Partial Molar Volume of Proteins Studied by the Three-Dimensional Reference Interaction Site Model Theory†

Takashi Imai

Department of Bioscience and Bioinformatics, Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan

Andriy Kovalenko

National Institute for Nanotechnology, National Research Council of Canada, W6-010, ECERF Building, University of Alberta, 9107 116th Street, Edmonton, Alberta T6G 2V4, Canada

Fumio Hirata*

Department of Theoretical Studies, Institute for Molecular Science, Okazaki, Aichi 444-8585, Japan

Received: September 24, 2004; In Final Form: November 24, 2004

The three-dimensional reference interaction site model (3D-RISM) theory is applied to the analysis of hydration effects on the partial molar volume of proteins. For the native structure of some proteins, the partial molar volume is decomposed into geometric and hydration contributions using the 3D-RISM theory combined with the geometric volume calculation. The hydration contributions are correlated with the surface properties of the protein. The thermal volume, which is the volume of voids around the protein induced by the thermal fluctuation of water molecules, is directly proportional to the accessible surface area of the protein. The interaction volume, which is the contribution of electrostatic interactions between the protein and water molecules, is apparently governed by the charged atomic groups on the protein surface. The polar atomic groups do not make any contribution to the interaction volume. The volume differences between low- and high-pressure structures of lysozyme are also analyzed by the present method.

1. Introduction

The partial molar volume (PMV) of protein has drawn much attention as one of the most fundamental thermodynamic quantities to characterize the conformational stability of a protein.[1] PMV has principal importance in the analysis of pressure-induced denaturation of protein from the viewpoint of Le Chatelier's law.[2] As pressure is applied, a protein changes its structure to one with a lesser PMV value.[3] Recently, Akasaka and co-workers[4−6] proposed the "volume theorem" of protein folding based on a high-pressure NMR measurement. They claim that the PMV of a protein generally decreases in parallel with the loss of its conformational order. In other words, PMV can be an order parameter in protein folding. This is a quite controversial statement, but it may put a new insight into the conformational stability of protein if it is fully investigated. Other important progress was made recently by Terazima and co-workers[7,8] concerning the solvation thermodynamics of protein including PMV. They could have realized the time-dependent solvation thermodynamics of protein in solution based on the transient grating spectroscopy. It means that one can measure PMV, for example, of a protein at each intermediate conformation in the course of unfolding or refolding. The latest developments in the experimental study of protein have refreshed the motivation for investigating the PMV of biomolecules as one of the central issues in biophysics.

The way most commonly used to analyze the relation between the PMV and the structure of a protein is to decompose the PMV into several contributions.[9,10] From Chalikian and Breslauer,[10] PMV is decomposed into five terms

$$\bar{V} = V_{id} + V_W + V_V + V_T + V_I \quad (1)$$

where V_{id} is the ideal volume contribution due to the translational degree of freedom of a protein molecule, V_W is the van der Waals volume, V_V is the volume of structural voids within the solvent-inaccessible core, V_T is the so-called thermal volume resulting from thermally induced molecular fluctuations between the solute and solvent molecules, and V_I is the interaction volume representing the change in the solvent volume associated with interactions between solvent molecules and charged and polar groups of the solute molecules. The decomposition provides a molecular picture of the PMV of a protein. However, the determination of each term from experimental data is impossible without rather simplified assumptions. Although V_W and V_V may be obtained from a conventional geometric calculation, the volumetric contributions of hydration, V_T and V_I, have been difficult to obtain from the geometric consideration. Chalikian and Breslauer[10] adopted the assumption that V_T and V_I are proportional to the accessible surface area (ASA) of a protein in order to analyze the volume changes accompanying conformational transitions of proteins. That sounds reasonable from a conceptual viewpoint but has never been confirmed. Reliable estimates of those components from the molecular simulation would be rather hopeless as well, because the quantities are essentially derivatives of the solvation free energy, the evaluation of which is already one of the most difficult tasks for the method. The statistical mechanics of liquids may be the only hope to quantitatively determine V_T and V_I.

* Author to whom correspondence should be addressed. E-mail: hirata@ims.ac.jp.

† Part of the special issue "David Chandler Festschrift".

10.1021/jp045667c CCC: $30.25 © 2005 American Chemical Society
Published on Web 01/15/2005

The statistical mechanics of liquids has a rather long history, but it was relatively recent that a tractable and reasonably rigorous framework of the theory appeared. The theory was based on the Ornstein−Zernike equation complemented with a variety of closure relations such as the Percus−Yevick (PY) and hypernetted chain (HNC) approximations.[11] The theory was successful for describing the structure of liquids in terms of the density pair correlation functions, which serve also as a weighting factor for a statistical average of mechanical quantities to produce thermodynamics. The application of the theory, however, had been largely limited to so-called "simple liquids", systems consisting of spherical molecules, so that the development was conceptual and physical rather than chemical. Due to the limitation, the theory failed to attract much attention from chemists who were interested in, for example, chemical reactions, solvatochromism, and protein stability. There are two ingredients to be considered for a successful theory in chemistry: molecular geometry and charge distribution. In 1972, Chandler and Andersen[12,13] made a crucial step toward the liquid state theory in chemistry presenting the reference interaction site model (RISM) theory, which enables the incorporation of molecular geometry. Nearly a decade later, the theory was extended by Hirata and Rossky[14] to account for the charge distribution of a molecules or the electrostatic interactions, which scored another milestone toward the liquid state theory in chemistry. The theory was applied to a variety of chemical processes in solution including chemical reactions, liquid dynamics, biomolecular solvation, and so on.[15,16] Among the applications of the theory, most challenging without doubt would be the solvation thermodynamics of biomolecules, in particular, protein. Protein has a large number of interaction sites or atoms, which are strongly correlated in terms of interactions with the solvent. The RISM theory with the standard closure relations, HNC or PY, often breaks down in describing the strong correlation, especially for the native state of a globular protein having a core region where the solvent is not accessible. The most conspicuous failure has been inspected in the PMV.

The RISM theory, coupled with the Kirkwood−Buff (KB) solution theory,[17] has been successfully applied to the PMV in a wide variety of small molecular systems, including ions[18,19] and hydrocarbons[20,21] in aqueous solution and other nonpolar solutions.[22,23] However, the RISM theory with the HNC closure has been found to underestimate the PMV of amino acids systematically.[24] It looked hopeless when the theory was applied to protein. We have observed that the PMV starts to decrease with an increasing number of amino acids at some size and even becomes negative in some cases.[25] The observation is unphysical since it is well-known experimentally that the PMV of a protein increases almost linearly with the molecular weight. The problem arises from a superposition approximation with respect to the site−site direct correlation function. An efficient bridge correction improves the estimation of the PMV values of the amino acids, but the correction is not sufficient for larger and bulkier peptides.[26] Recently, we showed that the three-dimensional (3D) generalizations[27−29] of the RISM theory provide a drastic improvement in the quantitative estimation of the PMVs of amino acids.[25] We have further applied the 3D-RISM theory to globular proteins and demonstrated that the PMV of proteins calculated by the theory is in quantitative agreement with the experimental data.[30] The 3D-RISM theory has also been used to decompose the PMV of oligopeptides into the volume components defined by eq 1 in order to investigate the volume change associated with the helix−coil transition.[31]

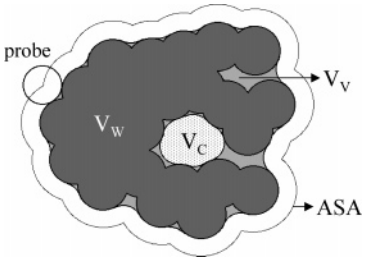

Figure 1. Two-dimensional illustration of the van der Waals volume (V_W), the void volume (V_V), the cavity volume (V_C), and the accessible surface area (ASA).

In this paper, we investigate the PMV of some peptides and proteins, including some proteins studied in our previous paper,[30] using the 3D-RISM-KB theory. PMV is decomposed into the geometric and thermodynamic volume contributions defined by eq 1 to investigate the hydration effects on the PMV. The hydration effects are also investigated in terms of the distribution function of the solvent water. The present method is applied to the analysis of the volume changes associated with a conformational transition from a low-pressure structure to a high-pressure structure as a test case.

2. Methods

2.1. Theoretical Definition of Volume Components. All of the volume components in eq 1 can be obtained from the 3D-RISM-KB theory and from the geometric calculation. They are defined as follows.[31] The ideal volume V_{id} is the ideal gas contribution to the PMV, $\bar{V}$, and is naturally included in the 3D-RISM-KB equation (eq 5, section 2.2). The van der Waals volume V_W is the volume occupied by the van der Waals spheres representing atoms. The void volume V_V is defined as the volume of void space inside or on the surface of the solute molecule to which the solvent probe cannot access. The two geometric volumes, V_W and V_V, are obtained from the geometric calculation (section 2.3). It should be noted that V_V is distinguished from the cavity volume V_C, which is the volume of "void" space to which the solvent cannot access from outside but which can hold the solvent. V_C does not contribute to PMV because the solvent is always distributed in the cavity in the 3D-RISM theory. A two-dimensional representation of these geometric volumes is illustrated in Figure 1. The thermal volume V_T is considered as the volume of voids around the solute molecule due to imperfect packing of the solvent. We define V_T as $\bar{V}_0 - (V_{id} + V_W + V_V)$. $\bar{V}_0$ is the PMV of the hypothetical molecule whose atomic charges are completely removed and thus is essentially the packing contribution of the PMV; the attractive part of the Lennard-Jones interaction makes only a minor contribution. The interaction volume V_I is defined as $\bar{V} - \bar{V}_0$, which is the contribution due to the electrostatic interactions between the solute and the solvent. $\bar{V}$ and $\bar{V}_0$ are calculated by the 3D-RISM-KB theory (section 2.2).

2.2. PMV Calculation by the 3D-RISM Theory. The PMV is calculated by the 3D-RISM[27−29] and Kirkwood−Buff[17] theories. In our procedure, the 3D correlation functions of the solvent around the protein are obtained first using the 3D-RISM theory from the intermolecular potential functions and the thermodynamic conditions such as temperature and solvent density. Then, PMV is calculated from the 3D correlation functions using the Kirkwood−Buff theory. The details have been shown in previous papers. Here, we provide only a brief outline of the theories.

For a solute−solvent system at infinite dilution, the 3D-RISM equation is written as[27−29]

$$h_\gamma(\mathbf{r}) = \sum_{\gamma'} c_{\gamma'}(\mathbf{r}) * (w^{vv}_{\gamma'\gamma}(r) + \rho h^{vv}_{\gamma'\gamma}(r)) \quad (2)$$

where $h_\gamma(\mathbf{r})$ and $c_\gamma(\mathbf{r})$ are, respectively, the 3D total and direct correlation functions of solvent site γ around the solute, $w^{vv}_{\gamma'\gamma}(r)$ and $h^{vv}_{\gamma'\gamma}(r)$ are, respectively, the site−site intramolecular and total correlation functions of solvent, ρ is the number density of solvent, and the asterisk denotes the convolution integral. $h^{vv}_{\gamma'\gamma}(r)$ is obtained independently from a single-component RISM theory. In this calculation, we adopt the DRISM theory[32,33] to ensure the dielectric consistency. The 3D-RISM equation is complemented by a closure equation. We employ the 3D-HNC closure including corrections for the supercell periodicity artifact to the 3D correlation functions.[29] The closure takes the form

$$h_\gamma(\mathbf{r}) = \exp(-\beta u_\gamma(\mathbf{r}) + h_\gamma(\mathbf{r}) - c_\gamma(\mathbf{r}) - \Delta Q_\gamma) + \Delta Q_\gamma - 1 \quad (3)$$

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

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

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

The PMV is calculated through the Kirkwood−Buff theory extended to the 3D-RISM description[25]

$$\bar{V} = k_B T \chi_T^0 (1 - \rho \sum_\gamma \int_{V_{\text{cell}}} c_\gamma(\mathbf{r})) \quad (5)$$

where χ_T^0 is the isothermal compressibility of pure solvent, which is obtained from the site−site direct correlation functions of the pure solvent.[24] The first term $k_B T \chi_T^0$ in eq 5 corresponds to the ideal volume V_{id}.

The 3D potential function $u_\gamma(\mathbf{r})$ was constructed from the site−site pair potential function, which consists of the Lennard-Jones 12−6 and the Coulomb potential terms. The protein−water Lennard-Jones potential parameters were estimated by the Lorentz−Berthelot mixing rules, using the AMBER (parm99/02) parameters[34,35] for the protein sites and the SPC/E parameters[36] for the water sites with a correction concerning the hydrogen Lennard-Jones parameter ($\sigma = 0.4$ Å and $\epsilon = 0.05$ kcal/mol). For the water−water interaction potential, we used the SPC/E model modified with the O−H repulsive term.[37] Other input data to the theory are the number density $\rho = 0.033329$ Å^{-3}, temperature $T = 298.15$ K, and dielectric constant $\epsilon = 78.38$ of ambient water.

The 3D-RISM/HNC equations were solved on a grid of 256^3 points in a cubic supercell of 128 Å^3. The supercell is sufficiently large to hold a protein molecule with enough hydration space. The grid space of 0.5 Å is fine enough to obtain the results without significant numerical errors. The modified direct inversion in the iterative subspace (MDIIS) method[38] was used to converge the equations. The convergence threshold was determined as the convergence error is less than 1 cm^3/mol for the PMV.

TABLE 1: Molecular Characteristics of the Proteins Studied

				fraction (%) of		
	PDB ID	N_{residue}	molwt	α-helix	β-sheet	q_{total}
melittin	2MLT[a]	26	2853	81	0	+6
BPTI	5PTI	58	6518	14	24	+6
erabutoxin B	3EBX	62	6863	0	44	+2
ubiquitin	1UBQ	76	8565	16	32	0
RNase A	8RAT	124	13694	18	33	+4
lysozyme	1HEL	129	14313	31	6	+8
β-lactoglobulin A	1BSY	162	18354	12	38	−9
α-chymotrypsinogen A	2CGA[a]	245	25661	7	29	+4

[a] Chain A used.

2.3. Geometric Volume Calculation. The geometric volumes, V_W and V_V, were calculated by considering Lennard-Jones atoms as hard spheres, with the Alpha Shapes program.[39] For the diameter of atoms (d_{atom}) used in the geometric calculation, we employed the Lennard-Jones σ used in the 3D-RISM calculation. For the diameter of the solvent probe, the Lennard-Jones σ of SPC/E-oxygen, 3.166 Å, was employed. Although the definition maintains consistency between the geometric calculation and the 3D-RISM calculation, the diameter is apparently too large for that of a hard-sphere model because the Lennard-Jones atoms can overlap due to the thermal effect. Therefore, we also calculated the geometric volume with an alternative definition of the diameter. Since the overlap hardly occurs within the distance where the potential energy is much larger than $k_B T$, we determined the diameter at the distance where the potential energy is $k_B T$. In the definition, the diameter of solvent probe is 2.928 Å, which is rather close to the conventional value of 2.8 Å. Hereafter, d_0 denotes the diameter by the former definition, which is, in other words, the distance where the potential energy is zero, and $d_{k_B T}$ denotes the diameter by the later definition.

2.4. Materials. We selected six proteins, bovine pancreatic trypsin inhibitor (BPTI), human erythrocyte ubiquitin, bovine pancreatic ribonuclease (RNase) A, hen egg-white lysozyme, bovine milk β-lactoglobulin A, and bovine pancreatic α-chymotrypsinogen A. Honey bee venom melittin and sea snake venom erabutoxin B, which are small peptides, were also chosen to make a connection between proteins and small peptides. The number of residues ranges widely from 26 to 245. There are variations of the secondary structure, α-helix, β-sheet, or both. The molecular characteristics are compiled in Table 1. The 3D coordinates were taken from the Protein Data Bank, and then hydrogen atoms were added at the proper positions. The 3D structures of the proteins are shown by the space-filling models in Figure 2.

3. Results and Discussion

3.1. Partial Molar Volume Calculated with the 3D-RISM-KB Theory. The PMV values for proteins calculated with the 3D-RISM-KB theory are given in Table 2. It has been known that the PMVs of proteins in their native states are approximately proportional to their molecular weights. To see the relationship, the theoretical values of the PMVs are plotted against the molecular weights, along with the experimental data available, in Figure 3. It is obvious that the PMV in fact has a linear relationship to the molecular weight, regardless of the variety of secondary structure characteristics. Figure 3 also demonstrates excellent agreement between the theoretical values and the experimental data, as has been demonstrated in our previous paper.[30]

A series of negative and negligible deviations from the experimental data arises mainly from the imperfections of

TABLE 2: Partial Molar Volume ($\bar{V}$, cm³/mol) of Proteins in Aqueous Solution at 25 °C and Volume Components[a]

	V	V_W	V_V	V_T	V_I	ASA
			$d_{atom} = d_0$			
melittin	2061	1640	359	52	9	2690
BPTI	4393	3479	980	11	−78	4121
erabutoxin B	4571	3612	987	22	−51	4284
ubiquitin	6082	4746	1470	1	−136	4785
RNase A	9300	7274	2194	−5	−164	6862
lysozyme	9769	7635	2318	−19	−166	6690
β-lactoglobulin A	13189	10210	3520	−203	−339	8248
α-chymotrypsinogen A	18222	14001	4706	−179	−307	10480
			$d_{atom} = d_{k_BT}$			
melittin	2061	1294	403	354	9	2612
BPTI	4393	2786	1152	532	−78	4132
erabutoxin B	4571	2905	1185	531	−51	4262
ubiquitin	6082	3776	1800	641	−136	4839
RNase A	9300	5843	2740	880	−164	6921
lysozyme	9769	6126	2938	870	−166	6746
β-lactoglobulin A	13189	8057	4205	1265	−339	8529
α-chymotrypsinogen A	18222	11173	5866	1489	−307	10787

[a] The van der Waals volume (V_W), void volume (V_V), thermal volume (V_T), interaction volume (V_I), and accessible surface area (ASA, Å²) are given. The ideal volume (V_{id}) is 1 cm³/mol at the temperature. The volume components obtained from the geometric calculation using the diameters d_0 and d_{k_BT} are given in the upper and lower rows, respectively.

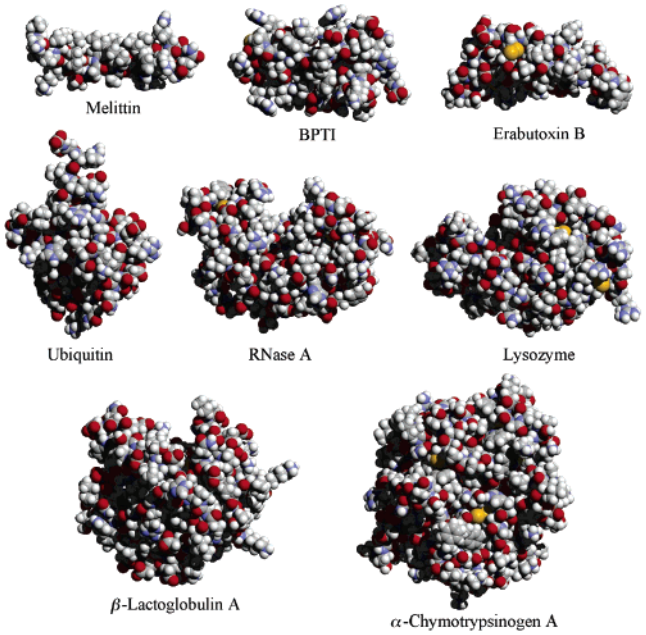

Figure 2. Three-dimensional structures of the proteins studied, represented by the standard space-filling model.

potential parameters and the disregard for intramolecular fluctuations of proteins. The former is not concerned with the theory. It is to be developed in the field of molecular simulation, independent of the theory of statistical mechanics. The later cannot be ignored theoretically since an experimentally observed value of the PMV involves the effects of intramolecular fluctuations. The observable PMV can be calculated by taking the ensemble average of the PMV along the conformational space of the protein molecule.[21] However, the negligible deviation implies that intramolecular fluctuation within the native state may be ignored practically in the analysis of the PMV.

3.2. Analyses of Volume Components. The volume components of proteins in eq 1 are also listed in Table 2, together with the ASA. The ideal term V_{id} is omitted from the table because it takes a constant value of 1 cm³/mol at a temperature of 25 °C irrespective of any properties of the protein. As has been described in section 2.3, we adopted two different diameters, d_0 and d_{k_BT}, as the atomic diameters used in the geometric calculations. By definition, the selection of the diameter does not concern the 3D-RISM calculation; therefore it does not affect the PMV $\bar{V}$ and the interaction volume V_I. Although the van der Waals volume V_W and the void volume V_V are not qualitatively affected by the diameter, the quantity and quality of the thermal volume V_T depend on the diameter, as detailed below.

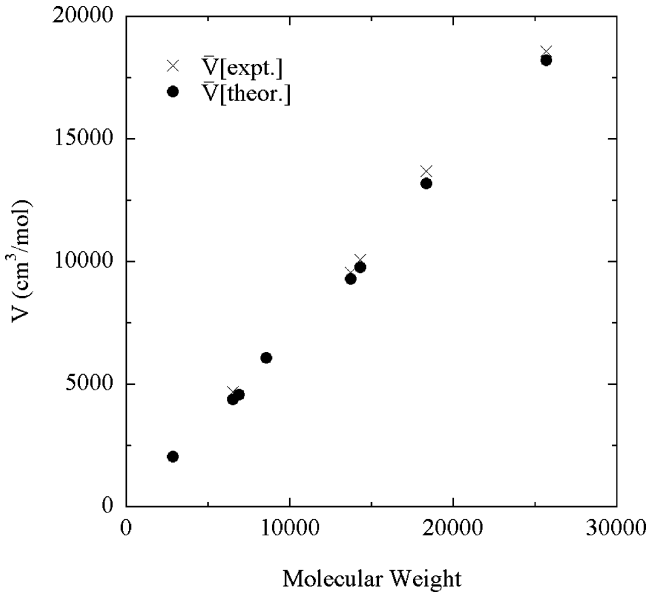

Figure 3. The partial molar volume ($\bar{V}$) calculated with the 3D-RISM-KB theory and the corresponding experimental data versus the molecular weight.

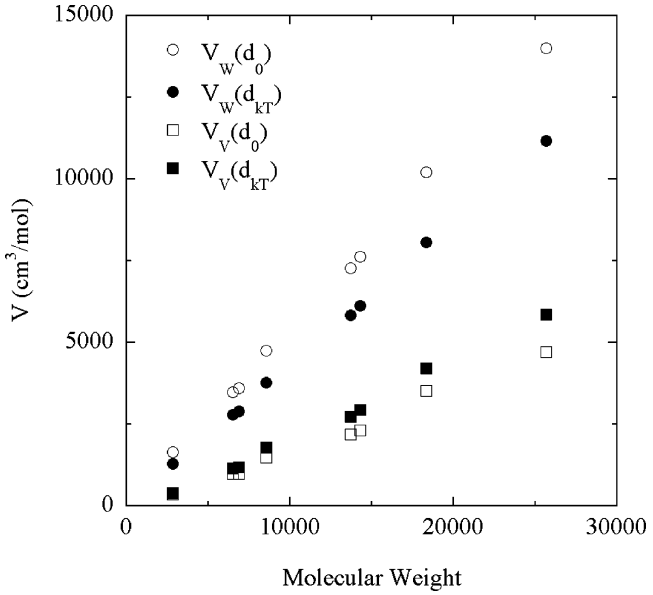

Figure 4. The van der Waals volume (V_W) and the void volume (V_V) of proteins versus the molecular weight: open marks, values calculated using d_0; filled marks, those using d_{k_BT}.

3.2.1. Geometric Volumes. Figure 4 shows the dependence of the van der Waals volume V_W on the molecular weight. It is quite natural that V_W has a linear relationship to the molecular weight because V_W is only constituted by the atomic volumes and is intrinsically independent of the protein structure. The linear relationship is irrespective of which definition of diameter

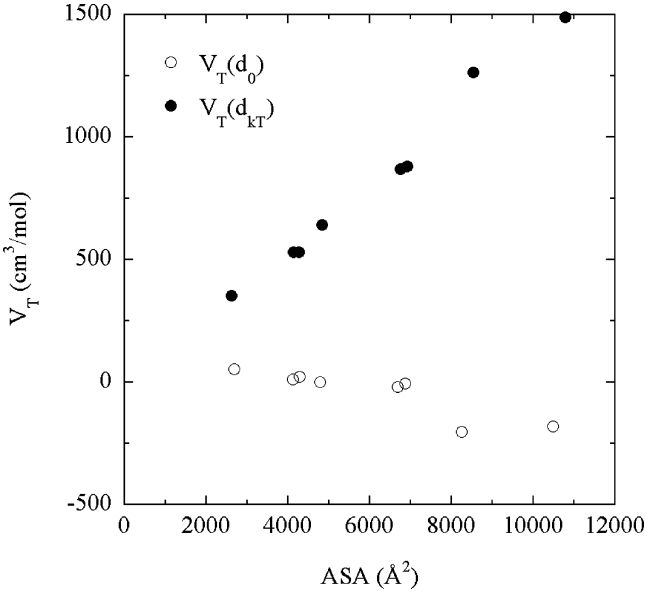

Figure 5. The thermal volume (V_T) of proteins versus the accessible surface area (ASA): open marks, values calculated using d_0; filled marks, those using d_{k_BT}.

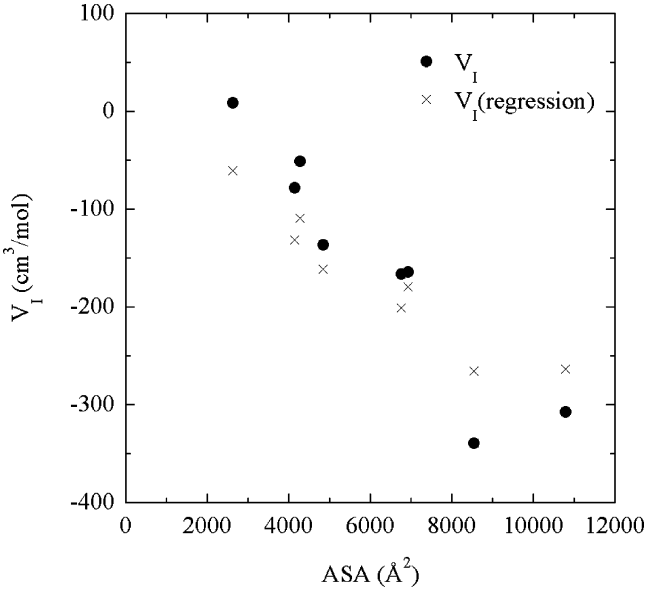

Figure 6. The interaction volume (V_I) of proteins versus the accessible surface area (ASA). Crosses indicate the values calculated by the linear multiple regression equation (see text).

is used. A similar result has also been observed using another parameter set.[40] The proportionality coefficients for using d_0 and d_{k_BT} are 0.544 and 0.434 cm^3/g, respectively.

The void volume V_V is also plotted against the molecular weight in Figure 4. V_V is also proportional roughly to the molecular weight, for both d_0 and d_{k_BT}, as well as for another parameter set.[40] That is however nontrivial because V_V presumably depends on structural characteristics of the protein. For instance, when a protein is completely extended, V_V becomes much smaller because the inside voids vanish.[31] The proteins employed in this study have different secondary structures, namely, α-helix, β-sheet, or both, but have a similar tertiary structure, "globule". The result therefore indicates that V_V does not much depend on the secondary structure, even though it does on the tertiary structure. For those globular proteins, our calculation shows the proportionality coefficients of 0.177 and 0.218 cm^3/g for d_0 and d_{k_BT}, respectively.

3.2.2. Thermal Volume. The thermal volume V_T is plotted against the ASA in Figure 5 to confirm the conventional assumption of a linear relationship between them.[10] V_T calculated using d_{k_BT} ($V_T(d_{k_BT})$) has a positive linearity, which agrees with the conventional assumption. However, $V_T(d_0)$ seems to have a slightly negative one. The difference in the relationship arises from the difference in the quality. In fact, $V_T(d_0)$ contains two different thermal effects on the volume. One is the volume of voids around the solute molecule created by thermal vibrations or fluctuations of solvent molecules, which is the original meaning of the thermal volume. The other is the volume of thermally induced overlaps between the solute and solvent molecules. The former gives a positive contribution to V_T, while the latter gives a negative contribution. Thus, the two opposite contributions are compensated with each other in $V_T(d_0)$. However, $V_T(d_{k_BT})$ excludes the overlap effect since the overlap hardly occurs within the distance at which the potential is higher than k_BT. Thus, $V_T(d_{k_BT})$ coincides with the original definition of the thermal volume. The analysis suggests that d_{k_BT} should be used as the atomic diameter when we examine the PMV based on the decomposition given by eq 1. (Our previous definition of d_{atom},[31] in which d_0 was used for a solute atom and a conventional value of 2.8 Å is used for the solvent probe, provided a worse result in the linearity. Data not shown.)

Finally, we conclude that V_T is directly proportional to the ASA when we adopt an appropriate definition of the atomic diameter. By use of d_{k_BT}, the proportionality coefficient is 0.226 Å. The linear relationship indicates that the complex shape of the protein does not complicate the thermal effect on the PMV or the details are averaged over a whole protein. If we completely accepted the molecular picture of V_T that the voids are thermally created between the protein molecule and the water molecules, then the thickness of the "empty layer" would be 0.226 Å. The value is on the same order of magnitude as a value, 0.5 Å, which has been obtained with a similar definition from experimental PMV data for small molecules.[41]

3.2.3. Interaction Volume. The interaction volume V_I is also plotted against the ASA in Figure 6. There seems to be a negative correlation between V_I and the ASA. However, it shows a much larger dispersion than that of V_T. Figure 6 also indicates that V_I is not governed by the total charge of the protein, which is given in Table 1. For instance, ubiquitin, which is a neutral protein, has a somewhat larger absolute value of V_I than those of the other charged proteins with the similar ASA values. It is, however, expected that charged groups have a significant influence on the V_I value through local-hydration-structure changes due to a kind of electrostriction, especially when they are located on the surface of the protein molecule.

To confirm the correlation between V_I and the chemical or electrostatic properties of a protein surface, we have calculated the ASA contributions of the charged (S_c), polar (S_p), and nonpolar (S_n) atomic groups. Here, the charged group includes the carboxyl group (COO^-) of Asp and Glu, the amino group (NH_3^+) of Lys, the guanidino group ($NHC(NH_2)_2^+$) of Arg, and the zwitterionic terminal groups (NH_3^+ and COO^-). The polar group contains the hydroxyl group (OH) of Ser, Thr, and Tyr, the carboxamide group ($CONH_2$) of Asn and Gln, the amide group (NH) in aromatic ring of His and Trp, and the main-chain peptide group (CONH). The other atoms compose nonpolar groups. The ASA contributions are listed in Table 3. It was found that the proteins having relatively larger absolute values of V_I than those with similar sizes show a relatively higher degree of exposure of charged groups. Uncharged groups do not apparently affect the properties of V_I. Assuming the linearity to each surface area, V_I may be expressed by the

TABLE 3: Accessible Surface Areas of the Charged (S_c), Polar (S_p), and Nonpolar (S_n) Atomic Groups[a]

	S_c		S_p		S_n	
melittin	341	(13)	501	(19)	1770	(68)
BPTI	1084	(26)	838	(20)	2210	(53)
erabutoxin B	636	(15)	1341	(31)	2285	(54)
ubiquitin	1354	(28)	1237	(26)	2248	(46)
RNase A	1051	(15)	2233	(32)	3638	(53)
lysozyme	1450	(21)	2142	(32)	3153	(47)
β-lactoglobulin A	2136	(25)	1793	(21)	4600	(54)
α-chymotrypsinogen A	1363	(13)	3706	(34)	5719	(53)

[a] The values in parentheses are the percentages of the areas in whole protein surface.

equation: $V_I(\text{Å}) = c_cS_c + c_pS_p + c_nS_n$.[40] We have tried calculating the coefficients by a linear multiple regression analysis using our data. The obtained values of c_c, c_p, and c_n are −0.128, −0.038, and −0.021 Å, respectively. The values calculated by the regression equation are also given in Figure 6. The equation cannot completely reproduce the original data, and therefore each value itself has no substantial meaning. Nevertheless, the analysis suggests the significance of charged groups on the surface but not of polar groups. In fact, the contribution of polar groups is equally negligible as that of nonpolar groups. That agrees with our previous results for smaller molecules.[24,31]

The contribution of a charged group actually exhibits more complicated effects than that considered above. The effect is reduced when charged groups with opposite signs are close to each other, while it is enhanced when charged groups with the same signs assemble, though such a structure is not very stable. If the protein solution included electrolytes, then they would decrease the effect of charged groups, depending on the concentration. Such complicated effects are hardly expressed by any empirical method. V_I obtained using the 3D-RISM theory includes the effects naturally.

3.3. Hydration Shell Analysis. The difference between charged and polar groups in their contributions to the interaction volume V_I is further investigated in terms of the distribution function of solvent water. PMV can be expressed using the radial distribution function (RDF) instead of the 3D direct correlation function, as follows

$$\bar{V} = k_B T \chi_T^0 - \int_0^\infty (g_{X\gamma}(r) - 1)\, d\mathbf{r} \qquad (6)$$

where $g_{X\gamma}(r)$ is the RDF between an arbitrary site (X) of the solute and an arbitrary site (γ) of the solvent. From the definition, V_I is given by

$$V_I = -\int_0^\infty \Delta g_{X\gamma}(r)\, d\mathbf{r} \qquad (7)$$

where Δg denotes the change in the RDF by the introduction of the atomic charges or the RDF difference between the "real" molecule and the hypothetical one without the atomic charges. To analyze the contribution from the local hydration shell,[19,42] we use the function defined by

$$V_I(\lambda) = -\int_0^\lambda \Delta g_{XO}(r)\, d\mathbf{r} \qquad (8)$$

where the water oxygen is adopted as a solvent site because the distribution of the water hydrogens does not give immediate effects on volumetric properties.

Figure 7a shows the RDFs, which are reproduced from the 3D distribution functions, of $X = N^\zeta$ (NH_3^+ group) in Lys26, C^δ (COO^- group) in Glu7, C^δ ($CONH_2$ group) in Gln31, and

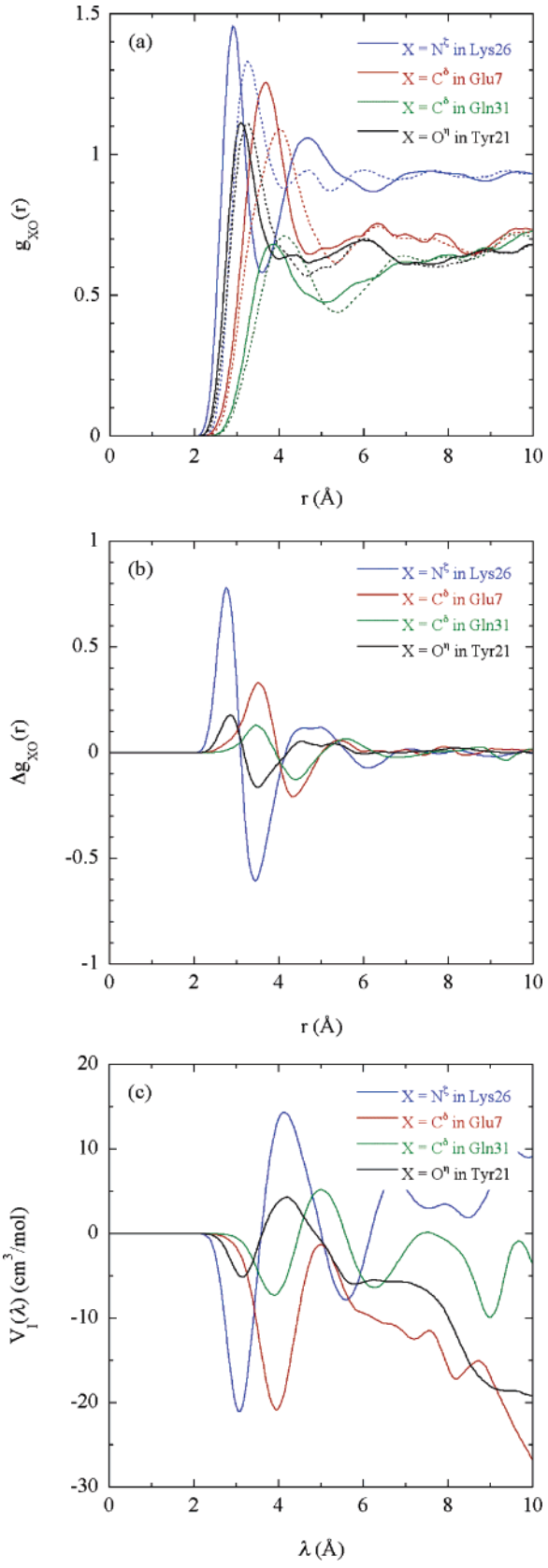

Figure 7. (a) Radial distribution functions (RDFs) of water oxygen around BPTI sites, reproduced from the 3D distribution function: solid line, the real molecule; dotted line, the hypothetical molecule without the atomic charges. (b) RDF changes by the atomic charges. Each of them is the difference between the solid and the dotted lines in part a. (c) Local hydration shell contributions to the interaction volume V_I (eq 8).

TABLE 4: Partial Molar Volume ($\bar{V}$, cm^3/mol) and Volume Components of Lysozyme for the Structures at Pressures of 30 and 2000 bar[a]

	PDB ID	$\bar{V}$	V_W	V_V	V_T	V_I	ASA
30 bar	lGXV	9416	6082	2581	809	−57	6494
2000 bar	lGXX	9296	6048	2493	803	−49	6500
difference		−120	−34	−88	−6	8	6

[a] The van der Waals volume (V_W), void volume (V_V), thermal volume (V_T), interaction volume (V_I), and accessible surface area (ASA, Å^2) are given. The ideal volume (V_{id}) is 1 cm^3/mol. The geometric volumes are calculated using the diameters d_{k_BT}.

O$^\eta$ (OH group) in Tyr21 for BPTI, as examples, together with corresponding the RDFs for the hypothetical BPTI without atomic charges. They are chosen because they are the atoms exposed most in each of the charged and polar groups in BPTI. The differences Δg are given in Figure 7b. By introduction of the atomic charges, the hydrating water molecules come closer to the protein sites for both the charged and the polar groups. The first peaks of the charged groups become much higher, whereas those of the polar groups do not change so much. As a result, the change in hydration in the immediate vicinity of the charged groups makes a large contribution to V_I, as shown in Figure 7c. The contribution from the hydration around the polar groups is similar in quality but is much less in magnitude. The conclusion from the hydration shell analysis is consistent with the result of the statistical analysis presented above.

3.4. Volume Changes Associated with Conformational Transition. In this subsection, we investigate the volume changes of lysozyme associated with a conformational transition induced by pressure as a test case. Recently, low- and high-pressure structures of the protein in solution have been determined by Refaee et al.[43] by means of a 2D-NMR analysis. The experiment naturally concludes that the PMV of the high-pressure structure (HPS) is less than that of the low-pressure structure (LPS), based on the thermodynamic relationship[2]

$$-k_BT\left(\frac{\partial \ln K}{\partial P}\right)_T = \Delta\bar{V} \tag{9}$$

where K is the equilibrium constant between HPS and LPS and $\Delta\bar{V}$ is the PMV difference between the two conformations. However, it is not a trivial task to confirm the conclusion by actually measuring the PMV, because HPS is not an equilibrium conformation at a low-pressure condition. In this respect, the theory has a definite advantage in determining the PMV of the hypothetical conformation, although it is a nontrivial question if the theory actually provides a lesser PMV for the HPS compared to that of the LPS.

The theoretical results for the PMV obtained from the 3D-RISM-KB method are 9416 and 9296 cm^3/mol, respectively, for LPS and HPS at atmospheric pressure. The PMV decreases by 120 cm^3/mol upon the transition from LPS to HPS, which is consistent with the experimental observation stated above. The result demonstrates the reliability of the theory for analyzing the conformational response of proteins to pressure.

The PMV is decomposed into the volume contributions using the method presented in the previous section. The values of volume components are listed in Table 4. The dominant component in the PMV change is the void volume V_V. Structural voids in the interior of lysozyme are compressed by pressure, and the volume is reduced. The van der Waals volume V_W also decreases to some extent. Although conformational changes by other perturbations, including temperature change, pH change, and the addition of denaturant, would not involve a change in V_W, pressure reduces V_W by making atoms overlap against the repulsive potential. The hydration effects, represented by the thermal volume V_T and the interaction volume V_I, make little contribution to the PMV change in the present systems. That is because the whole protein structure does not change much by the present pressure and only local compression takes place. Thus, the result shows the volume change at an initial step in pressure-induced denaturation processes. Applying higher pressure may cause more global changes in the protein structure. In such cases, it is more likely that the hydration contributions become significant. However, the problem cannot be investigated at this stage because such unfolded structures are not available in the database. It is our plan in the near future to study the problem by coupling the 3D-RISM method with molecular simulation techniques to sample various structures of proteins, including non-native structures.

4. Concluding Remarks

This study presents the first theoretical attempt to decompose the partial molar volume (PMV) of proteins into several contributions from molecular geometry and hydration using the 3D-RISM method. We have succeeded in decomposing the PMV into the ideal volume, the van der Waals volume, the void volume, the thermal volume, and the interaction volume. The van der Waals and the void volumes have linear relationships with the molecular weight. The thermal volume is directly proportional to the accessible surface area (ASA) of the protein, when we adopt an appropriate definition of the diameter of atoms used in the geometric consideration. We propose d_{k_BT}, the distance at which the potential energy is equal to k_BT, as the diameter. The interaction volume has a correlation with the ASA of the charged atomic groups. Polar groups make very little contribution to the interaction volume as nonpolar groups do. The hydration shell analysis also shows that the effect of the charged groups is much larger than that of polar groups.

The conformational response of lysozyme to pressure was successfully predicted by the theory in terms of the difference of the PMV between the low- and high-pressure structures. The theoretical decomposition of the PMV is even more useful to investigate the volume changes associated with structural changes and ligand binding. It is also interesting to clarify the relation between the volume changes and the hydration-structure changes accompanying such processes. Applications of the methods presented here to such issues are now in progress.

Acknowledgment. The authors thank Y. Maruyama for his help in computations. This work is supported in part by a grant from the NAREGI Nanoscience Project of the Japanese Ministry of Education, Culture, Sports, Science and Technology (MONBUKAGAKUSHO) and a Grant-in-Aid for Scientific Research on Priority Area of "Water and Biomolecules" from the ministry. T.I. is also grateful for the support by a Grant-in-Aid for Young Scientists from the ministry.

References and Notes

(1) Chalikian, T. V. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 207.
(2) Balny, C.; Masson, P.; Heremans, K. *Biochim. Biophys. Acta* **2002**, *1595*, 3.
(3) Royer, C. A. *Biochim. Biophys. Acta* **2002**, *1595*, 201.
(4) Kitahara, R.; Yamada, H.; Akasaka, K.; Wright, P. E. *J. Mol. Biol.* **2002**, *320*, 311.
(5) Kitahara, R.; Akasaka, K. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3167.
(6) Akasaka, K. *Biochemistry* **2003**, *42*, 10876.
(7) Takeshita, K.; Hirota, N.; Imamoto, Y.; Kataoka, M.; Tokunaga, F.; Terazima, M. *J. Am. Chem. Soc.* **2000**, *122*, 8524.

(8) Terazima, M. *Bull. Chem. Soc. Jpn.* **2004**, *77*, 23.
(9) Kauzmann, W. *Adv. Protein Chem.* **1959**, *14*, 1.
(10) Chalikian, T. V.; Breslauer, K. J. *Biopolymers* **1996**, *39*, 619.
(11) Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids*, 2nd ed.; Academic Press: London, 1986.
(12) Chandler, D.; Andersen, H. C. *J. Chem. Phys.* **1972**, *57*, 1930.
(13) Chandler, D. *J. Chem. Phys.* **1973**, *59*, 2749.
(14) Hirata, F.; Rossky, P. J. *Chem. Phys. Lett.* **1981**, *83*, 329.
(15) Hirata, F. *Bull. Chem. Soc. Jpn.* **1998**, *71*, 1483.
(16) *Molecular Theory of Solvation*; Hirata, F., Ed.; Kluwer: Dordrecht, 2003.
(17) Kirkwood, J. G.; Buff, F. P. *J. Chem. Phys.* **1951**, *19*, 774.
(18) Chong, S.-H.; Hirata, F. *J. Phys. Chem. B* **1997**, *101*, 3209.
(19) Imai, T.; Nomura, H.; Kinoshita, M.; Hirata, F. *J. Phys. Chem. B* **2002**, *106*, 7308.
(20) Leu, L.; Blankschtein, D. *J. Phys. Chem.* **1992**, *96*, 8582.
(21) Imai, T.; Hirata, F. *J. Chem. Phys.* **2003**, *119*, 5623.
(22) Ohba, M.; Kawaizumi, F.; Nomura, H. *J. Phys. Chem.* **1992**, *96*, 5129.
(23) Amakasu, Y.; Ohba, M.; Kawaizumi, F.; Nomura, H. *J. Phys. Chem.* **1995**, *99*, 9258.
(24) Imai, T.; Kinoshita, M.; Hirata, F. *J. Chem. Phys.* **2000**, *112*, 9469.
(25) Harano, Y.; Imai, T.; Kovalenko, A.; Kinoshita, M.; Hirata, F. *J. Chem. Phys.* **2001**, *114*, 9506.
(26) Kinoshita, M.; Imai, T.; Kovalenko, A.; Hirata, F. *Chem. Phys. Lett.* **2001**, *348*, 337.
(27) Beglov, D.; Roux, B. *J. Phys. Chem. B* **1997**, *101*, 7821.
(28) Kovalenko, A.; Hirata, F. *Chem. Phys. Lett.* **1998**, *290*, 237.
(29) Kovalenko, A.; Hirata, F. *J. Chem. Phys.* **2000**, *112*, 10391.
(30) Imai, T.; Kovalenko, A.; Hirata, F. *Chem. Phys. Lett.* **2004**, *395*, 1.
(31) Imai, T.; Harano, Y.; Kovalenko, A.; Hirata, F. *Biopolymers* **2001**, *59*, 512.
(32) Perkyns, J. S.; Pettitt, B. M. *Chem. Phys. Lett.* **1992**, *190*, 626.
(33) Perkyns, J. S.; Pettitt, B. M. *J. Chem. Phys.* **1992**, *97*, 7656.
(34) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049.
(35) Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27.
(36) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269.
(37) Pettitt, B. M.; Rossky, P. J. *J. Chem. Phys.* **1982**, *77*, 1451.
(38) Kovalenko, A.; Ten-no, S.; Hirata, F. *J. Comput. Chem.* **1999**, *20*, 928.
(39) Edelsbrunner, H.; Facello, M.; Fu, P.; Liang, J. In *Proceedings of the 28th Annual Hawaii International Conference on System Science*; IEEE Computer Society Press: Los Alamitos, 1995; p 256.
(40) Chalikian, T. V.; Totrov, M.; Abagyan, R.; Breslauer, K. J. *J. Mol. Biol.* **1996**, *260*, 588.
(41) Kharakoz, D. P. *J. Solution Chem.* **1992**, *21*, 569.
(42) Matubayasi, N.; Levy, R. M. *J. Phys. Chem.* **1996**, *100*, 2681.
(43) Refaee, M.; Tezuka, T.; Akasaka, K.; Williamson, M. P. *J. Mol. Biol.* **2003**, *327*, 857.