, A.; Hirata, F. *Chem. Phys. Lett.* **1998**, *290*, 237.
(22) Beglov, D.; Roux, B. *J. Phys. Chem.* **1997**, *101*, 7821.
(23) Yoshida, N.; Phongphanphanee, S.; Maruyama, Y.; Imai, T.; Hirata, F. *J. Am. Chem. Soc.* **2006**, *128*, 12042.
(24) Yoshida, N.; Phongphanphanee, S.; Hirata, F. *J. Phys. Chem. B* **2007**, *111*, 4588.
(25) Imai, T.; Hiraoka, R.; Seto, T.; Kovalenko, A.; Hirata, F. *J. Phys. Chem. B* **2007**, *111*, 11585.
(26) Kiyota, Y.; Hiraoka, R.; Yoshida, N.; Maruyama, Y.; Imai, T.; Hirata, F. *J. Am. Chem. Soc.* **2009**, *131*, 3852.
(27) Imai, T.; Hiraoka, R.; Kovalenko, A.; Hirata, F. *J. Am. Chem. Soc.* **2005**, *127*, 15334.
(28) Imai, T.; Hiraoka, R.; Kovalenko, A.; Hirata, F. *Proteins: Struct,. Funct., Bioinf.* **2007**, *66*, 804.

It should be emphasized here that the distribution functions obtained by the 3D-RISM theory naturally include not only the thermodynamic effects of hydration, such as preferential solvation,[29] but also the effects of water at atomic resolution, such as water bridging between ligand and protein. Moreover, the theory can deal with ligand-water mixtures of arbitrary concentrations at a constant computational cost. Thus, the 3D-RISM-based method realizes the entire identification of ligand-binding sites, or "ligand mapping," on a protein surface in a real solution system.

In this study, the 3D-RISM-based method is applied to the binding of four small organic-solvent molecules, isopropanol, acetone, acetonitrile, and phenol, to the enzyme protein thermolysin, because detailed MSCS experimental data[11,12] have been reported for this system. The theoretically predicted binding modes of these organic-solvent ligands are compared with the experimental data in order to validate this theoretical approach. The dependence of ligand-binding modes on the ligand concentration in solution, which is essentially untreatable by other computational methods, is also discussed.

Methods and Theory

Outline of the Mapping Procedure. The most probable positions and orientations were determined for four organic-solvent ligands, isopropanol, acetone, acetonitrile, and phenol, on the surface of thermolysin. The procedure consists of three steps as follows. (Hereafter, "solvent" represents both organic-solvent ligand and water, while "ligand" denotes the organic-solvent ligand only.)

(1) The 3D distribution functions $g_\gamma(\mathbf{r})$ of ligand sites γ on the 3D grid points were calculated by the 3D-RISM theory for systems of thermolysin in a 1 M ligand−aqueous solution at a temperature of 300 K.

(2) Highly probable binding modes (HPBMs) of the ligand were reconstructed from the 3D ligand-site distributions $g_\gamma(\mathbf{r})$ using the grid search algorithm described below. The HPBMs are defined as the modes, i.e., positions and orientations, whose "binding free energies" are less than a threshold value ϵ_{max}. In the calculation, the "site-integrated" potential of mean force (SI-PMF) described below (in the "Grid Search Algorithm" section) is considered as the binding free energy. The threshold was determined with respect to each ligand, so that the 30 most probable binding modes were finally obtained (at the next step): $\epsilon_{max} = -3.24$, -2.92, -3.97, and -2.64 kcal/mol for isopropanol, acetone, phenol, and acetonitrile, respectively.

(3) In general, HPBMs gather in several positions to form clusters. In this final step, the most probable binding mode (MPBM) was extracted from each cluster according to the following procedure. First, the HPBMs obtained are ranked according to their SI-PMF values. Binding modes are then picked up in sequence and subjected to the following selection: if the binding mode "overlaps" with any mode that is ranked above it, it is eliminated from the rank list. The "overlap" is defined as $U > k_BT$, where U is the intermolecular energy, the same as that used for the 3D-RISM calculation, between the molecules with the two modes, and k_B and T are the Boltzmann constant and temperature, respectively. The binding modes that finally remain in the list are identified as MPBMs.

Preparation of Thermolysin Structure. The X-ray crystal structure of ligand-free thermolysin (PDB (protein data bank) code 2tlx)[11] was used. All water molecules and additional substances were removed in advance of the 3D-RISM calculation, except for one zinc and four calcium ions. These ions were left unchanged, because the zinc ion is located at the active site and has a substantial role in the catalytic activity,[30] and the calcium ions are essential to maintain the protein structure.[31] The atomic coordinate with higher occupancy was adopted when there were two candidates. Protonation states of the histidine residues were determined using an empirical method, known as PROPKA,[32] except for H142, H146, and H231. This exception is because the current version of PROPKA ignores the zinc and calcium ions. Since H142 and H146 are bound to the zinc ion at the N^ϵ position, they were protonated at the N^δ position. H231 does not bind to the zinc ion, but locates near the zinc site. The N^δ position of H231 was found to be obviously protonated to make a hydrogen bond with the carboxyl group of D226; however, the protonation of N^ϵ was not trivial. Therefore, a free energy calculation was performed using the 3D-RISM theory, and the pK_a value at the N^ϵ position was estimated to be 11.6. A detailed description of the pK_a estimation is given in the Supporting Information. Based on the result, H231 was protonated at both the N^δ and N^ϵ positions. Finally, all the missing hydrogen atoms were added at the appropriate positions, followed by geometric optimization for only the hydrogen atoms.

3D-RISM Calculation. A general description of the 3D-RISM theory has been previously reported,[19−21] and is also provided in the Supporting Information. A brief outline of the calculations is presented here. First, the solvent site−site correlation functions in each mixture of ligand and water are calculated using the RISM theory coupled with the Kovalenko−Hirata (KH) closure approximation,[33] starting with the potential functions and the densities of ligand and water, and the temperature. The solvent-site distribution functions are then obtained on a 3D grid around the protein by solving the 3D-RISM/KH equations, using the solvent site−site correlation functions and the potential functions between the protein and solvents. The Amber99 parameter set[34] was used for the protein. The simple point charge (SPC) model[35] was adopted for water, using a correction with respect to the Lennard-Jones parameters of the hydrogen sites[36] ($\sigma = 0.4$ Å and $\varepsilon = 0.05$ kcal/mol). The potential parameters of the ligands were taken from the united-atom optimized potentials for liquid simulations (OPLS-UA). Those of isopropanol,[37] acetone,[38] and acetonitrile[39] were taken from the literature. The parameters of phenol were constructed by merging those of benzene[40] and methanol.[41] The 3D-RISM/KH equations were solved on a 3D uniform rectangular grid of $216 \times 216 \times 216$ points in a cubic supercell of size $108 \times 108 \times 108$ Å^3, which is large enough to accommodate the thermolysin molecule with sufficient solvation space (a margin of more than 20 Å on each side of the cube).

Grid Search Algorithm. The molecular distribution of the ligand in six dimensions is approximated, in a Kirkwood manner, by the product of the 3D ligand-site distribution functions,

(29) Imai, T.; Kovalenko, A.; Hirata, F.; Kidera, A. *Interdiscip. Sci. Comput. Life Sci.* **2009**, *1*, 156.
(30) Latt, S. A.; Holmquist, B.; Vallee, B. L. *Biochem. Biophys. Res. Commun.* **1969**, *37*, 333.
(31) Feder, J.; Garrett, L. R.; Wildi, B. S. *Biochemistry* **1971**, *9*, 2784.
(32) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins: Struct,. Funct., Bioinf.* **2005**, *61*, 704.
(33) Kovalenko, A.; Hirata, F. *J. Chem. Phys.* **1999**, *110*, 10095.
(34) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049.
(35) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Pullman, B., Ed; Reidel: Dordrecht, The Netherlands, 1981; p 331.
(36) Pettitt, B. M.; Rossky, P. J. *J. Chem. Phys.* **1982**, *77*, 1451.
(37) Jorgensen, W. L. *J. Phys. Chem.* **1986**, *90*, 1276.
(38) Jorgensen, W. L.; Briggs, J. M.; Contreras, L. *J. Phys. Chem.* **1990**, *94*, 1683.
(39) Jorgensen, W. L.; Briggs, J. M. *Mol. Phys.* **1988**, *63*, 547.
(40) Jorgensen, W. L.; Madura, J. D.; Swerson, C. J. *J. Am. Chem. Soc.* **1984**, *106*, 6638.
(41) Jorgensen, W. L.; Nguyen, T. B. *J. Comput. Chem.* **1993**, *14*, 195.

$$\bar{G}(\mathbf{R},\boldsymbol{\Omega}) = \prod_{\gamma} g_{\gamma}(\mathbf{R} + \mathbf{q}_{\gamma}(\boldsymbol{\Omega})) = \prod_{\gamma} g_{\gamma}(\mathbf{r}_{\gamma}(\mathbf{R},\boldsymbol{\Omega})) \tag{1}$$

where $\mathbf{R} = (x, y, z)$ and $\boldsymbol{\Omega} = (\varphi, \theta, \psi)$ are the translational (Cartesian) and rotational (Euler angle) coordinates of the ligand molecule, respectively, $\mathbf{q}_{\gamma}$ is the internal coordinate of ligand site γ, and $\mathbf{r}_{\gamma}(\mathbf{R},\boldsymbol{\Omega})$ represents the absolute coordinate of ligand site γ when the ligand molecule assumes position $\mathbf{R}$ and orientation $\boldsymbol{\Omega}$. The site distribution is calculated by

$$g_{\gamma}(\mathbf{r}_{\gamma}(\mathbf{R},\boldsymbol{\Omega})) = \int g_{\gamma}(\mathbf{r}) p_{\gamma}(\mathbf{r} - \mathbf{r}_{\gamma}(\mathbf{R},\boldsymbol{\Omega}))\, \mathrm{d}\mathbf{r} \tag{2}$$

where p_{γ} is the density field of ligand site γ, providing a geometry of the ligand molecule, which is generally represented by the delta function if the coordinate is a continuous variable. The distribution functions are specified only at the grid points in the 3D-RISM calculation; therefore, the convolution integral is implemented by discrete calculation,

$$g_{\gamma}(\mathbf{r}_{\gamma}(i\Delta x, j\Delta y, k\Delta z;\boldsymbol{\Omega})) = \sum_{i',j',k'} g_{\gamma}(i'\Delta x, j'\Delta y, k'\Delta z) \times p_{\gamma}(i'\Delta x - x_{\gamma}, j'\Delta y - y_{\gamma}, k'\Delta z - z_{\gamma};\boldsymbol{\Omega}) \tag{3}$$

where (i, j, k) indicates a 3D grid point, $(\Delta x, \Delta y, \Delta z)$ is the grid interval, and $(x_{\gamma}, y_{\gamma}, z_{\gamma}; \boldsymbol{\Omega})$ represents $\mathbf{r}_{\gamma}(i\Delta x, j\Delta y, k\Delta z; \boldsymbol{\Omega})$. A ligand site is not generally exactly on a grid point; therefore, p_{γ} is defined by a Gaussian-type function instead of the delta function,

$$p_{\gamma}(i\Delta x - x_{\gamma}, j\Delta y - y_{\gamma}, k\Delta z - z_{\gamma}) \approx \begin{cases} \dfrac{1}{C}\exp\left(-\dfrac{1}{2}\left(\dfrac{(i\Delta x - x_{\gamma})^2}{\sigma_x^2} + \dfrac{(j\Delta y - y_{\gamma})^2}{\sigma_y^2} + \dfrac{(k\Delta z - z_{\gamma})^2}{\sigma_z^2}\right)\right) & \text{at the 8 vertices of the unit cubic cell where site } \gamma \text{ is found} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where C is the normalization factor,

$$C = \sum_{i,j,k}^{8\ \text{vertexes}} \exp\left(-\frac{1}{2}\left(\frac{(i\Delta x - x_{\gamma})^2}{\sigma_x^2} + \frac{(j\Delta y - y_{\gamma})^2}{\sigma_y^2} + \frac{(k\Delta z - z_{\gamma})^2}{\sigma_z^2}\right)\right) \tag{5}$$

In eqs 4 and 5, the Gaussian half-width parameters σ are set to be half of the grid interval for sufficient damping of the terms from other grid points when the ligand site is very close to a grid point.

In the grid search calculation, the ligand molecular distribution for a given orientation $\boldsymbol{\Omega}$, which gives the relative probability of the ligand molecule placed in the arrangement, is calculated by

$$\bar{G}(i\Delta x, j\Delta y, k\Delta z;\boldsymbol{\Omega}) = \prod_{\gamma} g_{\gamma}(\mathbf{r}_{\gamma}(i\Delta x, j\Delta y, k\Delta z;\boldsymbol{\Omega})) \tag{6}$$

or equivalently, the "site-integrated" potential of mean force (SI-PMF) is defined by

$$\bar{W}(i\Delta x, j\Delta y, k\Delta z;\boldsymbol{\Omega}) = -k_{\mathrm{B}}T \ln \bar{G}(i\Delta x, j\Delta y, k\Delta z;\boldsymbol{\Omega}) = -k_{\mathrm{B}}T \sum_{\gamma} \ln g_{\gamma}(\mathbf{r}_{\gamma}(i\Delta x, j\Delta y, k\Delta z;\boldsymbol{\Omega})) \tag{7}$$

It should be noted that the discrete convolution in eq 3 can be efficiently implemented using FFT, reducing the computational effort from $O(N^6)$ to $O(N^3 \ln N^3)$, in general docking algorithms.[15,42,43] In this case, however, because p_{γ} takes the zero value at all the grid points except the 8 vertices, conventional grid search scaling as $O(M \times N^3)$ (M being the number of grid points with a nonzero value) is faster than FFT for $M < \ln N$.[3] Therefore, in this study, the conventional grid search was adopted rather than a FFT-based grid search.

For convenience of calculation, in eq 3 the same translational grid interval of 0.5 Å was used as in the 3D-RISM calculation. The rotational grid was discretized with an interval of 15° for Euler angles φ and θ; ψ was simply scaled by sin θ (i.e., $\Delta\psi = 15°/[\sin\theta]$) in order to avoid oversampling along the z-axis.

This grid search requires only a reasonable computational time, which is generally less than that for the 3D-RISM calculation to obtain the 3D-site distributions. For example, it takes about 4 h to solve the 3D-RISM/KH equations for thermolysin in an isopropanol−water mixture solvent on a 216 × 216 × 216 grid, using four cores of 2.6 GHz Opteron processor, while it takes under an hour to determine the MPBMs of isopropanol from the site distributions by the grid search algorithm on the same system.

Results and Discussion

Isopropanol on Thermolysin. Figure 1 provides a full illustration of the theoretical results and the X-ray crystallographic data for binding of isopropanol to thermolysin. The isosurface representation in Figure 1a shows the 3D distribution functions of the isopropanol sites on the protein surface (gray), $g_{\mathrm{CH3}} > 4$ (yellow), $g_{\mathrm{CH}} > 4$ (orange), $g_{\mathrm{O}} > 4$ (red), and $g_{\mathrm{H}} > 4$ (light blue), where an average of the two CH_3 sites is used for simplicity. A number of peaks for each site are found over the protein surface. Even though the distribution functions include all information on the binding modes of isopropanol, it is difficult to picture the binding modes explicitly from the appearance.

Figure 1b depicts HPBMs (represented in the ball-and-stick model: green, carbon; red, oxygen; white, hydrogen) reduced from the distribution functions by the grid search algorithm based on the Kirkwood superposition approximation. At this stage, the binding sites appear obviously more limited than in the distribution functions. Within the limited areas, HPBMs form some large or small clusters. Each of the clusters indicates a basin of the binding free energy with a certain resolution, and the size of each cluster signifies the broadness of the free energy basin. Significantly, the largest cluster is found in the center cleft, which corresponds to the active site of thermolysin. The MPBMs extracted from the clusters are shown in Figure 1c. It was found that the MPBMs in the active site are similar to the crystallographic positions of isopropanol[44] (ball-and-stick model: cyan, carbon; mauve, oxygen), which are shown in Figure 1d.

However, there are some differences between the theoretical and experimental results; some crystallographic molecules are not identified by the theoretical method and some theoretically predicted ones are not found in the X-ray crystal structure. In the former case, such molecules are located not in a cleft, but above the surface, and all of these are found only in the crystal soaked in 100% isopropanol solvent. In the latter case, many

(42) Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A. A.; Aflalo, C.; Vakser, I. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 2195.
(43) Gabb, H. A.; Jackson, R. M.; Sternberg, M. J. E. *J. Mol. Biol.* **1997**, *272*, 106.
(44) These were taken from the isopropanol-bound structure obtained in the pure isopropanol solvent (PDB code, 8tli).[11] We located them in the ligand-free structure by superimposing the two protein structures.

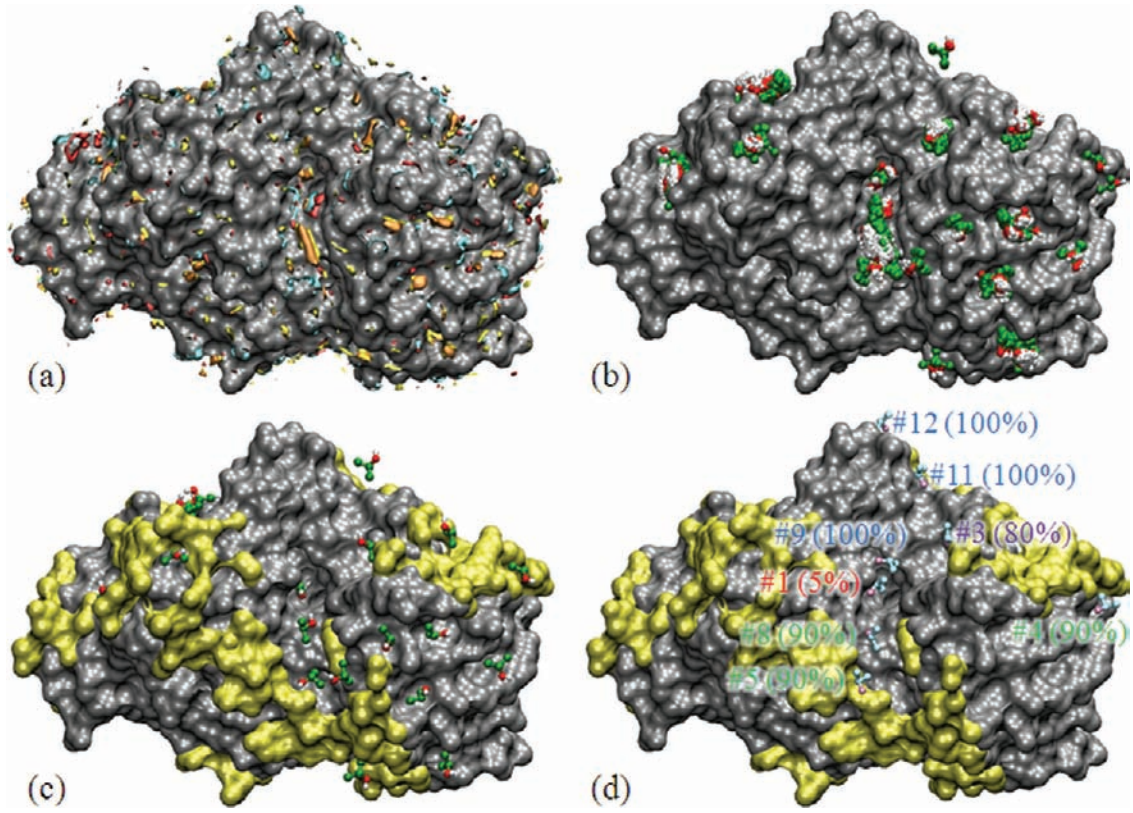

Figure 1. Binding modes of isopropanol on the surface of thermolysin in 1 M isopropanol−aqueous solution. (a) Isosurface representation of the 3D distribution functions of the isopropanol sites. The threshold $g(\mathbf{r}) > 4$ is used for all the atomic sites: yellow surfaces and spots represent the distribution of CH_3 (the average of two CH_3 sites is shown for simplicity); orange, CH; red, O; light blue, H. The protein surface is colored gray. (b) HPBMs reduced from the distribution functions using our grid search algorithm, represented by the ball-and-stick model: green atoms represent carbon; red, oxygen; white, hydrogen. (c) MPBMs, shown in the same representation as in (b). (d) X-ray crystallographic isopropanol molecules represented by the ball-and-stick model: cyan, carbon; mauve, oxygen. The molecules are numbered according to the original paper.[11] Values in parentheses indicate the minimum percent concentration above which the molecule is present. Surfaces colored yellow are of atoms that pass within 5 Å of the adjacent thermolysin molecules in the crystal.

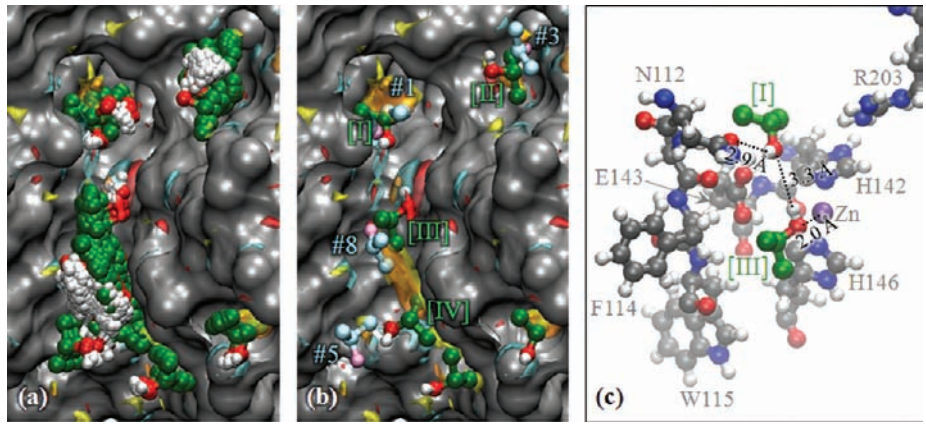

Figure 2. Binding modes of isopropanol in the active site of thermolysin in 1 M isopropanol−aqueous solution. (a) Isosurface representation of the 3D distribution functions, $g(\mathbf{r}) > 4$, of the isopropanol sites and HPBMs. (b) Isosurface representation of $g(\mathbf{r}) > 4$, MPBMs, and X-ray crystallographic isopropanol molecules. The model and color representations are the same as those in Figure 1. (c) Ball-and-stick model representation of MPBM-I and III and its surrounding residues (gray, carbon; red, oxygen; blue, nitrogen; white, hydrogen).

of the apparent positive faults can be explained by exclusion, due to adjacent protein molecules in the crystal. The contact regions on the protein surface are colored yellow in parts c and d of Figure 1. The theoretically predicted molecules in the upper left, upper right, and lower center parts of the protein are most likely to be excluded by adjacent protein molecules in the crystal. Thus, most of the discrepancies can be ascribed to crystal field influences.

Isopropanol in the Active Site. Figure 2 shows a magnified view around the active site. The MPBMs identified by our method are marked with Roman numerals, as specified in Figure 2b, while the crystallographic molecules are indicated by Arabic numerals according to the original paper.[11] The MPBMs are in good agreement with the crystallographic positions, except for MPBM-IV, as shown in Figure 2b. The root-mean-square deviations (RMSDs) between the MPBMs and the corresponding

Table 1. Comparison of the Binding Modes Identified by the 3D-RISM Theory with the X-ray Crystallographic Positions at the Active Site of Thermolysin, along with the Results of the Multiple Copy Simultaneous Search (MCSS) Method

	3D-RISM				MCSS[a]			
	Lowest[b]		Closest[c]		Lowest[b]		Closest[c]	
site[d]	rmsd[e]	SI-PMF[f]	rmsd[e]	SI-PMF[f]	rmsd[e]	Energy[g]	rmsd[e]	Energy[g]
				Isopropanol (IPA)				
IPA-1	1.44	−3.54	1.44	−3.54	2.7	−31.1	1.5	−25.2
IPA-3	2.41	−3.48	1.97	−3.38	3.7	−33.3	2.8	−22.6
IPA-5	4.02	−3.50	1.68	−3.29	4.5	−25.1	1.1	−23.5
IPA-8	1.96	−3.75	1.57	−3.27	0.9	−20.8	0.9	−20.8
				Acetone (ACN)				
ACN-1	0.52	−3.51	0.42	−3.13	1.2	−37.8	1.2	−21.7
ACN-6	3.23	−3.40	2.48	−2.98	3.4	−33.9	2.5	−21.7
				Phenol (IPH)				
IPH-1	0.41	−3.98	0.41	−3.98	2.2	−22.8	0.7	−21.2
				Acetonitrile (CCN)				
CCN-1	3.21	−2.79	1.77	−2.66	2.4	−28.8	2.4	−28.8

[a] Data from ref 12. [b] Lowest-energy (SI-PMF for 3D-RISM or interaction energy for MCSS) molecule within the nearest cluster to the X-ray position. [c] Closest molecule to the X-ray position. [d] X-ray crystallographic sites which are numbered according to ref 12 [e] Root mean square deviation (Å) between the theoretical and experimental positions (based on heavy atoms). [f] "Site-integrated" potential of mean force (kcal/mol). [g] Interaction energy between MCSS functional groups and protein sites (kcal/mol).

crystallographic molecules are 1.44, 2.41, and 1.96 Å, for MPBM-I, II, and III, respectively. (The rmsd values are summarized in Table 1.) MPBM-IV has only a small overlap with the proximal crystallographic molecule (IPA-5); the rmsd is 4.02 Å. It should be noted that some HPBMs which are found to be closer to IPA-5. The rmsd for the closest HPBM has a small value of 1.68 Å.

The discrepancy between MPBM-IV and IPA-5 can be explained as follows. Figure 2a shows a broad distribution of HPBMs around the MPBM-IV site, which implies that there are some broad SI-PMF minima. In such cases, the local minimum that determines the MPBM can be readily shifted by a change in the environment. Therefore, this discrepancy probably results from the difference in the environments created by the aqueous solution and crystal. MPBM-IV is found to be largely dependent on the ligand concentration, as described later ("Concentration Dependence" section). IPA-5 is also apparently affected by the crystal field, as described in the previous section (see also Figure 1d).

It would be informative to compare the current results with those using the MCSS method, which is one of the most prevalent approaches to mapping of small ligands on protein surfaces. The MCSS results are also compiled in Table 1. Based on the comparison, it is fair to conclude that the 3D-RISM-based method can predict the experimental positions with better accuracy. It may seem that the 3D-RISM result is worse than MCSS for the rmsd with IPA-8, but the MCSS energy is very high compared to the others, which implies that the binding mode is only a minor local minimum in the MCSS method. This comparison provides some implications for the significance of the effect of water, because water molecules play an important role, even in the crystal structure, whereas the effects of water are totally ignored in the MCSS calculation.[11] In that sense, it is reasonable that the 3D-RISM-based method produces results that are more consistent with the experimental results than the MCSS method does.

The chemical stabilities of MPBM-I and III are now discussed in detail. Figure 2c shows the interaction pattern among the MPBMs and the surrounding residues. MPBM-I is found to be stabilized by a weak hydrogen bond with the side-chain carbonyl oxygen of N112; the distance between the isopropanol hydrogen and the carbonyl oxygen (d_{OH}) is 2.9 Å. Some of the HPBMs in the same cluster to which MPBM-I belongs form hydrogen bonds with the side-chain guanidyl group of R203, as shown in Figure 2a. This implies that the isopropanol molecule is fluctuating between two hydrogen-bonding partners in the aqueous solution, even though the structure specified as MPBM-I is the most stable within the fluctuation. The stability of MPBM-I may be due to a weak hydrogen bond network ($d_{OH} = 3.3$ Å) with the other isopropanol molecule identified as MPBM-III. Moreover, MPBM-III is strongly bound to the zinc ion at a distance of 2.0 Å.

Concentration Dependence. It is generally assumed that ligand-binding modes do not change with variation in the ligand concentration. This assumption is premised on the insensitivity of the binding free energy surface to the solvent condition. It might be acceptable for a highly dilute solution mixture; however, ligand-binding modes can generally be modified to a certain extent with the ligand concentration. Here, the dependence of the MPBMs on the ligand concentration is investigated.

Figure 3a shows the MPBMs in pure isopropanol solvent. Although MPBM-I is found to be located at a position similar to that in the 1 M solution (Figure 2b), its orientation is changed significantly. In the pure solvent, the isopropanol molecule is most stabilized when it forms a hydrogen bond with R203, as depicted in Figure 3c, although it prefers the hydrogen bond with N112 in the 1 M solution (Figure 2c).

This orientational transition is explained in terms of the distribution of water molecules. With the current method, probable binding modes for water can be simultaneously determined using the same procedure as that for the ligand. Figure 4 shows the HPBMs and MPBMs of water molecules[45] in a 1 M isopropanol−aqueous solution. A water molecule can also bind to both N112 and R203 with a hydrogen bond, as shown in Figure 4a. (It should be noted that only the MPBM hydrogen bonding to N112 appears in Figure 4b, because both the binding modes belong to a single cluster.) Therefore, a subtle

(45) The SI-PDF threshold for water was determined to be−2.285 kcal/mol for the calculation in pure water so that the number of the MPBMs obtained was equal to that of the crystallographic water molecules found in the ligand-free structure. The threshold value was also used for the calculation in the 1 M solution.

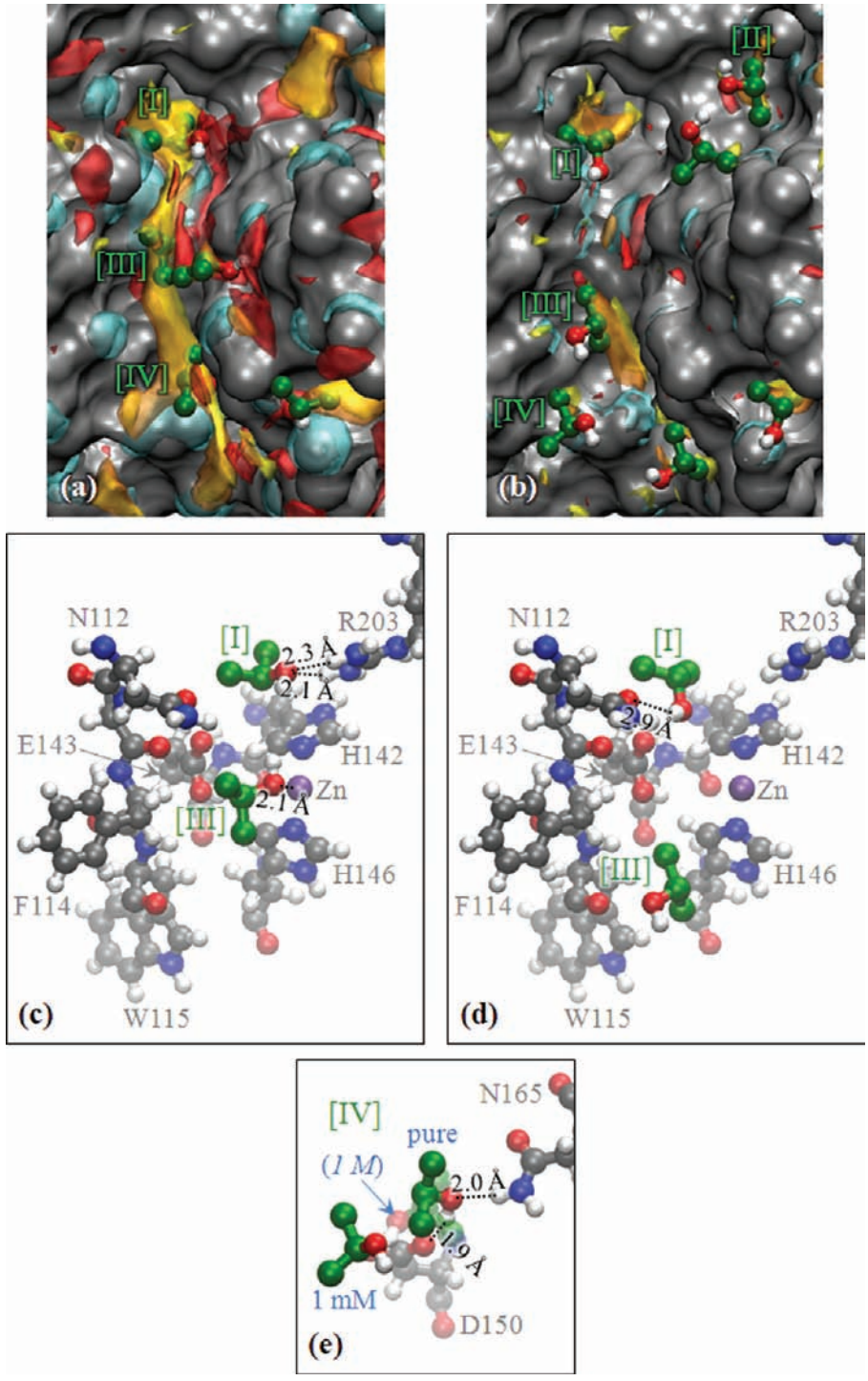

Figure 3. MPBMs of isopropanol in the active site of thermolysin at different concentrations: (a) in pure isopropanol and (b) in 1 mM isopropanol−aqueous solution. MPBM-I and III and its surrounding residues: (c) in pure isopropanol and (d) in 1 mM isopropanol−aqueous solution, and (e) MPBM-IV at different concentrations and its surrounding residues. The model and color representations are the same as those in Figures 1 and 2.

balance in the binding free energy between isopropanol and water determines which binding mode of isopropanol is more stable in the aqueous solution. On the other hand, it is governed only by the competition among the binding modes of isopropanol in the pure isopropanol solvent. In this case, the isopropanol molecule prefers to form hydrogen bonds with R203 rather than with N112 in the pure solvent; however, in aqueous solution it prefers to form a hydrogen bond with N112 than with R203 by competing with water for the guanidyl group of R203 as a hydrogen-bonding partner.

A similar orientational transition is observed for MPBM-IV. In the pure isopropanol solvent, the isopropanol molecule is strongly stabilized by two hydrogen bonds; one is created between the isopropanol hydrogen and one of the carboxyl oxygens of D150 at a distance of 1.9 Å, and the other is formed between the isopropanol oxygen and the amide hydrogen of N165 at 2.0 Å, as illustrated in Figure 3e. In the aqueous solution, water molecules replace these hydrogen bonds and push the isopropanol hydroxyl group toward the solution phase.

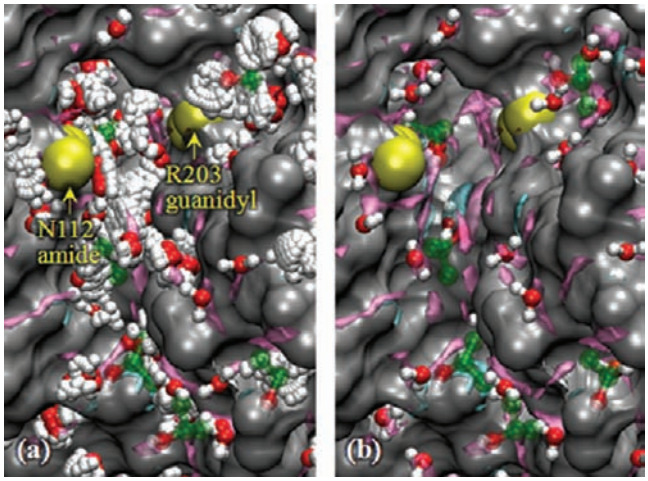

Figure 4. Binding modes of water in the active site of thermolysin in 1 M isopropanol−aqueous solution. (a) HPBMs (b) MPBMs represented by the ball-and-stick model (red, oxygen; white, hydrogen). Isosurfaces of the 3D distribution functions, $g(\mathbf{r}) > 4$, of the water sites; oxygen (mauve) and hydrogen (light blue), are also shown. The MPBMs of isopropanol are depicted together using a transparent ball-and-stick model (green, carbon; red, oxygen; white, hydrogen).

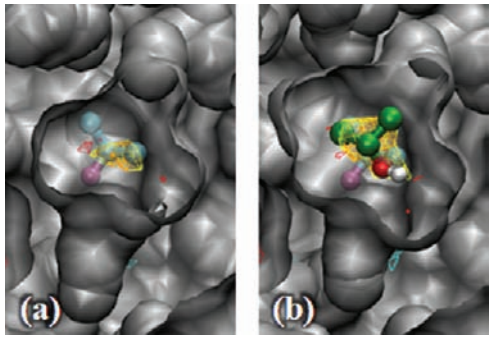

Figure 5. Binding modes of isopropanol in the buried site of thermolysin in 1 M isopropanol−aqueous solution, (a) using the isopropanol-free structure, and (b) using the isopropanol-bound structure with crystal soaked in pure isopropanol. The model and color representations are the same as those in Figure 1, except for the X-ray result represented by the transparent model.

In contrast, the stable interaction of MPBM-III with the zinc ion found in the pure solvent holds even in the 1 M aqueous solution. However, when the concentration is reduced to 1 mM, the isopropanol molecule concedes the interaction with the zinc ion to a water molecule, and then moves to another preferable position, as shown in images b and d of Figure 3. This translation is accompanied by a shift in the position of MPBM-IV to avoid the overlap with MPBM-III, as seen in images b and e of Figure 3.

Isopropanol in a Cavity. An isopropanol molecule is also found within a buried cavity in the crystals soaked in a solution with a concentration of more than 10%. The theoretical result for the cavity is shown in Figure 5a; no HPBMs are found in the cavity. This unsuccessful result is due to a rather apparent reason. In this calculation, the isopropanol-free structure of thermolysin was used. The cavity in the structure is not large enough to completely accommodate an isopropanol molecule. The isopropanol-site distributions obtained are only minor, due to van der Waals repulsion, and therefore the SI-PMFs of the corresponding binding modes are too high to be accepted in the grid search.

In order to investigate the structural dependence of the binding modes, the same calculation was performed using an isopropanol-bound structure (PDB code, 8tli) rather than the isopropanol-free structure, and the result is given in Figure 5b. The protein surface representation clearly shows a larger cavity in the isopropanol-bound structure, and in this case, an isopropanol-binding mode is found at almost the same position as the crystallographic position. Thus, the structural change induced by ligand-binding is crucial for the prediction of the ligand-binding site, especially when the ligand molecule binds into a buried cavity of the protein. On the other hand, the binding modes in the active site, MPBM-I, II, III, and IV, are not significantly affected by the choice of structure. (The result for the isopropanol-bound structure is given in Figure S1.) This is because only minor structural changes are induced by isopropanol-binding in the active site.

Acetone in the Active Site. Figure 6a shows the MPBMs of acetone in the active site. (HPBMs are shown in Figure S2a.) MPBM-I is in excellent agreement with the crystallographic acetone molecule (ACN-1).[46] The rmsd between the theoretical and experimental binding positions is 0.52 Å. The MCSS also predicts an energy-minimum structure close to ACN-1, although the rmsd value (1.2 Å) is larger than that predicted using the 3D-RISM-based method, as given in Table 1. The rmsd between MPBM-II and ACN-6 has a rather large value of 3.23 Å. Even the closest HPBM is separated from the crystallographic position by a rmsd of 2.48 Å. For this site, the MCSS energy-minimum position is also away from the crystallographic position in the same range, which implies that this discrepancy may arise from crystal-field effects. A third MPBM is found in our calculation, even though no corresponding molecule is found in the crystal structure. In the crystallographic study,[12] the maximum concentration for acetone was 70%. Considering that the isopropanol molecule at the corresponding position (IPA-8) was only found at concentrations of more than 90%, an acetone molecule would likely be observed in this position if crystallographic structures could be obtained at higher concentrations. As described later in detail ("Comparison with a Known Inhibitor" section), a carbonyl moiety of an inhibitor is bound to thermolysin at the position corresponding to MPBM-III.

Figure 6b shows the detailed binding pattern including MPBM-I and III and its surrounding residues. MPBM-I is stabilized by hydrogen bonds with two hydrogens in the guanidyl group of R203; one at a distance of 2.0 Å and the other at 2.8 Å. It is interesting to compare this result with the MPBM-I of isopropanol (Figure 2c). Although the isopropanol molecule intrinsically prefers binding to the guanidyl groups of R203, it is turned to the carbonyl oxygen of N112 in aqueous solution due to the effects of water. In contrast, the orientation of MPBM-I is maintained even in the aqueous solution, because an acetone molecule cannot make a hydrogen bond with the carbonyl oxygen, but it can form stable hydrogen bonds with the amino groups. MPBM-III forms a strong hydrogen bond with the main-chain peptide hydrogen of W115 at a distance of 2.0 Å, rather than interacting with the zinc ion. This binding mode is also determined by a free energy balance in the interactions between acetone, water, and the interaction sites of protein, as is shown for isopropanol MPBM-I.

Phenol in the Active Site. The MPBMs of phenol in the active site are shown in Figure 6c. (HPBMs are shown in Figure S2b.)

(46) The crystallographic acetone positions were determined by superimposing the acetone-bound structure obtained in the 70% acetone−aqueous solution (PDB code, 1fjq) onto the ligand-free structure.

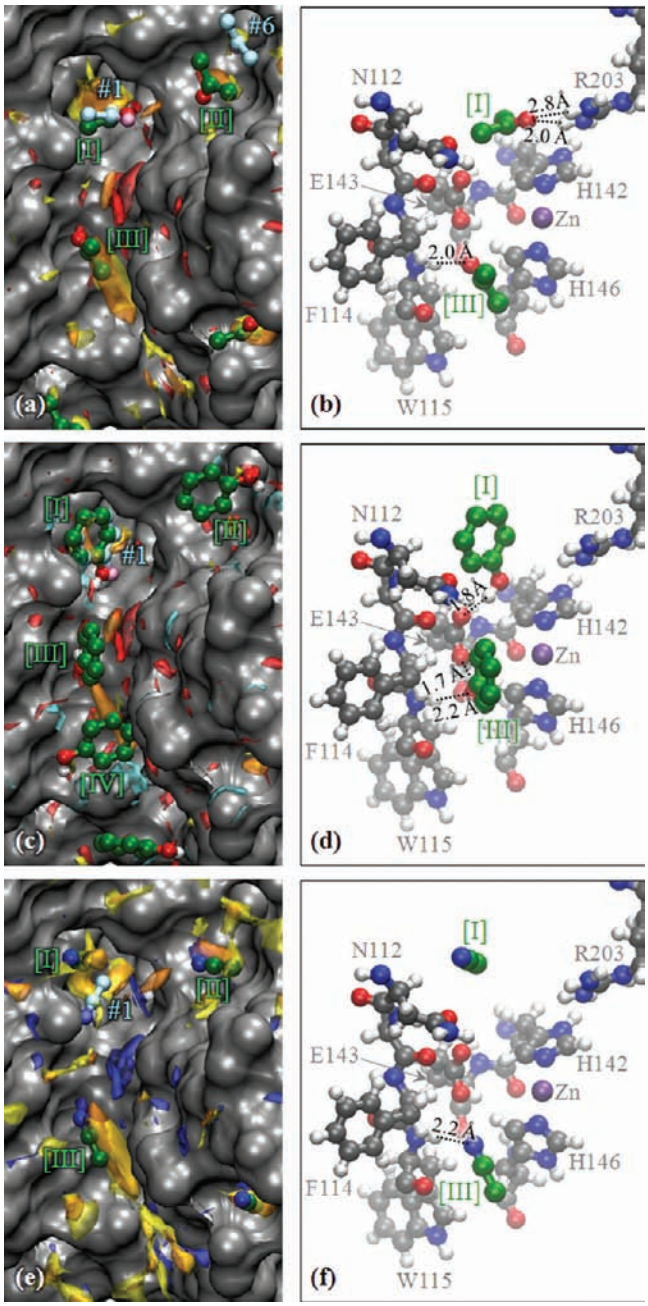

Figure 6. MPBMs of acetone, phenol, and acetonitrile in the active site of thermolysin in their respective 1 M aqueous solutions. (a) Isosurface representation of the 3D distribution functions, $g(\mathbf{r}) > 4$, of the acetone sites (yellow, CH_3; orange, C; red, O), MPBMs represented by the ball-and-stick model (green, carbon; red, oxygen), and X-ray crystallographic acetone molecules represented by the ball-and-stick model (cyan, carbon; mauve, oxygen), numbered according to the original paper.[12] (b) Acetone MPBM-I and III and its surrounding residues. (c) 3D distribution functions, $g(\mathbf{r}) > 4$, of the phenol sites (yellow, average for two *o*-CH sites, average for two *m*-CH sites, and *p*-CH; orange, C; red, O; light blue, H), MPBMs (green, carbon; red, oxygen; white, hydrogen), and the X-ray crystallographic phenol molecule (cyan, carbon; mauve, oxygen). (d) Phenol MPBM-I and III and its surrounding residues. (e) 3D distribution functions, $g(\mathbf{r}) > 4$, of the acetonitrile sites (yellow, CH_3; orange, C; blue, N), MPBMs (green, carbon; blue, nitrogen), and the X-ray crystallographic acetonitrile molecule (cyan, carbon; blue, nitrogen). (f) Acetonitrile MPBM-I and -III and surrounding residues.

MPBM-I is again quantitatively consistent with the crystallographic position (IPH-1),[47] as is indicated by the rmsd value of 0.41 Å. MPBM-I forms a strong hydrogen bond with one of the carboxyl oxygens of E143 at a distance of 1.8 Å, as shown in Figure 6d, in the same manner as the crystallographic phenol molecule. However, the MCSS predicts another binding mode which interact with the guanidyl group of R203,[12] which is separated from the crystallographic position by 2.2 Å in rmsd (Table 1). The misprediction by the MCSS is evidently due to the neglect of water molecules, because the vicinity of the guanidyl groups of R203 is a water-preferential region.

Although there is only the one phenol molecule within the active site in the crystal structure, our present method predicts three additional MPBMs that are located near the position of the isopropanol MPBMs, II, III, and IV. It should be noted that the crystal structure was obtained only in a dilute 0.4% phenol solution. Therefore, we can again expect that the corresponding molecules would be found there if the crystal structures could be obtained in much more concentrated solutions.

Apart from the experimental result, it is interesting to have some discussion regarding the additional MPBMs, in particular, MPBM-III and IV. MPBM-III has two hydrogen bonds with the protein; the hydroxyl hydrogen binds to the other carboxyl oxygen of E143, which is not bound to MPBM-I, at a distance of 1.7 Å, and the hydroxyl oxygen binds to the main-chain peptide hydrogen of W115 at 2.2 Å. Although phenol has a hydroxyl group similar to isopropanol, the binding mode is different, due to certain hydrophobic effects of the benzene ring. Interestingly, it is found that MPBM-IV again corresponds to the position of a benzyl moiety of an inhibitor; the detail is described in the "Comparison with a Known Inhibitor" section.

Acetonitrile in the Active Site. The MPBMs of acetonitrile in the active site are shown in Figure 6e. (HPBMs are shown in Figure S2c.) MPBM-I is located 3.21 Å (rmsd) away from the crystallographic position (CCN-1),[48] even though a few HPBMs are found at closer positions. In this case, each of the three positions identified by the 3D-RISM-based method, the MCSS method, and X-ray crystallography is several ångstroms different from the others. (See also Figure 4 in ref 12 for the MCSS position.) The binding mode predicted by the MCSS is again stabilized by hydrogen bonds with the guanidyl group of R203, whereas MPBM-I does not have any direct interactions with protein sites (Figure 6(f)), but it is stabilized by hydration. CCN-1 also accepts a hydrogen bond from a crystallographic water molecule rather than from functional groups of the protein, even though the binding mode is distinct from MPBM-I. In addition to MPBM-I, the 3D-RISM-based method predicts the other two MPBMs, II and III, which exhibit similar binding patterns to those of acetone.

Comparison with a Known Inhibitor. It has already been demonstrated in the MSCS experimental studies[11,12] that IPA-1, ACN-1, and IPH-1 correspond to binding modes of some moieties of known inhibitors. Since MPBM-I for each ligand identified in this study is essentially identical with the crystallographic binding mode, we will not repeat the discussion regarding the subsite. Instead, the other subsites that have been identified by the 3D-RISM-based method but that could not be found by the MSCS experiment are discussed.

We have checked a total of 30 inhibitor-bound structures of thermolysin resolved in the PDB to date. There are few

(47) The crystallographic phenol position was determined by superimposing the phenol-bound structure obtained in the 0.4% phenol−aqueous solution (PDB code, 1fjw) onto the ligand-free structure.

(48) The crystallographic acetonitrile position was determined by superimposing the acetonitrile-bound structure obtained in the 80% acetonitrile−aqueous solution (PDB code, 1fju) onto the ligand-free structure.

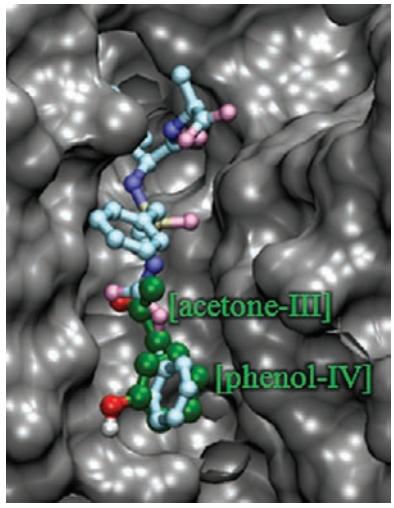

Figure 7. Comparison of two binding modes identified using the present method (MPBM-III of acetone and MPBM-IV of phenol) with the crystallographic structure of carbobenzoxy-L-*Phe*^{P}-L-Leu-L-Ala (ZF^{P}LA) in the active site of thermolysin. ZF^{P}LA is represented by the small ball-and-stick model (cyan, carbon; mauve, oxygen; blue, nitrogen; tan, phosphorus). The model and color representations for the MPBMs of acetone and phenol are the same as those in Figure 6.

inhibitors binding all along the active site including the subsites corresponding to MPBM-III and IV, because when a long peptide-like ligand binds along the active site, the peptide cleavage reaction proceeds at reaction sites including the zinc ion and the carboxyl group of E143.[49] Carbobenzoxy-L-*Phe*^{P}-L-Leu-L-Ala (ZF^{P}LA) is a phosphonamidate peptide inhibitor, which was designed to realize extensive binding to the active site without peptide cleavage by replacing the peptide linkage with the tetrahedral phosphorus of a phosphonamidate group.[50] Here, the binding modes of the organic-solvent ligands obtained using the 3D-RISM-based method are compared with the binding structure of ZF^{P}LA.

The crystal structure of ZF^{P}LA binding to the active site[51] is shown in Figure 7. The MPBM-III of acetone is found to overlap with the carbonyl moiety of ZF^{P}LA. The acetone molecule forms a hydrogen bond with the main-chain peptide hydrogen of W115. The carbonyl group of ZF^{P}LA also has the same hydrogen-bonding pattern to stabilize the ZF^{P}LA-thermolysin complex. Moreover, one of the benzene rings of ZF^{P}LA is found at the same subsite as the MPBM-IV of phenol. The small difference in orientation is probably due to the hydration effect of the phenol hydroxyl group, because the ZF^{P}LA benzyl group does not have a hydroxyl group. We have thus demonstrated that the 3D-RISM-based method can successfully predict the binding subsites of the inhibitor that the MSCS experimental method has failed to detect.

This result implies a possibility of application of the 3D-RISM-based method to FBDD. In a possible approach, first a set of fragment molecules including functional groups commonly found in drug molecules is prepared. The MPBMs of the fragments are then obtained for the target protein, using the 3D-RISM-based method. In a general case, the active site can be identified on the basis of clustering of the MPBMs. Furthermore, promising candidate compounds can be constructed by linking or merging a couple of selected fragment MPBMs in the active site. In practice, the MPBMs may also be used for modifications to a known drug or lead compound by linking or merging with the known structure. Development of a theoretical scheme for efficient linking or merging of the fragment MPBMs is an important issue to be resolved in the next stage.

Conclusions

A novel computational method was proposed for the mapping of small ligands on a protein surface. The method is based on the 3D-RISM theory, which can consider ligands and water at the same level in terms of the site distribution; therefore, ligand mapping is carried out within a realistic model including the coexistence of water, which is distinct from most of the other computational approaches. The 3D-RISM-based method was applied to the binding of several small organic solvents to thermolysin in order to evaluate its efficiency. The results demonstrated that the 3D-RISM-based method can reproduce major binding modes found in the X-ray crystallographic studies with sufficient accuracy. Furthermore, the method succeeded in identifying some binding modes associated with a known inhibitor, which could not be experimentally detected.

In addition to its predictive ability, it should be emphasized that the 3D-RISM-based method can reveal the dependence of ligand-binding modes on the ligand concentration. We have presented evidence to show that some binding modes are readily affected by variation in the ligand concentration, while other modes are hardly perturbed. These results are explained in terms of the distribution of water, also obtained from the 3D-RISM theory. Such concentration dependence has not been a subject of discussion to date, due to the limitations of other methods presently employed. However, the present results clearly indicate the significance of ligand-concentration effects associated with the effect of water.

Thus, the 3D-RISM-based method of ligand-mapping is a promising approach for the identification of binding modes for functional moieties of ligand molecules, such as inhibitors and drug molecules, particularly in the context of FBDD. Nevertheless, there is an important problem to be resolved at the next stage, namely protein structural change induced by ligand-binding. We have developed some methodologies applicable to address this issue. If the structural change is limited within a free energy basin, the combination of the 3D-RISM theory with the classical energy minimization method[25] can be employed. In this approach, the protein structure is optimized in order to minimize the intraprotein energy plus the solvation free energy obtained by the 3D-RISM theory, and then the optimized protein structure and the solvent distribution are simultaneously obtained. If the protein can be extensively deformed by ligand binding, then molecular simulations, such as molecular dynamics[52] and Monte Carlo,[53] should be adopted in combination with the 3D-RISM theory. Using this combination, the structural changes and fluctuations of protein forced by the ligand and water can be simulated, and the solvent distributions for the sampled protein structures can then be obtained, which produce their ensemble average. These advanced approaches would further improve 3D-RISM-based ligand-mapping.

(49) Hangauer, D. G.; Monzingo, A. F.; Matthews, B. W. *Biochemistry* **1984**, *23*, 5730.
(50) Bartlett, P. A.; Marlowe, C. K. *Biochemistry* **1987**, *26*, 8553.
(51) Holden, H. M.; Tronrud, D. E.; Monzingo, A. F.; Weaver, L. H.; Matthews, B. W. *Biochemistry* **1987**, *26*, 8542.
(52) Miyata, T.; Hirata, F. *J. Comput. Chem.* **2008**, *29*, 871.
(53) Mitsutake, A.; Maruyama, Y.; Imai, T.; Kinoshita, M.; Okamoto, Y.; Hirata, F. Unpublished.
(54) Humphrey, W.; Dalke, A.; Schulten, K. *Mol. Graphics* **1996**, *14*, 33.

Acknowledgment. T.I. is grateful to Prof. T. Ikura and Prof. T. Terada for stimulating and helpful discussions. This work was supported by Research and Development of the Next-Generation Integrated Simulation of Living Matter, a part of the Development and Use of the Next-Generation Supercomputer Project of the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. Part of the calculations were performed with the RIKEN Super Combined Cluster (RSCC). A.K. acknowledges support from the National Research Council (NRC) of Canada. Molecular graphics were prepared using VMD.[54]

Supporting Information Available: Detailed descriptions of the 3D-RISM theory and the estimation of pK_a for H231, the most probable binding modes of isopropanol in the active site when using the isopropanol-bound structure (Figure S1), the highly probable binding modes of acetone, phenol, and acetonitrile in the active site (Figure S2), and complete refs 6, 7, and 14. This material is available free of charge via the Internet at http://pubs.acs.org.

JA905029T