Roles of water in protein structure and function studied by molecular liquid theory

Takashi Imai

Research Program for Computational Science, RIKEN, Wako, Saitama 351-0198, Japan

TABLE OF CONTENTS

1. ABSTRACT

The roles of water in the structure and function of proteins have not been completely elucidated. Although molecular simulation has been widely used for the investigation of protein structure and function, it is not always useful for elucidating the roles of water because the effect of water ranges from atomic to thermodynamic level. The three-dimensional reference interaction site model (3D-RISM) theory, which is a statistical-mechanical theory of molecular liquids, can yield the solvation structure at the atomic level and calculate the thermodynamic quantities from the intermolecular potentials. In the last few years, the author and coworkers have succeeded in applying the 3D-RISM theory to protein aqueous solution systems and demonstrated that the theory is useful for investigating the roles of water. This article reviews some of the recent applications and findings, which are concerned with molecular recognition by protein, protein folding, and the partial molar volume of protein which is related to the pressure effect on protein.

2. INTRODUCTION

Water plays vital roles in stabilizing the native structure of proteins in aqueous solution. As the solvent medium, water stabilizes the hydrophobic core of the protein interior (1,2), which is a representative of the protein architecture. Water moderates intraprotein electrostatic interactions, such as hydrogen bonding and salt bridging, by the dielectric screening (2). If we alter the solvent properties, for example, by adding cosolvent, the native structure of proteins is destabilized or stabilized depending on the properties of the cosolvent (3-5). Changing the thermodynamic conditions also can change the stability of the native structure (5,6). Since an increase of the temperature directly destabilizes the protein structure by the effect of the structural entropy of the protein itself, the effect of water is not so apparent in the thermal denaturation. In contrast, it is evident that the protein denaturation by lowering the temperature is explained by the change in the solvent properties. It is also considered that the pressure-induced structural transition of proteins is

due to the effect of water on the basis of the fact that the application of pressure generally swells proteins to reduce the system volume rather than compresses them directly (7,8).

Water molecules buried in the cavities of a protein and trapped in the clefts are of significant importance in the structural stability of the protein at the atomic level. They maintain the specific local structure of the protein by mediating hydrogen bonds or simply filling void spaces in the protein (9,10). That is implied by the finding that a mutation with creating a hydrated cavity in a protein less destabilizes the native structure of the protein than a mutation with an empty cavity (11). It has also been found that the removal of internal water molecules by the reduction of the water content in the crystallization can modify the local structure of protein in the vicinity where the water molecules were located (12). The significance of such internal water molecules is also recognized by the fact that they are highly conserved in homologous protein families with the same structural motifs (13,14), as the key residues are.

Water plays crucial roles also in the chemical processes related to protein functions, especially in the recognition of other molecules. A ligand molecule is bound to the receptor protein not only by the direct ligand-protein interaction but also by the water-mediated interaction (15,16). For instance, it was found that a water molecule bridges the binding of certain HIV protease inhibitors to the receptor (17). The bridging water molecule was the key to designing new drug molecules. Even if there are no water molecules explicitly mediating a ligand-protein binding, water as the solvent medium modulates the binding through some non-local effects such as the hydrophobic interaction and the dielectric screening.

The effect of water is conceptually classified into two categories. One is the effect when water can be regarded as a continuous medium, which is represented by the hydrophobic effect and the dielectric screening. In considering this effect, we focus on the bulk properties of water, such as the density and the permittivity, more than the molecular properties at the atomic level, such as the ability as hydrogen-bonding donor and acceptor. The other is the effect which is assigned to individual water molecules, such as water molecules buried in the protein cavities and those bridging a ligand-protein binding. In this case, each of the water molecules acts as an integral component of the biomolecular systems. Both of the effects are equally important in considering the roles of water in protein structure and function.

There are mainly two different approaches to theoretically or computationally investigate the roles of water. One is based on the continuum solvent model (18). The generalized Born (GB) model (19) is a representative of the continuum model and is currently prevalent among computational biochemists and biophysicists. The GB model gives an approximate analytical expression of the electrostatic contribution to the solvation free energy (SFE) for a solute molecule immersed in a continuous medium.

An empirical method based on the surface area (SA) (20) is another approach to estimate the non-electrostatic contribution to the SFE. The SA method uses empirical relations between the SFE and the atomic SA of small organic compounds. In the method, the SFE of a large molecule is calculated with the empirical parameters which have been adjusted so as to reproduce experimental values of the SFE of small organic compounds. The GB and SA methods can be combined to obtain both of the electrostatic and non-electrostatic contributions. The GB/SA method or other continuum solvent approaches can approximately provide the SFE with minor computational cost, but it is not always successful in protein thermodynamics. In fact, the properties of water in the vicinity of a protein molecule can be largely different from those of bulk water. The effects are not described by the continuum nor empirical model because they are determined by cooperative many-body protein-water and water-water interactions. The most fatal drawback of these approaches is that they can never take into account the contributions from the individual water molecules which could have special roles as described above.

Molecular simulation with explicit water molecules can, in principle, incorporate the effects from the individual water molecules, if the simulation is performed correctly to sample the whole configuration space of water molecules. The SFE changes associated with chemical processes such as the structural transition of proteins and ligand-protein binding can be calculated accordingly, using advanced simulation techniques such as free energy perturbation and thermodynamic integration methods (18,21). However, such a free energy calculation requires very high computational expenses, so that the application is practically limited to a small perturbation of the chemical processes.

The integral equation theory of molecular liquids (22) is an alternative approach to investigate the solvent effects. In the theory, the density pair correlation functions such as the radial distribution function are calculated from given intermolecular interaction potentials by solving integral equations derived from statistical mechanics. The correlation functions describe the liquid structure and also serve as weighting factors for statistical averages of mechanical quantities to produce thermodynamic quantities. This means that the theory has the capability of describing the roles of water at the atomic and thermodynamic levels equally. The statistical mechanics of liquids has a rather long history, but it was quite recent that the theory was successfully applied to biomolecular solution systems. Originally, the application of the theory was limited to so-called simple liquids, which consist of monatomic, spherical particles. Later, the theory was extended so as to deal with molecular liquids by expressing the molecular correlation functions in terms of the atomic site-site correlation functions (23). It is known as the reference interaction site model (RISM) theory. The RISM theory was further extended to account for the charge distribution of molecules (24). The RISM theory was a possible candidate for a theoretical method to be used instead of molecular simulation.

The RISM theory has been successfully applied to a variety of chemical processes in solution (25,26). However, it was found that the theory failed in predicting the thermodynamic quantities when we apply the theory to the systems including very large molecules such as proteins, which have a large number of interaction atomic sites. The most conspicuous failure was inspected in the partial molar volume (PMV). We observed that the RISM theory systematically underestimates the PMV of amino acids and peptides (27), and even predicts negative values for the PMV of proteins, which is apparently unphysical. The problem is primary due to imperfect account of the solvent exclusion by the solute, which arises from partial neglect of the multi-site correlation effects in the site-site treatment.

More recently, the RISM theory was generalized to describe the three-dimensional (3D) correlation functions of the solvent around the solute (28,29). The 3D spatial generalization made possible the perfect account of the solvent exclusion effect by the solute of arbitrary shape. That was proved by the fact that the use of the 3D-RISM theory drastically improved the quantitative estimation of the PMV of amino acids and peptides (27). It was in 2004 that we succeeded in applying the 3D-RISM theory to proteins in aqueous solution (30). In that study, we demonstrated that the PMV values of proteins predicted by the 3D-RISM theory were in quantitative agreement with the corresponding experimental data. After that study, we have continuously applied the 3D-RISM theory to protein aqueous solution systems (31), and have succeeded in clarifying the roles of water in several chemical processes associated with protein structure and function at both the atomic and thermodynamic levels.

This article presents a selective review of our recent studies on the roles of water in protein structure and function by the 3D-RISM theory. First, some studies concerning molecular recognition by protein are presented. The studies demonstrated that the 3D-RISM theory has the ability to detect the binding sites of ligand molecules as well as water molecules in a protein. Next, the essential roles of water in protein folding are discussed. It is found that the entropic as well as energetic contributions of hydration are significant in the stabilization of the native structure of proteins. Finally, the molecular mechanism of the PMV change associated with the pressure-induced structural transition of a protein is discussed. A theoretical analysis indicates that the penetration of water into the protein is the key to the process. Before discussing those results, we shall briefly review the 3D-RISM theory in the following section.

3. THEORETICAL BACKGROUND

3.1. Brief outline of the 3D-RISM theory

For a solute-solvent system at infinite dilution, in which solvent may be mixture, the 3D-RISM integral equation is written as (26,28,29)

$$h_\gamma^{uv}(\mathbf{r}) = \sum_\eta \int c_\eta^{uv}(\mathbf{r}')\left(w_{\eta\gamma}^{vv}(|\mathbf{r}'-\mathbf{r}|) + \rho_\gamma h_{\eta\gamma}^{vv}(|\mathbf{r}'-\mathbf{r}|)\right)d\mathbf{r}', \quad (1)$$

where $h_\gamma^{uv}(\mathbf{r})$ and $c_\gamma^{uv}(\mathbf{r})$ are, respectively, the total and direct correlation functions of solvent site γ in the 3D space ($\mathbf{r} = (x,y,z)$) around the solute molecule; $w_{\eta\gamma}^{vv}(r)$ is the site-site intramolecular correlation function, which determines the geometry of the solvent molecules; $h_{\eta\gamma}^{vv}(r)$ is the site-site total correlation function, which is obtained in advance from the RISM calculation for bulk solvent, as described below; and ρ_γ is the number density of molecular species to which solvent site γ belongs. The 3D spatial distribution function $g_\gamma^{uv}(\mathbf{r})$ is given by $g_\gamma^{uv}(\mathbf{r}) = h_\gamma^{uv}(\mathbf{r}) + 1$. The 3D-RISM equation is complemented by a closure equation appropriate to the system. Conventionally, we choose the 3D spatial generalization of the hypernetted chain (HNC) approximation,

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

where $u_\gamma^{uv}(\mathbf{r})$ is the interaction potential between solvent site γ and the whole solute specified on the 3D grid, and $\beta = 1/(k_BT)$ is the inverse of the Boltzmann constant times temperature. We may adopt another closure equation proposed by Kovalenko and Hirata, the so-called KH closure (26,32). The KH equation is given by partial linearization of the HNC equation. Specifically, it combines the HNC approximation in the spatial regions of density depletion where $g_\gamma^{uv}(\mathbf{r}) < 1$, and the mean spherical approximation (MSA), $c_\gamma^{uv}(\mathbf{r}) = -\beta u_\gamma^{uv}(\mathbf{r})$, in the enrichment regions where $g_\gamma^{uv}(\mathbf{r}) > 1$. Since the MSA-type linearization prevents the artificial divergence in the correlation functions due to a collective effect of large interaction potentials, the KH closure is appropriate for mixed solvent systems with strongly attractive chemical functionalities.

The site-site total correlation functions of solvent $h_{\eta\gamma}^{vv}(r)$ used in the 3D-RISM calculation are calculated independently from the ordinary RISM theory for radial correlations. Similarly to the 3D-RISM calculation, the RISM equation,

$$h_{\alpha\gamma}^{vv}(r) = \sum_{\eta,v} \int w_{\alpha\eta}^{vv}(r')c_\eta^{uv}(|\mathbf{r}'-\mathbf{r}''|)$$
$$\left(w_{v\gamma}^{vv}(|\mathbf{r}''-\mathbf{r}|) + \rho_v h_{v\gamma}^{vv}(|\mathbf{r}''-\mathbf{r}|)\right)d\mathbf{r}'d\mathbf{r}'', \quad (3)$$

is solved simultaneously with the HNC closure for the radial correlations,

$$h_{\alpha\gamma}^{vv}(r) = \exp\left(-\beta u_{\alpha\gamma}^{vv}(r) + h_{\alpha\gamma}^{vv}(r) - c_{\alpha\gamma}^{vv}(r)\right) - 1, \quad (4)$$

where $u_{\alpha\gamma}^{vv}(r)$ is the pair interaction potential between solvent sites α and γ. We may also employ the KH approximation instead of the HNC, if it is more appropriate.

The 3D potential functions $u_\gamma^{uv}(\mathbf{r})$ are constructed from the site-site pair potential functions $u_{a\gamma}^{uv}(r)$ by

$$u_\gamma^{uv}(\mathbf{r}) = \sum_a u_{a\gamma}^{uv}(|\mathbf{r}_a - \mathbf{r}|), \qquad (5)$$

where the subscription a denotes an atomic interaction site of the solute molecule. The pair potential function can be taken from the standard model generally used in molecular simulation, which consists of the Lennard-Jones and electrostatic interaction terms. For example, the Amber or Charmm force field for protein and the SPC/E or TIP3P model for water are used. The 3D potential may be calculated using the minimum image convention and the Ewald summation methods. In that case, the special electrostatic correlations for the 3D supercell finiteness are added to Equation 2, in order to provide an accurate calculation for the solute containing site charges (33). Other input data for the 3D-RISM calculation are the temperature and the number density of each solvent species, which determine the thermodynamic condition of the system. There are no parameters particular to the theory.

In order to visually recognize the 3D distribution function, isosurface plots of the 3D distribution function of water oxygen $g_O(\mathbf{r})$ around hen egg-white lysozyme obtained by the 3D-RISM theory are given in Figure 1. The green surfaces show the areas where $g_O(\mathbf{r})$ is larger than 2, 4, and 8 in the left, center, and right pictures, respectively. The lysozyme molecule is represented by the standard space-filling model. As can be seen in the figure, the 3D distribution function provides the solvation structure around the solute at atomic resolution.

3.2. Calculation of thermodynamic quantities

The thermodynamic quantities are calculated by integration of the correlation functions. There are mainly two theoretical ways to obtain the thermodynamic quantities from the correlation functions. One is the Kirkwood-Buff (KB) solution theory (34), which is a general theory of solution and is independent of the approximation used in the theory. The other is the Singer-Chandler (SC) formulation (35), which is only valid under the HNC approximation and its certain modifications such as the KH approximation.

The solvation free energy (SFE) is calculated by the 3D generalization of the SC equation from the 3D spatial correlation functions obtained by the 3D-RISM/HNC calculation (26):

$$\Delta\mu_s = k_B T \sum_i \rho_i \sum_{\gamma_i} \int \left(\frac{1}{2} h_{\gamma_i}^{uv}(\mathbf{r})^2 - c_{\gamma_i}^{uv}(\mathbf{r}) - \frac{1}{2} h_{\gamma_i}^{uv}(\mathbf{r}) c_{\gamma_i}^{uv}(\mathbf{r}) \right) d\mathbf{r},$$
$$(6)$$

where the subscript i indicates the label of molecular species of solvent, and γ_i denotes the site label of the ith solvent molecule. The corresponding equation for the KH approximation has been also proposed. Similar equations

for the derivatives of SFE, such as the solvation entropy, can be obtained by the differentiation of the SC equation with respect to the corresponding thermodynamic quantities, such as the temperature (36,37).

The partial molar volume (PMV) can be also obtained from the differentiation of the SC equation, but it is generally calculated on the basis of the KB theory. Using the 3D-RISM expression, the KB equation for the PMV at infinite dilution for the solute molecule is written as (27,38)

$$\bar{V}^0 = \chi_T^0 k_B T \left(1 - \sum_i \rho_i \sum_{\gamma_i} \int c_{\gamma_i}^{uv}(\mathbf{r}) d\mathbf{r} \right), \qquad (7)$$

where χ_T^0 is the isothermal compressibility of the bulk solvent. It is obtained from the site-site correlation functions of solvent as follows (38):

$$\chi_T^0 = \left(k_B T \sum_i \rho_i \left(1 - \sum_j \rho_j \sum_{\alpha,\gamma} \int c_{\alpha_i \gamma_j}^{vv}(r) 4\pi r^2 dr \right) \right)^{-1}. \qquad (8)$$

4. MOLECULAR RECOGNITION BY PROTEIN

4.1. Water molecules in protein cavities

As described above, water molecules confined in restricted spaces within a protein play important roles in the structure and function of the protein. In a certain sense, the confinement of water molecules is the most fundamental process of the molecular recognition by protein. Therefore, prediction of the coordination structure of water molecules can be characterized as a starting point of the theoretical investigation for the molecular recognition by protein. Although the 3D-RISM theory had proven itself to be capable of predicting the hydration structure and thermodynamics at least qualitatively, it was not entirely sure that the theory can be applied to water molecules confined in small cavities of protein because statistical-mechanical theories including the 3D-RISM theory are not designed for such heterogeneous systems but have been developed for homogeneous fluids. However, the following result (39,40) unambiguously demonstrates that the 3D-RISM theory has the ability of detecting water molecules confined in a protein cavity at the atomic level.

Figure 2A shows isosurface representation of the 3D distribution functions $g(\mathbf{r})$ of water oxygen and hydrogen in one of the major internal cavities of lysozyme obtained by the 3D-RISM/HNC theory. The green and pink surfaces (or spots) indicate the major peaks of the distribution functions of water oxygen and hydrogen, respectively, which are defined by $g(\mathbf{r}) > 8$. There are four distinct peaks of oxygen and eight peaks of hydrogen, which are apparently assigned to four water molecules. Figure 2B presents the most probable model of the hydration structure reconstructed from the peaks of the distribution functions. The distribution predicted by the theory is in almost perfect correspondence with the crystallographic water sites found by X-ray analysis (41), which is shown in Figure 2C.
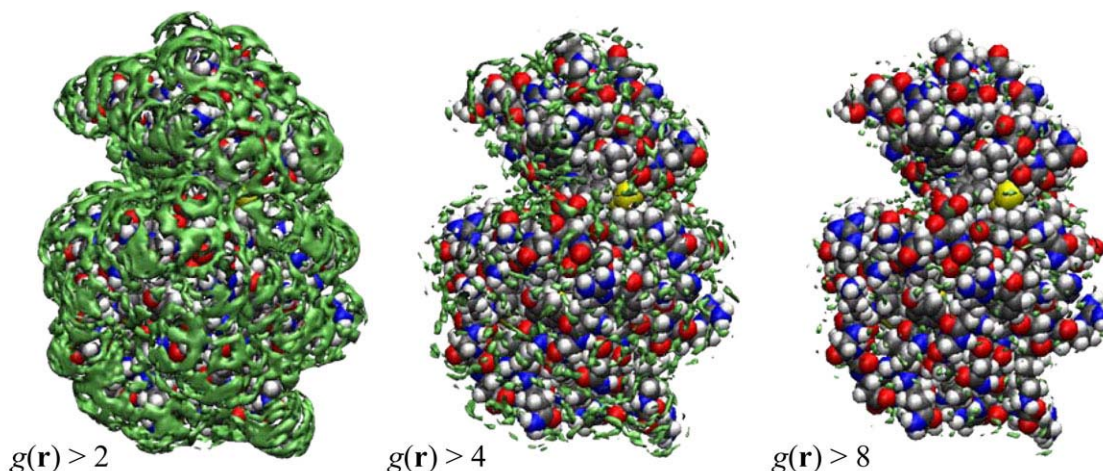
Figure 1. Isosurface representation of the three-dimensional distribution function, $g(\mathbf{r})$, of water oxygen around lysozyme calculated by the 3D-RISM theory. The green surfaces show the areas where the distribution function is larger than 2 (left), 4 (center), and 8 (right). The lysozyme molecule is represented by the standard space-filling model.
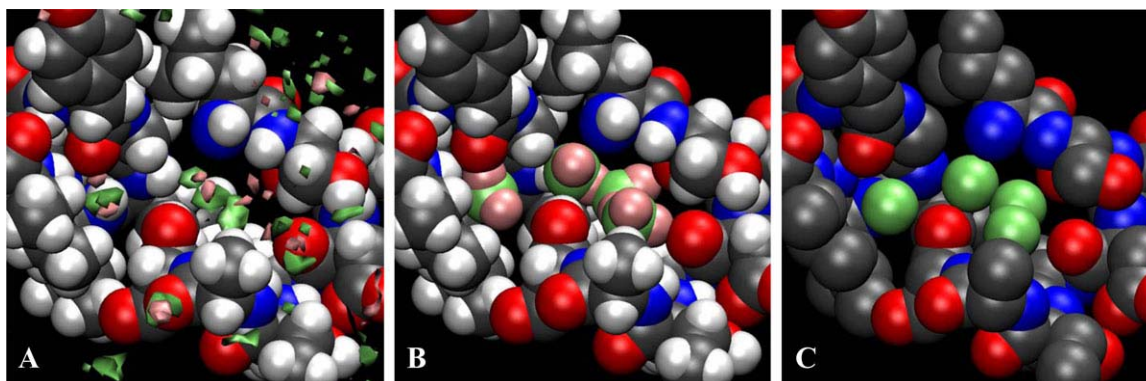

Figure 2. Water molecules in the largest cavity of lysozyme. Only the surrounding residues are displayed with the standard space-filling model. The residues in the front side are deleted to facilitate visualization of the inside of the cavity. A. Isosurface plots of the three-dimensional distribution functions of water oxygen (green) and hydrogen (pink) calculated by the 3D-RISM theory. The surfaces (or spots) show the areas where the distribution function is larger than 8. B. The most probable model of the hydration structure reconstructed from the three-dimensional distribution functions. C. Water oxygen atoms detected by X-ray crystallography.

The hydration number is as significant as the hydration structure, and prediction of the former is admittedly more difficult than that of the latter. The equilibrium number of water molecules in a cavity is not always equal to the number of binding sites found in the cavity. For example, Figure 2B represents only an average structure of water molecules, but it does not indicate that there are four water molecules in the cavity. The equilibrium number of water molecules in a cavity is calculated by integrating the distribution function within the cavity space. In this case, the equilibrium number of water molecules in the cavity was calculated to be 3.6, which is actually less than the number of water-binding sites, 4, detected by the 3D-RISM theory as well as X-ray analysis. The "reduction" in the hydration number was confirmed by molecular dynamics simulation with a special treatment. The simulation was started with the initial state in which four water molecules were placed at the four X-

ray water sites. In the simulation, we observed that the most and the second most outer water molecules sometimes leaved and entered the cavity in the nano-second time scale and the average number of water molecules which stayed in the cavity was 3.5, which is in good agreement with the theoretical prediction. Although the 3D-RISM theory takes no explicit account of the dynamics of the water molecules, it can provide the reasonable hydration number through the statistical-mechanical relations.

This result not only proves that the 3D distribution function obtained by the 3D-RISM theory correctly describes water molecules around a protein, but also implies that the thermodynamic quantities calculated from the 3D correlation functions properly reflect the effect of water at the atomic level. The roles of water in protein structure seen through thermodynamic quantities are discussed in Sections 5 and 6.
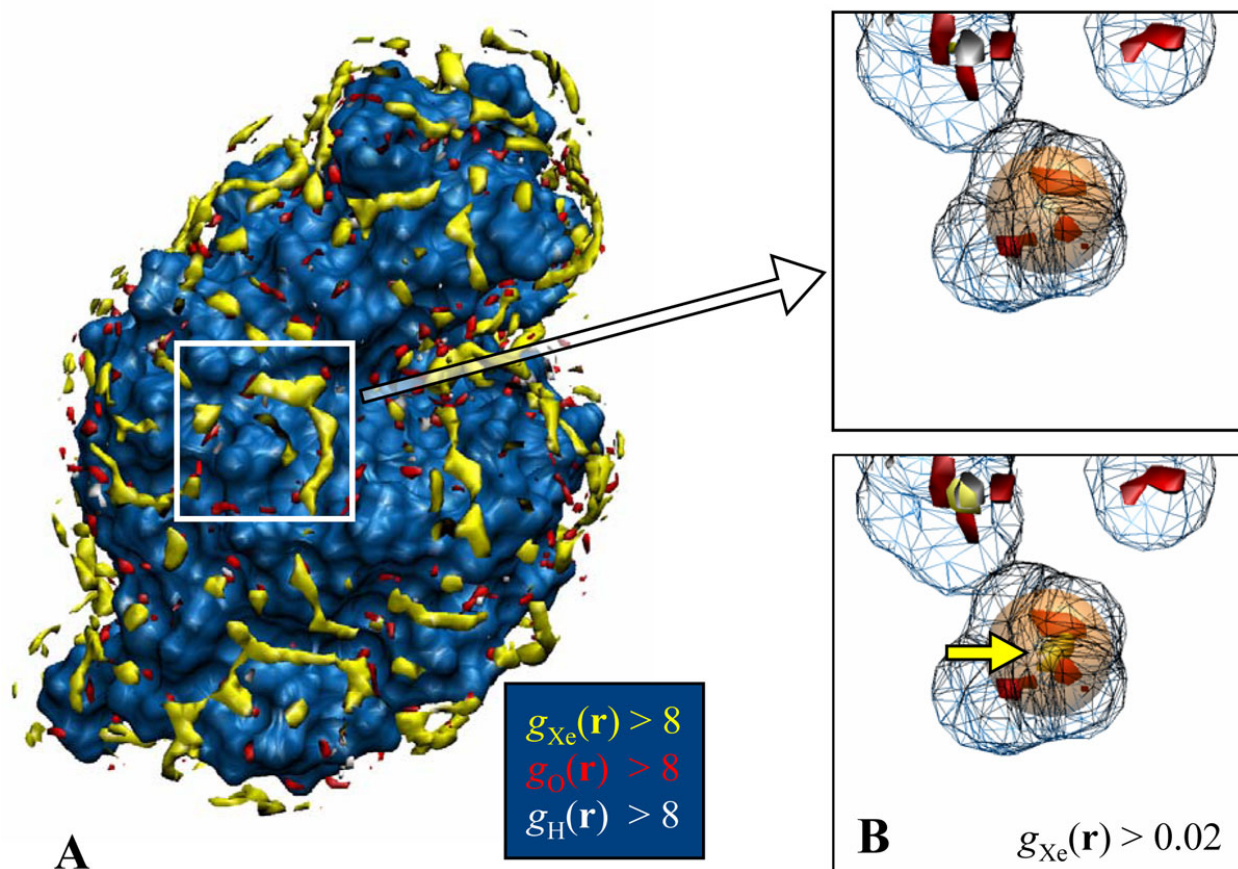
Figure 3. Isosurface plots of the three-dimensional distribution functions, $g(\mathbf{r})$, of xenon and water (oxygen and hydrogen) around lysozyme calculated by the 3D-RISM theory. The yellow, red, and white surfaces (or spots) show the areas where the distribution functions are greater than 8 for xenon, oxygen, and hydrogen, respectively. A. Overall view. The blue surface represents the molecular surface of lysozyme. B. The distributions at the internal xenon-binding site. In the lower panel, the isosurface of $g(\mathrm{r}) > 0.02$ instead of $g(\mathrm{r}) > 8$ is presented for xenon in order to show the minor peak in the region, which is indicated by the yellow arrow. The mesh surface is of the internal cavity. The orange sphere represents the xenon found in the X-ray structure.

4.2. Toward the theoretical prediction of ligand recognition by protein

The 3D-RISM theory is a promising method to investigate the recognition of ligand molecules by a protein. In order to apply the theory to the problem of ligand recognition, we have only to replace water solvent in the calculation described above by aqueous solution of ligand molecules. By analyzing the 3D distributions of ligand and water molecules around the protein, we can find distinct peaks of them at their binding sites. If, for example, the ligand-binding structure bridged by a water molecule is stable, we will find the corresponding peaks of the ligand and water molecules at the binding site. The 3D-RISM theory thus enables us to predict the ligand-binding and the effect of water with atomic resolution both for ligand and water. As the first step in this direction, we applied the theory to a feasible system including simple ligands, which is concerned with the binding of noble gases to lysozyme (42). The following are some results demonstrating the ability of the theory.

Figure 3A shows the 3D distribution functions of xenon and water (oxygen and hydrogen) around lysozyme, obtained by the 3D-RISM/HNC calculation for lysozyme in water-xenon mixture at the concentration of 0.001 M, which is in the same order of the solubility of xenon, 0.0044 M. The yellow, red, and white surfaces indicate the regions where $g(\mathbf{r}) > 8$ for xenon, oxygen, and hydrogen, respectively. Since the binding affinity of xenon is not so high, a number of peaks of xenon are found over the protein surface. Interestingly enough, the well-defined peaks of xenon are separated from those of water. In other words, the theory predicts the preferential binding of ligand or water on each site of the protein surface.

It is meaningful to compare the distribution of xenon obtained by the 3D-RISM theory with the xenon sites in the X-ray structure (43), even though their conditions are different; the former is an aqueous solution under atmospheric pressure, whereas the latter is crystal under xenon gas pressure of 1.2 MPa. Here, we focus our attention on one of the internal cavities, which is a xenon
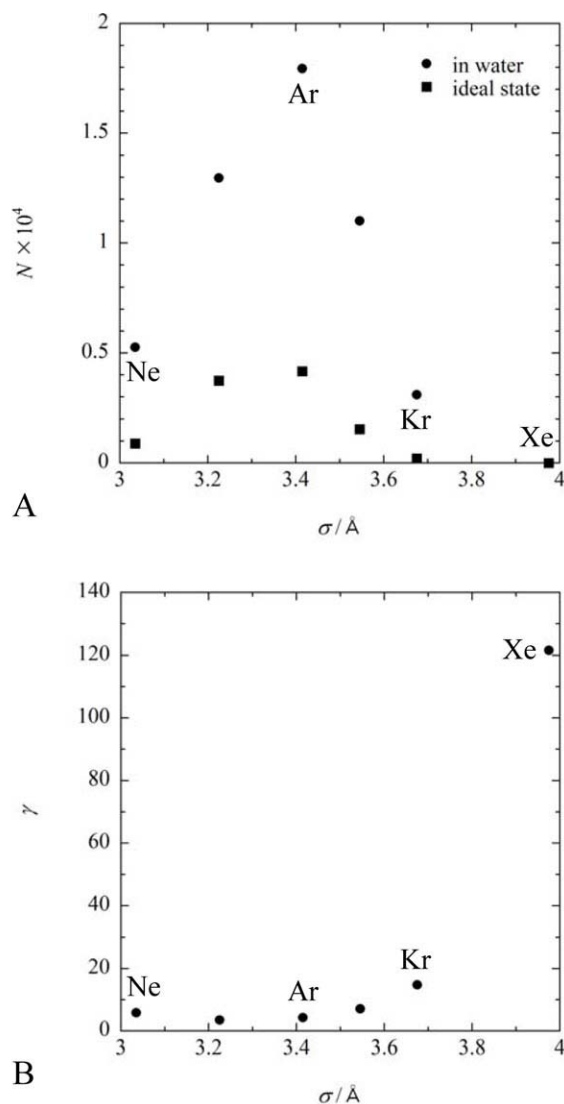
Figure 4. A. Coordination numbers of noble gases at the internal cavity of lysozyme in the 0.001 M aqueous solution and in vacuum (ideal state), plotted against the atomic diameter (Lennard-Jones σ) of the gases. B. The water-effect coefficient γ, which is defined by the ratio of the coordination number in the aqueous solution to that in the ideal state.

binding site found by X-ray analysis. Figure 3B compares the theoretical result of the 3D distribution of xenon with the X-ray xenon site in the cavity. In fact, the xenon peak found there is a minor one probably because of the difference in the system condition; nevertheless, the location is in complete agreement with the X-ray site. It should be noted that the peaks of water are shifted away from the xenon binding site. (It was found that an outstanding peak of xenon distribution at the substrate-binding site was also in good correspondence with the other X-ray xenon site.)

It is more interesting to investigate the size dependence of the coordination number, N, of noble gases at the binding site, because the ligand-size dependence of binding affinity is regarded as the most essential origin of molecular recognition by protein. The coordination numbers of some noble gases are plotted against their atomic diameter (the Lennard-Jones σ parameter) in Figure 4A. The result shows that the coordination number becomes larger with increasing the gas size up to $\sigma \approx 3.4$ Å, while it decreases in the region where $\sigma > 3.4$ Å. As a result, argon has the strongest binding affinity to the internal cavity among the noble gases.

The binding affinity obtained by the 3D-RISM theory includes not only the contribution from the direct interaction between the ligand and protein but also the effect of water. To evaluate the effect of water, here we define a coefficient given by

$$\gamma = \frac{N}{N^{id}} = \frac{\rho_{lig} \int_{V \in \text{binding site}} g_{lig}(\mathbf{r})d\mathbf{r}}{\rho_{lig} \int_{V \in \text{binding site}} g_{lig}^{id}(\mathbf{r})d\mathbf{r}}, \tag{9}$$

where N^{id} and $g_{lig}^{id}(\mathbf{r})$ are, respectively, the coordination number and the distribution function of noble gas (ligand) in the ideal state where no water exists. The ideal distribution function is calculated from the direct interaction potential between the ligand and protein $u_{lig}(\mathbf{r})$ by $g_{lig}^{id}(\mathbf{r}) = \exp\left(-\beta u_{lig}(\mathbf{r})\right)$. Thus, the coefficient γ indicates how significant the effect of water is in increasing ($\gamma > 1$) or decreasing ($\gamma < 1$) the coordination number relative to the ideal state. The size dependence of the ideal coordination number, N^{id}, and the water-effect coefficient, γ, are shown in Figure 4. The result indicates that the coordination number is increased by water and the effect is more enhanced by an increase of the gas size. Since the cavity is hydrophobic, it is reasonable to consider this effect as the hydrophobic effect. The noble gases larger than argon are destabilized by some van der Waals overlap with the protein in the ideal state. Water moderates the destabilization by the hydrophobic effect. As a result, the effect of water shifts the peak of the binding affinity to the larger gas, increasing the overall binding affinity.

Although the present study gives only a simple demonstration, the present approach can be used for other general problems of molecular recognition by protein. Actually, the 3D-RISM theory has succeeded in describing selective ion-binding to mutant human lysozyme molecules (44,45) and also preferential binding of some organic compounds such as urea to staphylococcal nuclease (46). As an advanced application, we can, in theory, deal with the drug design taking the effect of water into account, by replacing noble gases with drug molecules and lysozyme with the receptor protein. However, calculation for a mixture including such large molecules would not be as straightforward in practice and would require further improvement of the site-site description in RISM.

5. PROTEIN FOLDING

5.1. Controversial issue on the roles of water in protein folding

The molecular mechanism of protein folding, especially on the roles of water, has not been completely elucidated. Since a polypeptide chain of protein folds into a unique native structure from a large number of random coil structures, it is evident that the protein undergoes a large loss of the conformational entropy through the folding. Therefore, there must be a certain free energy factor prevailing over the entropy loss to drive protein folding. It was considered that the entropic penalty can be compensated by an energy gain through the formation of intramolecular interactions, such as hydrogen bonding, salt bridging, and van der Waals attraction (2). However, the reduction of protein surface by folding reduces the intermolecular interactions between solvent water and the protein. This gives a large loss in the hydration energy, which competes against the gain in the intramolecular interaction energy. In a typical case, the hydrogen bonds between protein and water are partly replaced with the intraprotein and water-water hydrogen bonds. Thus, it is now considered that the total energetic contribution is almost unchanged upon protein folding (47-49).

The hydrophobic effect was proposed as another possible candidate for a factor suppressing the conformational entropy loss (1,2,50). The water molecules in the vicinity of hydrophobic groups are entropically unfavorable because they are restricted in the rotational motion. When a protein folds, the hydrophobic groups are buried and such entropically unfavorable water molecules are released. It is generally considered that this process can lead to a sufficiently large entropic gain, which compensates for the large loss of the conformational entropy. This is a conventional idea of the hydrophobic effect which is currently prevalent but has not been proved.

Recently, Harano and Kinoshita proposed a new concept for protein folding (51,52). They argued that the translational motion of water molecules drives a protein to fold by lowering the free energy of water. When the folding process is considered under the isochoric (constant volume) condition, the free energy lowering arises from the translational entropy gain for water which is attributable primarily to the decrease in the excluded volume generated by the protein. They evaluated the translational entropy gain upon the folding using the hard-body model for protein and the hard-sphere model for solvent water in order to exclusively investigate the entropic term. Their result demonstrated that the native structure is stabilized by the translational entropy gain for water against the conformational entropy loss for the protein. The change in the translational entropy of water is a major constituent of the hydration entropy. On the other hand, the hydrophobic effect is also related to the hydration entropy. However, the concept proposed by Harano and Kinoshita is substantially different from the traditional idea of the hydrophobic effect. In the former the hydration entropy gain upon the folding stems from the increase in the total volume available to the translational motion of water molecules in the whole system, whereas in the latter it is ascribed mainly to the reduction of the rotational entropy loss for water in the vicinity of nonpolar groups of the protein.

The elucidation of the roles of water in protein folding is thus a controversial issue. Both the energetic and entropic contributions of hydration are still unresolved. Specifically, the following questions should be answered: (i) whether the intramolecular energy gain is canceled by the loss of the hydration energy through the folding process, (ii) whether the hydration entropy is a substantial driving force of the folding, and (iii) what is the physical origin of the hydration entropy. The use of the 3D-RISM theory realizes the direct analysis of the changes in the thermodynamic quantities upon protein folding, which gives the answers to these questions (53).

5.2. Revealing the essential roles of water in protein folding

To answer the above questions, here we consider the free energy change in the transition from random coil (unfolded) structures to the native (folded) structure of protein G, which is a model protein of 56 amino acid residues. The Helmholtz free energy of the protein with a fixed conformation in aqueous solution is given by

$$F = E + \Delta\mu , \qquad (10)$$

where E is the intramolecular interaction energy of protein itself and $\Delta\mu$ is the hydration free energy (HFE). The free energy difference between the native and random coil conformers is expressed by

$$\Delta F = \Delta E - T\Delta S + \Delta\Delta\mu , \qquad (11)$$

where ΔE, ΔS, and $\Delta\Delta\mu$ are the differences in the protein intramolecular energy, the protein conformational entropy, and the HFE, respectively.

The protein intramolecular energy is readily calculated by the ordinary molecular mechanics with a proper force field. The HFE is obtained from the 3D-RISM calculation described above. However, the protein conformational entropy is not treated directly by this theory and therefore should be estimated from somewhere else. Nevertheless, this is not so serious because the entropic contribution $(-T\Delta S)$ is obviously positive as described above and our aim here is primarily to find the free energy factors competing against the positive contribution of the conformational entropy. Therefore, instead of Equation 11, hereafter let us consider the free energy difference excluding the conformational entropy:

$$\Delta F = \Delta E + \Delta\Delta\mu . \qquad (12)$$

The HFE is further decomposed into the protein-water interaction energy $\Delta\varepsilon^{uv}$, the water reorganization energy $\Delta\varepsilon^{vv}$, and the hydration entropy terms Δs:

$$\Delta\mu = \Delta\varepsilon^{uv} + \Delta\varepsilon^{vv} - T\Delta s . \qquad (13)$$
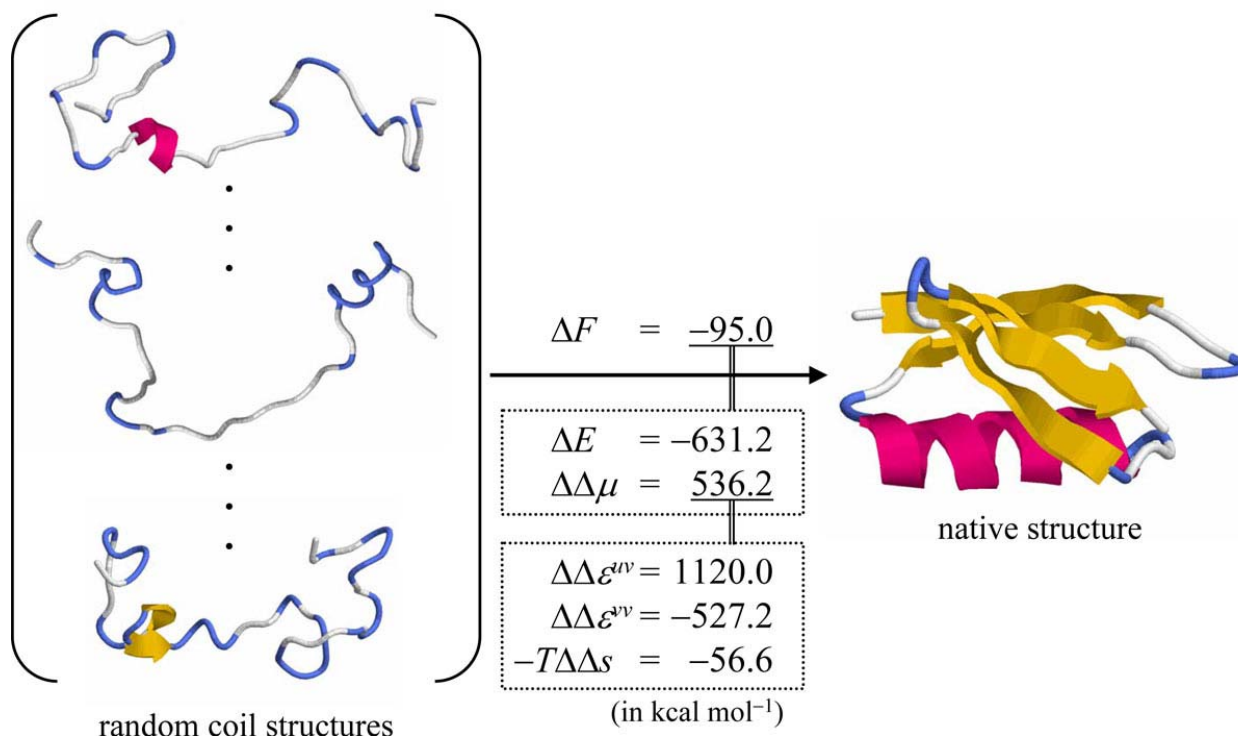
Figure 5. Changes in the free energy components upon the transition from coil to native state of protein G in aqueous solution. The value of each component for coil state is the average value for 32 random coil model structures. F, free energy excluding conformational entropy ($F = E + \Delta\mu$); E, protein intramolecular energy; $\Delta\mu$, hydration free energy ($\Delta\mu = \Delta\varepsilon^{uv} + \Delta\varepsilon^{vv} - T\Delta s$); $\Delta\varepsilon^{uv}$, protein-water interaction energy; $\Delta\varepsilon^{vv}$, water reorganization energy; and Δs, hydration entropy.

Each of these terms is also obtained within the 3D-RISM framework. The hydration entropy is calculated by the temperature derivative of the HFE at constant density of water:

$$\Delta s = -\left(\frac{\partial \Delta\mu}{\partial T}\right)_\rho . \qquad (14)$$

This derivative may be calculated using the first-order finite difference. The protein-water interaction energy is calculated from the 3D potential and distribution function by

$$\Delta\varepsilon^{uv} = \rho \sum_\gamma \int u_\gamma^{uv}(\mathbf{r})g_\gamma^{uv}(\mathbf{r})d\mathbf{r} . \qquad (15)$$

The water reorganization energy is defined by

$$\Delta\varepsilon^{vv} = \frac{\rho^2}{2} \sum_{\alpha,\gamma} \int_0^\infty u_{\alpha\gamma}^{vv}(r)\delta g_{\alpha\gamma}^{vv}(r)4\pi r^2 dr , \qquad (16)$$

where $\delta g_{\alpha\gamma}^{vv}(r)$ is the first-order perturbation to the radial distribution function of water $g_{\alpha\gamma}^{vv}(r)$ due to the insertion of the protein at infinite dilution. This term is the energy arising from the structural changes induced in solvent water. In practice, as the total HFE $\Delta\mu$ is calculated by

Equation 6, the water reorganization energy is more conveniently calculated from the decomposition $\Delta\varepsilon^{vv} = \Delta\mu - (\Delta\varepsilon^{vv} - T\Delta s)$.

Figure 5 illustrates the structural change of protein G along with the changes in the free energy components calculated by the 3D-RISM/HNC theory (53). In the calculation, it was assumed that the values for the native structure are represented by those for a single conformation based on the PDB structure (2GB1) and the values for the random coil structures are approximated by the average values for 32 energy-minimized random coil structures. The total free energy change $\Delta F = -95.0$ kcal mol^{-1} is reasonable because it is comparable in magnitude with the experimental and empirical estimations of the conformational entropy loss of the protein, $-T\Delta S = 41.9$ kcal mol^{-1} (54) and 110.5 kcal mol^{-1} (55,56).

The gain in the protein intramolecular energy $\Delta E = -631.2$ kcal mol^{-1} is largely canceled by the loss of the hydration energy $\Delta\varepsilon = \Delta\varepsilon^{uv} + \Delta\varepsilon^{vv} = 592.8$ kcal mol^{-1}. This result is consistent with the intuitive idea that the protein-water interactions are partly replaced with the intraprotein and water-water interactions through the folding process. However, the energy balance is not complete. The folding results in an energy gain of -38.4 kcal mol^{-1}. Thus, the energy terms as a whole stimulate the protein to fold, even through the effect is much less than expected in vacuum. This is the answer to the first question.
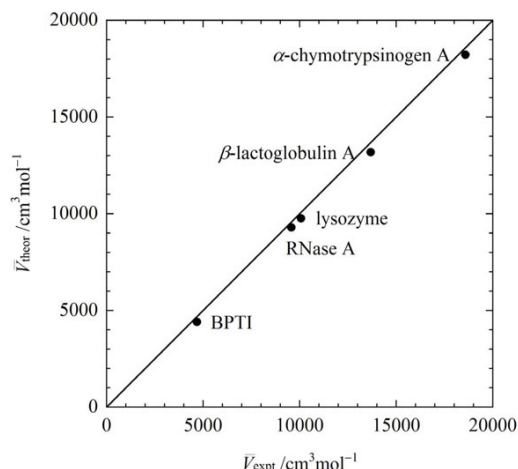
Figure 6. Partial molar volumes of five representative proteins, bovine pancreatic trypsin inhibitor (BPTI), bovine pancreatic ribonuclease (RNase) A, hen egg-white lysozyme, bovine milk β-lactoglobulin A, and bovine pancreatic α-chymotrypsinogen A, calculated by the 3D-RISM theory ($\bar{V}_{theor}$), plotted against the corresponding experimental data ($\bar{V}_{expt}$). The solid line marks the exact coincidence.

The result of the free energy decomposition indicates that the hydration entropy also drives the protein to fold: $-T\Delta s = -56.6$ kcal mol^{-1}. Interestingly enough, the hydration entropy gain is comparable with or even larger than the total energy gain, -38.4 kcal mol^{-1}. This result answers the second question; that is, the hydration entropy has a substantial significance in protein folding.

It is actually not so straightforward to answer the third question. However, the following analysis provides us with a certain way to the answer. Here, let us further decompose the hydration entropy into nonelectrostatic and electrostatic contributions. In the decomposition, first the nonelectrostatic contribution Δs_0 is calculated using a hypothetical protein whose partial charges are completely removed. Then, the electrostatic contribution is obtained by $\Delta s_q = \Delta s - \Delta s_0$. The decomposition shows that the change in the hydration entropy term, $-T\Delta\Delta s = -56.6$ kcal mol^{-1}, is determined mainly by the nonelectrostatic contribution, $-T\Delta\Delta s_0 = -43.9$ kcal mol^{-1}, but much less includes the electrostatic contribution, $-T\Delta\Delta s_q = -12.6$ kcal mol^{-1}. According to previous studies (57,58), the nonelectrostatic contribution $-T\Delta s_0$ of a sufficiently large solute molecule is ascribed to the translational entropy loss for water originating mainly from a reduction of the configurational phase space for water molecules due to the volume exclusion by the solute, rather than the rotational entropy loss due to the reorganization of solvation structure in the vicinity of the solute. On the other hand, the electrostatic contribution $-T\Delta s_q$ is substantially due to the electrostatic confinement or reorganization of hydration structure near the protein. Thus, the result implies that the gain in the hydration entropy is mainly due to a reduction of the excluded volume of the protein through the folding.

In conclusion, a protein folds to reduce the translational entropy loss for water by decreasing the excluded volume of the protein. This mechanism is different from the traditional idea based on the hydrophobic effect. The balance between the intraprotein, protein-water, and water-water interaction energies also results in stabilizing the folded structure of the protein. The energetic contribution is comparable with or somewhat less than the entropic contribution. However, the entropic contribution becomes much more significant as protein is larger, because the entropic contribution originating from the translational entropy loss for water monotonically increases with the protein size or volume, whereas the energetic contribution does not have such the nature of the monotonic dependence on the protein size. The dependence of the protein size on the balance between the energetic and entropic contributions needs to be investigated in the future work.

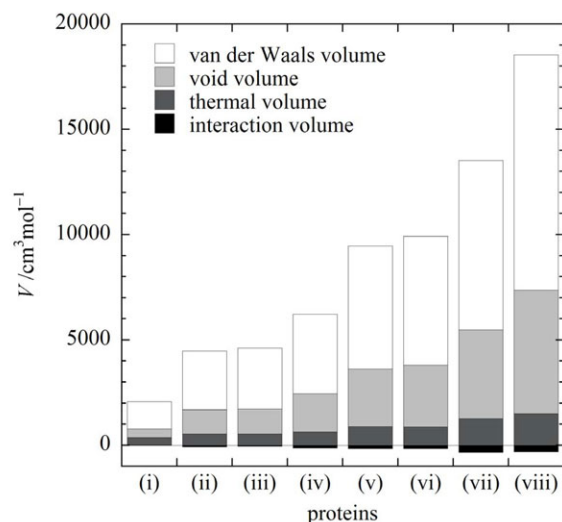## 6. PARTIAL MOLAR VOLUME OF PROTEIN AND PRESSURE-INDUCED STRUCTURAL TRANSITION

6.1. Partial molar volume of protein

The partial molar volume (PMV) of a protein is of principal importance in the analysis of the pressure effects on the structure and function of the protein (59,60), such as pressure-induced denaturation (structural transition) and pressure control of enzyme reaction. That is expressed by the following equation:

$$\left(\frac{\partial \ln K}{\partial p}\right)_T = -\frac{\Delta\bar{V}}{k_B T}, \tag{17}$$

where K is the equilibrium constant and $\Delta\bar{V}$ is the difference in PMV between reactant and product, namely the reaction volume. In the case of protein denaturation, reactant and product correspond to native and denatured states, respectively. According to the relation, the question, "Why and how does pressure denaturalize a protein?" can be converted into "Why and how is the PMV of the denatured structure of the protein smaller than that of the native structure?"

In this context, the problem is how we obtain the PMV of a protein for a given state (structure). Since the PMV of a solute is defined as the volume change of solution when the solute is immersed into the solution, the PMV is affected not only by the intrinsic volume of the solute molecule but also by the solute-solvent interaction or solvation. Therefore, the geometric volume computation (61-63) is not sufficient for the PMV calculation, though it has been conventionally used. As described above, the PMV can be calculated by the 3D-RISM theory coupled with the KB theory (Equation 7). In the calculation, the effect of solvation is naturally included through the correlation functions obtained from the 3D-RISM theory. It has been demonstrated that the 3D-RISM theory can quantitatively reproduce the experimental data of the PMV for biomolecules of various sizes, from amino acids to proteins (27,30,64). Figure 6 provides a comparison between theoretical results and experimental data for the

Figure 7. Volume components in the partial molar volume of eight representative peptides and proteins, (i) honey bee venom melittin, (ii) sea snake venom erabutoxin B, (iii) bovine pancreatic trypsin inhibitor, (iv) human erythrocyte ubiquitin, (v) bovine pancreatic ribonuclease A, (vi) hen egg-white lysozyme, (vii) bovine milk β-lactoglobulin A, and (viii) bovine pancreatic α-chymotrypsinogen A, calculated by the 3D-RISM theory in combination with the geometric volume computation.

PMV of several proteins, exhibiting an almost perfect correspondence of the theory to experiments.

It is essential for understanding the PMV change to extract the effect of water from the PMV. In our analysis, the effect of water is obtained from a theoretical decomposition of the PMV. The decomposition is given by (60,64-66)

$$\bar{V} = V_W + V_V + V_T + V_I + \chi_T^0 k_B T ,\qquad (17)$$

where the volume components are defined and calculated as follows. The van der Waals volume V_W is the volume occupied by the atomic spheres of solute. The void volume V_V is defined as void space inside of the solute and on its surface that the solvent probe cannot access. The sum of them corresponds to the so-called molecular volume of the solute. The two geometric volume terms are obtained from the conventional volume computation, using the atomic diameters converted from the LJ parameters employed in the 3D-RISM calculation (64). The so-called thermal volume V_T is a volume contribution of solvation, which is regarded as average empty space around the solute due to thermally-induced imperfect packing of solvent. In the calculation, it is defined by $V_T = \bar{V}_0 - (V_W + V_V) - \chi_T^0 k_B T$, where $\bar{V}_0$ is the PMV of a hypothetical molecule whose atomic charges are completely removed. The so-called interaction volume V_I is the change in the solvent volume induced by the electrostatic interaction between the solute and solvent. It is obtained by $V_I = \bar{V} - \bar{V}_0$. The PMVs $\bar{V}$

and $\bar{V}_0$ are calculated from the 3D-RISM theory as described above. The last term, $\chi_T^0 k_B T$, is the ideal contribution only from the translational degrees of freedom of the solute. The ideal term is naturally included in the 3D-KB equation (Equation 7).

Figure 7 shows the decomposition of the PMV of several peptides and proteins of different sizes. The ideal term is omitted from the figure because it takes a small constant value of 1 cm^3 mol^{-1} under the thermodynamic condition of ambient water, irrespective of the chemical properties of the solute (38). It is obvious that the van der Waals volume accounts for the most part of the PMV. The void volume also makes up a significant part of the PMV. In contrast, the solvation components, especially the interaction volume, make only a minor contribution to the PMV. However, this result never implies that the solvation is insignificant in analyzing the PMV of proteins. When we consider the volume change upon a conformational transition such as discussed below, it is most likely that the change in the van der Waals volume is negligible unless unusual van der Waals overlaps occur. In such cases, the solvation components become much more significant (66-68).

6.2. Volume change upon the pressure-induced structural transition of protein

The pressure-induced structural transition of proteins has been continuously attracting the attention of biochemists and biophysicists (7,8,59). As described above, the molecular mechanism can be revealed by analyzing the PMV change upon the transition through Equation 17. If both of the low- and high-pressure structures (LPS and HPS) are available, we can readily calculate and analyze the PMV change associated with the structural change by the 3D-RISM theory. There are, however, only a few proteins whose 3D structural data in aqueous solution under high pressure have been resolved. One of them is ubiquitin. The 3D atomic coordinates of the LPS and HPS under 3 MPa and 300 MPa, which are shown in Figure 8A, have been determined by high-pressure NMR technique (69). We have recently applied the volume analysis based on the 3D-RISM theory to the system of ubiquitin, and found out a significant role of water in the PMV change upon the structural transition (68).

The PMV values of the LPS and HPS of ubiquitin calculated by the 3D-RISM/HNC theory were 5788.4 cm^3 mol^{-1} and 5741.2 cm^3 mol^{-1}, respectively. It should be noted that the PMVs were calculated under the ambient pressure of 0.1 MPa both for LPS and HPS in order to consider the two-state LPS-HPS equilibrium on the basis of Equation 17. The PMV decreases by −47.2 cm^3 mol^{-1} upon the structural transition. The volume decomposition analysis indicated that the changes in the van der Waals, void, thermal, and interaction volume components were $\Delta V_W = -3.6$ cm^3 mol^{-1}, $\Delta V_V = -101.3$ cm^3 mol^{-1}, $\Delta V_T = 34.9$ cm^3 mol^{-1}, and $\Delta V_I = 22.8$ cm^3 mol^{-1}. The result demonstrates that the total PMV reduction is primarily caused by the decrease in the void volume, which is partially cancelled by the increases in the

$$\Delta \overline{V} = -47.2$$

$$\Delta V_W = -3.6$$
$$\Delta V_V = -101.3$$
$$\Delta V_T = 34.9$$
$$\Delta V_I = 22.8$$

(in cm³mol⁻¹)
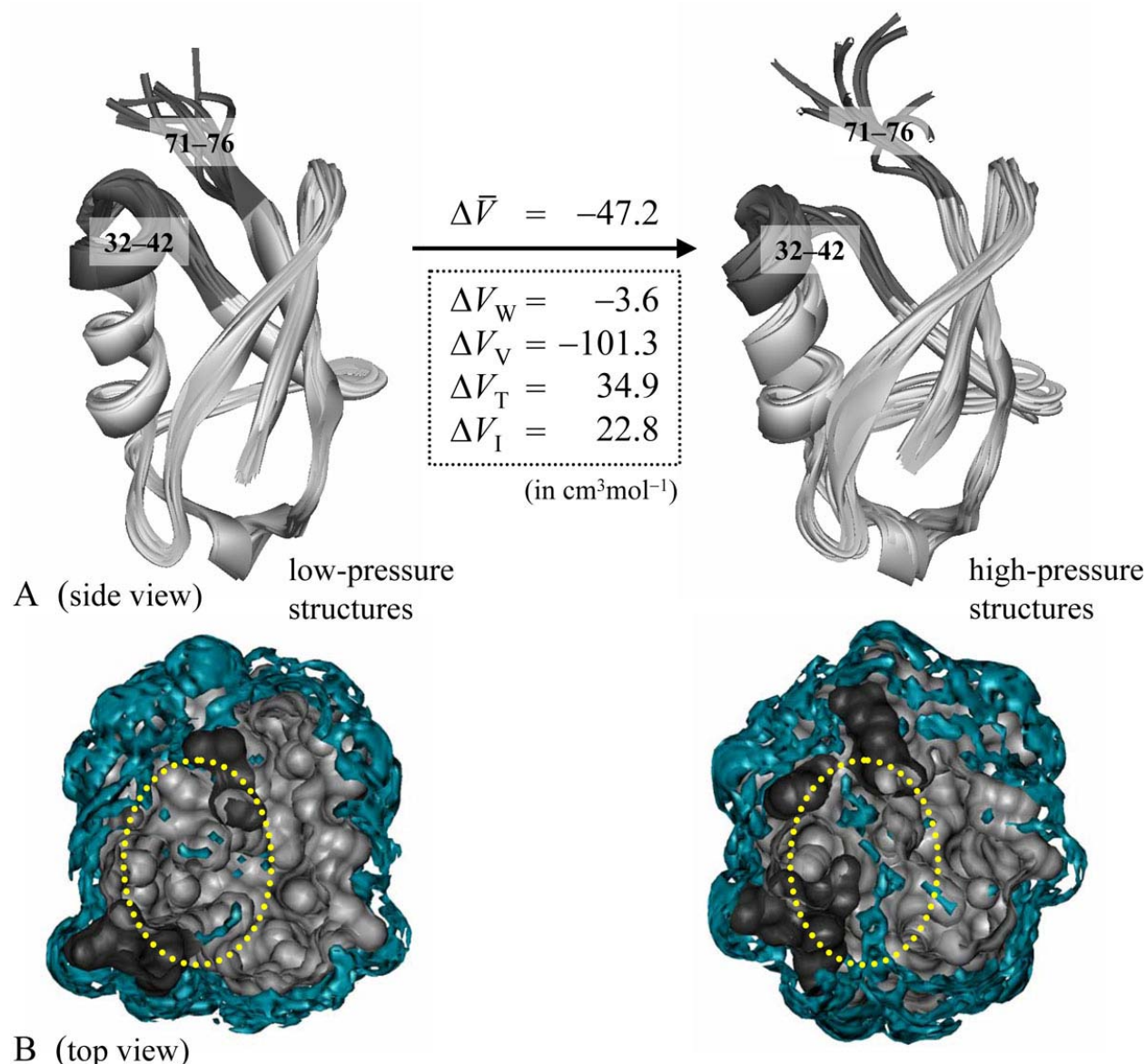
Figure 8. Changes in the structure, in the volume components, and in the distribution of internal water associated with the pressure-induced structural transition of ubiquitin. A. Solid ribbon representation of low-pressure (3 MPa) and high-pressure (300 MPa) structures. Each 10 superimposed structures included in PDB files 1V80 (3 MPa) and 1V81 (300 MPa) are displayed. The dark gray residues are the segments that undergo the largest (71-76) and second largest (32-42) displacement through the transition. The data given are the total change in the partial molar volume ($\overline{V}$) and the changes in the van der Waals (V_W), void (V_V), thermal (V_T), and interaction (V_I) volumes. B. Isosurface representation of the three-dimensional distribution function of water oxygen. The blue surfaces show the areas where the distribution function is larger than 2. The dark gray and light gray surfaces indicate the molecular surfaces of residues 32-42 and the others, respectively. The upper parts (the front parts in the figure) are clipped to bring the internal water distribution (marked by the yellow dotted circle) into view.

two hydration contributions. These volume changes can be related to the changes in the protein structure as follows. The decrease in the void volume indicates partial disappearance of the structural voids in the protein. The increase in the thermal volume implies the generation of additional empty space around the protein, primarily due to the extension of the protein surface (64,66). On the basis of

those finding, we concluded that the PMV reduction was caused by the penetration of water molecules into the protein interior. That is because the water penetration can eliminate the void space within the protein, and at the same time it can expand the protein surface. It is an important point that the former effect makes the total PMV decrease, suppressing the volume increase due to the latter effect.

Furthermore, we estimated the changes in the volume components of some fragments of the protein in order to find which part of the protein exerts the most substantial effect on the volume changes. The analysis indicated that the volume changes were primarily determined by a specific small part of the protein, which corresponds to the segment which has undergone the secondary outstanding displacement through the structural transition, namely the segment of residues 32-42, which is designated in Figure 8A. It follows that the water molecules penetrated into that small area to reduce the PMV of the protein. The water penetration was confirmed in terms of the water distribution. Figure 8B shows isosurface plots of the 3D distribution functions of water oxygen for LPS and HPS, which is also obtained by the 3D-RISM theory. It is apparent that a structural channel is created in the aforementioned part of HPS so that the water distribution is enhanced there. This change in the water distribution obviously indicates the water penetration predicted above by the volume analysis.

In conclusion, the PMV reduction upon the pressure-induced structural transition of ubiquitin is caused by the penetration of water into a specific internal region of the protein, which eliminates the structural voids in the region more than the volume expansion due to the additional hydration. This implies that the application of pressure generally stabilizes the protein structure swelled by the water penetration, which would be the very "pressure-denatured structure" or its precursor. Our conclusion supports the earlier model that the water penetration is a driving force of the pressure-induced denaturation of proteins (7,8,59,70-75).

It should be emphasized that it was essential for the conclusion to connect the water distribution analysis at the atomic level and the volume analysis at the thermodynamic level within the single framework of the 3D-RISM theory. If we observed only the water penetration in HPS, we could not conclude whether the water penetration was the cause or the result of the structural transition; in other words, whether the water penetration induces the structural transition or the structural transition (by other effects) leads to the water penetration. In the present study, we found that the water penetration (atomic-level observation) indeed causes the reduction of the PMV of the protein (thermodynamic-level observation), which induces the structural transition.

7. PERSPECTIVE

We have reviewed some recent applications of the 3D-RISM theory to protein aqueous solution systems. These studies demonstrate that the 3D-RISM theory is useful for understanding the roles of water in protein structure and function, which should be sometimes explained at the atomic level and sometimes at the thermodynamic level. Such a multilevel function of water is hardly approached by the conventional molecular simulation as well as experiments. Therefore, the roles of water have not been sufficiently understood, and what is even worse, it is sometimes misunderstood, even though

the importance has been well recognized since the early stages of protein science. The studies presented here are only the first step toward the complete elucidation of the water-involved molecular mechanisms underlying protein structure and function by the 3D-RISM theory. In the near future, the theory will reveal the roles of water in biological molecular recognition processes such as the binding of drug molecules to the receptor protein, ion permeation in ion channels, and anesthetic action at the molecular level. The findings can be directly applied to pharmaceutical and biomedical studies such as drug design. The entire understanding of the roles of water in the conformational stability of proteins will shed new light on the longstanding issue on protein folding and protein structure prediction.

8. ACKNOWLEDGEMENTS

9. REFERENCES

1. W. Kauzmann: Some factors in the interpretation of protein denaturation. *Adv Protein Chem* 14, 1-63 (1959)

2. K. A. Dill: Dominant forces in protein folding. *Biochemistry* 29, 7133-7155 (1990)

3. K. D. Collins, M. W. Washabaugh: The Hofmeister effect and the behaviour of water at interfaces. *Q Rev Biophys* 18, 323-422 (1985)

4. S. N. Timasheff: The control of protein stability and association by weak interactions with water: How do solvents affect these processes? *Annu Rev Biophys Biomol Struct* 22, 67-97 (1993)

5. C. Scharnagl, M. Reif, J. Friendrich: Stability of proteins: Temperature, pressure and the role of the solvent. *Biochim Biophys Acta* 1749, 187-213 (2005)

6. G. I. Makhatadze, P. L. Privalov: Energetics of protein structure. *Adv Protein Chem* 47, 307-425 (1995)

7. J. L. Silva, G. Weber: Pressure stability of proteins. *Annu Rev Phys Chem* 44, 89-113 (1993)

8. F. Meersman, C. M. Dobson, K. Heremans: Protein unfolding, amyloid fibril formation and configurational energy landscapes under high pressure conditions. *Chem Soc Rev* 35, 908-917 (2006).

9. A. A. Rashin, M. Iofin, B. Honig: Internal cavities and buried waters in globular proteins. *Biochemistry* 25, 3619-3625 (1986)

10. M. A. Williams, J. M. Goodfellow, J. M. Thornton: Buried waters and internal cavities in monomeric proteins. *Protein Sci* 3, 1224-1235 (1994)

11. K. Takano, Y. Yamagata, K. Yutani: Buried water molecules contribute to the conformational stability of a protein. *Protein Eng* 16, 5-9 (2003)

12. H. G. Nagendra, N. Sukumar, M. Vijayan: Role of water in plasticity, stability, and action of proteins: The crystal structure of lysozyme at very low levels of hydration. *Proteins Struct Funct Genet* 32, 229-240 (1998)

13. J. S. Finer-Moore, A. A. Kossiakoff, J. H. Hurley, T. Earnest, R. M. Stroud: Solvent structure in crystals of trypsin determined by X-ray and neutron diffraction. *Proteins Struct Funct Genet* 12, 203-222 (1992)

14. U. Screenivasan, P. H. Axelsen: Buried water in homologous serine proteases. *Biochemistry* 31, 12785-12791 (1992)

15. J. E. Ladbury: Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chem Biol* 3, 973-980 (1996)

16. Z. Li, T. Lazaridis: Water at biomolecular binding interfaces. *Phys Chem Chem Phys* 9, 573-581 (2007)

17. P. Y. S. Lam, P. K. Jadhav, C. J. Eyermann, C. N. Hodge, Y. Ru, L. T. Bacheler, J. L. Meek, M. J. Otto, M. M. Rayner, Y. N. Wong, C.-H. Chang, P. C. Weber, D. A. Jackson, T. R. Sharpe, S. Erickson-Viitanen: Rational design of potent, bioabailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* 263, 380-384 (1994)

18. M. Orozco, F. J. Luque: Theoretical methods for the description of the solvent effect in biomolecular systems. *Chem Rev* 100, 4187-4225 (2000)

19. W. C. Still, A. Tempczyk, R. C. Hawley, T. Hendrickson, Semianalytical treatment of solvation for molecular mechanics and dynamics. *J Am Chem Soc* 112, 6127-6129 (1990)

20. T. Ooi, M. Oobatake, G. Némethy, H. A. Scheraga: Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. *Proc Natl Acad Sci USA* 84, 3086-3090 (1987)

21. W. Wang, O. Donini, C. M. Reyes, P. A. Kollman: Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annu Rev Biophys Biomol Struct* 30, 211-243 (2001)

22. J.-P. Hansen, I. R. McDonald: Theory of simple liquids, 3rd ed, Academic Press, London (2006)

23. D. Chandler, H. C. Andersen: Optimized cluster expansions for classical fluids. II. Theory of molecular liquids. *J Chem Phys* 57, 1930-1937 (1972)

24. F. Hirata, P. J. Rossky: An extended RISM equation for molecular polar fluids. *Chem Phys Lett* 83, 329-334 (1981)

25. F. Hirata: Chemical processes in solution studied by an integral equation theory of molecular liquids. *Bull Chem Soc Jpn* 71, 1483-1499 (1998)

26. F. Hirata (Ed.): Molecular theory of solvation, Kluwer, Dordrecht (2003)

27. Y. Harano, T. Imai, A. Kovalenko, M. Kinoshita, F. Hirata: Theoretical study for the partial molar volume of amino acids and polypeptides by the three-dimensional reference interaction site model. *J Chem Phys* 114, 9506-9511 (2001)

28. D. Beglov, B. Roux: An integral equation to describe the solvation of polar molecules in liquid water. *J Phys Chem B* 101, 7821-7826 (1997)

29. A. Kovalenko, F. Hirata: Three-dimensional density profiles of water in contact with a solute of arbitrary shape: A RISM approach. *Chem Phys Lett* 290, 237-244 (1998)

30. T. Imai, A. Kovalenko, F. Hirata: Solvation thermodynamics of protein studied by the 3D-RISM theory. *Chem Phys Lett* 395, 1-6 (2004)

31. T. Imai, A. Kovalenko, F. Hirata: Hydration structure, thermodynamics, and functions of protein studied by the 3D-RISM theory. *Mol Simul* 32, 817-824 (2006)

32. A. Kovalenko, F. Hirata: Self-consistent description of a metal-water interface by the Kohn-Sham density functional theory and the three-dimensional reference interaction site model. *J Chem Phys* 110, 10095-10112 (1999)

33. A. Kovalenko, F. Hirata: Potentials of mean force of simple ions in ambient aqueous solution. I. Three-dimensional reference interaction site model approach. *J Chem Phys* 112, 10391-10402 (2000)

34. J. G. Kirkwood, F. P. Buff: The statistical mechanical theory of solutions. I. *J Chem Phys* 19, 774-777 (1951)

35. S. J. Singer, D. Chandler: Free energy functions in the extended RISM approximation. *Mol Phys* 55, 621-625 (1985)

36. H.-A. Yu, M. Karplus: A thermodynamic analysis of solvation. *J Chem Phys* 89, 2366-2379 (1988)

37. H.-A. Yu, B. Roux, M. Karplus: Solvation thermodynamics: An approach from analytic temperature derivatives. *J Chem Phys* 92, 5020-5033 (1989)

38. T. Imai, M. Kinoshita, F. Hirata: Theoretical study for partial molar volume of amino acids in aqueous solution:

Implication of ideal fluctuation volume. *J Chem Phys* 112, 9469-9478 (2000)

39. T. Imai, R. Hiraoka, A. Kovalenko, F. Hirata: Water molecules in a protein cavity detected by a statistical-mechanical theory. *J Am Chem Soc* 127, 15334-15335 (2005)

40. T. Imai, R. Hiraoka, A. Kovalenko, F. Hirata: Locating missing water molecules in protein cavities by the three-dimensional reference interaction site model theory of molecular solvation. *Proteins Struct Funct Bioinformat* 66, 804-813 (2007)

41. K. P. Wilson, B. A. Malcom, B. M. Matthews: Structural and thermodynamic analysis of compensating mutations within the core of chicken egg white lysozyme. *J Biol Chem* 267, 10842-10849 (1992)

42. T. Imai, R. Hiraoka, T. Seto, A. Kovalenko, F. Hirata: Three-dimensional distribution function theory for the prediction of protein-ligand binding sites and affinities: Application to the binding of noble gases to hen egg-white lysozyme in aqueous solution. *J Phys Chem B* 111, 11585-11591 (2007)

43. T. Prangé, M. Schiltz, L. Permot, N. Colloc'h, S. Longhi, W. Bourguet, R. Fourme: Exploring hydrophobic sites in proteins with xenon or krypton. *Proteins Struct Funct Genet* 30, 61-73 (1998)

44. N. Yoshida, S. Phongphanphanee, Y. Maruyama, T. Imai, F. Hirata: Selective ion-binding by protein probed with the 3D-RISM theory. *J Am Chem Soc* 128, 12042-12043 (2006)

45. N. Yoshida, S. Phongphanphanee, F. Hirata: Selective ion binding by protein probed with the statistical mechanical integral equation theory. *J Phys Chem B* 111, 4588-4595 (2007)

46. T. Yamazaki, T. Imai, F. Hirata, A. Kovalenko: Theoretical study of the cosolvent effect on the partial molar volume change of staphylococcal nuclease associated with pressure denaturation. *J Phys Chem B* 111, 1206-1212 (2007)

47. B. Honig, A.-S. Yang: Free energy balance in protein folding. *Adv Protein Chem* 46, 27-58 (1995)

48. T. Lazaridis, G. Archontis, M. Karplus: Enthalpic contribution to protein stability: Insights from atom-based calculations and statistical mechanics. *Adv Protein Chem* 47, 231-306 (1995)

49. R. L. Baldwin: In search of the energetic role of peptide hydrogen bonds. *J Biol Chem* 278, 17581-17588 (2003)

50. C. Tanford: Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. *J Am Chem Soc* 84, 4240-4247 (1962)

51. Y. Harano, M. Kinoshita: Large gain in translational entropy of water is a major driving force in protein folding. *Chem Phys Lett* 399, 342-348 (2004)

52. Y. Harano, M. Kinoshita: Translational-entropy gain of solvent upon protein folding. *Biophys J* 89, 2701-2710 (2005)

53. T. Imai, Y. Harano, M. Kinoshita, A. Kovalenko, F. Hirata: Theoretical analysis on changes in thermodynamic quantities upon protein folding: Essential role of hydration. *J Chem Phys* 126, 225102 (2007)

54. J. Fitter: A measure of conformational entropy change during thermal protein unfolding using neutron spectroscopy. *Biophys J* 84, 3924-3930 (2003)

55. J. A. D'Aquino, J. Gómez, V. J. Hilser, K. H. Lee, L. M. Amzel, E. Freire: The magnitude of the backbone conformational entropy change in protein folding. *Proteins Struct Funct Genet* 25, 143-156 (1996)

56. A. J. Doig, M. J. E. Sternberg: Side-chain conformational entropy in protein folding. *Protein Sci* 4, 2247-2251 (1995)

57. M. Kinoshita, N. Matubayasi, Y. Harano, N. Nakahara: Pair-correlation entropy of hydrophobic hydration: Decomposition into translational and orientational contributions and analysis of solute-size effects. *J Chem Phys* 124, 024911 (2006)

58. T. Imai, Y. Harano, M. Kinoshita, A. Kovalenko, F. Hirata: A theoretical analysis on hydration thermodynamics of proteins. *J Chem Phys* 125, 024911 (2006)

59. C. Balny, P. Masson, K. Heremans: High pressure effects on biological macromolecules: from structural changes to alteration of cellular processes. *Biochim Biophys Acta* 1595, 3-10 (2002)

60. T. Imai: Molecular theory of partial molar volume and its applications to biomolecular systems. *Condens Matter Phys* 10, 343-361 (2007)

61. F. M. Richards: Areas, volumes, packing, and protein structure. *Annu Rev Biophys Bioeng* 6, 151-176 (1977)

62. M. L. Connolly: Computation of molecular volume. *J Am Chem Soc* 107, 1118-1124 (1985)

63. J. Liang, H. Edelsbrunner, P. Fu, P. V. Sudhakar, S. Subramaniam: Analytical shape computation of macromolecules: I. Molecular area and volume through alpha shape. *Proteins Struct Funct Genet* 33, 1-17 (1998)

64. T. Imai, A. Kovalenko, F. Hirata: Partial molar volume of proteins studied by the three-dimensional

reference interaction site model theory. *J Phys Chem B* 109, 6658-6665 (2005)

65. T. V. Chalikian, K. J. Breslauer: On volume changes accompanying conformational transitions of biopolymers. *Biopolymers* 39, 619-626 (1996)

66. T. Imai, Y. Harano, A. Kovalenko, F. Hirata: Theoretical study for volume changes associated with the helix-coil transition of peptides. *Biopolymers* 59, 512-519 (2001)

67. T. Imai, T. Takekiyo, A. Kovalenko, F. Hirata, M. Kato, Y. Taniguchi: Theoretical study of volume changes associated with the helix-coil transition of an alanine-rich peptide in aqueous solution. *Biopolymers* 79, 97-105 (2005)

68. T. Imai, S. Ohyama, A. Kovalenko, F. Hirata: Theoretical study of the partial molar volume change associated with the pressure-induced structural transition of ubiquitin. *Protein Sci* 16, 1927-1933 (2007)

69. R. Kitahara, S. Yokoyama, K. Akasaka: NMR snapshots of a fluctuating protein structure: Ubiquitin at 30 bar-3 kbar. *J Mol Biol* 347, 277-285 (2005)

70. G. Hummer, S. Grade, A. E. García, M. E. Paulaitis, L. R. Pratt: The pressure dependence of hydrophobic interactions is consistent with the observed pressure denaturation of proteins. *Proc Natl Acad Sci USA* 95, 1552-1555 (1998)

71. H. Lesch, J. Schlichter, J. Friedrich, J. M. Vanderkooi: Molecular probes: What is the range of their interaction with the environment? *Biophys J* 86, 467-472 (2004)

72. A. Paliwal, D. Asthagiri, D. P. Bossev, M. E. Paulaitis: Pressure denaturation of staphylococcal nuclease studied by neutron small-angle scattering and molecular simulation. *Biophys J* 87, 3479-3492 (2004)

73. M. D. Collins, G. Hummer, M. L. Quillin, B. W. Matthews, S. M. Gruner: Cooperative water filling of a nonpolar protein cavity observed by high-pressure crystallography and simulation. *Proc Natl Acad Sci USA* 102, 16668-16671 (2005)

74. Y. Harano, M. Kinoshita: Crucial importance of translational entropy of water in pressure denaturation of proteins. *J Chem Phys* 125, 024910 (2006)

75. R. Day, A. E. García: Water penetration in the low and high pressure native states of ubiquitin. *Proteins Struct Funct Bioinformat* 70, 1175-1184 (2008)

Send correspondence to: Takashi Imai, Molecular-scale Research and Development Team, Research and Development Group for Next-generation Integrated Living Matter Simulation, Research Program for Computational Science, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan, Tel: 81-48-462-1458, Fax: 81-48-467-4532, E-mail: takashi.imai@riken.jp