Morphometric approach to thermodynamic quantities of solvation of complex molecules: Extension to multicomponent solvent

Ryota Kodama,[1] Roland Roth,[2] Yuichi Harano,[3] and Masahiro Kinoshita[4,a)]
[1]*Graduate School of Energy Science, Kyoto University, Uji, Kyoto 611-0011, Japan*
[2]*Institut für Theoretische Physik, Universität Erlangen-Nürnberg, Staudtstrasse 7, 91058 Erlangen, Germany*
[3]*Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan*
[4]*Institute of Advanced Energy, Kyoto University, Uji, Kyoto 611-0011, Japan*

(Received 23 May 2011; accepted 1 July 2011; published online 29 July 2011)

The morphometric approach (MA) is a powerful tool for calculating a solvation free energy (SFE) and related quantities of solvation thermodynamics of complex molecules. Here, we extend it to a solvent consisting of m components. In the integral equation theories, the SFE is expressed as the sum of m terms each of which comprises solute-component j correlation functions ($j = 1, \ldots, m$). The MA is applied to each term in a formally separate manner: The term is expressed as a linear combination of the four geometric measures, excluded volume, solvent-accessible surface area, and integrated mean and Gaussian curvatures of the accessible surface, which are calculated for component j. The total number of the geometric measures or the coefficients in the linear combinations is $4m$. The coefficients are determined in simple geometries, i.e., for spherical solutes with various diameters in the same multicomponent solvent. The SFE of the spherical solutes are calculated using the radial-symmetric integral equation theory. The extended version of the MA is illustrated for a protein modeled as a set of fused hard spheres immersed in a binary mixture of hard spheres. Several mixtures of different molecular-diameter ratios and compositions and 30 structures of the protein with a variety of radii of gyration are considered for the illustration purpose. The SFE calculated by the MA is compared with that by the direct application of the three-dimensional integral equation theory (3D-IET) to the protein. The deviations of the MA values from the 3D-IET values are less than 1.5%. The computation time required is over four orders of magnitude shorter than that in the 3D-IET. The MA thus developed is expected to be best suited to analyses concerning the effects of cosolvents such as urea on the structural stability of a protein. © *2011 American Institute of Physics*. [doi:10.1063/1.3617247]

I. INTRODUCTION

The structures of polyatomic solutes stabilized in solvent are enormously influenced by the solvent properties. One of the most important quantities in accounting for the solvent effects is the solvation free energy (SFE). However, it is very difficult to calculate the SFE of a large, complex solute molecule like a protein. In the argument of the structural stability of a protein, it is often required that the SFE be evaluated for a huge number of different structures given. Therefore, it is crucial to develop an approach which allows us to calculate the SFE for each structure with sufficient accuracy as well as with minor computational effort. To this end, in an earlier work[1] we developed the so-called morphometric approach (MA).

The idea of the MA is to express a solvation quantity Z by the linear combination of only four geometric measures of a solute molecule[1,2] (V, A, X, and Y),

$$Z = C_1 V + C_2 A + C_3 X + C_4 Y. \quad (1)$$

Here, V is the excluded volume (the volume of the space which the centers of solvent molecules cannot enter), A is the solvent-accessible surface area, and X and Y are the integrated mean and Gaussian curvatures of the accessible surface, respectively. In the approach, the solute shape enters Z only via the four geometric measures. Therefore, the four coefficients ($C_1 - C_4$) can be determined in simple geometries. They are calculated from the values of Z for spherical solutes with various diameters immersed in a model solvent. The radial-symmetric integral equation theory[3] is a useful tool in the calculation.

A solute insertion under isochoric condition gives rise to a decrease in the total volume available to the translational displacement of solvent molecules (i.e., an increase in the solvent crowding), leading to a loss of the solvent entropy. This is a primary origin of the solvophobicity. Upon protein folding in water (protein folding occurs with almost constant system volume and pressure[4,5]), for example, a large gain in the water entropy occurs.[4–7] An important point is that the translational-entropy gain predominates over the rotational-entropy gain.[5] Further, in many cases the water-entropy change can be analyzed by modeling water as hard spheres.[6,7] An exception is the case where the temperature dependence of the entropic effect plays an important role. For example, the increase in solubility of methane observed at low temperatures and cold denaturation of a protein[8,9] cannot be elucidated by the

a)Author to whom correspondence should be addressed. Electronic mail: kinoshit@iae.kyoto-u.ac.jp.

hard-sphere model for water. Here, an interesting question arises: How is this entropic effect influenced by the presence of cosolvents?

The MA or similar approaches have been applied to a variety of problems where complexly shaped, solvophobic solutes are treated,[10–13] or where the solvophobic effect plays important roles. In particular, we have been quite successful in elucidating folding/unfolding mechanisms of proteins,[5–9,14–17] discriminating the native fold of a protein from a number of misfolded decoys,[18–20] and uncovering the rotation mechanism of a motor protein.[21] In these problems, the solvation entropy is the key quantity calculated by modeling a protein with a fixed structure as a set of fused hard spheres. The solvent is formed either by hard spheres or water molecules modeled as hard spheres in which point multipoles[22–25] are embedded.

In studies on the unfolding mechanisms, the SFE or related quantities of solvation thermodynamics must be calculated for a number of different structures of a protein, and the use of an efficient method, such as the MA, is highly necessitated. We have developed physical pictures of pressure,[14,15] cold,[8,9] and thermal[16,17] denaturating by employing the MA. The remaining problem is the denaturation caused by the addition of cosolvents such as urea. To tackle this type of problem, the MA that has been developed for pure solvent needs to be extended to a multicomponent solvent. In the present study, we perform such an extension.

The basic idea of the extension is first described, and then it is illustrated for calculating the SFE of a model protein immersed in a binary-mixture solvent. The protein is modeled as a set of fused hard spheres and the binary-mixture solvent comprises hard spheres with two different diameters. For a variety of structures of protein G (the number of residues is 56; the Protein Data Bank Code is 2GB1), we calculate the values of the SFE using the extended MA and the direct application of the three-dimensional integral equation theory (3D-IET) (Refs. 26–30) to the large, polyatomic solute molecule. The SFE values calculated via the two routes are compared with the result that the deviations of the values obtained by the morphometric approach from those by the 3D-IET are less than 1.5%. The computation time required in the MA is over four orders of magnitude shorter than that in the 3D-IET. Moreover, the MA does not suffer the drawback of the 3D-IET, namely, a large amount of computer storage requirements.

II. MODEL AND THEORY

A. Solvent and protein models

The solvent comprises m components. It is formed by hard spheres with m different diameters. In the present study, m is set at 2 and binary hard-sphere mixtures are considered. We test the six systems whose specifications are given in Table I. Systems 0–1 and 0–2 are pure-solvent systems. The packing fraction 0.3831 and the molecular diameter 0.28 nm are the values for water at ambient temperature and pressure. Binary hard-sphere mixtures with different compositions, whose diameter ratios are 1.5 and 2.0, are considered as systems 1 through 4. The packing fraction of one of the components is roughly in the range 0.05−0.35.

TABLE I. Six systems considered: d_j is the molecular diameter of the jth component of the solvent, $\rho_j d^3$ ($d = 0.28$ nm) is the reduced number density of the jth component, and η_j is the packing fraction of the jth component defined as $\eta_j = \pi \rho_j d_j^3/6$.

System	d_1 [nm]	d_2/d_1	$\rho_1 d^3$	$\rho_2 d^3$	η_1	η_2	$\eta_1+\eta_2$
0–1	0.28	...	0.7317	...	0.3831	...	0.3831
0–2	0.42	...	0.2168	...	0.3831	...	0.3831
1	0.28	1.50	0.5000	0.0800	0.2618	0.1414	0.4032
2	0.28	2.00	0.6000	0.0250	0.3142	0.1047	0.4189
3	0.28	1.50	0.2700	0.1482	0.1414	0.2618	0.4032
4	0.28	1.50	0.1000	0.1985	0.0524	0.3508	0.4032

The protein with a fixed structure is modeled as a set of fused hard spheres. The (x, y, z)-coordinates of all the protein atoms are used as part of the input data to account for the polyatomic characteristics of each structure on the atomic level. The diameter of each atom is set at the σ-value of the Lennard-Jones potential parameters of CHARMM22.[31] We consider 30 different structures of protein G which were taken from local-minimum states of the energy function found in a replica-exchange molecular dynamics simulation using all-atom potentials.[32]

B. Three-dimensional integral equation theory

The integral equation theory is a statistical-mechanical theory which is popular in liquid state physics. It was originally developed for a spherically symmetric system. The 3D-IET theory we employ is an extension to general systems described using the cartesian coordinate system.[26–30] The great advantage of the 3D version is that details of the polyatomic structure of a solute molecule can explicitly be taken into account.[26,29]

Solute I of arbitrary geometry is immersed at infinite dilution in mixtures of spheres with diameters d_j ($j = 1, \ldots, m$) forming the solvent. The Ornstein-Zernike (OZ) equation in the Fourier space is expressed by

$$W_{Ij}(k_x, k_y, k_z) = \sum_{n=1}^{m} \rho_n C_{In}(k_x, k_y, k_z) H_{nj}(k), \quad j = 1, \ldots, m \tag{2}$$

and the hypernetted-chain (HNC) closure equation[3] is written as

$$c_{Ij}(x, y, z) = \exp\{-u_{Ij}(x, y, z)/(k_B T)\} \exp\{w_{Ij}(x, y, z)\} - w_{Ij}(x, y, z) - 1, \quad j = 1, \ldots, m. \tag{3}$$

Here, $w = h - c$, c is the direct correlation function, h is the total correlation function, u is the potential, ρ_j is the bulk density of solvent spheres of diameter d_j, k_B is Boltzmann's constant, T is the absolute temperature. The capital letters (C, H, and W) represent the Fourier transforms. $H_{lj}(k)$ ($l = 1, \ldots, m$; $j = 1, \ldots, m$; $k^2 = k_x^2 + k_y^2 + k_z^2$) calculated using the radial-symmetric HNC theory for spherical particles is part of the input data. We emphasize that the OZ equation is *exact* while the bridge function is neglected in the HNC closure equation.

In the present study, solute I is a protein molecule with a prescribed structure. In order to eliminate all the unreasonable overlaps of the constituent atoms, they are moved to the locally optimized coordinates by employing a standard energy-minimization method with the all-atom potentials. After this treatment, the protein molecule is switched to a set of fused hard spheres. There are no unreasonable overlaps of the constituent atoms.

Equations (2) and (3) are numerically solved on a cubic grid. The numerical procedure is briefly summarized as follows: (1) $u_{Ij}(x, y, z)$ ($j = 1, \ldots, m$) is calculated at each 3D grid point; (2) $w_{Ij}(x, y, z)$ is initialized to zero; (3) $c_{Ij}(x, y, z)$ is calculated from Eq. (3), and $c_{Ij}(x, y, z)$ is transformed to $C_{ij}(k_x, k_y, k_z)$ using the 3D fast Fourier transform (3D-FFT); (4) $W_{Ij}(k_x, k_y, k_z)$ is calculated from Eq. (2) and $W_{Ij}(k_x, k_y, k_z)$ is inverted to $w_{Ij}(x, y, z)$ using the 3D-FFT; and (5) steps (3) and (4) are repeated until the input and output functions for $w_{Ij}(x, y, z)$ become identical within convergence tolerance. On grid points where a solvent particle and the solute overlap, $\exp\{-u_{Ij}(x, y, z)/(k_BT)\}$ is zero. On those where a solvent particle is in contact with the solute, it is set at 0.5, and otherwise it is unity. The grid spacing ($\triangle x$, $\triangle y$, and $\triangle z$) is set at $0.2d$ ($d = 0.28$ nm), and the grid resolution ($N_x \times N_y \times N_z$) is 256 $\times$ 256 $\times$ 256. It has been verified that the spacing is sufficiently small and the box size ($N_x\triangle x$, $N_y\triangle y$, $N_z\triangle z$) is large enough for the correlation functions at the box surfaces to be essentially zero.

A great advantage of the HNC approximation is that the SFE of solute I, $\Delta\mu$, is obtained from the simple integration of the direct and total correlation functions expressed by

$$\Delta\mu = \sum_{j=1}^{m} \Delta\mu_j, \tag{4}$$

$$\Delta\mu_j/(k_BT) = \rho_j \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{h_{Ij}(x, y, z)^2/2 - c_{Ij}(x, y, z) - h_{Ij}(x, y, z)c_{Ij}(x, y, z)/2\} dx dy dz, \quad j = 1, \ldots, m. \tag{5}$$

Equation (5) is obtained as an extension of the Morita-Hiroike formula[33,34] to the 3D system. We note that the SFE is expressed as the sum of m terms each of which comprises solute-component j correlation functions ($j = 1, \ldots, m$). The SFE is "the excess free energy of the solvent in which solute I is immersed" minus "the excess free energy of a pure solvent."

C. Morphometric approach

We apply the idea of the MA (Refs. 1 and 2) to each component of the solvent

$$\Delta\mu_j/(k_BT) = C_{1j}V_j + C_{2j}A_j + C_{3j}X_j + C_{4j}Y_j, \quad j = 1, \ldots, m. \tag{6}$$

Here, V is the excluded volume, A is the solvent-accessible surface area, and X and Y are the integrated mean and Gaussian curvatures of the accessible surface, respectively. The subscript "j" represents that the value is for the solvent whose molecular diameter is d_j. Therefore, the solute shape is represented in terms of $4m$ geometric measures. The $4m$ coefficients (C_{1j}–C_{4j}, $j = 1, \ldots, m$), which are independent of the solute shape, can be determined in simple geometries. They are calculated from the values of the SFE for hard-sphere solutes with various diameters (d_U: $d_U/d = 0.02, 0.04, 0.08, 0.16, 0.32, 0.64, 1.0, 2.0, 4.0, 8.0, 16$, and 30, $d = 0.28$ nm) immersed in the model solvent. The radial-symmetric integral equation theory for spherical particles[3,32] with the HNC approximation is employed in the calculation. The SFE of a hard-sphere solute is given by the Morita-Hiroike formula,[33,34]

$$\Delta\mu_{sp} = \sum_{j=1}^{m} \Delta\mu_{spj}, \tag{7}$$

$$\Delta\mu_{spj}/(k_BT) = 4\pi\rho_j \int_0^{\infty} r^2\{h_j(r)^2/2 - c_j(r) - h_j(r)c_j(r)/2\} dr, \quad j = 1, \ldots, m. \tag{8}$$

The subscript "sp" represents that the value is for a "spherical" solute. The $4m$ coefficients are determined by the least square fitting applied to the following equation for hard-sphere solutes:

$$\Delta\mu_{spj}/(k_BT) = C_{1j}\left(4\pi R_j^3/3\right) + C_{2j}\left(4\pi R_j^2\right) + C_{3j}(4\pi R_j) + C_{4j}(4\pi), \quad R_j = (d_U + d_j)/2, \quad j = 1, \ldots, m, \tag{9}$$

where $\Delta\mu_{spj}/(k_BT)$ is a function of d_U (the values of $\Delta\mu_{spj}/(k_BT)$ are prepared for the 12 different values of d_U). Once the determination is accomplished, $\Delta\mu$ of a protein with a fixed structure is obtained from Eq. (4) in which $\Delta\mu_j$ is calculated from Eq. (6) using V_j, A_j, X_j, and Y_j ($j = 1, \ldots, m$).

The coefficients determined by the least square fitting for the six systems are collected in Table II. The sum of $[\{(\Delta\mu_{sp})^{MA} - (\Delta\mu_{sp})^{IET}\}/(\Delta\mu_{sp})^{IET}]^2$ ($(\Delta\mu_{sp})^{MA}$ is the SFE calculated by the MA and $(\Delta\mu_{sp})^{IET}$ that by the 3D-IET) over the 12 different values of d_U is minimized in the fitting. The solvent diameter of system 0–2 is 1.5 times larger than that of system 0–1. The four coefficients for system 0–2 are then obtained by dividing them for system 0–1 by 1.5^3, 1.5^2, 1.5^1, and 1.5^0, respectively. As a result, the coefficients of V and A for system 0–2 are much smaller than those for system 0–1. This is indicative that a solvent with a larger molecular diameter has smaller solvation effects. The least square fitting is achieved almost perfectly: As illustrated in Table III for system 1, the deviations for the 12 different values of d_U are considerably less than 0.2% (this is also true for the other 5 systems).

For checking the accuracy of the MA, the calculation of the SFE for a model protein (calculation I) and that for spherical solutes with various sizes (calculation II) must be performed using the same method. If a molecular dynamics (MD) simulation is employed in calculation I, for example, calculation II must also be made using the same MD simulation. If an IET is applied to calculation I with a particular closure (e.g., the HNC closure adopted in the present study), calculation II is to be carried out using the IET with the same closure.

TABLE II. Coefficients determined by the least square fitting for the six systems.

System	C_{11} [1/Å^3]	C_{21} [1/Å^2]	$4\pi C_{31}$ [1/Å]	$4\pi C_{41}$ [−]	C_{12} [1/Å^3]	C_{22} [1/Å^2]	$4\pi C_{32}$ [1/Å]	$4\pi C_{42}$ [−]
0–1	0.2698	−0.2141	2.3337	−0.6087	...	...	...	...
0–2	0.0799	−0.0950	1.5443	−0.5944	...	...	...	...
1	0.1701	−0.1297	1.3847	−0.3604	0.0629	−0.0934	1.8447	−0.9543
2	0.2238	−0.1750	1.8917	−0.5004	0.0436	−0.0977	2.8206	−2.1553
3	0.0746	−0.0512	0.4975	−0.1099	0.0900	−0.1262	2.3826	−1.1699
4	0.0232	−0.0140	0.1158	−0.0155	0.0967	−0.1263	2.2404	−1.0057

As for the reliability of *the IET result itself*, the IET with the HNC closure was shown to give quite satisfactory results even in a quantitative sense for the following examples: the potential of mean force (PMF) between large spherical solutes immersed in pure solvent calculated by the radial-symmetric IET;[35] the PMF between large spherical and nonspherical solutes immersed in pure solvent calculated by the 3D-IET;[27] and the PMF between large spherical solutes immersed in a multicomponent solvent calculated by the radial symmetric IET.[36] The PMF represents "the SFE of two solutes separated by a distance" minus "the SFE of two solutes whose separation is infinitely large." The reliability test was performed by comparing the IET results obtained by the HNC closures with those either from the density functional theory,[27,36] which was shown to give the results indistinguishable from those by a MD simulation for rigid-body systems, or from the IET with the *exact* bridge functions.[27,35]

It is important to note that the MA can be extended to the case where a model water is considered as the solvent. In fact, we have already performed such extension.[5–9,14,16–21] The radial-symmetric IET is replaced by the angle-dependent IET[22–25] for molecular liquids applied to a multipolar model for water. A feature of this theory is that the water-water and solute-water orientational correlations are explicitly taken into account. It has been shown to give a quantitatively accurate value of the hydration free energy of a nonpolar solute.[25] Spherical ions (e.g., Na^+ and Cl^-) can also be included in the model water without difficulty.[22,24,37,38] Further, we believe that the solute-solvent van der Waals and electrostatic interactions can be incorporated in the MA, as long as the ions are present and their concentration is high enough to screen the electrostatic interaction (this is the case for aqueous solution under physiological conditions). We are now investigating this incorporation.

TABLE III. Performance of the least square fitting for system 1: $\beta = 1/(k_B T)$, $d = 0.28$ nm, and d_U is the diameter of the hard-sphere solute. The deviation D is defined by $D\ [\%] = 100\{(\Delta\mu_{sp})^{MA} - (\Delta\mu_{sp})^{IET}\}/(\Delta\mu_{sp})^{IET}$, where $(\Delta\mu_{sp})^{MA}$ is the SFE calculated by the morphometric approach and $(\Delta\mu_{sp})^{IET}$ that by the three-dimensional integral equation theory.

d_U/d	$\beta(\Delta\mu_{sp})^{MA}$	$\beta(\Delta\mu_{sp})^{IET}$	D [%]
0.02	5.600×10^{-1}	5.592×10^{-1}	0.16
0.04	6.029×10^{-1}	6.030×10^{-1}	−0.02
0.08	6.992×10^{-1}	7.004×10^{-1}	−0.16
0.16	9.362×10^{-1}	9.370×10^{-1}	−0.09
0.32	1.608×10^{0}	1.607×10^{0}	0.09
0.64	3.965×10^{0}	3.963×10^{0}	0.05
1.0	8.792×10^{0}	8.785×10^{0}	0.08
2.0	4.033×10^{1}	4.036×10^{1}	−0.06
4.0	2.373×10^{2}	2.377×10^{2}	−0.15
8.0	1.618×10^{3}	1.619×10^{3}	−0.07
16	1.193×10^{4}	1.193×10^{4}	0.05
30	7.570×10^{4}	7.561×10^{4}	0.12

III. RESULTS AND DISCUSSION

A. Accuracy of $\Delta\mu_1$, $\Delta\mu_2$, and $\Delta\mu$ for the native structure and an unfolded structure

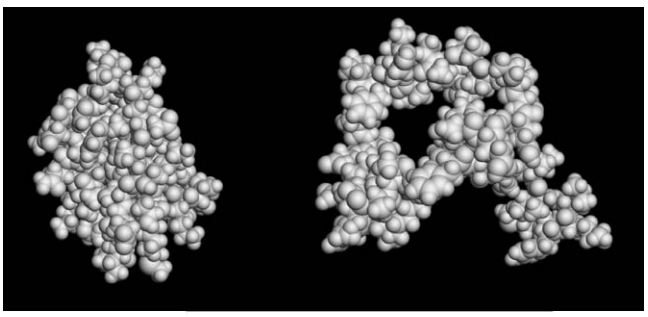
First, we consider the native structure and the unfolded structure shown in Fig. 1(a) and in Fig. 1(b), respectively. The SFE obtained from the MA $(\Delta\mu)^{MA}$ is compared with that calculated by the 3D-IET $(\Delta\mu)^{IET}$ for the native structure in Table IV and for the unfolded structure in Table V. The comparison is also made in terms of $\Delta\mu_1$ and $\Delta\mu_2$. The six systems are considered. The deviation D is defined as either

$$D\ [\%] = 100\{(\Delta\mu)^{MA} - (\Delta\mu)^{IET}\}/(\Delta\mu)^{IET} \tag{10}$$

or

$$D\ [\%] = 100\{(\Delta\mu_j)^{MA} - (\Delta\mu_j)^{IET}\}/(\Delta\mu_j)^{IET}, \quad j = 1, 2. \tag{11}$$

The superscripts, "MA" and "IET," represent that the quantities are calculated by the MA and by the 3D-IET, respectively. In pure solvents, the deviation for $\Delta\mu$ is less than 0.15%. In the mixture solvents, the deviations for $\Delta\mu_1$ and $\Delta\mu_2$ are larger: They become larger as η_1 and η_2 decrease, respectively. The minor component (i.e., the component whose number density is smaller) undergoes a larger



(a) (b)

FIG. 1. Two representative structures of protein G considered in Tables IV and V. They are drawn by Accelrys Discovery Studio 3.0.0. (a) The native structure with $R_g = 1.064$ nm considered in Table IV (R_g is the radius of gyration). (b) An unfolded structure with $R_g = 1.445$ nm considered in Table V.

TABLE IV. Deviations (see Eqs. (10) and (11)) of $\Delta\mu_1$, $\Delta\mu_2$, and $\Delta\mu$ calculated by the morphometric approach from those by the three-dimensional integral equation theory, respectively. The native structure of protein G shown in Fig. 1(a) is considered. $\Delta\mu$ is the solvation free energy, and its two terms are denoted by $\Delta\mu_1$ and $\Delta\mu_2$: $\beta = 1/(k_BT)$. The superscripts, "MA" and "IET," represent that the quantities are calculated by the MA and by the 3D-IET, respectively.

	$\Delta\mu_1$			$\Delta\mu_2$			$\Delta\mu$		
System	$\beta(\Delta\mu_1)^{MA}$	$\beta(\Delta\mu_1)^{IET}$	D [%]	$\beta(\Delta\mu_2)^{MA}$	$\beta(\Delta\mu_2)^{IET}$	D [%]	$\beta(\Delta\mu)^{MA}$	$\beta(\Delta\mu)^{IET}$	D [%]
0–1	2400.2	2404.6	−0.14	...	...	...	2400.2	2404.6	−0.14
0–2	810.2	811.4	−0.15	...	...	...	810.2	811.4	−0.15
1	1531.5	1539.3	−0.51	579.3	562.2	3.04	2110.7	2101.4	0.44
2	2000.2	2007.2	−0.35	407.2	381.8	6.65	2407.4	2389.0	0.77
3	690.4	699.5	−1.29	853.4	834.4	2.27	1543.8	1533.9	0.65
4	220.7	226.6	−2.61	944.3	933.0	1.21	1165.0	1159.6	0.47

deviation. There is a trend that the deviation for $\Delta\mu_2$ becomes larger as the molecular-diameter ratio increases. (The solvent diameter of component 1 is smaller than that of component 2.) The maximum deviation observed exceeds 8%. However, the deviations for $\Delta\mu_1$ and $\Delta\mu_2$ are always negative and positive, respectively. Significant cancellation of the deviations occurs when the summation, $\Delta\mu_1+\Delta\mu_2$, is taken with the result that the deviation for $\Delta\mu$ is only in the range 0.4%−1.0%. Since the important quantity to be calculated accurately is $\Delta\mu$, we can conclude that the accuracy of the MA is fairly high.

To examine the effects of the grid size employed in the 3D-IET on the deviation for $\Delta\mu$, we perform additional 3D-IET calculations for the two representative structures shown in Fig. 1. The grid spacing ($\triangle x$, $\triangle y$, and $\triangle z$) and the grid resolution ($N_x \times N_y \times N_z$) are set at $0.1d$ ($d = 0.28$ nm) and $512 \times 512 \times 512$, respectively. In system 3, for example, the deviation for the native structure shown in Fig. 1(a) changes from 0.65% to 0.82% and that for the unfolded structure shown in Fig. 1(b) changes from 0.86% to 1.04%. Thus, we find a finer grid size leads to no improvement of the agreement between the SFE-values from the two different sources.

B. Accuracy of $\Delta\mu_1$, $\Delta\mu_2$, and $\Delta\mu$ for a variety of structures

To verify the broad applicability of the MA, we consider a variety of structures of protein G. The deviations for $\Delta\mu_1$, $\Delta\mu_2$, and $\Delta\mu$ in system 1 are plotted against the radius of gyration R_g in Fig. 2. Those in systems 2, 3, and 4 exhibit qualitatively the same characteristics. The deviations for $\Delta\mu_1$ and $\Delta\mu_2$ are always negative and positive, respectively. However, they are somewhat cancelled out when the summation, $\Delta\mu_1+\Delta\mu_2$, is taken with the result that the deviation for $\Delta\mu$ becomes considerably small. The deviation for $\Delta\mu$ is plotted against R_g in Fig. 3. The systems 1 through 4 (i.e., the mixture solvents) are considered. Despite that the range of the R_g-values covered is quite wide (1.0 nm−2.8 nm), the deviation is almost independent of the structure characteristics and its maximum value is only ∼1.5%.

C. Physical origin of deviations

Figure 4 shows a pair of isolated large spheres (a) and fused large spheres (b) immersed in a binary mixture solvent comprising small spheres and medium-sized spheres. There are spaces which the centers of small spheres cannot enter (s-spaces) and those which medium-sized spheres cannot enter (m-spaces). The overlap of s-spaces occurs in (b), and the total volume of s-spaces are smaller than in (a) by the volume of the overlapped space V_s. The same can be mentioned for m-spaces, and the volume of the overlapped space is denoted by V_m. An important point is that there are spaces which *the centers of small spheres can enter but those of medium-sized spheres cannot* (ms-spaces). The overlap of ms-spaces occurs in (b), and the total volume of ms-spaces are smaller than in (a) by "$V_m-V_s-V_{green}$" where V_{green} is the volume of the space marked in green in Fig. 4(b). It follows from Eq. (6) that

TABLE V. Deviations (see Eqs. (10) and (11)) of $\Delta\mu_1$, $\Delta\mu_2$, and $\Delta\mu$ calculated by the morphometric approach from those by the three-dimensional integral equation theory, respectively. The unfolded structure of protein G shown in Fig. 1(b) is considered. $\Delta\mu$ is the solvation free energy, and its two terms are denoted by $\Delta\mu_1$ and $\Delta\mu_2$: $\beta = 1/(k_BT)$. The superscripts, "MA" and "IET," represent that the quantities are calculated by the MA and by the 3D-IET, respectively.

	$\Delta\mu_1$			$\Delta\mu_2$			$\Delta\mu$		
System	$\beta(\Delta\mu_1)^{MA}$	$\beta(\Delta\mu_1)^{IET}$	D [%]	$\beta(\Delta\mu_2)^{MA}$	$\beta(\Delta\mu_2)^{IET}$	D [%]	$\beta(\Delta\mu)^{MA}$	$\beta(\Delta\mu)^{IET}$	D [%]
0–1	2552.2	2553.8	−0.06	...	...	...	2552.2	2553.8	−0.06
0–2	896.2	896.7	−0.05	...	...	...	896.2	896.7	−0.05
1	1634.3	1642.3	−0.49	620.9	598.2	3.81	2255.2	2240.5	0.66
2	2129.2	2136.6	−0.34	436.5	403.7	8.13	2565.7	2540.3	1.00
3	743.6	754.4	−1.43	923.7	898.7	2.79	1667.3	1653.1	0.86
4	239.9	246.7	−2.78	1032.3	1017.5	1.46	1272.2	1264.2	0.63

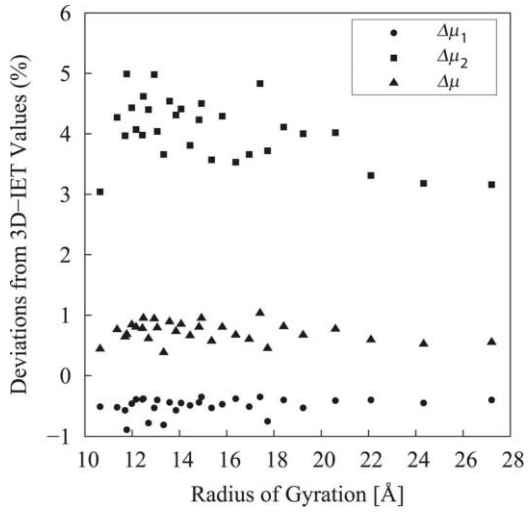


FIG. 2. Deviations (see Eqs. (10) and (11)) of $\Delta\mu_1$, $\Delta\mu_2$, and $\Delta\mu$ calculated by the morphometric approach from those by the three-dimensional integral equation theory (3D-IET), respectively. System 1 and 30 different structures of protein G are considered. $\Delta\mu$ is the solvation free energy, and its two terms are denoted by $\Delta\mu_1$ and $\Delta\mu_2$.

s-spaces and m-spaces are considered but ms-spaces are not in the MA. Therefore, V_s and V_m are taken into account while V_{green} is neglected (i.e., V_{green} is set at zero). Consequently, the total volume of ms-spaces is overestimated by the neglect in (b): The restriction of the translational freedom caused by the insertion of a pair of fused large spheres is underestimated for small spheres, whereas it is overestimated for medium-sized spheres. (The above argument is made by assuming that the accuracy of the MA is influenced by the presence of ms-spaces primarily through the excluded volume.)

A protein is geometrically a set of fused spheres. The sizes of the spheres are not large, but there are a number of such spheres. In the MA where Eq. (6) is adopted and the coefficients are determined for isolated spheres, $\Delta\mu_1$ is underestimated while $\Delta\mu_2$ is overestimated because the spaces like ms-spaces are not taken into consideration. This is why the deviations for $\Delta\mu_1$ and $\Delta\mu_2$ are always negative and positive,

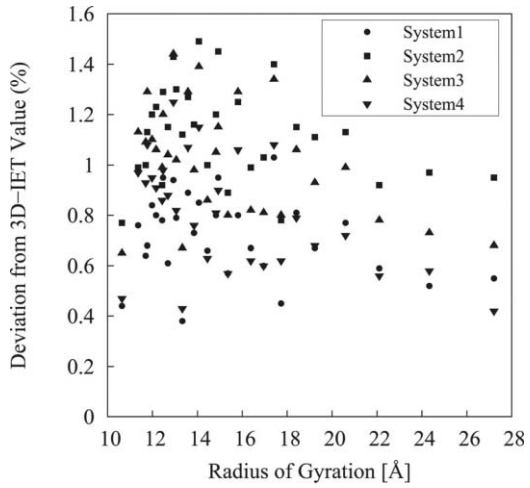


FIG. 3. Deviation (see Eq. (10)) of the solvation free energy ($\Delta\mu$) calculated by the morphometric approach from that by the 3D-IET. Systems 1 through 4 and 30 different structures of protein G are considered.

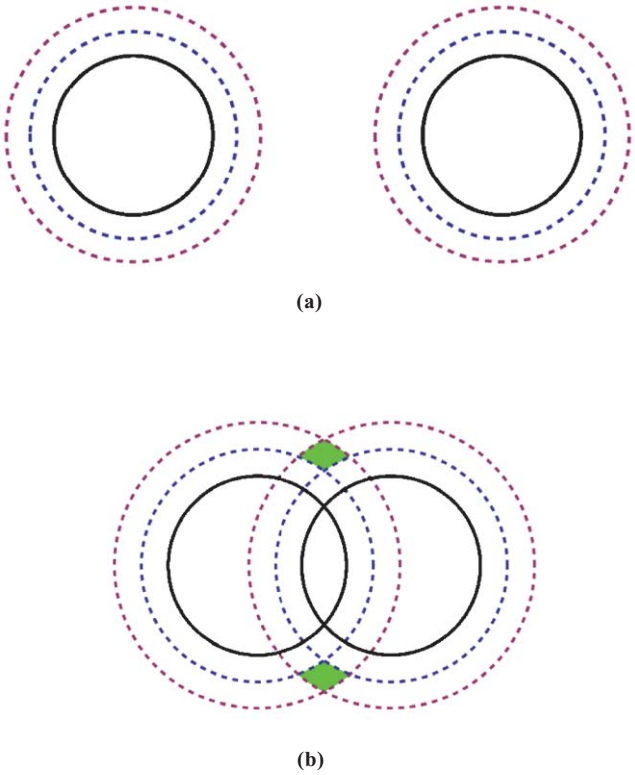


FIG. 4. Excluded spaces which the centers of small and medium-sized spheres cannot enter in the presence of a pair of isolated (a) and fused (b) large spheres. Black solid line: large sphere. Blue broken line: excluded space for small spheres. Red broken line: excluded space for medium-sized spheres.

respectively. The incorporation of the spaces like ms-spaces (their four geometric measures) cannot readily be carried out. However, the deviations are fortunately cancelled out to a significantly large extent when the summation, $\Delta\mu_1+\Delta\mu_2$, is taken with the result that the deviation for $\Delta\mu$ becomes considerably small.

D. Brief discussion on solvent effects

We have been quite successful in elucidating pressure,[14,15,39,40] thermal,[16,17] and cold[8,9] denaturation with the emphasis on the entropic components of the system free energy. Our principal concern in the next stage is the elucidation of protein denaturation caused by the cosolvent (e.g., urea) addition on the basis of the entropic components. At low pressures, a folded conformation is more favored than an unfolded state in terms of the solvent entropy (this is not true at high pressures[14,15,39,40]). However, the former is less favored in terms of the protein intramolecular conformational entropy. If the solvent-entropy effect, which drives a protein to fold, becomes sufficiently less powerful by the cosolvent addition, the intramolecular-conformational entropy dominates, giving rise to unfolding or denaturation.

It is observed from Tables IV and V that with respect to $(\Delta\mu)^{MA}$ (i.e., contribution from the solvent entropy) the native structure is more stable than the unfolded structure by $-152k_BT$, $-86k_BT$, $-144k_BT$, $-158k_BT$, $-123k_BT$, and $-107k_BT$ in systems 0–1, 0–2, 1, 2, 3, and 4, respectively. With respect to $(\Delta\mu)^{IET}$ the numbers change to $-149k_BT$,

$-85k_BT$, $-139k_BT$, $-151k_BT$, $-119k_BT$, and $-105k_BT$, respectively, but the order remains exactly the same. In our earlier work,[5] we proposed a method for estimating the minimum and maximum values of the loss of the intramolecular conformational entropy upon folding (i.e., upon the transition from an unfolded state to the native structure). For protein G, the minimum and maximum values are $70k_BT$ and $218k_BT$, respectively, in terms of the free-energy increase upon folding: Here, we assume that the free-energy increase is roughly $\sim 144k_BT$ ($144 = (70 + 218)/2$). That is, the native structure is less stable than the unfolded structure by $\sim 144k_BT$ with respect to the intramolecular conformational entropy. As regards the total free energy incorporating the solvent entropy as well, the native structure is more stable in systems 0–1 and 2 while the unfolded structure is more stable in systems 0–2, 3, and 4. The two structures are almost equally stable in system 1.

The solvent diameter of component 2 in system 2 is larger than in systems 1, 3, and 4, but the total packing fraction $(\eta_1+\eta_2)$ of system 2 is the highest. As a consequence, component 2 acts as a stabilizer of the native structure in system 2. The relative stability between two structures of a protein is determined by complicated interplay of the details of the protein structures and the solvent specifications (in particular, the molecular diameter of component 2 and the total packing fraction). A variety of structures should be considered in a theoretical analysis on the effects due to a cosolvent, and we intend to perform this type of analysis in the next stage.

IV. CONCLUDING REMARKS

We have extended the MA, which is a powerful tool for calculating a SFE and related quantities of solvation thermodynamics of complex molecules, so that it can be applied to a multicomponent solvent. The extended version is illustrated for a protein (protein G with 56 residues) modeled as a set of fused hard spheres immersed in a binary mixture of hard spheres. A total of 30 structures of the protein with a variety of radii of gyration are considered for the illustration purpose. The molecular-diameter ratio d_2/d_1 of the mixture solvent is set at 1.5 or 2.0, and three different compositions are tested for $d_2/d_1 = 1.5$. The SFE obtained through the extended MA is compared with that calculated by the direct application of the 3D-IET to the protein. The deviations are less than 1.5%. The computation time required is over four orders of magnitude shorter than that in the 3D-IET. The MA does not suffer the disadvantage of the 3D-IET, namely, a large amount of computer storage requirements. Further, in the MA the decomposition of a thermodynamic quantity of solvation into the four terms provides physical insights into the microscopic mechanisms of protein folding and unfolding.[5,8,9,14,15]

The MA thus developed is expected to be best suited to analyses concerning the effects of cosolvents on the structural stability of a protein. A protein can be denatured by the addition of a denaturant such as urea. However, the mechanism of urea-induced denaturation is not well understood yet. A possible interpretation is that urea has high affinity with the protein surface and the presence of urea in aqueous solution makes the protein take an unfolded structure with a larger surface area. Another possibility is that urea-induced changes in the solvophobic effect or in the solvation entropy play essential roles. A clue to the mechanism is that a protein is denatured only when the urea concentration becomes quite high (as high as ~8M). We are inclined to think that the weakening of the solvent-entropy effect is an important cause of the denaturation. On the other hand, it is known that the addition of sucrose can stabilize the native structure of a protein. As discussed in Sec. III D, the second component acts as either a stabilizer or a destabilizer of the native structure by complicated interplay of the solvent diameter of the second component and the total packing fraction of the mixture solvent. In any case, we believe that as in the case of pressure denaturation of a protein,[14,15,39,40] the stabilization/destabilization mechanism can be elucidated not by considering small molecules[41–43] but by directly treating a protein with a variety of structures, i.e., with a variety set of geometric measures.

ACKNOWLEDGMENTS

This work was supported by Grants-in-Aid for Scientific Research on Innovative Areas (Grant No. 20118004) and that on (B) (Grant No. 22300100) from the Ministry of Education, Culture, Sports, Science and Technology of Japan, by the Grand Challenges in Next-Generation Integrated Nanoscience, MEXT, Japan, and by Kyoto University Global Center of Excellence (GCOE) of Energy Science.

[1] R. Roth, Y. Harano, and M. Kinoshita, Phys. Rev. Lett. **97**, 078101 (2006).
[2] P. M. König, R. Roth, and K. R. Mecke, Phys. Rev. Lett. **93**, 160601 (2004).
[3] J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, 3rd ed. (Academic, London, 2005).
[4] N. Baden, S. Hirota, T. Takabe, N. Funasaki, and M. Terazima, J. Chem. Phys. **127**, 175103 (2007).
[5] T. Yoshidome, M. Kinoshita, S. Hirota, N. Baden, and M. Terazima, J. Chem. Phys. **128**, 225104 (2008).
[6] M. Kinoshita, Front. Biosci. **14**, 3419 (2009).
[7] M. Kinoshita, Int. J. Mol. Sci. **10**, 1064 (2009).
[8] T. Yoshidome and M. Kinoshita, Phys. Rev. E **79**, 030905(R) (2009).
[9] H. Oshima, T. Yoshidome, K. Amano, and M. Kinoshita, J. Chem. Phys. **131**, 205102 (2009).
[10] K. Mecke and C. H. Arns, J. Phys. Condens. Matter **17**, S503 (2005).
[11] H. Hansen-Goos, R. Roth, K. Mecke, and S. Dietrich, Phys. Rev. Lett. **99**, 128101 (2007).
[12] L.-T. Cheng, J. Dzubiella, J. A. McCammon, and B. Li, J. Chem. Phys. **127**, 084503 (2007).
[13] L.-T. Cheng, Z. Wang, P. Setny, J. Dzubiella, B. Li, and J. A. McCammon, J. Chem. Phys. **131**, 144102 (2009).
[14] Y. Harano, T. Yoshidome, and M. Kinoshita, J. Chem. Phys. **129**, 145103 (2008).
[15] T. Yoshidome, Y. Harano, and M. Kinoshita, Phys. Rev. E **79**, 011912 (2009).
[16] K. Amano, T. Yoshidome, Y. Harano, K. Oda, and M. Kinoshita, Chem. Phys. Lett. **474**, 190 (2009).
[17] K. Oda, R. Kodama, T. Yoshidome, M. Yamanaka, Y. Sambongi, and M. Kinoshita, J. Chem. Phys. **134**, 025101 (2011).
[18] Y. Harano, R. Roth, Y. Sugita, M. Ikeguchi, and M. Kinoshita, Chem. Phys. Lett. **437**, 112 (2007).
[19] T. Yoshidome, K. Oda, Y. Harano, R. Roth, Y. Sugita, M. Ikeguchi, and M. Kinoshita, Proteins **77**, 950 (2009).
[20] S. Yasuda, T. Yoshidome, Y. Harano, R. Roth, H. Oshima, K. Oda, Y. Sugita, M. Ikeguchi, and M. Kinoshita, Proteins **79**, 2161 (2011).
[21] T. Yoshidome, Y. Ito, M. Ikeguchi, and M. Kinoshita, J. Am. Chem. Soc. **133**, 4030 (2011).
[22] P. G. Kusalik and G. N. Patey, J. Chem. Phys. **88**, 7715 (1988).
[23] P. G. Kusalik and G. N. Patey, Mol. Phys. **65**, 1105 (1988).

[24]M. Kinoshita and D. R. Bérard, J. Comput. Phys. **124**, 230 (1996).
[25]M. Kinoshita, J. Chem. Phys. **128**, 024507 (2008).
[26]M. Ikeguchi and J. Doi, J. Chem. Phys. **103**, 5011 (1995).
[27]M. Kinoshita, J. Chem. Phys. **116**, 3493 (2002).
[28]M. Kinoshita, Chem. Phys. Lett. **387**, 47 (2004).
[29]Y. Harano and M. Kinoshita, Biophys. J. **89**, 2701 (2005).
[30]M. Kinoshita, Chem. Eng. Sci. **61**, 2150 (2006).
[31]A. D. MacKerell Jr., D. Bashford, M. Bellott, R. L. Dunbrack Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, J. Phys. Chem. B **102**, 3586 (1998).
[32]Y. Okamoto, Recent Res. Dev. Pure Appl. Chem. **2**, 1 (1998).
[33]T. Morita, Prog. Theor. Phys. **23**, 829 (1960).
[34]T. Morita and K. Hiroike, Prog. Theor. Phys. **25**, 537 (1961).
[35]M. Kinoshita, S. Iba, K. Kuwamoto, and M. Harada, J. Chem. Phys. **105**, 7177 (1996).
[36]R. Roth and M. Kinoshita, J. Chem. Phys. **125**, 084910 (2006).
[37]M. Kinoshita, S. Iba, and M. Harada, J. Chem. Phys. **105**, 2487 (1996).
[38]D. R. Bérard, M. Kinoshita, N. M. Cann, and G. N. Patey, J. Chem. Phys. **107**, 4719 (1997).
[39]Y. Harano and M. Kinoshita, J. Phys. Condens. Matter **18**, L107 (2006).
[40]Y. Harano and M. Kinoshita, J. Chem. Phys. **125**, 024910 (2006).
[41]Y. Nozaki, and C. Tanford, J. Biol. Chem. **238**, 4074 (1963).
[42]M. Roseman, and W. P. Jencks, J. Am. Chem. Soc. **97**, 631 (1975).
[43]M. Ikeguchi, S. Nakamura, and K. Shimizu, J. Am. Chem. Soc. **123**, 677 (2001).