Free-Energy Analysis of Solvation with the Method of Energy Representation

Nobuyuki Matubayasi

Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011 Japan

TABLE OF CONTENTS

- 1. Abstract
- 2. Introduction
- 3. Free-Energy Perturbation and Thermodynamic Integration Methods
- 4. Distribution Function in Solution
- 5. Density-Functional Theory
- 6. Radial Distribution Function and Reference Interaction Site Model
- 7. Method of Energy Representation
- 8. Application to Molecular Binding into Micelle and Membrane Systems
- 9. Perspective
- 10. Acknowledgments
- 11. References

1. ABSTRACT

A new theory of solutions, the method of energy representation, is introduced by adopting the solutesolvent interaction energy as the coordinate of distribution functions. The density-functional theory is formulated over the energy coordinate, and an approximate functional for the solvation free energy is given in terms of energy distribution functions in the solution and reference solvent systems. The method of energy representation greatly expands the scope of solution theory and is amenable to supercritical fluid, flexible molecules with intramolecular degrees of freedom, inhomogeneous system, and quantum-(QM/MM) mechanical/molecular-mechanical system. Through the combination with molecular simulation, the functional for the solvation free energy is demonstrated to perform well for nonpolar, polar, and ionic solutes in water over a wide range of thermodynamics conditions, with drastic reduction of the computational demand compared to the standard free-energy perturbation and thermodynamic integration methods. As an application to inhomogeneous system involving flexible species, the molecular binding into micelle and membrane is analyzed by treating micelle and membrane as a mixed solvent system consisting of water and amphiphilic molecule.

2. INTRODUCTION

The most fundamental quantity to describe a process in solution is the free energy (change). Indeed, it governs the equilibrium and rate constants of the process. The free-energy change corresponding to the insertion process of a solute in solution is the chemical potential (solvation free energy). Once the chemical potentials are known for the species present in the initial and final states of a process of interest, the free-energy change for the process can be readily evaluated. Therefore, it is of primary importance in statistical mechanics of solutions to establish a scheme to determine the chemical potential (solvation free energy) of a solute in solution.

In the present review, we describe the scheme of computing the solvation free energy. The free energy in solution is notorious for its heavy computational demand. We approach this difficulty by combining the molecular simulation and statistical-mechanical theory of solutions. Since the target of solution chemistry expands and now includes supercritical fluid, quantum-classical coupled system, and nano-organized systems such micelle, membrane, and protein, the theory of solutions also needs to be (re-)formulated to treat these frontline subjects. The



Figure 1. A schematic distribution of ΔU value in the particle insertion method.

purpose of this review is to introduce a new theory of solutions which is amenable to diverse areas of applications. We first describe the standard scheme of freeenergy calculation. We then present the concept of distribution function and the density-functional theory connecting the distribution function and free energy. Finally, we formulate a new method of solutions, the method of energy representation, and show its application to molecular binding into micelle and membrane.

3. FREE-ENERGY PERTURBATION AND THERMODYNAMIC INTEGRATION METHODS

Let H_0 and H_1 be respectively the Hamiltonians at the initial and final states of a process in solution. When the solvation process is concerned, the initial and final states are typically the pure solvent and solution systems of interest. The corresponding free-energy change ΔF is given by

$$\exp(-\beta\Delta F) = \frac{\int d\Gamma \exp(-\beta H_1)}{\int d\Gamma \exp(-\beta H_0)}$$
(3.1)

where β is the inverse of the product of the Boltzmann constant k_B and the temperature T and Γ is the (collective) coordinate for the phase space. When the classical statistical mechanics is adopted and the Hamiltonian change between the initial and final states does not involve the kinetic part, Eq. (3.1) reduces to

$$\exp(-\beta\Delta F) = \frac{\int d\mathbf{X} \exp(-\beta U_1)}{\int d\mathbf{X} \exp(-\beta U_0)}$$
(3.2)

where **X** is the (collective) coordinate for the configuration and U_0 and U_1 are the potential energies of the system at the initial and final states, respectively. Equation (3.2) is the starting point of our development. It should be noted that Eq. (3.2) cannot be used when a quantum fluid is to be treated. When U_1 - U_0 is denoted by ΔU , Eq. (3.2) is rewritten as

$$\exp(-\beta\Delta F) = \langle \exp(-\beta\Delta U) \rangle_0 \tag{3.3}$$

where $< ... >_0$ is the ensemble average taken with respect to the initial state whose configuration is sampled according to the potential energy U_0 . Equation (3.3) shows that the free-energy change ΔF can be calculated, in principle, by performing only the simulation for the initial state and averaging the Boltzmann factor of the relevant energy change ΔU . Indeed, Eq. (3.3) is the basis of the particle insertion method for evaluating the solvation free energy of a solute in solution (1,2,3). In the particle insertion method, the pure solvent is simulated and the solute molecule of interest is inserted randomly into the pure solvent. The solvation free energy is then obtained from

$$\exp(-\beta\Delta F) = \int d(\Delta U) \exp(-\beta\Delta U) f(\Delta U)$$
(3.4)

where $f(\Delta U)$ is the probability distribution function of ΔU in the pure solvent system. The particle insertion method is convenient and fast since only the pure solvent configurations need to be prepared and the free energy is calculated from a one-step insertion process of the solute. As is well documented, however, the particle insertion method is applicable only for a small and weakly interacting solute. See Figure 1. When the solute is large, it almost always overlaps with solvent molecules upon insertion and $f(\Delta U)$ is well sampled only toward large (repulsive) value of ΔU . The Boltzmann factor exp (- $\beta \Delta U$) increases steeply, on the other hand, toward small (attractive) ΔU . The small ΔU region, which is often ill sampled, makes a significant contribution in Eq. (3.4). Thus, Eq. (3.4) is not computationally useful and the particle insertion method cannot be used for most of "interesting" systems. Actually, the calculation of the average of the exponential of the energy change is often prohibitive unless the energy change is small in magnitude.

The standard and often used methods to circumvent the difficulty associated with the form of Eq. (3.4) are the free-energy perturbation and thermodynamic integration methods (1,2,3). These methods are generally applicable to free-energy evaluation. In the present review, we restrict our development to the solvation process; the initial state is the pure solvent and the final state is the solution system of interest.

The free-energy perturbation method utilizes intermediate states connecting the initial and final states of the process of interest. Let V_i (i = 0, ..., N) be a sequence of potential energies where the initial and final ones V_0 and V_N are taken to be the potential functions U_0 and U_1 for the initial and final states, respectively. For an arbitrary (set of) V_i , Eq. (3.2) can be expressed as

$$\exp\left(-\beta\Delta F\right) = \prod_{i=0}^{N-1} \left\langle \exp\left(-\beta\left(V_{i+1}-V_i\right)\right) \right\rangle_i$$
(3.5)

where $< ... >_i$ is the ensemble average taken with respect to the potential function V_i . Equation (3.5) shows that ΔF is given as the sum of the free-energy change accompanying the energy change from V_i to V_{i+1} (i = 0, ..., N-1). The states corresponding to V_i (i = 1, ..., N-1) are called intermediate states. The free energy is a state function and does not depend on the choice of the intermediate states in principle. For the computational implementation, however, the point is to "select" the set of V_i so that the change from V_i to V_{i+1} is "small" in magnitude. When V_i and V_{i+1} are "similar" and the energy change is small, the difficulty encountered in the particle insertion method can be circumvented and the calculation of the free-energy change becomes feasible. The drawback is that a number of intermediate states need to be prepared and that the computational cost is enhanced accordingly.

In the thermodynamic integration method, the intermediate states are introduced with respect to the coupling parameter λ ($0 \le \lambda \le 1$). The potential function at the coupling parameter of λ is denoted by U_{λ} and satisfies $U_{\lambda} = U_0$ and $U_{\lambda} = U_1$ at the initial and final states ($\lambda = 0$ and 1), respectively. The intermediate states correspond to $0 < \lambda < 1$. The form of averaging-the-exponential is then avoided by rewriting Eq. (3.2) as

$$\Delta F = \int_{0}^{1} d\lambda \left\langle \frac{\partial U_{\lambda}}{\partial \lambda} \right\rangle_{\lambda}$$
(3.6)

where $< ... >_{\lambda}$ is the ensemble average when the potential energy is U_{λ} . As is the case of the free-energy perturbation method, ΔF value calculated by Eq. (3.6) is independent of the choice of the intermediate states in principle. The integrand of Eq. (3.6) is a preferable average from the computational viewpoint. The exponential average is not involved any more. In practice, the integral of Eq. (3.6) is replaced by a discretized sum and a finite number of intermediate states are to be treated explicitly. Since a systematic error is introduced by the discretization, a large number of intermediate states need to be prepared and the computational demand increases correspondingly.

In both the free-energy perturbation and thermodynamic integration methods, the key to the computational accuracy and efficiency is the choice of the intermediate states as a function of the coupling parameter λ . Note that the intermediate states adopted in the free-energy perturbation method can be considered a finite subset of the intermediate states introduced continuously over $0 < \lambda < 1$. A straightforward implementation of the intermediate states is possible by varying the system potential energy linearly. When the solvation is concerned and the solute-solvent interaction is expressed as the sum of Lennard-Jones and Coulombic terms, the linear variation is realized by the intermediate solute-solvent interaction given by

$$\lambda \sum_{i,j} \left(4\varepsilon_{ij} \left\{ \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right\} + \frac{q_{i}q_{j}}{r_{ij}} \right)$$
(3.7)

where *i* and *j* refer to solute and solvent interaction sites, respectively. The first term in the sum expresses the Lennard-Jones interaction at the distance r_{ij} between the solute and solvent sites, and ε_{ij} and σ_{ij} are the energy and length parameters, respectively. The second term in the sum corresponds to the Coulombic interaction, and q_i and q_j are the charges on the solute and solvent sites, respectively. The linear scaling of the solute-solvent interaction with Eq.

(3.7) is often ill-behaved numerically around $\lambda = 0$. This is related to the appearance of r = 0 singularity at $\lambda = 0$. To alleviate the problem, a nonlinear scaling, which is obtained by replacing λ with λ^n ($n \ge 2$) in Eq. (3.7), can be used (4,5). The calculation becomes more stable near λ = 0. The r = 0 singularity is still present in the nonlinear scaling, however, and careful analysis is necessary to determine the λ values actually sampled. Another choice of the intermediate states is provided by

$$4\lambda\varepsilon \left(\left(\frac{\sigma^2}{r^2 + (1 - \lambda)\delta} \right)^6 - \left(\frac{\sigma^2}{r^2 + (1 - \lambda)\delta} \right)^3 \right)$$
(3.8)

or similar expressions for the Lennard-Jones part (4,6). The r = 0 singularity is then removed and the free-energy calculation on the basis of Eq. (3.8) is numerically stable.

Although the free-energy perturbation and thermodynamic integration methods are exact under a given set of potential functions in principle, they are not free from systematic errors in practice. The systematic error most often encountered in the free-energy perturbation method is the non-coincidence of the free-energy changes ΔF calculated from the forward variation of the coupling parameter λ from 0 to 1 and the backward variation from 1 to 0. The common practice is to average the ΔF from the forward and backward calculations. It is pointed out, however, that the simple averaging is itself a source of systematic error (7). To achieve the accuracy, the use of Bennett's weighting function is recommended (7,8). In the thermodynamic integration method, a systematic error is inevitable when the integral over λ in Eq. (3.6) is discretized. A careful examination of discretization is necessary, especially when the integrand of Eq. (3.6) exhibits a non-monotonic dependence on λ and/or varies steeply over some range of λ .

4. DISTRIBUTION FUNCTION IN SOLUTION

A molecular picture of solutions is established through distribution (correlation) functions. Correspondingly, a molecular description of the solvation free energy can be implemented by formulating a functional which expresses the solvation free energy in terms only of distribution functions in the solution and pure solvent systems of interest. An approximate functional needs to be constructed in practice, however, since the exact functional involves an infinite series of many-body distribution functions (9). The theories introduced in Secs. 6 and 7 are formulated to provide the solvation free energy with simple distribution functions in closed form. In this section, a general description of distribution function is provided.

The system of our interest is a dilute solution containing a single solute molecule. Even when the solute concentration is finite, our development is valid by viewing one of the solute molecules as the "solute" and the others as part of mixed solvent. To describe completely the configuration of a solvent molecule relative to the solute, the position and orientation need to be specified



Figure 2. O-O and O-H radial distribution functions g(r) of water at 1 g/cm³ and 25 °C.

simultaneously. The complete set of the position and orientation is called the full coordinate and is denoted collectively by \mathbf{x} . If the solute and/or solvent are flexible, the intramolecular degrees of freedom are also incorporated into \mathbf{x} . In the full coordinate representation, the

instantaneous distribution $\hat{\rho}^{f}$ is introduced as

$$\hat{\rho}^{f}(\mathbf{x}) = \sum_{i} \delta(\mathbf{x} - \mathbf{x}_{i})$$
(4.1)

where \mathbf{x}_i is the full coordinate of the *i*th solvent molecule and the sum is taken over all the solvent molecules. The superscript *f* is attached to emphasize that Eq. (4.1) is in the full coordinate representation. The distribution functions are generated from the averages of products of $\hat{\rho}^f$ in the

are generated from the averages of products of p^{-1} in the system of interest.

When the distribution function is generated from $\hat{\rho}^{f}$, a multidimensional description is inevitable. Indeed, the full coordinate \mathbf{x} is 6-dimensional (5 for linear molecule) for rigid species and involves more for flexible species. To implement the full coordinate representation, the expansion in terms of spherical harmonics can be (1). employed However, the calculation of multidimensional distribution functions is often slow in molecular simulation and the numerical realization is not straightforward. When the coordinate has too "fine" information content, the corresponding distribution function needs additional methodology for handling. Too much information is not desirable both from the computational viewpoint (large memory and slow convergence) and the conceptual viewpoint (unclear perception in mind).

It is then useful to reduce the information content by introducing a "projected" coordinate. With projection, some information of \mathbf{x} is retained, while the others are disregarded. When the projection is implemented with respect to a function $P(\mathbf{x})$, the corresponding distribution functions are generated from the instantaneous distribution given by



Figure 3. The distribution function $\rho^{\ell}(\varepsilon)$ of the pair interaction energy ε of water at 1 g/cm³ and 25 °C. ρ^{n} is the number density of bulk water. The graduation of the ordinate is changed at 100 Å³ mol/kcal.

$$\sum \delta(p - P(\mathbf{x}_i)) \tag{4.2}$$

where p is the value of $P(\mathbf{x})$ and serves as the coordinate for the distribution function.

A typical choice of $P(\mathbf{x})$ is the radial distance between the atomic sites (interaction sites) of the molecule. When a pair of atomic sites in the solute and solvent molecules is picked up, the histogram of its radial distance is averaged with an appropriate normalization to give the site-site radial distribution function. For example, when the solute and solvent is both H₂O (when one of the molecules in pure water is viewed as the "solute" and the others as the "solvent"), the O-O, O-H, and H-H radial distribution functions are generated by the projections onto O-O, O-H, and H-H distances, respectively. It should be noted that the radial distribution functions do not represent a simultaneous distribution of a set of site-site radial distances. In the case of water, the O-O radial distribution function specifies only the O-O distance and the other distance information such as those for O-H and H-H is disregarded. Similarly, the O-H radial distribution function does not contain explicit information about the O-O and H-H distances. In Figure 2, we show the O-O and O-H radial distribution functions of water at 1 g/cm3 and 25 °C. It is seen from the O-O radial distribution function that the neighboring water molecules stay in the distance of ~ 2.8 Å. The O-H radial distribution function shows that the intermolecular hydrogen-bonding distance is ~ 2 Å. The second peak of the O-O radial distribution function is characteristic of water. In simple liquid, the second peak appears at about twice the distance for the first peak. In water, the second-peak position is ~1.6 times of the firstpeak position. This provides a view that the ice-like structure persists even in liquid water.

Another useful distribution function is for the pair interaction energy. An example is shown in Figure 3

for water at 1 g/cm³ and 25 °C. The peak at \sim -6 kcal/mol corresponds to the intermolecular hydrogen bonding of water in the liquid state. In the high-energy regime, the distribution function vanishes. This reflects the excluded volume effect and is consistent with the fact the radial distribution functions vanish at short distances.

Of course, the choice of the projecting function $P(\mathbf{x})$ of Eq. (4.2) is not unique. The choice depends on the purpose. For example, when the angle of the hydrogen bonding is of interest, it is most useful to adopt the hydrogen-bonding angle itself as $P(\mathbf{x})$ (10). In general, no *a priori* criterion is present for preferable projection. The desirable form of projection can be based only upon the target quantity to be investigated.

5. DENSITY-FUNCTIONAL THEORY

The solvation free energy can be evaluated using the Kirkwood charging formula (1,2,3). It introduces a set of intermediate states connecting the initial and final states of the gradual insertion process of the solute; the initial and final states correspond to the pure solvent system without the solute and the solution system of interest, respectively. The free-energy perturbation and thermodynamic integration methods are based on the Kirkwood charging formula, and in principle provide the "exact" solvation free energy under a given set of potential functions. These methods are, however, expensive in computation and practical only to small molecules. For the purpose of analyzing the solvation free energy on the molecular level, furthermore, it is necessary to express the solvation free energy in terms only of distribution functions in the solution and pure solvent systems of interest. Within the framework of the Kirkwood charging formula, the intermediate states are actually arbitrary and are employed for the convenience of formulation and computation. They are not of physical significance since the free energy is a state function.

A molecular description of the solvation free energy can be implemented by formulating a functional which expresses the solvation free energy in terms of distribution functions in the solution and pure solvent systems. The exact functional is not useful, however, since it is an infinite series of many-body distribution functions (9). In practice, an approximate but accurate functional needs to be constructed which is expressed with few-body distribution functions in closed form. When such a functional is formulated and the distribution functions constituting the approximate functional are readily obtained by computer simulation, the solvation free energy can be determined and analyzed with reasonable computational load in terms of exact, microscopic information of the systems of interest.

It is a statistical-mechanical theory of solutions to express the solvation free energy as a functional of distribution functions. Traditionally, the theory of solutions is formulated with a diagrammatic approach (1), in which an approximation is provided through a two-step procedure. In the first step, the free energy and/or distribution function is expanded with respect to the solute-solvent interaction potential function or its related function as an infinite, perturbation series. In the second step, a renormalization scheme is applied; a set of functions are defined through partial summation of the series and are employed for substitution to make the infinite series more tractable. An approximation is typically introduced by neglecting diagrams of ill character.

We adopt an alternative route to the distribution function theory. The approach is based upon the densityfunctional theory. In this approach, the change of variables is conducted through Legendre transform from the solutesolvent interaction potential function to the solute-solvent distribution function or the solvent density around the solute. The (solvation) free energy is then expressed approximately by expanding the corresponding Legendretransformed function with respect to the distribution function to some low order.

The target quantity of the development is the solvation free energy. The solvation free energy $\Delta \mu$ is the free-energy change corresponding to the gradual insertion process of the solute molecule. In $\Delta \mu$, only the contribution from the potential energy is involved and the ideal (kinetic) contribution is excluded. When the intramolecular energy of the solute is $\Psi(\psi)$ with the solute configuration ψ and the total solvent-solvent interaction energy is $U(\mathbf{X})$, $\Delta \mu$ is expressed as

$$\exp(-\beta\Delta\mu) = \frac{\int d\psi d\mathbf{X} \exp(-\beta \langle \Psi(\psi) + \sum_{i} v(\psi, \mathbf{x}_{i}) + U(\mathbf{X}) \rangle)}{\int d\psi d\mathbf{X} \exp(-\beta \langle \Psi(\psi) + U(\mathbf{X}) \rangle)}$$
(5.1)

where **X** represents the solvent configuration collectively and β is the inverse of the product of the Boltzmann constant k_B and the temperature *T*. A restriction of attention to a certain set of solute intramolecular state can be made simply by the corresponding alteration of the domain of integration over ψ .

The starting point of the density-functional treatment is the Kirkwood charging formula. When **x** is the full coordinate of a (single) solvent molecule relative to the solute as denoted in Sec. 4 and the solute-solvent interaction potential of interest is $v(\mathbf{x})$, the intermediate states are described as $u_{\lambda}(\mathbf{x})$, where λ is the coupling parameter to identify the state. When $\lambda = 0$, the system is the pure solvent system and $u_0(\mathbf{x}) = 0$ (no solute-solvent interaction). When $\lambda = 1$, the solute interacts with the solvent at full coupling and $u_1(\mathbf{x}) = v(\mathbf{x})$. The form of $u_{\lambda}(\mathbf{x})$ at $0 < \lambda < 1$ is arbitrary. The Kirkwood charging formula is an integration over λ and is expressed as

$$\Delta \mu = \int_{0}^{1} d\lambda \int d\mathbf{x} \frac{\partial u_{\lambda}(\mathbf{x})}{\partial \lambda} \rho^{f}(\mathbf{x}; u_{\lambda})$$
(5.2)

where $\rho(\mathbf{x};u_{\lambda})$ is the ensemble average of Eq. (4.1) in the presence of the solute-solvent interaction u_{λ} . The superscript *f* means that the function is represented over the full coordinate **x**. The partial integration then provides

$$\Delta \mu = \int d\mathbf{x} v(\mathbf{x}) \rho^{f}(\mathbf{x}; v) - \int_{0}^{1} d\lambda \int d\mathbf{x} u_{\lambda}(\mathbf{x}) \frac{\partial \rho^{f}(\mathbf{x}; u_{\lambda})}{\partial \lambda}$$

$$\equiv \int d\mathbf{x} v(\mathbf{x}) \rho^{f}(\mathbf{x}) - F^{f}[\rho^{f}(\mathbf{x})]$$
(5.3)

where the density-functional F^{f} is defined in the second equation. $\Delta \mu$ and F^{f} are related to each other with Legendre transform since the map is proved to be one-to-one from the solute-solvent interaction potential to the distribution function (1). An approximation can be devised by introducing the indirect part ω^{f} of the potential of mean force as

 $\rho^{f}{}_{\lambda}(\mathbf{x}) = \rho^{f}{}_{0}(\mathbf{x})\exp\left(-\beta\left(u(\mathbf{x};\rho^{f}{}_{\lambda}) + \omega^{f}(\mathbf{x};\rho^{f}{}_{\lambda})\right)\right)_{(5.4)}$

Actually, the dependence is now written in terms of the distribution function β_{λ} , instead of the potential. This is possible due to the property of one-to-one correspondence. When the solvent-solvent correlation is absent (low-density limit), σ is zero. In other words, all the "complicated" solvent-solvent correlations are put into ω Equations (5.3) and (5.4) lead exactly to

$$F^{f}[\rho^{f}(\mathbf{x})] = k_{B}T \int d\mathbf{x} \begin{bmatrix} \left(\rho^{f}(\mathbf{x}) - \rho^{f}_{0}(\mathbf{x})\right) - \rho^{f}(\mathbf{x})\log\left(\frac{\rho^{f}(\mathbf{x})}{\rho^{f}_{0}(\mathbf{x})}\right) \\ -\beta\left(\rho^{f}(\mathbf{x}) - \rho^{-0}_{0}(\mathbf{x})\right)_{0}^{1} d\lambda \omega^{f}(\mathbf{x};\rho^{f}_{\lambda}) \end{bmatrix}$$
(5.5)

when u_{λ} is taken so that ϕ_{λ} varies linearly against λ . Equation (5.5) is exact, and an approximation is introduced to the λ integral of ϕ^{f} When ϕ^{f} is taken to vary linearly with λ , the HNC (hypernetted-chain) approximation is obtained. When exp $(-\beta \phi^{f})$ -1 is set to be linear, it is the PY (Percus-Yevick) approximation.

The above is a brief introduction to the densityfunctional theory of solutions. The mathematical development is quite straightforward. The numerical implementation is difficult, however, in the full coordinate representation. As noted in Sec. 4, the full coordinate is multidimensional; the solute-solvent distribution is a function over high-dimensional configuration space and cannot be implemented in practice. To overcome the problem of dimensionality, it is necessary to introduce a projected coordinate. In Sec. 7, we introduce the energy representation and formulate the density-functional theory in the energy representation.

6. RADIAL DISTRIBUTION FUNCTION AND REFERENCE INTERACTION SITE MODEL

The method of reference interaction site model (RISM) is based on the site-site radial distribution functions. It introduces the "direct correlation functions" through the inverses of the correlation matrices for the sitesite distance and formulates an approximate set of integral equation for the site-site radial distribution functions by adopting "closure" relationships between the radial distribution functions and direct correlation functions (1,11,12,13,14). Compared to the molecular simulation method, the method of integral equation is much faster. The speed is achieved by restricting the attention only to the radial distribution functions and adopting approximate closures. Furthermore, the solvation free energy is expressed in closed form for some types of closure relationships (13,14). In this case, no reference to the intermediate states of the solute insertion process is required and the solvation free energy can be evaluated directly from the radial distribution functions obtained from the integral equation. When a closed-form functional for the solvation free energy is given in terms of distribution functions, the functional not only provides an efficient route of computation, but also sets a basis for the molecular understanding with respect to the distribution functions.

A drawback is present, of course, in any approximate method of solutions. Under a given set of potential functions, the molecular simulation gives the exact distribution functions when it is done long enough. In contrast, since the closure relationship is approximate, the radial distribution function obtained from the integral equation method is approximate. The solvation free energy calculated from the integral equation theory has two sources of errors. One is due to the approximate nature of the potential functions (force field), and the other comes from the approximation involved in the integral equation.

The drawbacks characteristic of RISM and its variants are related to the fact that they do not treat the whole molecule as a single unit and view a molecule as a collection of interaction sites. The method is thus applicable only when the potential function is of sitesite form. As a consequence, the electronic distribution cannot be treated in the cloud-like form as implemented in quantum theories, but needs to be contracted into a set of point charges. In addition, the integral equation is ill-behaved unless all of the interaction sites carry the repulsive core explicitly. For example, many of potential functions of H₂O do not assign a repulsive core at the H-site. This is simply because the repulsive core for the O-site is large enough that other molecules cannot come too close to the H-site. The repulsive core of the H-site is buried within the O-site core and is not necessary to be treated explicitly at the level of potential functions. In RISM, however, a core parameter needs to be assigned to the H-site, too. The core parameter actually dictates the resulting solvation free energy sensitively and acts as an adjustable parameter in the method. Another, related problem is so-called the "problem of auxiliary site". The solution to the RISM integral equations exhibits unphysical dependence on the presence of auxiliary sites which simply label points in a molecule and make no contribution to the intermolecular interaction. This type of difficulty is absent when the whole molecule is treated as a single unit. The difficulty arises when a molecule is treated as a collection of sites. In RISM, the correlation between a pair of sites is described at the two-body level for both the intramolecular and intermolecular ones. Since the sites in a molecule are tightly bound with one another, a partial incorporation of the intramolecular correlation is not desirable. This point is exemplified when the density is low. The RISM integral equations are not exact in the limit of zero solvent density and are not suitable to evaluate the solvation free energy in a low-density fluid. It is wellknown in this instance, too, that the low-density limit is given exactly when the whole molecule is treated as a single unit in the full coordinate representation (1). Finally, since the molecular structure is an input in the RISM approach, an additional scheme needs to be devised to deal with flexible molecules.

In the commonly used RISM approach, the solvation free energy is often expressed in closed form in terms of radial distribution functions. An improvement of the approach may then be possible through combination with the molecular simulation; the radial distribution functions are exact under the used set of potential functions when they are calculated from the molecular simulation, instead of the integral equation. This line of approach is developed by Kast and Truong (15,16,17), and the computational efficiency is achieved compared to the free-energy perturbation and thermodynamic integration methods. A special procedure is needed, however, to handle the intramolecular correlation matrices in the range of small reciprocal vector (large distance). When the reciprocal vector approaches zero. the intramolecular correlation matrices become illconditioned and the calculation procedure suffers from numerical instability.

The above drawbacks of RISM and its variants are well documented since their first formulations (11,12). They are all related to the point that a molecule is treated as a collection of sites. In the method of energy representation introduced next, each of the solute and solvent molecules is taken to be a single unit as a whole, and those drawbacks vanish.

7. METHOD OF ENERGY REPRESENTATION

In the method of energy representation, the projecting function $P(\mathbf{x})$ of Eq. (4.2) is taken to be the solute-solvent pair interaction energy. Figure 3 is an example of the distribution function in the energy representation. To introduce the energy representation, it is necessary to specify the solute-solvent interaction potential v of interest. Of course, v is a function of the solute configuration ψ and the solvent configuration \mathbf{x} . The

instantaneous distribution $\hat{\rho}^e$ is defined as

$$\hat{\rho}^{e}(\varepsilon) = \sum_{i} \delta(v(\psi, \mathbf{x}_{i}) - \varepsilon)$$
(7.1)

where the sum is taken over the solvent molecules and a superscript e is attached to emphasize that a function is represented over the energy coordinate. The distribution functions in the energy representation are generated from

the averages of products of $\hat{\rho}^e$ in the system of interest.

In the energy representation, the densityfunctional theory can be formulated by restricting the set of solute-solvent interaction potentials $u_{\lambda}(\psi, \mathbf{x})$ to those which are constant over an equienergy surface of $v(\psi, \mathbf{x})$. In this case, when the value of $v(\psi, \mathbf{x})$ is denoted as ε , the intermediate states can be written as $u_{\lambda}(\varepsilon)$. At the end points ($\lambda = 0$ and 1), $u_0(\varepsilon) = 0$ and $u_1(\varepsilon) = \varepsilon$ since $v(\psi, \mathbf{x})$ itself is the potential function in the solution system of interest. It is then possible to show that the Kirkwood charging formula is given by

$$\Delta \mu = \int_{0}^{1} d\lambda \int d\varepsilon \, \frac{\partial u_{\lambda}(\varepsilon)}{\partial \lambda} \rho^{e}(\varepsilon; u_{\lambda})$$
(7.2)

where $\rho^{\ell}(\varepsilon, u_{\lambda})$ is the ensemble average of Eq. (7.1) in the presence of the solute-solvent interaction u_{λ} . The superscript *e* is attached to mean the representation over the energy coordinate ε . The Legendre transform is also possible as

$$\Delta \mu = \int d\varepsilon u_1(\varepsilon) \rho^{\varepsilon}(\varepsilon; v) - \int_0^1 d\lambda \int d\varepsilon u_{\lambda}(\varepsilon) \frac{\partial \rho^{\varepsilon}(\varepsilon; u_{\lambda})}{\partial \lambda}$$

= $\int d\varepsilon \varepsilon \rho^{\varepsilon}(\varepsilon) - F^{\varepsilon} \left[\rho^{\varepsilon}(\varepsilon) \right]$ (7.3)

and the indirect part ω^e of the potential of mean force in the energy representation can be introduced correspondingly simply by rewriting **x** of Eq. (5.4) with ε as

$$\rho^{e_{\lambda}}(\varepsilon) = \rho^{e_{0}}(\varepsilon) \exp\left(-\beta\left(u(\varepsilon;\rho^{e_{\lambda}}) + \omega^{e}(\varepsilon;\rho^{e_{\lambda}})\right)\right) \quad (7.4)$$

The density-functional is then expressed exactly as

$$F^{e}[\rho^{e}(\varepsilon)] = k_{B}T \int d\varepsilon \begin{bmatrix} \left(\rho^{e}(\varepsilon) - \rho^{e}_{0}(\varepsilon)\right) - \rho^{e}(\varepsilon)\log\left(\frac{\rho^{e}(\varepsilon)}{\rho^{e}_{0}(\varepsilon)}\right) \\ -\beta\left(\rho^{e}(\varepsilon) - \rho^{e}_{0}(\varepsilon)\right) \int_{0}^{1} d\lambda \omega^{e}(\varepsilon; \rho^{e}_{\lambda}) \end{bmatrix}$$
(7.5)

when u_{λ} is taken so that ρ_{λ}^{e} varies linearly against λ . Note the parallelism of Eqs. (7.2)-(7.5) to Eqs. (5.2)-(5.5). In Eqs. (7.4) and (7.5), u and ω are written to depend on the distribution function ρ_{λ}^{e} , instead of the potential u_{λ} , by virtue of the property of one-to-one correspondence (18). In the energy representation, the HNC-type and PY-type approximations are obtained by assuming the linear dependencies of ω^{e} and exp $(-\beta\omega^{e})$ -1 on λ , respectively. Although Eq. (5.5) is hard to implement due to the highdimensionality of **x**, Eq. (7.5) is straightforward to handle since ε is one-dimensional.

In the currently used version of the method of energy representation (19,20,21), the solvation free energy $\Delta\mu$ is approximately expressed in terms of distribution functions constructed from $\hat{\rho}^e$ in the solution and reference solvent systems. In our treatments, the solution system refers to the system in which the solute molecule interacts with the solvent under the solute-solvent interaction v of interest at full coupling. In the solution, the average distribution ρ^e of the v value is relevant in the approximate construction of $\Delta\mu$ and is given by

$$\rho^{e}(\varepsilon) = \left\langle \hat{\rho}^{e}(\varepsilon) \right\rangle \tag{7.6}$$

where $\langle ... \rangle$ represents the ensemble average in the solution system of interest. On the other hand, the reference solvent system denotes the system in which the solute does not interact with the solvent and the solvent configuration is generated only by the solvent-solvent interaction. At an instantaneous configuration of the reference solvent system, $\hat{\rho}^e$ is constructed by placing the solute molecule in the system as a test particle without disturbing the solvent configuration. The average distribution ρ_0^e and the correlation matrix χ_0^e then appear in the approximate functional for $\Delta \mu$ and are expressed, respectively, as

$$\rho_0^e(\varepsilon) = \left\langle \hat{\rho}^e(\varepsilon) \right\rangle_0 \tag{7.7}$$

and

$$\chi_{0}^{e}(\varepsilon,\eta) = \left\langle \hat{\rho}^{e}(\varepsilon) \hat{\rho}^{e}(\eta) \right\rangle_{0} - \left\langle \hat{\rho}^{e}(\varepsilon) \right\rangle_{0} \left\langle \hat{\rho}^{e}(\eta) \right\rangle_{0}$$
(7.8)

where $\langle ... \rangle_0$ represents the ensemble average in the reference solvent system. In the sampling corresponding to $\langle ... \rangle_0$, the solute and solvent degrees of freedom are uncoupled from each other in the probability distribution.

An approximate functional for $\Delta\mu$ is derived in Ref. (19). The functional is constructed by adopting the PY-type approximation in the unfavorable region of the solute-solvent interaction and the HNC-type approximation in the favorable region. $\Delta\mu$ is then given by a set of definitions and equations listed as

$$w^{e}(\varepsilon) = -k_{B}T \log\left(\frac{\rho^{e}(\varepsilon)}{\rho_{0}^{e}(\varepsilon)}\right) - \varepsilon$$
(7.9)

$$w_0^e(\varepsilon) = -k_B T \int d\eta \left(\frac{\delta(\varepsilon - \eta)}{\rho_0^e(\varepsilon)} - \left(\chi_0^e\right)^{-1}(\varepsilon, \eta) \right) \left(\rho^e(\eta) - \rho_0^e(\eta) \right)$$
(7.10)

 $\Delta \mu = \int d\varepsilon \ \varepsilon \rho^e(\varepsilon)$

$$-k_{B}T\int d\varepsilon \begin{bmatrix} \left(\rho^{e}(\varepsilon)-\rho_{0}^{e}(\varepsilon)\right)-\rho^{e}(\varepsilon)\log\left(\frac{\rho^{e}(\varepsilon)}{\rho_{0}^{e}(\varepsilon)}\right)\\ -\left\{\alpha(\varepsilon)F(\varepsilon)+(1-\alpha(\varepsilon))F_{0}(\varepsilon)\right\}\left(\rho^{e}(\varepsilon)-\rho_{0}^{e}(\varepsilon)\right)\end{bmatrix}$$
(7.11)

$$F(\varepsilon) = \begin{cases} \beta w^{\varepsilon}(\varepsilon) + 1 + \frac{\beta w^{\varepsilon}(\varepsilon)}{\exp(-\beta w^{\varepsilon}(\varepsilon)) - 1} & \text{(when } w^{\varepsilon}(\varepsilon) \le 0) & (7.12) \end{cases}$$

$$\left(\frac{-2}{2}\beta w^{\varepsilon}(\varepsilon) \right) \quad (\text{when } w^{\varepsilon}(\varepsilon) \ge 0)$$

$$F_{0}(\varepsilon) = \begin{cases} -\log(1 - \beta w_{0}^{\varepsilon}(\varepsilon)) + 1 + \frac{\log(1 - \beta w_{0}^{\varepsilon}(\varepsilon))}{\beta w_{0}^{\varepsilon}(\varepsilon)} & \text{(when } w_{0}^{\varepsilon}(\varepsilon) \le 0) \end{cases}$$
(7.13)

$$\alpha(\varepsilon) = \begin{cases} 1 & (\text{when } \nu_0(\varepsilon) \ge 0) \\ 1 - \left(\frac{\rho^e(\varepsilon) - \rho^e_0(\varepsilon)}{\rho^e(\varepsilon) + \rho^e_0(\varepsilon)}\right)^2 & (\text{when } \rho^e(\varepsilon) \le \rho^e_0(\varepsilon)) \end{cases}$$
(7.14)

When Eqs. (7.9)-(7.14) are used to evaluate the solvation free energy, the inputs are the three energy distribution functions ρ^e , ρ^e_0 , χ^e_0 given by Eqs. (7.6)-(7.8), respectively. ρ^e is obtained from a simulation of the solution system of interest, and ρ^e_0 and χ^e_0 are constructed with a simulation of the reference solvent system. The approximate scheme provided by Eqs. (7.6)-(7.14) does not require simulations at the intermediate states of solute insertion. This leads to substantial reduction of computational load in free-energy calculation. On the other hand, when Eqs. (7.9)-(7.14) are used with ρ^e , ρ^e_0 , χ^e_0 obtained from molecular simulation, a possible artifact in simulation arising, for example, from the use of periodic boundary condition and/or the handling scheme of electrostatic interaction (cutoff, Ewald, particlemesh Ewald, and so on), is generally retained.

It should be noted that when the solute molecule is inserted into the reference solvent system, it often overlaps with solvent molecules. The overlapping configurations contribute to ρ_0^e and χ_0^e at large energy coordinates and accounts for the excluded-volume effect in the solvation free energy. In addition, Eqs. (7.9)-(7.14) show that the ε value itself acts only as an index for the distribution functions ρ_0^e and χ_0^e when ε is large enough and the corresponding ρ^e is (essentially) zero. The ε value in the excluded-volume region is not used in the calculation of $\Delta \mu$, except in the construction of ρ_0^e and χ_0^e ; $\Delta \mu$ calculation in the large ε region is not affected by the change in the ε value as far as ε varied corresponds to the excluded volume. This property is of course satisfied for the exact functional for the solvation free energy, and is kept valid for the approximate one given by Eqs. (7.9)-(7.14).

In Figure 4, the approximate values of the solvation free energy $\Delta \mu$ for typical solute molecules in solvent water are compared to the corresponding exact values obtained from the free-energy perturbation method. The (solvent) density of 1.0 g/cm³ and the temperature of 25 °C is an ambient state, and the densities of 1.0, 0.6, and 0.2 g/cm³ at 400 °C correspond to high-, medium-, and low-density supercritical states. The good agreement is observed between the approximate and exact values. The agreement is particularly notable at the medium- and lowdensity states of 0.6 and 0.2 g/cm³ and 400 °C. When the solute is ionic, the density at the state of 0.2 g/cm³ and 400 °C is not yet low in the sense, for example, that the hydration number at that state is comparable to the numbers at ambient states (22,23). Even in this case, our approximate procedure is effective in determining the solvation free energy. The solvation free energies of water at 1.0 g/cm³ and 400 °C and of methanol and ethanol at 0.6 g/cm³ and 400 °C are rather small in magnitude. These behaviors are caused by the balance between the favorable and unfavorable contributions of the solute-solvent interactions, and are well reproduced by our approximate method. Therefore, the single functional given by Eqs. (7.6)-(7.14) provides an accurate and efficient route to the solvation free energy for various types of solutes over a wide range of thermodynamic conditions.

By virtue of Eq. (5.1), the average sum $\langle u \rangle$ of the solute-solvent interaction energy in the solution system of interest is smaller than or equal to $\Delta \mu$. This means that the density-functional F^e is always non-positive for any solute-solvent distribution function. Actually, the densityfunctional is a measure of the "difference" between ρ^e in the solution and ρ_0^e in the reference solvent. It is zero only when $\rho^e = \rho^e_0$. The density-functional is expected to be more negative when ρ^e and ρ^e_0 appear more differently. On the other hand, the first term of Eq. (7.3) is equal to $\langle u \rangle$, and is more negative when ρ^e is more populated in the lowenergy region of ε . A typical behavior is that ρ^{e_0} reduces monotonically toward the low-energy tail. Thus, the first term of Eq. (7.3) is more negative when ρ^e and ρ^e_0 are more different. This indicates that the first and second terms of Eq. (7.3) fluctuate to the same direction through the variation of ρ^{e} . It is then expected that $\Delta \mu$ of Eq. (7.3) converges faster in molecular simulation than its components expressed as the first and second terms of Eq. (7.3). Indeed, usual experience is that when (an approximate form of) Eq. (7.5) is employed, the solvation

Method of energy representation



Figure 4. The solvation free energy $\Delta \mu$ calculated by the approximate method of energy representation and the exact, free-energy perturbation method. The thermodynamic state is specified by the solvent density and temperature.

free energy $\Delta \mu$ converges faster than the average sum $\langle u \rangle$ of the solute-solvent interaction energy in the solution.

In the energy representation, the interaction energy between the solute and solvent is adopted for the one-dimensional coordinate of the distribution functions, and a functional for the solvation free energy is constructed from energy distribution functions in the solution and reference solvent systems of interest. The introduction of the energy coordinate for distribution functions is a kind of coarse-graining procedure for reducing the information content of the solute-solvent configuration. Since the set of configurations (structures) with the equal solute-solvent interaction energies are grouped into a unit in the energy representation, any approximate functional built in terms of energy distribution functions cannot violate the statisticalmechanical principle that the configurations with the same solute-solvent interaction energies contribute to the solvation free energy with the equal weights.

As seen in Eq. (7.1), each of the solute and solvent molecules is taken as a single unit in the energy representation. The molecule is treated as a whole, while the coordinate for the distribution functions is onedimensional. No explicit reference is made to the detail of the molecular structure by focusing on the interaction energy. The following advantages then emerge in the method of energy representation.

Firstly, the method is straightforwardly applicable to molecules with intramolecular flexibility. The implementation is indifferent whether the molecule is rigid or flexible. The information of structural fluctuation of the molecule is adsorbed when the energy coordinate is introduced by Eq. (7.1). For large molecules constituting micellar, membrane, and protein systems, it is not allowed to neglect the molecular flexibility. In the method of energy representation, an additional and/or separate scheme is not necessary to be formulated for large, flexible species.

Secondly, the treatment of inhomogeneous system and clusters is straightforward. So far, the formulation does not assume the system homogeneity and the thermodynamic limit. The application to inhomogeneous and/or finite systems is then possible without modification. The binding of a molecule to such nanoscale structures as micelle, membrane, and protein can be viewed as a solvation in an inhomogeneous and finite, mixed solvent (24,25). The method of energy representation can thus be a useful approach to intermolecular correlation and association important in biological and interface sciences.

Thirdly, an accurate treatment is possible for supercritical fluid. In supercritical fluid, the solvent density and temperature can be varied over wide range and the solvent effect may act as a key to control a chemical process. It is well-known that supercritical fluid can be described accurately when the whole molecule is treated as a single unit (1). A multidimensional representation is necessary, however, in the usual coordinate space. By introducing the energy as the coordinate for distribution functions, the whole molecule can be taken as a single unit with keeping the description one-dimensional. The approximate functional given by Eqs. (7.6)-(7.14) incorporates the intermolecular correlation at the two-body level. The solvation free energy obtained is then exact to second order in the solvent density. Since the method is exact in the low-density regime, a formulation of a good approximation in the high-density regime leads to an accurate description over a wide range of solvent density.

Finally, the combination with the quantummechanical/molecular-mechanical (QM/MM) methodology can be performed. In QM/MM calculation, the many-body effect is introduced for the solute-solvent interaction and is beyond the applicability of conventional theories of solutions. In the method of energy representation, the fluctuation of the electronic state in response to the environment is viewed as a fluctuation of intramolecular degrees of freedom of the QM solute. The evaluation becomes feasible for the free energy for the many-body effect of electronic fluctuation. In addition, Eq. (7.1) makes no reference to the functional form of the potential function. It refers only to the value of the potential energy, and there is no need for deterioration or modification of the electronic-state calculation. Thus, the treatment is possible for an arbitrary distribution of charges. The contraction to a set of point charges is not necessary, and the effect of the diffuse (cloud-like) nature of electronic distribution can now be determined. The detail of the combination with the QM/MM methodology is given in Ref. (21).

8. APPLICATION TO MOLECULAR BINDING INTO MICELLE AND MEMBRANE SYSTEMS

Micelle is a self-assembled aggregate of amphiphilic surfactants in water. It involves a hydrophobic core and provides a favorable environment for organic compounds. Upon formation of a micelle, the solubility is often enhanced for an organic compound which is insoluble or sparingly soluble in water. This phenomenon is called solubilization, and is a most important function of a micelle.

Membrane is a soft, self-organizing aggregate of amphiphilic lipids in aqueous solution. It distinguishes one side of the solution from the other, and plays important roles in distribution and transport of a molecule. The key process toward the membrane function is the binding of a molecule into membrane.

On the microscopic level, the solubilization and membrane-binding are common in that they correspond to the transfer of a molecule from bulk water to the inside (or surface) of a nanoscale, self-organizing structure in solution. In both cases, the process is quantified by the (standard) free energy of binding. The binding free energy is determined by the cooperation and/or competition of the interactions among the surfactant or lipid, the solute molecule to be bound, and the solvent (water and cosolvents if present). A unified framework for the free- energy analysis at atomic resolution is thus desirable to be developed for the molecular understanding and control of the binding in micelle and membrane.

The basic idea of our scheme is to view a micellar or membrane solution as a mixed solvent. The surfactant or lipid molecules are treated not as solute species, but as part of the mixed solvent system. The molecule to be bound into the micelle or membrane is the only species regarded as the solute. In this view, the solvation free energy of the solute molecule in a micellar or membrane solution denotes the free-energy change for turning on the interactions of the solute with the solvent water, surfactants or lipids, and counterions if present. The binding is described by the difference in the solvation free energy between the micellar or membrane system and the neat solvent (pure water) system. When the host (micelle or membrane) is dilute, in particular, it is natural to set the origin of the solution at (or near) the center of the host. The mixed solvent system then consists of the surfactant or lipid



Figure 5. (a) The structure of the dodecyl sulfate anion. (b) The structure of DMPC. The molecule is divided into three portions of the hydrophobic tail, glycerol backbone, and hydrophilic headgroup. In the present work, the division is done as denoted in the figure, though the diving scheme is always of some ambiguity.

molecules forming the host, water, and distributed counterions (when the surfactant or lipid is ionic). Water is deficient in the interior of the host, and the surfactant or lipid molecules belonging to the host are localized around the origin. The mixed solvent is thus inhomogeneous even before the solute insertion. When the host is micelle, furthermore, the number of surfactants is finite since the aggregation number is typically 10^2 - 10^4 . The micellar solution is partially finite in the sense that the number is finite for the surfactant (and counterion) involved as solvent species in the dilute micellar solution. The point of our scheme is to evaluate the solvation free energy of a solute in an inhomogeneous and partially finite, mixed solvent system.

Even when the system is inhomogeneous and partially finite, a formally exact free-energy calculation is possible by the standard free energy perturbation and thermodynamic integration methods (1,2,3). These methods are notorious for the heavy computational demand, however, since an explicit reference needs to be made to the intermediate states of the gradual variation process of the system. We combine the large-scale molecular simulation with the method of energy representation to evaluate the binding free energy.

The surfactant employed is sodium dodecyl sulfate (SDS) and the lipid is 1,2-dimyristoyl-*sn*-glycero-3-phosphatidylcholine (DMPC). Their structures are illustrated in Figure 5. The solute molecule is methane,

benzene, and ethylbenzene in the micelle simulation, and is CO, CO₂, benzene, and ethylbenzene in the membrane simulation. The computational procedures are fully described in Refs. (24,25).

In Figure 6, we show the density profile of SDS micelle and DMPC bilayer. Figure 6(a) gives the densities of the hydrophobic tail, the headgroup, and water as functions of the distance r from the micellar center identified as the center of mass of the dodecyl sulfate anions. Since the eccentricity of the SDS micelle is small, it is natural to divide the micellar system into a set of regions on the basis of Figure 6(a). We introduce 6 regions in terms of the center-of-mass distance r of the solute from the micellar center by concentrically dividing the domain of r <30 Å with an equal interval of 5 Å. In the following, the regions are numbered I ... VI from the micelle inside to outside. According to Figure 6(a), the regions I-III corresponds to the hydrophobic core, where water is scarcely present. The region IV refers to the headgroup region in contact with water, and is the transition region from the hydrophobic core to the aqueous region outside the micelle represented by the regions V and VI.

In the membrane system simulated, the z-axis is taken to be normal to the plane of the DMPC bilayer and z = 0 is set to the z-coordinate of the center of mass of the DMPC molecules. In Figure 6(b), we show the densities of the hydrophobic tail, the glycerol backbone, the hydrophilic headgroup, and water obtained from the simulation of



Figure 6. (a) The densities of the hydrophobic tail, the headgroup, and water in the SDS micellar system as functions of the distance *r* from the center of mass of the dodecyl sulfate anions forming the micelle. The density of the hydrophobic tail refers to the sum of the (number) densities of the methyl and methylene groups of the dodecyl sulfate anion, and the density of the headgroup is the sum (number) density of the sulfur and oxygen atoms. The water density is expressed with respect to the center of mass of the water molecule. Six regions are introduced by dividing the domain of r < 30 Å with an interval of 5 Å, and are numbered I ... VI from the micelle inside to outside. (b) The densities of the hydrophobic tail, the glycerol backbone, the hydrophilic headgroup, and water in the DMPC membrane system as functions of the separation *z* along the *z*-axis from the averaged density is shown against the positive *z*-abscissa. The densities of the tail and glycerol refer to the sums of the (number) densities of the carbons, oxygens, nitrogen, and phosphorus in the corresponding portion. The water density is expressed with respect to the center of mass of the (number) densities of the carbons, oxygens, nitrogen, and phosphorus in the corresponding portion. The water density is expressed with respect to the center of mass of the water molecule. Six regions are introduced by dividing the domain of z < 30 Å with an interval of 5 Å, and are numbered I ... VI from the membrane inside to outside.

DMPC and water without solute. Since the (converged) distribution of particles is symmetric against the reflection at z = 0, the density averaged over the positive and negative z-domains is displayed only against the positive z-abscissa. As done for the SDS micelle, we introduce 6 regions of z by dividing the domain of |z| < 30 Å with an equal interval of 5 Å. The regions are numbered I ... VI from the membrane inside to outside. According to Figure 6(b), the regions I and II correspond to the hydrophobic interior, where water is scarcely present. The glycerol and headgroup are mainly located in the regions III-V, where the water density gradually varies. The density profile recovers the bulk behavior in the region VI, and the thickness of (one leaflet of) the DMPC layer is ~20 Å.

In Figure 7, we show the solvation free energy $\Delta\mu$ of the solute in the regions I-VI and in the bulk water. The bulk denotes the region far from the host (micelle or membrane), and $\Delta\mu$ in the bulk is the solvation free energy in neat water. Evidently, each hydrophobic solute is free-energetically stabilized in the inside of the host.

When the host is micelle, the distribution within the hydrophobic interior (regions I-III) is rather

diffuse. For benzene and ethylbenzene in the micelle, the $\Delta \mu$ difference between the regions I-III and IV shows that these solutes are localized in the hydrophobic core of the micelle. When the solute is methane and is bound into the SDS micelle, in contrast, the $\Delta \mu$ difference between the regions I-III and IV is small and the probability of finding the solute is appreciable also in the headgroup region. Thus, the solubilization of benzene and ethylbenzene is more sharply characterized as a transfer from the aqueous to hydrophobic environment. For all the solutes examined, the distribution is negligible in the aqueous regions V-VI outside the micelle. This is a support to the pseudophase model, and the support is stronger for benzene and ethylbenzene.

When the host is membrane, Figure 7(b) shows that $\Delta \mu$ varies by $\sim k_B T$ within the regions I-IV. A strong tendency of localization is not observed, and the hydrophilic portion of DMPC also interacts with the hydrophobic solute. In the SDS micelle, benzene and ethylbenzene are more strongly localized in the hydrophobic interior. This contrast corresponds, as evidenced in the molecular structure of Figure 5, to the fact that DMPC involves methylene and methyl groups in the hydrophilic segment, unlike SDS.



Figure 7. The solvation free energy $\Delta \mu$ in the regions I-VI and in the bulk for (a) the SDS micellar and (b) the DMPC membrane systems. The bulk denotes the region far from the host, and $\Delta \mu$ in the bulk is equal to the solvation free energy of the solute inserted into neat water. The error bar is expressed at 95% confidence level, and is smaller than the size of the corresponding data symbol when it is not shown. The lines connecting the data are drawn for eye guide.

9. PERSPECTIVE

In the present review, we introduce a new theory of solutions. The theory is formulated to satisfy the requirements for applications to diverse systems listed as (a) supercritical fluid covering wide range of density from gas-like, low density to liquid-like, high density

(b) flexible molecule (and ion)

(c) inhomogeneous system

(d) QM/MM system involving many-body interactions

The treatment of these systems becomes possible by adopting the solute-solvent interaction energy as the coordinate of distribution functions and formulating the density-functional theory over the energy coordinate. The distribution functions are obtained from molecular simulation, and an efficient and accurate free-energy calculation is now possible with an (approximate) functional.

According to Eqs. (7.9)-(7.14), a numerical

difficulty arises when ρ^e is non-zero and ρ_0^e is zero in some region of ε (there is no problem, on the other hand, when ρ^e is zero). This is encountered when the solute or solvent changes its (intramolecular) structure significantly through solute-solvent interaction. When the variation of the solute structure is large, the difficulty can actually be circumvented by modifying the sampling scheme with knowledge of computational results in solution (20). The functional for the solvation free energy is then modified in response to the sampling scheme. A challenge is present when the (intramolecular) structure changes drastically for the solvent. When the protein and water are viewed as a mixed solvent, the induced fit corresponds to such category. The concept of energy coordinate greatly expands the applicability of solution theory. A particular example is the combination with the QM/MM method. In the QM/MM method, the solute-solvent interaction is not pairwise additive any more and is beyond the applicability of conventional theory of solutions. With the energy concept, furthermore, micelle and membrane are now within the unified scope of the solution theory. The extensions to protein system and the connection to coarsegraining description are awaited.

10. ACKNOWLEDGMENTS

The author is grateful to Professors M. Nakahara and M. Kinoshita of Kyoto University, Professor H. Takahashi of Osaka University, and Professor M. Ikeguchi of Yokohama City University for valuable discussions and collaboration concerning the development and application of the methodology. This work is supported by the Grantsin-Aid for Scientific Research (Nos. 11740322, 13640509, 15205004, and 18350004) from the Japan Society for the Promotion of Science and by the Grantsin-Aid for Scientific Research on Priority Areas (Nos. 15076205, 20038034, 20118002) and the and Nanoscience Program of the Next-Generation Supercomputing Project from the Ministry of Education, Culture, Sports, Science, and Technology. The author further acknowledges the JST (Japan Science and Technology Agency)-CREST grant, a grant from the Association for the Progress of New Chemistry, a grant from the Suntory Institute for Bioorganic Research, and the Supercomputer Laboratory of Institute for Chemical Research, Kyoto University, for generous allocation of computation time.

11. REFERENCES

1. J. P. Hansen and I. R. McDonald: Theory of Simple Liquids, 2nd ed., *Academic Press* London (1986)

2. M. P. Allen and D. J. Tildesley: Computer Simulation of Liquids, *Oxford University Press* Oxford (1987)

3. D. Frenkel and B. Smit: Understanding Molecular Simulation, From Algorithms to Applications, 2nd ed., *Academic Press* London (2002)

4. T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber and W. F. van Gunsteren: Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem Phys Lett* 222, 529-539 (1994)

5. H. Resat and M. Mezei: Studies on free energy calculations. II. A theoretical approach to molecular solvation. *J Chem Phys* 101, 6126-6140 (1994)

6. M. Zacharias, T. P. Straatsma and J. A. McCammon: Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. *J Chem Phys* 100, 9025 (1994)

7. N. Lu, J. K. Singh and D. A. Kofke: Appropriate methods to combine forward and reverse free-energy perturbation averages. *J Chem Phys* 118, 2977-2984 (2003)

8. M. R. Shirts and V. S. Pande: Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration. *J Chem Phys* 122, 144107 (16 pages) (2005)

9. N. Matubayasi, L. H. Reed and R. M. Levy: Thermodynamics of Hydration Shell. 1. Excess Energy of a Hydrophobic Solute. *J Phys Chem* 98, 10640-10649 (1994)

10. N. Matubayasi: Matching-mismatching of Water Geometry and Hydrophobic Hydration. *J Am Chem Soc* 116, 1450-1456 (1994)

11. D. Chandler and H. C. Andersen: Optimized Cluster Expansions for Classical Fluids. II. Theory of Molecular Liquids. *J Chem Phys* 57, 1930-1937 (1972)

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

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

14. A. Kovalenko and 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)

15. S. M. Kast: Combinations of simulation and integral equation theory. *Phys Chem Chem Phys* 3, 5087-5092 (2001)

16. H. Freedman and T. N. Truong: A coupled RISM/MD or MC simulation methodology for solvation free energies. *Chem Phys Lett* 381, 362-367 (2003)

17. H. Freedman and T. N. Truong: Coupled reference interaction site model'simulation approach for thermochemistry of solvation: Theory and prospects. *J Chem Phys* 121, 2187-2198 (2004)

18. N. Matubayasi and M. Nakahara: Theory of solutions in the energetic representation. I. Formulation. *J Chem Phys* 113, 6070-6081 (2000)

19. N. Matubayasi and M. Nakahara: Theory of solutions in the energy representation. II. Functional for the chemical potential. *J Chem Phys* 117, 3605-3616 (2002); erratum, *J Chem Phys* 118, 2446 (2003)

20. N. Matubayasi and M. Nakahara: Theory of solutions in the energy representation. III. Treatment of the molecular flexibility. *J Chem Phys* 119, 9686-9702 (2003)

21. H. Takahashi, N. Matubayasi, M. Nakahara and T. Nitta: A quantum chemical approach to the free energy calculations in condensed systems: The QM/MM method combined with the theory of energy representation. *J Chem Phys* 121, 3989-3999 (2004)

22. P. B. Balbuena, K. P. Johnston and P. J. Rossky: Computer Simulation Study of an sN2 Reaction in Supercritical Water. *J Phys Chem* 99, 1554-1565 (1995)

23. P. B. Balbuena, K. P. Johnston and P. J. Rossky: Molecular Dynamics Simulation of Electrolyte Solutions in Ambient and Supercritical Water. 1. Ion Solvation. *J Phys Chem* 100, 2706-2715 (1996)

24. N. Matubayasi, K. K. Liang and M. Nakahara: Freeenergy analysis of solubilization in micelle. *J Chem Phys* 124, 154908 (13 pages) (2006)

25. M. Matubayasi, W. Shinoda and M. Nakahara: Freeenergy analysis of the molecular binding into lipid membrane with the method of energy representation. *J Chem Phys* 128, 195107 (13 pages) (2008)

Key Words: Free Energy, Solvation, Distribution Function, Energy Representation, Solution, Water, Micelle, Membrane, Biomolecular Hydration, Supercritical Fluid, Review

Send correspondence to: Nobuyuki Matubayasi, Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011 Japan, Tel: 81-774-38-3071, Fax: 81-774-38-3076, Email: nobuyuki@scl.kyoto-u.ac.jp

http://www.bioscience.org/current/vol14.htm