

Cavity Particle in Aqueous Solution with a Hydrophobic Solute: Structure, Energetics, and Functionals

Published as part of The Journal of Physical Chemistry virtual special issue "Peter J. Rossky Festschrift".

Bin W. Zhang, Nobuyuki Matubayasi,* and Ronald M. Levy*



Cite This: <https://dx.doi.org/10.1021/acs.jpcb.0c02721>



Read Online

ACCESS |



Metrics & More



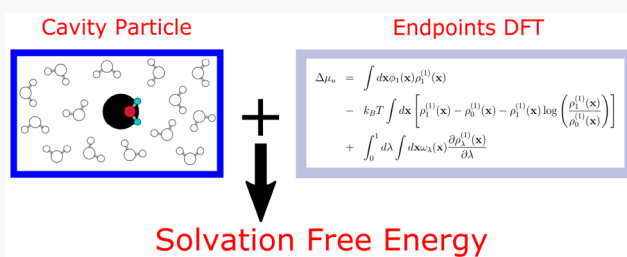
Article Recommendations



Supporting Information

ABSTRACT: Endpoints density functional theory (DFT) provides a framework for calculating the excess chemical potential of a solute in solution using solvent distribution functions obtained from both physical endpoints of a hypothetical charging process which transforms the solvent density from that of the pure liquid to the solution state. In this work, the endpoints DFT equations are formulated in terms of the indirect (solvent-mediated) contribution $\omega(\mathbf{x})$ to the solute–solvent potential of mean force, and their connections are established with the conventional DFT expressions which are based on the use of direct correlation functions. ω

actually corresponds to the free-energy cost to move a cavity particle (a tagged solvent molecule which interacts with the other solvent molecules but not the solute) from the bulk to the configuration $\mathbf{x}$ of a solvent molecule relative to the solute and is a suitable variable to describe the solvent effects on the solute–solvent interactions. HNC and PY type approximations are then used to integrate the DFT charging integral involved in the exact expression for the excess chemical potential. With these approximations, molecular simulations are to be performed at the two endpoints of solute insertion: pure solvent without the solute and the solution system with the fully coupled solute–solvent interaction. An endpoints method thus utilizes the ensembles of intermolecular configurations of physical interest, which are often readily accessible with MD simulations given the present computational power. To illustrate properties of the formulation, we perform simulations of model systems consisting of a cavity particle in an aqueous solution containing a spherical hydrophobic solute of three different sizes from which $\omega(\mathbf{x})$ and the solute chemical potential can be calculated using endpoints DFT expressions. These are compared with corresponding results obtained using the approximations needed in order to evaluate the endpoints DFT charging integral when cavity particle simulation data is not available. We analyze a new approximation (two-points quadratic HNC) to the DFT charging integral which captures the correct behavior of the cavity distributions at both endpoints of the solute insertion. The behavior of the cavity particle in simple and complex liquids plays an important role in various theoretical treatments of the solute chemical potential. For pure Lennard-Jones fluids, the free energy to bring a cavity particle from the bulk to the center of a fluid particle is negative. However, for solutes of varying size, this is not generally true for Lennard-Jones fluids or the systems studied in this work. We carry out energetic and structural analyses of the cavity particle in aqueous solution with hydrophobic solutes of varying size and discuss the results in the context of the hydrophobic effect.



1. INTRODUCTION

Classical density functional theory (DFT) provides a framework for studying the effects of solvation on a rich variety of phenomena which are of current interest in chemistry, biophysics, and materials science. While the classical DFT equations are usually formulated in terms of direct correlation functions, it is possible to cast the DFT equations instead in terms of the indirect part of the solute–solvent potential of mean force (indirect PMF) $\omega(\mathbf{x})$, where $\mathbf{x}$ is the configuration of a solvent molecule relative to the solute.^{1–4} The indirect term $\omega(\mathbf{x})$ has a physical interpretation as a free-energy difference that can be calculated by simulation in several different ways. The cavity distribution function $y(\mathbf{x})$ is the distribution function of a tagged solvent molecule in a solution

which interacts with all the other untagged solvent molecules but does not interact with the solute. The relationship between $\omega(\mathbf{x})$ and $y(\mathbf{x})$ is given by $\omega(\mathbf{x}) = -kT \log y(\mathbf{x})$. The properties of the cavity distribution function have been well-studied for Lennard-Jones fluids and solutions,^{5–7} but they are much less well-studied for aqueous solutions.^{8–10} The behavior of the

Received: March 27, 2020

Revised: May 28, 2020

Published: May 29, 2020

indirect PMF, $\omega(\mathbf{x})$ (or the direct correlation function $c(\mathbf{x})$), inside the excluded volume of the solute plays an important role in the determination of the solute chemical potential by DFT. Therefore, analysis of the behavior of $y(\mathbf{x})$ inside the solute excluded volume region by simulation can be used as a benchmark to compare with the predictions of approximate free-energy functional expressions for the solute chemical potential and also to guide the development of improved functionals.

In 2000 one of us (N.M.) proposed an approximate functional for calculating the excess chemical potential of a solute in solution which uses solvent–solvent distribution functions obtained from the pure liquid and the solute in solution (“endpoints DFT”) to construct the functional.^{1–3} The endpoints DFT functional can be constructed using one-dimensional solute–solvent interaction energy distributions or mixed four-dimensional (energy, position) distributions; this is a significant computational advantage as compared with using full six-dimensional (position, orientation) distributions.¹¹ In recent years endpoints DFT has been applied to a diverse array of problems in biophysics, polymers, and interfaces.^{12,13}

In this work, we provide the theoretical foundation of the endpoints DFT method. The key role in our formulation is played by the indirect PMF $\omega(\mathbf{x})$. This is a distinct feature which is different from the classical DFT formulation that employs direct correlation functions.^{14,15} The connection between the two formulations will be established. Although the formulation of the endpoints functional with the indirect PMF was presented in the literature,^{2,11,13} its relationship to the classical formulation was not described in an explicit and systematic manner previously. In fact, $\omega(\mathbf{x})$ (or the cavity distribution function) describes the collective effects of solvent on the solute–solvent correlation. A formulation relying upon $\omega(\mathbf{x})$ can thus be suitable for free energetics of solvation, in which the solvent effects are to be pursued. The exact functional for the excess chemical potential then involves an integration over the gradual insertion process of the solute, and approximations of HNC and PY type are introduced to express the functional in terms of distribution functions only at the endpoints of solute insertion. The endpoints are the pure solvent without solute and the solution system with the solute–solvent interaction at full coupling. They are of physical interest and can often be simulated readily with the computation power available at present. Typically, the classical DFT approximation scheme provides solvation thermodynamic properties using distribution functions only in pure solvent. In the endpoints scheme employed here, on the other hand, solvation thermodynamic properties are obtained from a density functional using distribution functions constructed from both endpoint states; the pure solvent and the solution of interest. The endpoints method thus utilizes the DFT formalism, together with solute–solvent distribution functions at the physical endpoints, to efficiently and accurately evaluate the excess chemical potential, which requires substantially more computational effort using standard methods like thermodynamic integration and free-energy perturbation.^{16,17}

On the basis of the formulation described in the preceding paragraph, we use computer simulations to analyze how $\omega(\mathbf{x})$ behaves within the solute core for a series of three methane-like hydrophobic solutes of increasing size in water. In the limit of very small solute size, at the center of the solute, it is well-known that $\omega(\mathbf{x}) < 0$ for solvation in the LJ solvent,^{5,7} while as the size of the solute grows large, $\omega(\mathbf{x}) > 0$ and approaches $-\mu$

(where μ is the solvation free energy of the solvent in the solution), both for the LJ solvent and for water. We contrast the qualitative behavior of $\omega(\mathbf{x})$ for the three methane-like solutes of different sizes in water and comment on the implications for theories of hydrophobic solvation. We also analyze how the hydrophobic solute–water interfacial structure is distorted as the cavity particle passes through the interface.^{8,18–35} In fact, the indirect PMF (the PMF of a cavity particle around a solute) in aqueous solutions has not been studied extensively.^{7–10} More in-depth studies have been restricted to the properties of cavity particles in simple liquids.^{5,6,15,36} Our purpose in examining a set of model solutes in water is to catalogue the collective effects of water on the cavity particle distribution around a hydrophobic solute, with the specific aim of evaluating quantities which are central to the construction of free-energy functionals within the framework of endpoints DFT, and to explore a possible way to improve the approximations.

In endpoints DFT, the excess contribution to the solvent reorganization free energy associated with linearly charging the density from that of the pure liquid to the solution distribution is estimated by a functional constructed using the simulations of the two physical endpoints. Within the solute core, however, the functional is constructed using only the pure water simulation data by inverting the pure water–water fluctuation correlation function. For larger solutes, it is possible to improve the approximation of the functional within the solute core by incorporating the results of cavity particle simulations. We describe a way to incorporate this information which leads to better treatment of the excluded volume effect for larger solutes within the endpoints DFT framework.

The organization of this paper is as follows. In section 2, we derive the formulas which relate the indirect PMF $\omega(\mathbf{x})$, which is the central focus of endpoints DFT, to the direct correlation function $c(\mathbf{x})$. In section 3, we describe the simulation methods, including a description of how the potential of mean force is constructed as a function of the distance from the hydrophobic solute. We present the results in section 4 and discussion in section 5, with conclusions in section 6.

2. THEORY

This work is concerned with the chemical potential of hydrophobic solutes in aqueous solution from the perspective of density functional theory. As described below, the endpoints DFT expression for the solute chemical potential is derived by combining an expression for the solvent density in the inhomogeneous solution based on the (Widom) potential distribution theorem with a solvent density charging formula that is constructed by performing a partial integration of the Kirkwood charging formula for the solvation free energy.^{1,2,13,37,38} This transforms the charging process from one which charges up the solute–solvent interaction potential to one which charges up the solvent density from the initial pure liquid state to the final inhomogeneous solution state. In this formulation of the problem, the key quantity we focus on is the indirect part of the solute–solvent potential of mean force $\omega(\mathbf{x})$. The endpoints DFT equations can be derived in the canonical or the grand canonical ensemble. The starting point for the standard DFT formulation of this problem is the grand potential.^{14,15} The functional derivative of the excess part of the grand potential is the one-point direct correlation function $c^{(1)}(\mathbf{x})$. Integration of the functional derivative leads to a Kirkwood-type charging formula of the direct correlation

function to obtain the excess free-energy contribution to the chemical potential of the solute. In this section, we show that the two expressions for the chemical potential of the solute in solution, namely the endpoints DFT expression and the standard DFT expression, are equivalent.

We adopt the grand canonical ensemble in the following. The solute molecule is fixed at the origin, and the configuration of a solvent molecule is expressed as $\mathbf{x}$. The position and orientation are collectively represented by $\mathbf{x}$, and the intramolecular degrees of freedom are also incorporated into $\mathbf{x}$ if the molecule is flexible. We treat a family of solute–solvent interactions $\phi_\lambda(\mathbf{x})$, where λ is the coupling parameter with $0 \leq \lambda \leq 1$. At $\lambda = 0$, the solute–solvent interaction is absent and $\phi_0 = 0$. ϕ_1 is the solute–solvent interaction of interest, and ϕ_λ in $0 < \lambda < 1$ describes the alchemical intermediate states connecting the endpoint states of $\lambda = 0$ and 1. $\rho_\lambda^{(1)}(\mathbf{x})$ further denotes the one-body distribution function of the solvent around the solute. It is the density of the solvent at configuration $\mathbf{x}$ and is the ensemble average with solute–solvent interaction potential ϕ_λ of the instantaneous distribution $\hat{\rho}(\mathbf{x})$ defined as

$$\hat{\rho}(\mathbf{x}) = \sum_i \delta(\mathbf{x} - \mathbf{x}_i) \quad (1)$$

where $\mathbf{x}_i$ is the (snapshot) configuration of the i th solvent molecule and the sum is taken over all the solvent molecules. It should be noted that the averaging is done for $\rho_\lambda^{(1)}$ in the presence of the solute–solvent interaction ϕ_λ ; this is the reason why $\rho_\lambda^{(1)}$ depends on λ , while $\hat{\rho}$ refers only to a snapshot configuration and is a quantity “before” the averaging procedure. $\rho_0^{(1)}$ is then the density in the absence of the solute and is equal to the bulk density of the solvent when the system is homogeneous. Note that our developments are valid also when the system is inhomogeneous.

The one-body direct correlation function $c_\lambda^{(1)}(\mathbf{x})$ is introduced as^{14,15,39}

$$\rho_\lambda^{(1)}(\mathbf{x}) = z \exp(-\beta \phi_\lambda(\mathbf{x}) + c_\lambda^{(1)}(\mathbf{x})) \quad (2)$$

where $\beta = 1/k_B T$ with the Boltzmann constant k_B and the temperature T , and z is given by

$$z = \frac{\exp(\beta \mu_v)}{\Lambda^3} \quad (3)$$

with the chemical potential μ_v and the thermal de Broglie length Λ of the solvent. μ_v does not vary with the coupling parameter λ in the grand canonical ensemble, and correspondingly, z does not depend on λ . $c_\lambda^{(1)}(\mathbf{x})$ vanishes in the ideal gas without solvent–solvent interactions. It thus describes the effects of solvent–solvent correlations at the level of one-body distribution of solvent. The indirect part of the solute–solvent potential of mean force $\omega_\lambda(\mathbf{x})$ is defined, on the other hand, as^{15,38}

$$\rho_\lambda^{(1)}(\mathbf{x}) = \rho_0^{(1)}(\mathbf{x}) \exp(-\beta \phi_\lambda(\mathbf{x}) - \beta \omega_\lambda(\mathbf{x})) \quad (4)$$

This definition uses the pure liquid distribution function ($\lambda = 0$) as the reference. Because $\phi_0(\mathbf{x}) = 0$ and z is constant against the variation of λ , eq 2 can be rewritten through $\rho_0^{(1)}(\mathbf{x}) = z \exp(c_0^{(1)}(\mathbf{x}))$ as

$$\rho_\lambda^{(1)}(\mathbf{x}) = \rho_0^{(1)}(\mathbf{x}) \exp(-\beta \phi_\lambda(\mathbf{x}) + c_\lambda^{(1)}(\mathbf{x}) - c_0^{(1)}(\mathbf{x})) \quad (5)$$

so that the definition also refers to the distribution function at $\lambda = 0$. By comparing eqs 4 and 5, we come to the conclusion that $c_\lambda^{(1)}$ and ω_λ are related to each other through

$$-\beta \omega_\lambda(\mathbf{x}) = c_\lambda^{(1)}(\mathbf{x}) - c_0^{(1)}(\mathbf{x}) \quad (6)$$

$\exp(-\beta \omega_\lambda(\mathbf{x}))$ is called the cavity distribution function. It describes the probability of favorable and unfavorable configurations of a “cavity” particle, which interacts with the other solvent molecules but not the solute. The molecular structure of the cavity particle is identical to that of the solvent, and the only difference is that the interaction with the solute is turned off. $c_\lambda^{(1)}(\mathbf{x})$ is connected through the potential distribution theorem to the excess chemical potential or the free energy of transferring a cavity particle from vacuum to configuration $\mathbf{x}$.^{15,39} The ideal-gas state is thus the reference for $c_\lambda^{(1)}$, and the effects of solvent–solvent correlations on the one-particle distribution function $\rho_\lambda^{(1)}$ are quantified there. ω_λ reflects the changes in solvent–solvent correlations due to the introduction of the solute–solvent interaction ϕ_λ , on the other hand, and $\omega_0 = 0$, in particular. ω_λ thus focuses more on the solute–solvent effects than $c_\lambda^{(1)}$.

When the ensemble is grand canonical, the solvation free energy, $\Delta\mu_u$, of the solute is equal to the difference of the grand potential at $\lambda = 1$ from that at $\lambda = 0$. In the density functional formalism, the grand potential Ω_λ is expressed as^{14,15}

$$\Omega_\lambda = F_\lambda^{\text{id}} + F_\lambda^{\text{ex}} + \int d\mathbf{x} \phi_\lambda(\mathbf{x}) \rho_\lambda^{(1)}(\mathbf{x}) - \mu_v \int d\mathbf{x} \rho_\lambda^{(1)}(\mathbf{x}) \quad (7)$$

where $(F_\lambda^{\text{id}} + F_\lambda^{\text{ex}})$ is called intrinsic free energy and F_λ^{id} and F_λ^{ex} are the ideal and excess parts, respectively. $\rho_\lambda^{(1)}(\mathbf{x})$ is provided by eq 2. F_λ^{id} is written explicitly with $\rho_\lambda^{(1)}$ as

$$F_\lambda^{\text{id}} = k_B T \int d\mathbf{x} \rho_\lambda^{(1)}(\mathbf{x}) (\log(\Lambda^3 \rho_\lambda^{(1)}(\mathbf{x})) - 1) \quad (8)$$

and F_λ^{ex} is connected to $c_\lambda^{(1)}(\mathbf{x})$ through

$$-k_B T c_\lambda^{(1)}(\mathbf{x}) = \frac{\delta F_\lambda^{\text{ex}}}{\delta \rho_\lambda^{(1)}(\mathbf{x})} \quad (9)$$

F_λ^{ex} vanishes for the ideal gas, and it is a notable feature of the DFT formalism that the ideal part is provided in closed form in terms of the one-body distribution function $\rho_\lambda^{(1)}(\mathbf{x})$. According to eq 7, $\Delta\mu_u$ is given by

$$\begin{aligned} \Delta\mu_u = & (F_1^{\text{id}} - F_0^{\text{id}}) - \mu_v \int d\mathbf{x} (\rho_1^{(1)}(\mathbf{x}) - \rho_0^{(1)}(\mathbf{x})) \\ & + (F_1^{\text{ex}} - F_0^{\text{ex}}) + \int d\mathbf{x} \phi_1(\mathbf{x}) \rho_1^{(1)}(\mathbf{x}) \end{aligned} \quad (10)$$

The first term reflects the change in the one-body distribution of the solvent due to the introduction of the solute–solvent interaction $\phi_1(\mathbf{x})$, and the second term is present because the number of solvent molecules in a grand canonical system can change with the interaction with the solute. The third term corresponds to the solute-induced effects on solvent–solvent correlations, and the last term is simply the sum of the solute–solvent interaction energy at $\lambda = 1$. The first two terms (those involving the ideal part and μ_v) on the right-hand side of eq 10 can be written by virtue of eqs 2, 3, and 8 as

$$-k_{\text{B}}T \int d\mathbf{x} (\rho_1^{(1)}(\mathbf{x}) - \rho_0^{(1)}(\mathbf{x})) + k_{\text{B}}T \int d\mathbf{x} \rho_1^{(1)}(\mathbf{x}) \log \left(\frac{\rho_1^{(1)}(\mathbf{x})}{\rho_0^{(1)}(\mathbf{x})} \right) + k_{\text{B}}T \int d\mathbf{x} c_0^{(1)}(\mathbf{x}) (\rho_1^{(1)}(\mathbf{x}) - \rho_0^{(1)}(\mathbf{x})) \quad (11)$$

because

$$k_{\text{B}}T c_0^{(1)}(\mathbf{x}) = k_{\text{B}}T \log(\Lambda^3 \rho_0^{(1)}(\mathbf{x})) - \mu_{\text{v}} \quad (12)$$

Of course $c_0^{(1)}(\mathbf{x})$ and $\rho_0^{(1)}(\mathbf{x})$ are independent of $\mathbf{x}$ for a homogeneous fluid as they correspond to the pure liquid state $\lambda = 0$. Equation 9 leads further to

$$F_1^{\text{ex}} - F_0^{\text{ex}} = -k_{\text{B}}T \int_0^1 d\lambda \int d\mathbf{x} c_{\lambda}^{(1)}(\mathbf{x}) \frac{\partial \rho_{\lambda}^{(1)}(\mathbf{x})}{\partial \lambda} \quad (13)$$

and by noting

$$\rho_1^{(1)}(\mathbf{x}) - \rho_0^{(1)}(\mathbf{x}) = \int_0^1 d\lambda \frac{\partial \rho_{\lambda}^{(1)}(\mathbf{x})}{\partial \lambda} \quad (14)$$

we obtain

$$F_1^{\text{ex}} - F_0^{\text{ex}} + k_{\text{B}}T \int d\mathbf{x} c_0^{(1)}(\mathbf{x}) (\rho_1^{(1)}(\mathbf{x}) - \rho_0^{(1)}(\mathbf{x})) = -k_{\text{B}}T \int_0^1 d\lambda \int d\mathbf{x} (c_{\lambda}^{(1)}(\mathbf{x}) - c_0^{(1)}(\mathbf{x})) \frac{\partial \rho_{\lambda}^{(1)}(\mathbf{x})}{\partial \lambda} \quad (15)$$

and

$$\Delta\mu_{\text{u}} = \int d\mathbf{x} \phi_1(\mathbf{x}) \rho_1^{(1)}(\mathbf{x}) - k_{\text{B}}T \int d\mathbf{x} \left[\rho_1^{(1)}(\mathbf{x}) - \rho_0^{(1)}(\mathbf{x}) \right] - \rho_1^{(1)}(\mathbf{x}) \log \left(\frac{\rho_1^{(1)}(\mathbf{x})}{\rho_0^{(1)}(\mathbf{x})} \right) - k_{\text{B}}T \int_0^1 d\lambda \int d\mathbf{x} (c_{\lambda}^{(1)}(\mathbf{x}) - c_0^{(1)}(\mathbf{x})) \frac{\partial \rho_{\lambda}^{(1)}(\mathbf{x})}{\partial \lambda} \quad (16)$$

Therefore, eq 10 can be rewritten by using eq 6 as

$$\Delta\mu_{\text{u}} = \int d\mathbf{x} \phi_1(\mathbf{x}) \rho_1^{(1)}(\mathbf{x}) - k_{\text{B}}T \int d\mathbf{x} \left[\rho_1^{(1)}(\mathbf{x}) - \rho_0^{(1)}(\mathbf{x}) - \rho_1^{(1)}(\mathbf{x}) \log \left(\frac{\rho_1^{(1)}(\mathbf{x})}{\rho_0^{(1)}(\mathbf{x})} \right) \right] + \int_0^1 d\lambda \int d\mathbf{x} \omega_{\lambda}(\mathbf{x}) \frac{\partial \rho_{\lambda}^{(1)}(\mathbf{x})}{\partial \lambda} \quad (17)$$

and this equation is indeed the expression derived in refs 2, 13, 37, and 38. The derivation started from the Kirkwood charging formula there and is valid in any statistical ensemble. Equation 17 is the basic equation of endpoints DFT. A particularly important feature of eq 17 is that different representations can be used to evaluate this expression using simulation data at the endpoints (pure liquid and solution). In the energy representation, $\mathbf{x}$ is reduced to a one-dimensional variable corresponding to the solute–solvent interaction energy (see section 5.3).

The preceding paragraph showed the equivalence between the density-functional formalism based on the direct correlation function and that on the indirect part of the potential of mean force. The key is provided by the correspondence expressed as eq 6. $\omega_{\lambda}(\mathbf{x})$ is actually a free-energy difference quantifying the solvent-mediated effect on

the solute–solvent correlations, and a wealth of numerical methods may be employed to compute it.^{16,17} The focus of the present work is the indirect part of the solute–solvent potential of mean force, and we will examine aqueous solutions of hydrophobic solutes with variable sizes.

The parallelism is also evident between the approximations for $c_{\lambda}^{(1)}$ and ω_{λ} . The hypernetted-chain (HNC) approximation for $c_{\lambda}^{(1)}$ is the first-order expansion in terms of $(\rho_{\lambda}^{(1)} - \rho_0^{(1)})$ and is expressed as^{14,15}

$$c_{\lambda}^{(1)}(\mathbf{x}) - c_0^{(1)}(\mathbf{x}) = \int d\mathbf{y} \frac{\delta c_{\lambda}^{(1)}(\mathbf{x})}{\delta \rho_{\lambda}^{(1)}(\mathbf{y})} \bigg|_{\lambda=0} (\rho_{\lambda}^{(1)}(\mathbf{y}) - \rho_0^{(1)}(\mathbf{y})) \quad (18)$$

In this equation, the integrand involves the two-body direct correlation function $c_0^{(2)}(\mathbf{x}, \mathbf{y})$. The two-body function is defined at $0 \leq \lambda \leq 1$ as

$$c_{\lambda}^{(2)}(\mathbf{x}, \mathbf{y}) = \frac{\delta c_{\lambda}^{(1)}(\mathbf{x})}{\delta \rho_{\lambda}^{(1)}(\mathbf{y})} \quad (19)$$

and is related to the (two-body) susceptibility $\chi_{\lambda}(\mathbf{x}, \mathbf{y})$ through

$$\chi_{\lambda}(\mathbf{x}, \mathbf{y}) = \langle \hat{\rho}(\mathbf{x}) \hat{\rho}(\mathbf{y}) \rangle_{\lambda} - \langle \hat{\rho}(\mathbf{x}) \rangle_{\lambda} \langle \hat{\rho}(\mathbf{y}) \rangle_{\lambda} = -k_{\text{B}}T \frac{\delta \rho_{\lambda}^{(1)}(\mathbf{x})}{\delta \phi_{\lambda}(\mathbf{y})} \quad (20)$$

$$\int d\mathbf{z} \left(\frac{\delta(\mathbf{x} - \mathbf{z})}{\rho_{\lambda}^{(1)}(\mathbf{z})} - c_{\lambda}^{(2)}(\mathbf{x}, \mathbf{z}) \right) \chi_{\lambda}(\mathbf{z}, \mathbf{y}) = \delta(\mathbf{x} - \mathbf{y}) \quad (21)$$

where the instantaneous distribution $\hat{\rho}(\mathbf{x})$ was introduced by eq 1; $\phi_{\lambda}(\mathbf{x})$ is the solute–solvent interaction potential, and $\langle \dots \rangle_{\lambda}$ denotes the ensemble average at λ . Equation 21 is equivalent to the Ornstein–Zernike equation, and with eqs 19 and 21, eq 18 is rewritten as

$$c_{\lambda}^{(1)}(\mathbf{x}) - c_0^{(1)}(\mathbf{x}) = \frac{\rho_{\lambda}^{(1)}(\mathbf{x}) - \rho_0^{(1)}(\mathbf{x})}{\rho_0^{(1)}(\mathbf{x})} - \int d\mathbf{y} (\chi_0^{-1}(\mathbf{x}, \mathbf{y}) (\rho_{\lambda}^{(1)}(\mathbf{y}) - \rho_0^{(1)}(\mathbf{y}))) \quad (22)$$

Using eqs 4, 6, and 20, eq 22 can be written in the following form:

$$\omega_{\lambda}(\mathbf{x}) = -k_{\text{B}}T \frac{\rho_{\lambda}^{(1)}(\mathbf{x}) - \rho_0^{(1)}(\mathbf{x})}{\rho_0^{(1)}(\mathbf{x})} - \int d\mathbf{y} \frac{\delta \phi_{\lambda}(\mathbf{x})}{\delta \rho_{\lambda}^{(1)}(\mathbf{y})} \bigg|_{\lambda=0} (\rho_{\lambda}^{(1)}(\mathbf{y}) - \rho_0^{(1)}(\mathbf{y})) = \int d\mathbf{y} \frac{\delta \omega_{\lambda}(\mathbf{x})}{\delta \rho_{\lambda}^{(1)}(\mathbf{y})} \bigg|_{\lambda=0} (\rho_{\lambda}^{(1)}(\mathbf{y}) - \rho_0^{(1)}(\mathbf{y})) \quad (23)$$

and the equivalence of eqs 18 and 23 is proved. The HNC approximation is the first-order expansion with $(\rho_{\lambda}^{(1)} - \rho_0^{(1)})$ for $c_{\lambda}^{(1)}$, and it is so for ω_{λ} .

A similar argument is possible for the Percus–Yevick (PY) approximation. It is the first-order expansion of $\exp(c_{\lambda}^{(1)})$ and is written as

$$\begin{aligned}\exp(c_\lambda^{(1)}(\mathbf{x})) &= \exp(c_0^{(1)}(\mathbf{x})) + \int d\mathbf{y} \frac{\delta \exp(c_\lambda^{(1)}(\mathbf{x}))}{\delta \rho_\lambda^{(1)}(\mathbf{y})} \bigg|_{\lambda=0} \\ &\quad (\rho_\lambda^{(1)}(\mathbf{y}) - \rho_0^{(1)}(\mathbf{y})) \\ &= \exp(c_0^{(1)}(\mathbf{x})) \left[1 + \int d\mathbf{y} \frac{\delta c_\lambda^{(1)}(\mathbf{x})}{\delta \rho_\lambda^{(1)}(\mathbf{y})} \bigg|_{\lambda=0} \right. \\ &\quad \left. (\rho_\lambda^{(1)}(\mathbf{y}) - \rho_0^{(1)}(\mathbf{y})) \right]\end{aligned}\quad (24)$$

By noting the modification of the right-hand side of eq 18 to those of eqs 22 and 23, eq 24 reduces by virtue of eq 6 to

$$\begin{aligned}\exp(-\beta\omega_\lambda(\mathbf{x})) - 1 &= -\beta \int d\mathbf{y} \frac{\delta \omega_\lambda(\mathbf{x})}{\delta \rho_\lambda^{(1)}(\mathbf{y})} \bigg|_{\lambda=0} (\rho_\lambda^{(1)}(\mathbf{y}) - \rho_0^{(1)}(\mathbf{y})) \\ &= \int d\mathbf{y} \frac{\delta \exp(-\beta\omega_\lambda(\mathbf{x}))}{\delta \rho_\lambda^{(1)}(\mathbf{y})} \bigg|_{\lambda=0} (\rho_\lambda^{(1)}(\mathbf{y}) - \rho_0^{(1)}(\mathbf{y}))\end{aligned}\quad (25)$$

This equation shows that the PY approximation is also the first-order expansion of $\exp(-\beta\omega_\lambda)$ with respect to $(\rho_\lambda^{(1)} - \rho_0^{(1)})$.

Equations 16 and 17 are simplified when the solute–solvent interaction potential ϕ_λ is taken so that the one-body distribution function $\rho_\lambda^{(1)}$ varies linearly with λ through^{2,11,13,15,37,38}

$$\rho_\lambda^{(1)}(\mathbf{x}) = \lambda \rho_1^{(1)}(\mathbf{x}) + (1 - \lambda) \rho_0^{(1)}(\mathbf{x}) \quad (26)$$

With this linear form for $\rho_\lambda^{(1)}(\mathbf{x})$, eq 17 reduces to

$$\begin{aligned}\Delta\mu_u &= \int d\mathbf{x} \phi_1(\mathbf{x}) \rho_1^{(1)}(\mathbf{x}) \\ &\quad - k_B T \int d\mathbf{x} \left[\rho_1^{(1)}(\mathbf{x}) - \rho_0^{(1)}(\mathbf{x}) - \rho_1^{(1)}(\mathbf{x}) \right. \\ &\quad \left. \log \left(\frac{\rho_1^{(1)}(\mathbf{x})}{\rho_0^{(1)}(\mathbf{x})} \right) \right] \\ &\quad + \int d\mathbf{x} (\rho_1^{(1)}(\mathbf{x}) - \rho_0^{(1)}(\mathbf{x})) \int_0^1 d\lambda \omega_\lambda(\mathbf{x}) \\ &= -k_B T \int d\mathbf{x} \left[(\rho_1^{(1)}(\mathbf{x}) - \rho_0^{(1)}(\mathbf{x})) + \beta \omega_1(\mathbf{x}) \rho_1^{(1)}(\mathbf{x}) \right. \\ &\quad \left. - \left(\beta \int_0^1 d\lambda \omega_\lambda(\mathbf{x}) \right) \times (\rho_1^{(1)}(\mathbf{x}) - \rho_0^{(1)}(\mathbf{x})) \right]\end{aligned}\quad (27)$$

Equation 27 is exact with the particular choice of eq 26, and $\Delta\mu_u$ is obtained when the integration over λ is conducted for ω_λ . When eq 26 is adopted, eqs 18, 23, 24, and 25 are further expressed as

$$c_\lambda^{(1)}(\mathbf{x}) = c_0^{(1)}(\mathbf{x}) + \lambda \frac{\partial c_\lambda^{(1)}(\mathbf{x})}{\partial \lambda} \bigg|_{\lambda=0} \quad (28)$$

$$\omega_\lambda(\mathbf{x}) = \lambda \frac{\partial \omega_\lambda(\mathbf{x})}{\partial \lambda} \bigg|_{\lambda=0} \quad (29)$$

$$\exp(c_\lambda^{(1)}(\mathbf{x})) = \exp(c_0^{(1)}(\mathbf{x})) + \lambda \frac{\partial \exp(c_\lambda^{(1)}(\mathbf{x}))}{\partial \lambda} \bigg|_{\lambda=0} \quad (30)$$

$$\begin{aligned}\exp(-\beta\omega_\lambda(\mathbf{x})) &= 1 + \lambda \frac{\partial \exp(-\beta\omega_\lambda(\mathbf{x}))}{\partial \lambda} \bigg|_{\lambda=0} \\ &= 1 - \lambda \beta \frac{\partial \omega_\lambda(\mathbf{x})}{\partial \lambda} \bigg|_{\lambda=0}\end{aligned}\quad (31)$$

respectively. These equations show that the HNC and PY approximations are expressed as the linear dependencies on λ : $c_\lambda^{(1)}$ and ω_λ for HNC or $\exp(c_\lambda^{(1)})$ and $\exp(-\beta\omega_\lambda)$ for PY. When ω_λ is provided explicitly as a function of λ in terms of its values (or derivatives) at $\lambda = 0$ and 1 (the endpoint states of solute insertion), the solvation free energy $\Delta\mu_u$ is expressed in terms only of the variables at the endpoints.^{2,13,37,40} The $\Delta\mu_u$ computation is called endpoints DFT when an approximate functional for $\Delta\mu_u$ is constructed from distribution functions obtained at molecular simulations at $\lambda = 0$ and 1. The endpoints computation is described in section 5.3 in more detail.

With the definition of the two-body direct correlation function $c_\lambda^{(2)}(\mathbf{x}, \mathbf{y})$ by eq 19 and its connection to the susceptibility through eq 21, the derivatives of $c_\lambda^{(1)}$ and ω_λ are determined from

$$\begin{aligned}-k_B T \frac{\partial c_\lambda^{(1)}(\mathbf{x})}{\partial \lambda} &= \frac{\partial \omega_\lambda(\mathbf{x})}{\partial \lambda} = -k_B T \int d\mathbf{y} c_\lambda^{(2)}(\mathbf{x}, \mathbf{y}) \frac{\partial \rho_\lambda^{(1)}(\mathbf{y})}{\partial \lambda} \\ &= -k_B T \frac{1}{\rho_\lambda^{(1)}(\mathbf{x})} \frac{\partial \rho_\lambda^{(1)}(\mathbf{x})}{\partial \lambda} \\ &\quad + k_B T \int d\mathbf{y} (\chi_\lambda)^{-1}(\mathbf{x}, \mathbf{y}) \frac{\partial \rho_\lambda^{(1)}(\mathbf{y})}{\partial \lambda}\end{aligned}\quad (32)$$

When the one-body distribution function $\rho_\lambda^{(1)}$ is taken so that it varies linearly with λ , the derivative of ω_λ at the pure solvent state is

$$\begin{aligned}\sigma_0(\mathbf{x}) &= \frac{\partial \omega_\lambda(\mathbf{x})}{\partial \lambda} \bigg|_{\lambda=0} \\ &= -k_B T \frac{\rho_1^{(1)}(\mathbf{x}) - \rho_0^{(1)}(\mathbf{x})}{\rho_0^{(1)}(\mathbf{x})} \\ &\quad + k_B T \int d\mathbf{y} (\chi_0)^{-1}(\mathbf{x}, \mathbf{y}) (\rho_1^{(1)}(\mathbf{y}) - \rho_0^{(1)}(\mathbf{y})) \\ &= -k_B T \int d\mathbf{y} \left(\frac{\delta(\mathbf{x} - \mathbf{y})}{\rho_0^{(1)}(\mathbf{x})} - (\chi_0)^{-1}(\mathbf{x}, \mathbf{y}) \right) (\rho_1^{(1)}(\mathbf{y}) - \rho_0^{(1)}(\mathbf{y}))\end{aligned}\quad (33)$$

The susceptibility χ_λ describes the two-point correlation between a pair of solvent molecules and needs to be inverted in eqs 22 and 32. It should be noted, in fact, that χ_λ is not invertible when the ensemble is canonical or isothermal–isobaric and the number of solvent molecules is constant. This reflects the fact that a shift in the solute–solvent potential by an additive constant does not affect the distribution function; a single distribution function corresponds to a number of potential functions that are different among one another by additive constants independent of the solute–solvent configurations. A valid version of the Ornstein–Zernike equation in the canonical ensemble has been sought in connection with the Kirkwood–Buff integral,^{41–46} and the one-to-one correspondence between the potential and distribution can be restored by

fixing the additive constant so that the potential vanishes at large distances between the solute and solvent.^{3,13,37} With the solute–solvent potential $\phi_\lambda(\mathbf{x})$, the last term of eq 32 is rewritten as

$$\frac{\partial \rho_\lambda^{(1)}(\mathbf{x})}{\partial \lambda} = -\beta \int d\mathbf{y} \chi_\lambda(\mathbf{x}, \mathbf{y}) \frac{\partial \phi_\lambda(\mathbf{y})}{\partial \lambda} \quad (34)$$

$(\partial \phi_\lambda(\mathbf{x})/\partial \lambda)$ is the solution to eq 34, and our convention is to adopt the one which vanishes at far separation of solute and solvent. The expression of $\phi_\lambda(\mathbf{x})$ with $(\partial \phi_\lambda/\partial \lambda)$ satisfying eq 34 is then unique, and the $(\chi_\lambda)^{-1}$ procedure in eqs 32 and 33 should be understood with the above convention. Actually, when $\phi_\lambda(\mathbf{x})$ is not zero at far separation, the solvation free energy for the solute with that potential is not an intensive quantity and involves a term that grows with the system size. Our scheme can thus be rephrased as a requirement that the solvation free energy is an intensive quantity over the whole $0 \leq \lambda \leq 1$.

We comment here on the use of endpoints DFT in the canonical versus grand canonical ensembles. In the DFT method described above, the difference among the statistical ensembles lies in the treatment of the susceptibility χ_λ in eq 32. In the grand canonical ensemble, χ_λ is positive definite and can be straightforwardly inverted. In the canonical or isothermal–isobaric ensemble, the pseudoinversion procedure in the preceding paragraph is employed. The solute–solvent interaction potential ϕ_λ that is the solution to eq 34 is independent of the ensemble, though. This is because ϕ_λ (is enforced to) vanish at far separation in all of the ensembles, and as a result, the solvation free energy does not depend on the ensemble. In fact, molecular simulations are to be conducted at $\lambda = 0$ and 1 when the endpoints DFT scheme is employed. The solvation free energy computed with the pseudoinversion in the canonical or isothermal–isobaric ensemble is still the same as the value that would be obtained in a grand-canonical simulation. This is a notable feature of the endpoints DFT scheme because molecular simulations are commonly performed in the canonical or isothermal–isobaric ensemble.

In this section, we described the endpoints DFT formalism. As seen with eqs 17 and 27, the indirect PMF $\omega(\mathbf{x})$ plays the key role in determining the solvation free energy of the solute $\Delta\mu_u$. $\omega(\mathbf{x})$ quantifies the collective effects of solvent on the solute–solvent distribution, and our formalism provides direct connection between the solvent effects at the molecular scale and the solvation energetics at the thermodynamic scale. The third terms of eqs 17 and 27 can be approximated through the HNC and PY forms to the endpoints expression as further noted in section 5.3. The endpoints method utilizes MD outputs at $\lambda = 0$ and 1, which is in harmony with the present status of computational chemistry and biophysics that a nanoscale MD is readily possible. The performance of the method is governed by the approximation adopted for ω_λ , and in the following sections, we will examine the indirect PMF for hydrophobic solutes in water.

3. SIMULATION METHOD

3.1. Cavity Particle. Consider an MD simulation of a solution with a fixed solute and one tagged solvent molecule performed in the canonical ensemble. The total potential energy of the system is

$$U(\mathbf{x}, \{\mathbf{x}\}) = \phi_1(\mathbf{x}) + U_{uv}(\{\mathbf{x}\}) + U_{vv}(\mathbf{x}, \{\mathbf{x}\}) + U_{vv}(\{\mathbf{x}\}) \quad (35)$$

where ϕ_1 (ϕ_λ at the $\lambda = 1$ state) is the interaction energy between the tagged solvent molecule and the solute; U_{vv} is the interaction energy between the tagged solvent molecule and the other solvent molecules; U_{uv} is the interaction energy between the solute and the other solvent molecules, and U_{vv} is the total interaction energy of all the other solvent molecules. $\mathbf{x}$ and $\{\mathbf{x}\}$ represent the coordinates of the tagged solvent molecule and the other solvent molecules, respectively. The density of one tagged water molecule is normalized to unity instead of the total number of water molecules, N . Namely, $N\rho_1^{(w)}(\mathbf{x}) = \rho_1^{(1)}(\mathbf{x})$, where $\rho_1^{(w)}(\mathbf{x})$ is the density of the tagged water molecule in the solution, and $N\rho_1^{(w)}(\infty) = \rho_1^{(1)}(\infty) = \rho_0^{(1)}(\mathbf{x}) + O(1/N)$ in the canonical ensemble, where $\mathbf{x} = \infty$ denotes that the water molecule is in the bulk.

Consider a different simulation state at which the interaction between the tagged solvent molecule and the solute is turned off. Such tagged solvent molecule will be termed a cavity particle in this study. The total potential energy of this system is

$$U_{\text{cav}}(\mathbf{x}, \{\mathbf{x}\}) = U_{uv}(\{\mathbf{x}\}) + U_{vv}(\mathbf{x}, \{\mathbf{x}\}) + U_{vv}(\{\mathbf{x}\}) \quad (36)$$

The density of the cavity particle in the solution, $\rho_1^{(c)}(\mathbf{x})$, and the density of a tagged water molecule, $\rho_1^{(w)}(\mathbf{x})$, described above satisfy a simple relationship:

$$\rho_1^{(c)}(\mathbf{x}) = \exp[\beta\phi_1(\mathbf{x})]\rho_1^{(w)}(\mathbf{x}) \quad (37)$$

Note that eq 37 actually holds with a multiplicative factor that is independent of $\mathbf{x}$ and reduces to unity in the limit of large system size (see the Supporting Information).³⁶ Combining eqs 4 and 37 yields

$$\omega(\mathbf{x}) = -k_B T \log \left(\frac{\rho_1^{(c)}(\mathbf{x})}{\rho_0^{(w)}(\mathbf{x})} \right) = -k_B T \log \left(\frac{\rho_1^{(c)}(\mathbf{x})}{\rho_0^{(c)}(\mathbf{x})} \right) \quad (38)$$

which shows that the indirect part of the solute–solvent potential of mean force, $\omega(\mathbf{x})$, can be estimated by the cavity particle density.

3.2. Simulation Details. We ran simulations for a solution that contains a methane-like hydrophobic solute represented by a fixed carbon atom, 699 TIP3P water molecules,⁴⁷ and a water cavity particle by using the GROMACS version 2016/3.⁴⁸ The simulation box is cubic, and the length of each side is 2.78 nm. The system was modeled by the OPLS-AA force field.⁴⁹ Before the NVT production simulations, the system was equilibrated by energy minimization and a 1 ns long NPT simulation. For the NPT simulations during the equilibration period, the pressure was kept at 1 bar by the Parrinello–Rahman pressure barostat with a pressure relaxation time of 2.0 ps;⁵⁰ the temperature was kept constant at 300 K by the velocity-rescaling with a time constant $\tau_t = 0.1$ ps. For the NVT simulations during the production period, the temperature was maintained at 300 K by the leapfrog stochastic dynamics integrator (Langevin dynamics) with a time constant $\tau_t = 2.0$ ps.⁵¹ The time step for the MD simulations was 2 fs; the water molecules were kept rigid with SETTLE.⁵² The long-range electrostatic interaction was treated using the smooth particle-mesh Ewald approach with a real-space cutoff of 1.0 nm and a spline order of 4.⁵³ To create a cavity particle, we chose one

water molecule and added an artificial bond between each atom of this water molecule and the solute by using the type 5 bond function in GROMACS, which creates a bond-like connection between two atoms without creating an interaction. This makes it possible to put the nonbonded interactions between the specified water (the cavity particle) and the solute onto the list of exclusions.⁴⁸ Therefore, there are no interactions between the chosen water molecule, namely, the cavity particle, and the solute in the simulations.

The charges, masses, and parameters of Lennard-Jones interactions of different atoms in this system are listed in Table 1.

Table 1. Force Field Parameters Used in MD Simulations

atom	C	O	H
charge (<i>e</i>)	0.0	−0.834	0.417
mass (u)	12.01100	15.99940	1.00800
ϵ_{ii} (kJ/mol)	0.276144	0.636386	0
σ_{ii} (nm)	0.350000	0.315061	0

1. The standard formula of the LJ potential V_{LJ} between two atoms is

$$V_{LJ}(r_{ij}) = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (39)$$

where $\sigma_{ij} = \sqrt{\sigma_i \sigma_j}$ and $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$. Because a cutoff of van der Waals interactions (1.0 nm) was applied, the actual LJ potential between the solute and a water molecule in our simulations is evaluated by

$$V_{LJ}^{(sw)}(r_{co}) = \begin{cases} 4\epsilon_{co} \left[\left(\frac{\sigma_{co}}{r_{co}} \right)^{12} - \left(\frac{\sigma_{co}}{r_{co}} \right)^6 \right] - V_{\text{shift}} & r_{co} < 1.0 \text{ nm} \\ 0 & r_{co} \geq 1.0 \text{ nm} \end{cases} \quad (40)$$

where r_{co} is the distance between the solute and the oxygen atom of the water molecule and V_{shift} is the van der Waals interaction energy between a carbon atom and an oxygen atom calculated using eq 39 when their distance equals the cutoff (1.0 nm). The LJ potential between the solute and a water molecule is plotted in the Figure S1a. The LJ potential

$V_{LJ}^{(sw)}(r_{co})$ is zero when the distance r_{co} is 0.332 nm and reaches its minimum −0.100 kcal/mol when the distance r_{co} is 0.373 nm. Because the hydrogen site in the water molecule does not carry LJ parameters, in addition, the interaction potential of eq 40 depends only on r_{co} . Because of this property, we can discuss the cavity distribution only over the coordinate of r_{co} .

We doubled and tripled the value of σ of the carbon atom to make larger solutes, which will be termed the C×2 and C×3 solutes in this article, respectively. Note that the radius of solvent exclusion is larger by factors of $\sqrt{2}$ and $\sqrt{3}$ for the C×2 and C×3 solutes, respectively, because of the geometric combination rule adopted. The LJ potential between the C×2 solute and a water molecule is plotted in Figure S1b. The LJ potential $V_{LJ}^{(sw)}(r_{co})$ is zero when the distance r_{co} is 0.470 nm and reaches its minimum, −0.096 kcal/mol, when the distance r_{co} is 0.527 nm. See the Supporting Information for the discussion about the LJ potential for the C×3 solute.

To obtain better sampling in the repulsive region of the solute, we applied an umbrella sampling algorithm, which includes 21 independent simulations (21 umbrella sampling windows). For the *i*th simulation ($0 \leq i \leq 20$), the parabolic restraint potential is

$$U_u(i) = \frac{1}{2} k_u (r_{co} - r_0(i))^2 \quad (41)$$

where $k_u = 1000$ kJ/mol is the force constant and $r_0(i) = 0.05i$ nm is the minimum position of the restraint potential. Each independent simulation lasted 200 ns. Additionally, we ran one simulation of the solution without cavity particles and another simulation of the solution that contains a free cavity particle. Each additional simulation lasted 300 ns. The position of the cavity particle (or the tagged water molecule) was recorded every 1 ps. The raw data were analyzed by UWHAM.⁵⁴

4. RESULTS

4.1. Potential of Mean Force of the Cavity Particle.

4.1.1. Methane-like Solute: C×1. Figure 1a shows the densities of the cavity particle and a tagged water molecule for the methane-like solute. Both densities converge to the density of water in pure solvent, $\rho_0^{(1)}(\mathbf{x})/N$ (or the density of the cavity particle in the bulk, $\rho_1^{(c)}(\infty)$), when the distance between the oxygen atom of the water molecule and the solute, r_{co} , approaches the simulation cutoff distance 1 nm. (Note that we ignore the small factor $O(1/N)$ that distinguishes the bulk

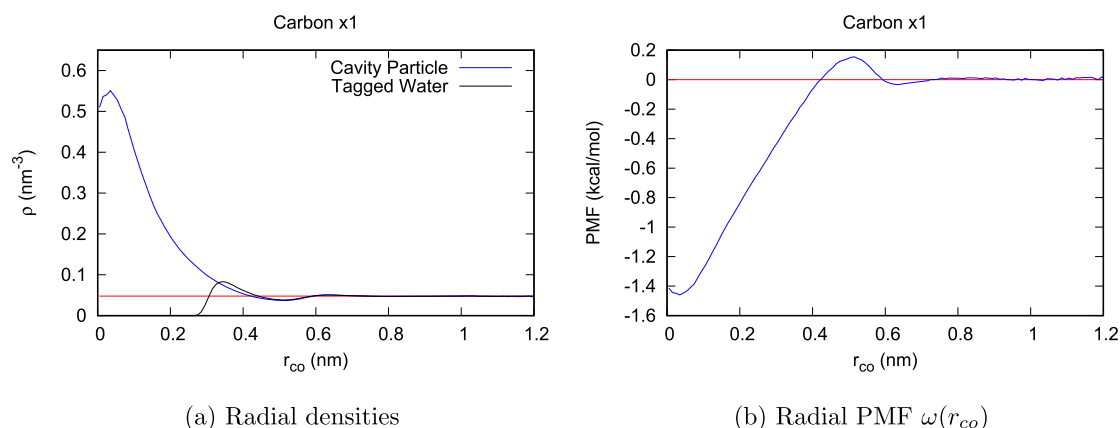


Figure 1. (a) Densities of the cavity particle and a tagged water molecule in a solution that contains a methane-like solute. The red line shows the density of pure solvent. (b) Radial PMF of the cavity particle $\omega(r_{co})$.

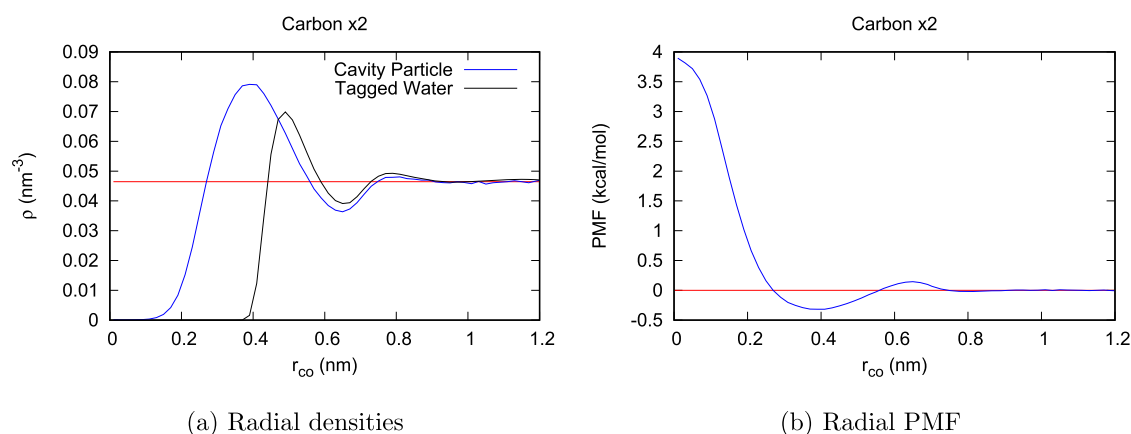


Figure 2. (a) Densities of the cavity particle and a tagged water molecule in a solution that contains a Cx2 solute. The red line is the density of pure solvent. (b) Radial PMF of the cavity particle.

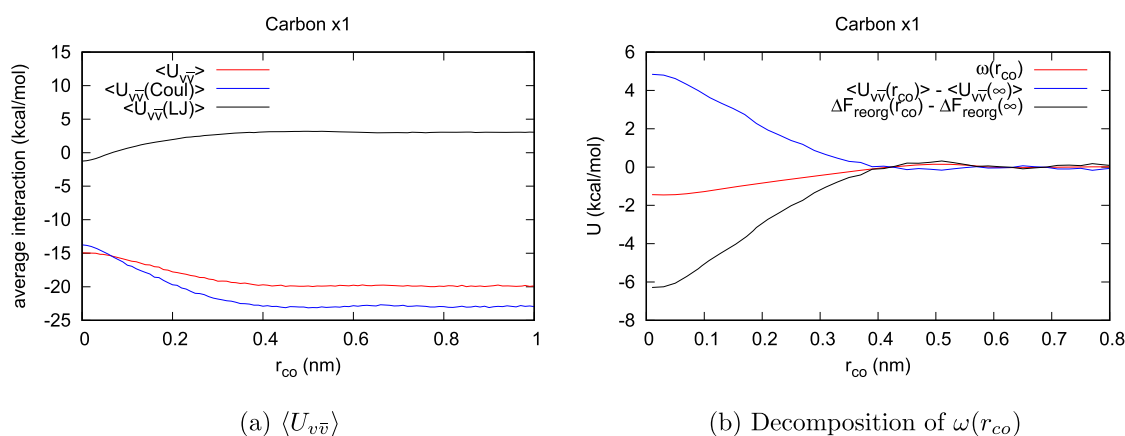


Figure 3. (a) Dependence of the average interaction energy between the cavity particle and the solvent on the distance between the cavity particle and the solute, r_{co} , in a solution that contains a methane-like solute. (b) Decomposition of PMF $\omega(r_{co})$ and the dependence of the components on the distance between the cavity particle and the solute, r_{co} .

density from the pure liquid density in the canonical ensemble.) The density of the tagged water molecule peaks at $r_{co} = 0.345$ nm where the LJ potential between the solute and the water molecule is approximately zero (see Figure S1a). When the distance r_{co} decreases further, the density of the tagged water molecule decreases rapidly approaching zero around $r_{co} = 0.275$ nm, whereas the density of the cavity particle keeps increasing and reaches its maximum when r_{co} equals 0.035 nm. Figure 1b shows the radial potential of mean force (PMF) of the cavity particle calculated by

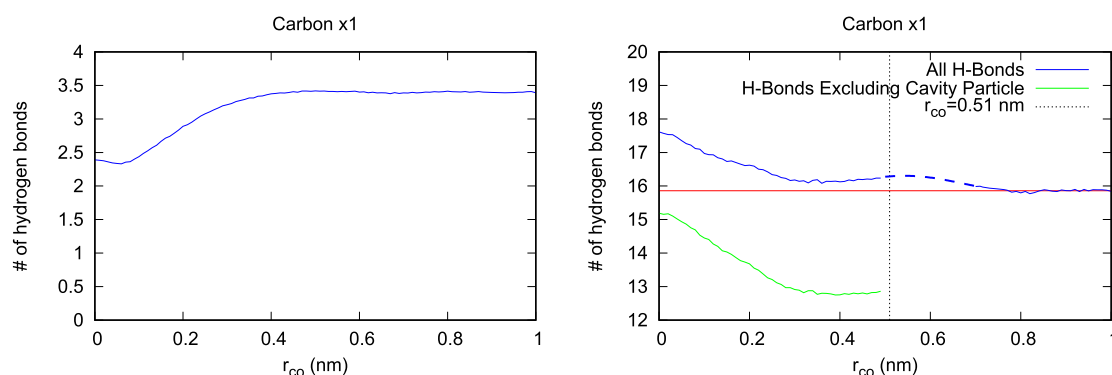
$$\omega(r_{co}) = -k_B T \ln \frac{\rho_1^{(c)}(r_{co})}{\rho_1^{(c)}(\infty)} \quad (42)$$

which represents the free-energy change of moving the cavity particle from the bulk to the position where the distance between the oxygen atom of the cavity particle and the solute is r_{co} . The indirect PMF ω is -1.41 ± 0.02 kcal/mol at the center of the solute ($r_{co} = 0$).

4.1.2. Hydrophobic Solute: Cx2 and Cx3. The densities of the cavity particle and the tagged water molecule for the solution containing the larger hydrophobic solute, Cx2, are plotted in Figure 2a. The density curve of the tagged water molecule maintains a similar shape compared with the case in which the solute is a carbon atom. (Note the difference in the scale of the vertical axes.) The density of the tagged water

molecule peaks at $r_{co} = 0.485$ nm, where the LJ potential between the solute and the water molecule is approximately zero (see Figure S1b). When the tagged water molecule moves closer to the solute, the density decreases rapidly and approaches zero at $r_{co} \approx 0.385$ nm. The density of the cavity particle reaches its maximum inside the repulsive core close to the distance at which the density of the tagged water approaches zero. However, unlike for the methane size solute, although there is no interaction between the cavity particle and the solute, the radial probability density of the cavity particle approaches zero at the center of the solute. In Figure 2b, we plot the radial PMF of the cavity particle. The radial PMF $\omega(r_{co})$ switches from negative to positive when r_{co} is ~ 0.270 nm and continues increasing when the distance r_{co} decreases. The radial PMF $\omega(r_{co}) = +3.89 \pm 0.03$ kcal/mol at the center of the solute ($r_{co} = 0$). We also analyzed the potential of mean force for a cavity particle in a solution with an even larger hydrophobic solute (Cx3). The results are presented in the Supporting Information (see Figure S2). For the case of the Cx3 solute, the maximum PMF ω is $+5.16 \pm 0.02$ kcal/mol at $r_{co} = 0$.

4.2. Interaction between the Cavity Particle and the Solvent. **4.2.1. Methane-like Solute: Cx1.** Figure 3a shows the change of the average interaction energy between the cavity particle and the solvent $\langle U_{vv} \rangle$ when the cavity particle is moved from the bulk to the center of the methane-like solute.



(a) Number of H-Bonds formed with the cavity particle

(b) Number of H-Bonds inside the first shell

Figure 4. (a) Dependence of the number of hydrogen bonds formed with the cavity particle on the distance between the cavity particle and the solute, r_{co} , in a solution that contains a methane-like solute. (b) Dependence of the number of hydrogen bonds inside the first solvation shell on the distance between the cavity particle and the solute, r_{co} . The radius of the sphere that encloses the solute and the first solvation shell is 0.51 nm. The red line shows the average number of hydrogen bonds within the first solvation shell when the cavity particle is in the bulk. The blue curve represents the total number of hydrogen bonds within the first solvation shell including the cavity particle; and the green curve represents the total number of hydrogen bonds formed by water molecules excluding the cavity particle within the first solvation shell.

The average total interaction energy $\langle U_{v\bar{v}} \rangle$ changes from -19.78 to -14.93 kcal/mol when the distance between the cavity particle and the solute r_{co} decreases from 1.00 nm to zero. We separated $\langle U_{v\bar{v}} \rangle$ into the Coulomb part and the LJ part. In the bulk, the Lennard-Jones interaction between the cavity particle and other water molecules is positive, but it switches to negative when the cavity particle moves into the solute. The magnitude of the change of the LJ interaction is 4.30 kcal/mol when r_{co} decreases from 1.00 nm to zero. The Coulomb interaction between the cavity particle and other water molecules is always negative. The Coulomb interaction becomes less attractive by +9.15 kcal/mol (from -22.89 to -13.74 kcal/mol) when r_{co} decreases from 1.00 nm to zero.

The free-energy change to insert a cavity particle in the solution at position r_{co} can be decomposed into two components, namely

$$\Delta F(r_{co}) = \langle U_{v\bar{v}}(r_{co}) \rangle + \Delta F_{\text{reorg}}(r_{co}) \quad (43)$$

where $\langle U_{v\bar{v}} \rangle$ is the average interaction energy between the cavity particle and the solvent and ΔF_{reorg} is the free energy required to reorganize the solvent water molecule around the inserted cavity particle. Similarly, we can decompose the PMF $\omega(r_{co})$ into two components

$$\begin{aligned} \omega(r_{co}) &= \Delta F(r_{co}) - \Delta F(\infty) \\ &= (\langle U_{v\bar{v}}(r_{co}) \rangle - \langle U_{v\bar{v}}(\infty) \rangle) + (\Delta F_{\text{reorg}}(r_{co}) - \Delta F_{\text{reorg}}(\infty)) \end{aligned} \quad (44)$$

where $(\langle U_{v\bar{v}}(r_{co}) \rangle - \langle U_{v\bar{v}}(\infty) \rangle)$ and $(\Delta F_{\text{reorg}}(r_{co}) - \Delta F_{\text{reorg}}(\infty))$ represent the changes of the average interaction energy between the cavity particle and the solvent and the reorganization free energy, respectively, when the cavity particle is moved from the bulk to the position where the distance between the oxygen atom of the cavity particle and the solute is r_{co} . The dependence of these two energy terms on r_{co} is shown in Figure 3b. When the cavity particle is in the bulk, the average interaction energy between the cavity particle and the solvent is negative, and the reorganization free energy is positive. Figure 3b shows that the magnitudes of both terms decrease when the cavity particle is moved from the bulk to the center of the carbon atom. For the methane-like solute, the

change of the average interaction energy $\langle U_{v\bar{v}} \rangle$ is +4.84 kcal/mol when r_{co} is reduced from 1.00 nm to zero, and the change of the reorganization free energy is -6.28 kcal/mol with the same variation of r_{co} .

We examined the change of the number of hydrogen bonds that are formed between the cavity particle and the solvent when the distance between the cavity particle and the solute r_{co} changes. The “hbond” function in the GROMACS software package was applied to analyze hydrogen bonds.⁴⁸ A geometrical criterion is used to determine whether a hydrogen bond exists: the cutoff distance is 0.35 nm, and the cutoff angle is 30°. The result is shown in Figure 4a. The average number of hydrogen bonds between the cavity particle and the solvent is ~ 3.4 when the cavity particle is in the bulk, which is the correct value for the TIP3P water model in water,⁴⁷ and begins to decrease when the cavity particle enters the first hydration shell of the methane-like solute at ~ 0.4 nm. The cavity particle and the solvent form ~ 2.4 hydrogen bonds when the cavity particle is located at the center of the methane-like solute ($r_{co} = 0$). The blue curve in Figure 4b shows how the total number of hydrogen bonds inside the first solvation shell ($r_{\text{first}} = 0.51$ nm) of the solute depends on the distance between the cavity particle and the solute r_{co} . As mentioned in section 4.1.1, Figure 1 shows the density of a tagged water molecule peaks at $r_{co} = 0.345$ nm in the first hydration shell of the methane-like solute. The radius of the first solvation shell is determined by the position of the nearest local minimum following the first maximum in the radial density distribution function. When measuring the total number of hydrogen bonds within the first solvation shell, only those hydrogen bonds whose donor and acceptor are both within the shell are considered. Therefore, the total number of hydrogen bonds inside the first solvation shell includes hydrogen bonds made between untagged pairs of water molecules, both of which must be inside the shell, and hydrogen bonds made between the cavity particle and untagged water molecules, both of which must also be inside the shell. As the cavity particle is moved into the first hydration shell, one water molecule is gradually pushed out of the shell. The total number of hydrogen bonds within the first solvation shell measured by the “hbond” function of GROMACS

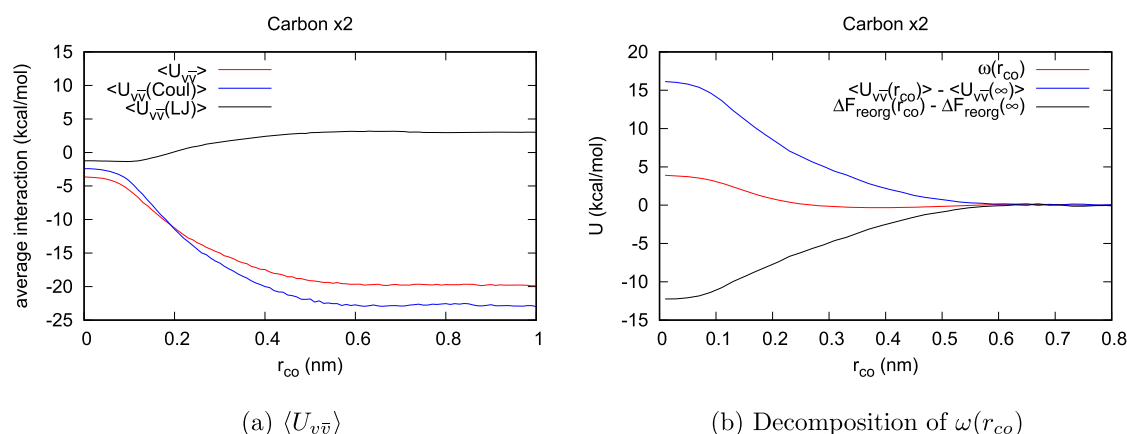


Figure 5. (a) Dependence of the average interaction energy between the cavity particle and the solvent on the distance between the cavity particle and the solute, r_{co} , in a solution containing a CX2 solute. (b) Decomposition of PMF ω and the dependence of the components on the distance between the cavity particle and the solute, r_{co} .

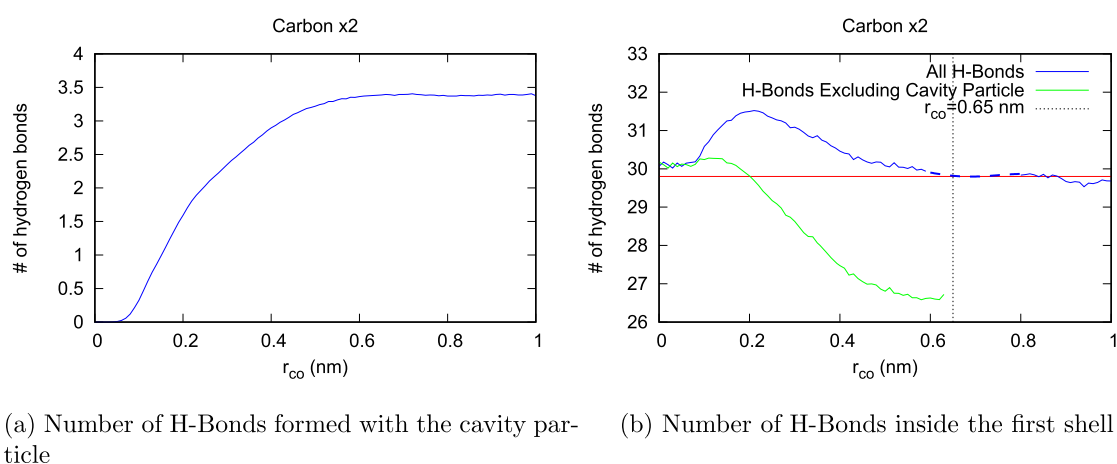


Figure 6. (a) Dependence of the number of hydrogen bonds formed with the cavity particle on the distance between the cavity particle and the solute, r_{co} , in a solution that contains a CX2 solute. (b) Dependence of the number of hydrogen bonds inside the first solvation shell on the distance between the cavity particle and the solute, r_{co} . The radius of the sphere that encloses the solute and the first solvation shell is 0.65 nm. The red line shows the average number of hydrogen bonds within the first solvation shell when the cavity particle is in the bulk. The blue curve represents the total number of hydrogen bonds within the first solvation shell including the cavity particle, and the green curve represents the total number of hydrogen bonds formed by water molecules excluding the cavity particle within the first solvation shell.

exhibits a jump around $r_{co} = 0.51$ nm. For visual clarity, in Figure 4b, we replaced the jump with the values of a polynomial function fitting the data points without the jump, which are shown by the blue dashed curve. As can be seen, once the cavity particle begins to penetrate the solute, the number of hydrogen bonds inside the first solvation shell keeps increasing when the distance between the cavity particle and the solute r_{co} decreases as the cavity particle approaches the center of the solute. The green curve in Figure 4b shows the change of the total number of hydrogen bonds formed only by water molecules (excluding the cavity particle) within the first solvation shell. When the cavity particle is located at the center of the solute ($r_{co} = 0$), the total number of hydrogen bonds formed by the water molecules in the first solvation shell is smaller by about one hydrogen bond than the number when the cavity particle is in the bulk because the (untagged) water molecules in the hydration shell are still slightly perturbed by the cavity particle at the center of this methane-like (CX1) solute.

4.2.2. Hydrophobic Solutes: CX2 and CX3. Figure 5a shows the change of the average interaction energy between

the cavity particle and the solvent $\langle U_{vv} \rangle$ when the cavity particle is moved from the bulk to the center of the CX2 solute. The average interaction energy $\langle U_{vv} \rangle$ changes from -19.78 to -3.64 kcal/mol when the distance between the cavity particle and the solute r_{co} decreases from 1.00 nm to zero. Similarly, the Lennard-Jones interaction between the cavity particle and other water molecules changes from positive to negative when the cavity particle moves into the solute. The magnitude of the change of the LJ interaction is 4.30 kcal/mol, which is approximately the same as the change when the solute is a carbon CX1 atom. However, the magnitude of the change of the Coulomb interaction for the CX2 solute, 20.31 kcal/mol, is much larger compared with the case for the methane-like (CX1) solute. Figure 5b shows the decomposition of the PMF $\omega(r_{co})$. The change of the average interaction energy $\langle U_{vv} \rangle$ is 16.22 kcal/mol when r_{co} is zero, and the change of the reorganization free energy is -12.33 kcal/mol when r_{co} is zero. The ranges of variation are thus larger than those for the CX1 solute.

Figure 6a shows the change of the number of hydrogen bonds that are formed between the cavity particle and the

solvent when the distance between the cavity particle and the solute r_{co} changes. The average number of hydrogen bonds between the cavity particle and the solvent begins to decrease when r_{co} is about 0.5 nm and decreases to zero when r_{co} is about 0.05 nm. The blue curve in Figure 6b shows the dependence of the total number of hydrogen bonds including the cavity particle inside the first shell ($r_{\text{first}} = 0.65$ nm) around the solute on the distance between the cavity particle and the solute r_{co} . Similarly, the total number of hydrogen bonds begins to increase when the cavity particle begins to overlap with the solute where $r_{\text{co}} \approx 0.5$ nm and reaches the maximum number (~ 31.5) when $r_{\text{co}} \approx 0.2$ nm. However, when the cavity particle moves deeper into the solute, the total number of hydrogen bonds within the first shell decreases. The green curve in Figure 6b shows the change of total number of hydrogen bonds formed by water molecules in the first shell. When r_{co} is smaller than 0.1 nm, the number of hydrogen bonds formed by water molecules in the first shell approximately equals the number when the cavity particle is in the bulk.

The same analyses were performed for the C×3 solute. The results, which are similar to those for the C×2 solute, are shown in the Supporting Information (see Figures S3 and S4).

5. DISCUSSION

5.1. Sign of Indirect PMF at $r_{\text{co}} = 0$ and Competition between $\langle U_{\text{vw}} \rangle$ and ΔF_{reorg} . When the solute is a carbon atom, which serves as a proxy for methane in water, the cavity particle has more favorable free energy when it overlaps with the solute (see Figure 1) and the indirect PMF $\omega(r_{\text{co}})$ is always negative inside the repulsive region. However, as the size of the hydrophobic solute increases to C×2, the sign of the indirect PMF $\omega(r_{\text{co}})$ changes and at $r_{\text{co}} = 0$ is positive. The magnitude of ω at $r_{\text{co}} = 0$ increases from $+3.89 \pm 0.03$ to $+5.16 \pm 0.02$ kcal/mol when the solute changes from C×2 to C×3, which agrees with the statement in the Introduction that $\omega(r_{\text{co}} = 0)$ approaches the negative of the solvation free energy of the solvent in the solution if the solute is very large. To validate the values of the indirect PMF $\omega(r_{\text{co}})$ at the center of the solute that we measured for three different solutes using umbrella sampling, we examined the following thermodynamic cycle: the free-energy change of moving a cavity particle from the vacuum to the center of the solute ($r_{\text{co}} = 0$) equals the sum of free-energy changes for moving a cavity particle from vacuum to the bulk and moving a cavity particle from the bulk to the center of the solute. Namely

$$\Delta F_{\text{cav}}(\text{vacuum} \rightarrow r_{\text{co}} = 0) = \Delta F_{\text{cav}}(\text{vacuum} \rightarrow \text{bulk}) + \omega(r_{\text{co}} = 0) \quad (45)$$

The comparisons between $\Delta F_{\text{cav}}(\text{vacuum} \rightarrow r_{\text{co}} = 0)$ and the sum of $\omega(r_{\text{co}} = 0)$ and $\Delta F_{\text{cav}}(\text{vacuum} \rightarrow \text{bulk})$ are shown in Figure S5. As can be seen, they agree with each other very well for all three solutes.

Similar results, the indirect PMF ω at the center of the solute changing from positive to negative when the size of the solute decreases, have been reported in the study of cavity correlation functions for Lennard-Jones fluids with solutes of varying size.⁷ Reference 7 also shows that the indirect PMF ω attains a minimum at the center of the solute for one-component systems. In our study, the methane-like solute is slightly larger than the oxygen atom of a water molecule (see Table 1), and as a result, the indirect PMF ω for the C×1

solute attains a minimum at $r_{\text{co}} = 0.035$ nm instead of the center of the solute.

Furthermore, for Lennard-Jones fluids, because of the absence of strong Coulomb interactions, the indirect PMF $\omega(r_{\text{co}})$ at the center of the solute ($r_{\text{co}} = 0$) increases more slowly than for aqueous solutions as the size of the solute increases. For the C×2 solute, the repulsive core region of the solute is about ~ 1.5 times the size of a solvent molecule. The indirect PMF $\omega(r_{\text{co}} = 0)$ is $+3.89$ kcal/mol, and Figure 2a shows that the radial density of the cavity particle is approximately zero at $r_{\text{co}} = 0$. In contrast, for a Lennard-Jones fluid with a solute whose repulsive core region is about ~ 2.5 times the size of a solvent molecule, the indirect PMF $\omega(r_{\text{co}} = 0)$ is about $1.0 k_{\text{B}}T$ (~ 0.6 kcal/mol when the temperature is 300 K); the radial density of the cavity particle at $r_{\text{co}} = 0$ is about $1/e \approx 37\%$ of the solvent density in the bulk.⁷ This indicates that when the cavity particle is located in the repulsive core of a hydrophobic solute immersed in an aqueous solvent, it is much more strongly attracted to the solute–solvent interface than is the case when the solvent is a Lennard-Jones fluid.

Equation 43 shows the decomposition of the free-energy change for moving a cavity particle from the vacuum to the solution at position r_{co} into a sum of two terms: the average interaction energy between the cavity particle and the solvent and the reorganization free energy of the solvent. The free-energy change for inserting a cavity particle from the vacuum into the bulk ($r_{\text{co}} = \infty$) is -6.29 ± 0.01 kcal/mol. Figures 3a and 5a show that the average interaction energy between the cavity particle and the solvent $\langle U_{\text{vw}} \rangle$ is ~ -19.78 kcal/mol in the bulk.⁴⁷ According to eq 43, the reorganization free energy, ΔF_{reorg} for the insertion of a cavity particle in the bulk is ~ 13.49 kcal/mol.

Furthermore, we decompose the indirect PMF $\omega(r_{\text{co}})$ into the sum of the change of the average interaction energy, $\langle U_{\text{vw}} \rangle$, and the change of the reorganization free energy, ΔF_{reorg} , when the cavity particle is moved from the bulk to position r_{co} . The results are shown in Figures 3b and 5b. Note that the contribution of $\Delta \langle U_{\text{vw}} \rangle$ to $\omega(r_{\text{co}})$ is positive and the contribution of $\Delta \Delta F_{\text{reorg}}$ is negative and that the magnitudes of both terms, $\langle U_{\text{vw}} \rangle$ and ΔF_{reorg} , become smaller when the distance between the cavity particle and the solute, r_{co} , decreases. The indirect PMF $\omega(r_{\text{co}})$ is the result of the competition between the changes of these two terms. If the change of $\langle U_{\text{vw}} \rangle$ is smaller than the change of ΔF_{reorg} when the cavity particle is moved from the bulk to the center of the solute, the indirect PMF $\omega(r_{\text{co}})$ at $r_{\text{co}} = 0$ is negative. Namely, the cavity particle prefers overlapping with the solute to staying in the bulk, as is the case for the methane-like solute (C×1). In contrast, if the change of $\langle U_{\text{vw}} \rangle$ is larger than the change of ΔF_{reorg} , the indirect PMF $\omega(r_{\text{co}})$ at $r_{\text{co}} = 0$ is positive, as is the case for the larger solutes (C×2 or C×3). As mentioned in section 4.1.2, for the C×2 and the C×3 solutes, the radial density of the cavity particle $\rho_1^{(\text{c})}(r_{\text{co}})$ reaches its maximum inside the repulsive core of the solute when the density of the tagged water molecule approaches zero. As the cavity particle penetrates the repulsive core of the solute, at first the cost to reorganize the hydration shell decreases more rapidly than the loss of the interaction energy between the cavity particle and the solvent. Therefore, for solutes that are larger than water molecules, the radial density of cavity particles reaches its maximum at a distance that is inside the solute but is adjacent to the solute–solvent interface.

5.2. Structural Features of the Cavity Particle Interacting with the Solvation Shell. The tagged water molecule (or cavity particle) forms an average of 3.4 hydrogen bonds with the solvent molecules in the bulk. Figures 4a and 6a show that the total number of hydrogen bonds formed with the cavity particle begins to decrease once the cavity particle begins to penetrate the repulsive core of the solute. For the methane-like solute (CX1), the cavity particle can still form 2.4 hydrogen bonds with the solvent molecules when the cavity particle is located at the center of the solute ($r_{co} = 0$). Figure 3a shows that the average Coulomb interaction energy between the cavity particle and the solvent is attractive when the cavity particle is located at the center of the methane-like solute with substantial magnitude (-13.74 kcal/mol). However, for the larger solutes (CX2 or CX3), the cavity particle cannot form any hydrogen bonds with the solvent molecules when it is located at the center of the solute. Figure 5a shows that the average Coulomb interaction energy between the cavity particle and the solvent (CX2) is as weak as -2.43 kcal/mol when r_{co} is zero.

The blue curves in Figures 4b and 6b show the changes of the total number of hydrogen bonds formed within the first solvation shell when the cavity particle is moved from the bulk to the center of the solute, which can be divided into four stages. In the first stage, the cavity particle is located outside of the first solvation shell. The total number of hydrogen bonds within the first solvation shell is constant in this region. In the second stage, the cavity particle enters the first solvation shell but does not overlap with the solute. The total number of hydrogen bonds within the first solvation shell remains the same as the number when the cavity particle is in the bulk because the cavity particle simply replaces a regular untagged water molecule in the first solvation shell. In the third stage, the cavity particle begins to penetrate the repulsive core of the solute, while the total number of hydrogen bonds within the first solvation shell continues increasing. An explanation for this is that one more water molecule enters the first solvation shell as the cavity particle moves inside the repulsive core of the solute. In the last stage, the cavity particle totally overlaps with the solute and moves deeper into the solute. Now the total number of hydrogen bonds within the first solvation shell continues decreasing and approaches the value at large r_{co} because the cavity particle is located far from the hydrophobic interface. Note the last stage does not exist when the solute is CX1 because of the small size of this methane-like solute.

For all three hydrophobic solutes (CX1, CX2, and CX3), as the cavity particle penetrates the repulsive core, it acts as a perturbation on the untagged water molecules in the hydrophobic hydration shell. However, the decrease in the number of hydrogen bonds between untagged water in the first shell is more than compensated by the hydrogen bonding between the cavity particle and the (untagged) water in the first shell. In other words, the hydration shell water molecules are able to maintain their hydrogen bonding to a large extent while also hydrogen bonding to the cavity particle inside the repulsive core. The ability of waters in the first shell to adapt in this way to the perturbation by the cavity particle is related to the enhanced fluctuations of water at hydrophobic interfaces.^{33,55,56}

The green curves in Figures 4b and 6b show the changes of the total number of hydrogen bonds formed only by water molecules (excluding the cavity particle) within the first solvation shell. The comparisons between the values at $r_{co} = 0$

and $r_{co} = 1.0$ nm indicate the disturbance to the first solvation shell caused by the cavity particle that is located at the center of the solute. For the methane-like solute, the total number of hydrogen bonds formed by the water molecule within the first solvation shell with the cavity particle located at the center of the solute is smaller than the number when the cavity particle is in the bulk. Figure 3b shows that the reorganization free energy is about ~ 7.21 kcal/mol to insert a cavity particle at the center of the solute, which is 6.28 kcal/mol smaller compared with the value of inserting a cavity particle in the bulk, 13.49 kcal/mol. However, for the CX2 and CX3 solutes, the total number of hydrogen bonds formed by the water molecule within the first solvation shell with the cavity particle located at the center of the solute is approximately the same as the total number when the cavity particle is in the bulk. This suggests that the water molecules within the first shell recover their unperturbed structure when the cavity particle penetrates deeper into the solute. Figure 5b shows that the reorganization free energy is only ~ 1.26 kcal/mol to insert a cavity particle at the center of the CX2 solute, which is much smaller compared with the value of the reorganization free energy of inserting a cavity particle in the bulk, 13.49 kcal/mol.

5.3. Solvation Free Energy from the Endpoints DFT Method. The DFT formalism in section 2 provides a connection between the solvation free energy of the solute, $\Delta\mu_w$ and the indirect part of the solute–solvent potential of mean force (indirect PMF), $\omega_\lambda(\mathbf{x})$, where $\mathbf{x}$ is the configuration (position and orientation) of a solvent molecule relative to the solute. A difficulty in practical implementation of DFT is the high dimensionality of $\mathbf{x}$. It is six-dimensional for the water molecule, for example. To circumvent the problem of dimensionality within the DFT framework, the method of energy representation was proposed.^{1,2,11,13,37} In this method, the solute–solvent distribution is described with respect to the solute–solvent pair interaction $v(\mathbf{x})$ of interest as

$$\hat{\rho}(\epsilon) = \sum_i \delta(\epsilon - v(\mathbf{x}_i)) \quad (46)$$

$\hat{\rho}(\epsilon)$ corresponds to the instantaneous histogram of the value ϵ of $v(\mathbf{x})$ in a snapshot configuration. The solute–solvent potential $\phi_\lambda(\mathbf{x})$ is then set to a composite form of $\phi_\lambda(v(\mathbf{x}))$, and the one-body distribution function $\rho_\lambda^{(1)}(\epsilon)$ can be expressed as the projection:

$$\rho_\lambda^{(1)}(\epsilon) = \int d\mathbf{x} \delta(\epsilon - v(\mathbf{x})) \rho_\lambda^{(1)}(\mathbf{x}) \quad (47)$$

The indirect part of the potential of mean force is introduced similarly to eq 4 as

$$\rho_\lambda^{(1)}(\epsilon) = \rho_0^{(1)}(\epsilon) \exp(-\beta\phi_\lambda(\epsilon) - \beta\omega_\lambda(\epsilon)) \quad (48)$$

and $\Delta\mu_u$ is expressed exactly as

$$\Delta\mu_u = -k_B T \int d\epsilon \left[(\rho_1^{(1)}(\epsilon) - \rho_0^{(1)}(\epsilon)) + \beta\omega_1(\epsilon)\rho_1^{(1)}(\epsilon) - \left(\beta \int_0^1 d\lambda \omega_\lambda(\epsilon) \right) \times (\rho_1^{(1)}(\epsilon) - \rho_0^{(1)}(\epsilon)) \right] \quad (49)$$

when $\rho_\lambda^{(1)}$ is taken to vary linearly with λ . The correspondence of eq 49 with eq 27 is striking. A valid formulation of DFT can be obtained by reducing the dimensionality of $\mathbf{x}$ to ϵ , and with the one-dimensional ϵ , the DFT treatment of $\Delta\mu_u$ is numerically much more feasible even for complicated solutes

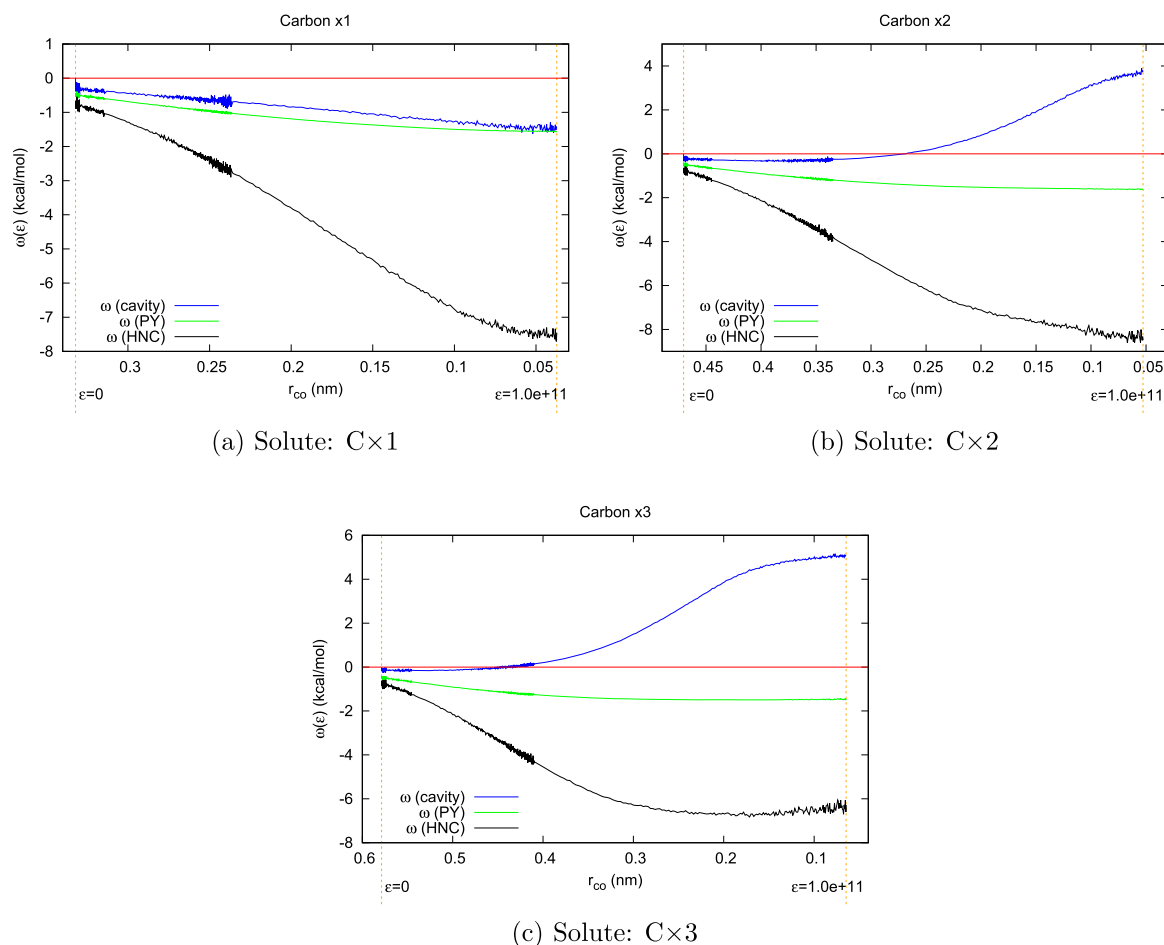


Figure 7. Comparison between the indirect PMF in the repulsive region estimated by the density of the cavity particle obtained from explicit simulations of the cavity particle in solution with the hydrophobic solute at the $\lambda = 1$ state and those from the HNC and PY approximations of the indirect PMF through eqs 50 and 51, respectively.

and/or solvent molecules. In eq 49, the integrand is written as a sum of three terms. The contribution of the first term to the solvation free energy is zero when the simulations are performed in the canonical ensemble. The second term does not depend on the indirect PMF $\omega_1(\epsilon)$ in the repulsive region of the solute because the density $\rho_1^{(1)}(\epsilon)$ in the repulsive region decreases rapidly to zero. Within the current implementation (called ERmod software),³⁷ the value of $\omega_\lambda(\epsilon)$ in the third term is approximately estimated (in the repulsive region) by combining the HNC-type form of

$$\omega_\lambda(\epsilon) = \lambda \sigma_0(\epsilon) \quad (50)$$

and the PY-type form of

$$-\beta \omega_\lambda(\epsilon) = \log(1 - \lambda \beta \sigma_0(\epsilon)) \quad (51)$$

where

$$\sigma_0(\epsilon) = \left. \frac{\partial \omega_\lambda(\epsilon)}{\partial \lambda} \right|_{\lambda=0} = -k_B T \int d\eta \left(\frac{\delta(\epsilon - \eta)}{\rho_0^{(1)}(\epsilon)} - (\chi_0)^{-1}(\epsilon, \eta) \right) (\rho_1^{(1)}(\eta) - \rho_0^{(1)}(\eta)) \quad (52)$$

and $\chi_0(\epsilon, \eta)$ is the correlation matrix in the energy representation at the pure solvent state (see eqs 29, 31, and 33).

In the following, ω denotes ω_1 at $\lambda = 1$ for notational brevity. Figure 7 shows the comparisons between the indirect PMF ω measured by the cavity particle density and the HNC and the PY approximations (through eqs 50 and 51, respectively, with $\lambda = 1$) in the solute repulsive core where the interaction energy ϵ is positive. For a methane-like solute, the PY approximation shows good agreement with the benchmark, the indirect PMF ω measured by using the density distribution of the cavity particle. The difference between the PY approximation and the benchmark is smaller than 0.2 kcal/mol in the whole repulsive region where $0 < \epsilon$. The difference between the HNC and the benchmark increases when the interaction energy between the tagged water molecule and the solute ϵ increases as the cavity particle penetrates deeper into the core. The difference is ~ 6 kcal/mol at $r_{co} = 0.0527$ nm (where ϵ is $1.0e + 11$ kcal/mol). For the Cx2 and the Cx3 solutes, neither the PY nor the HNC provides a good approximation of the indirect PMF ω in the deep-core region. When the solute is large, the indirect PMF ω will switch from negative to positive when the distance between the cavity particle and the solute, r_{co} , decreases. However, both the PY and the HNC approximations are

always negative in the repulsive region and fail to catch the sign change of the indirect PMF ω for large solutes.

In Table 2, we compare the solvation free energies of these three different solutes estimated by ERmod and the bench-

Table 2. Estimates of Solvation Free Energies (kcal/mol)^a

solute	FEP	ERmod	ERmod(2P-QHNC)
C×1	3.39 ± 0.01	3.39 ± 0.01	3.81 ± 0.02
C×2	8.44 ± 0.02	10.46 ± 0.01	9.68 ± 0.04
C×3	12.45 ± 0.04	19.21 ± 0.02	13.86 ± 0.18

^aThe benchmarks are obtained by free-energy perturbation (FEP) simulations.

marks that were estimated by the free-energy perturbation (FEP) algorithm. The volumes of the C×2 and C×3 solutes are ~ 3 and ~ 5 times that of the C×1 (methane-like) solute. For the C×1 and C×2 solutes, their solvation free energies are roughly proportional to their volumes. However, the ratio of solvation free energies of the C×3 solute and the C×1 solute ~ 3.7 fall between their surface area ratio ~ 3 and their volume ratio ~ 5 , and it is in the crossover region.²⁴ The solvation free energy of the methane-like solute estimated by ERmod matches the benchmark. However, for the C×2 and the C×3 solutes, the solvation free energy values estimated by ERmod are more positive compared to those of the benchmarks.

Equation 49 is an exact expression for the solvation free energy $\Delta\mu_w$ and in the endpoints DFT scheme, the integration of ω_λ in the repulsive core over $0 \leq \lambda \leq 1$ can be approximated through eqs 50 and 51. Let $\sigma_0(\epsilon)$ and $\omega(\epsilon)$ denote $(\partial\omega_\lambda(\epsilon)/\partial\lambda)_{\lambda=0}$ and $\omega_\lambda(\epsilon)$ at $\lambda = 1$, respectively. $\sigma_0 = \omega$ holds in the HNC relationship, and their difference corresponds to the bridge function. Figure 7 provides the numerical values of both σ_0 and ω , and it is observed that ω is more positive than σ_0 in the repulsive region of the solute–solvent interaction. This indicates that ω_λ in the repulsive region is convex downward with respect to the λ dependence, and the PY relationship of eq 51 is considered as a scheme to mimic the convexity. It is not “convex enough”, however, when the solute is as large as C×2 or C×3 or larger. In the deep-core region, ω_λ from the PY approximation stays negative over $0 < \lambda \leq 1$ while ω at $\lambda = 1$ is positive according to Figure 7. Actually, ω is equal to the negative of the solvation free energy of the solvent in its bulk when the solute is large enough and the cavity particle is located deeply in the solute core. This is because $\omega_\lambda(\epsilon)$ is the transfer free energy of the cavity particle from the bulk to the location (over the energy coordinate) ϵ and the cavity particle is isolated from the other solvent molecules when ϵ corresponds to the deep core of a large enough solute. The positive ω at $\lambda = 1$ in the deep cores of C×2 or C×3 reflects the above property, though the solute is still of molecular scale. The PY approximation is adopted in the repulsive region of the

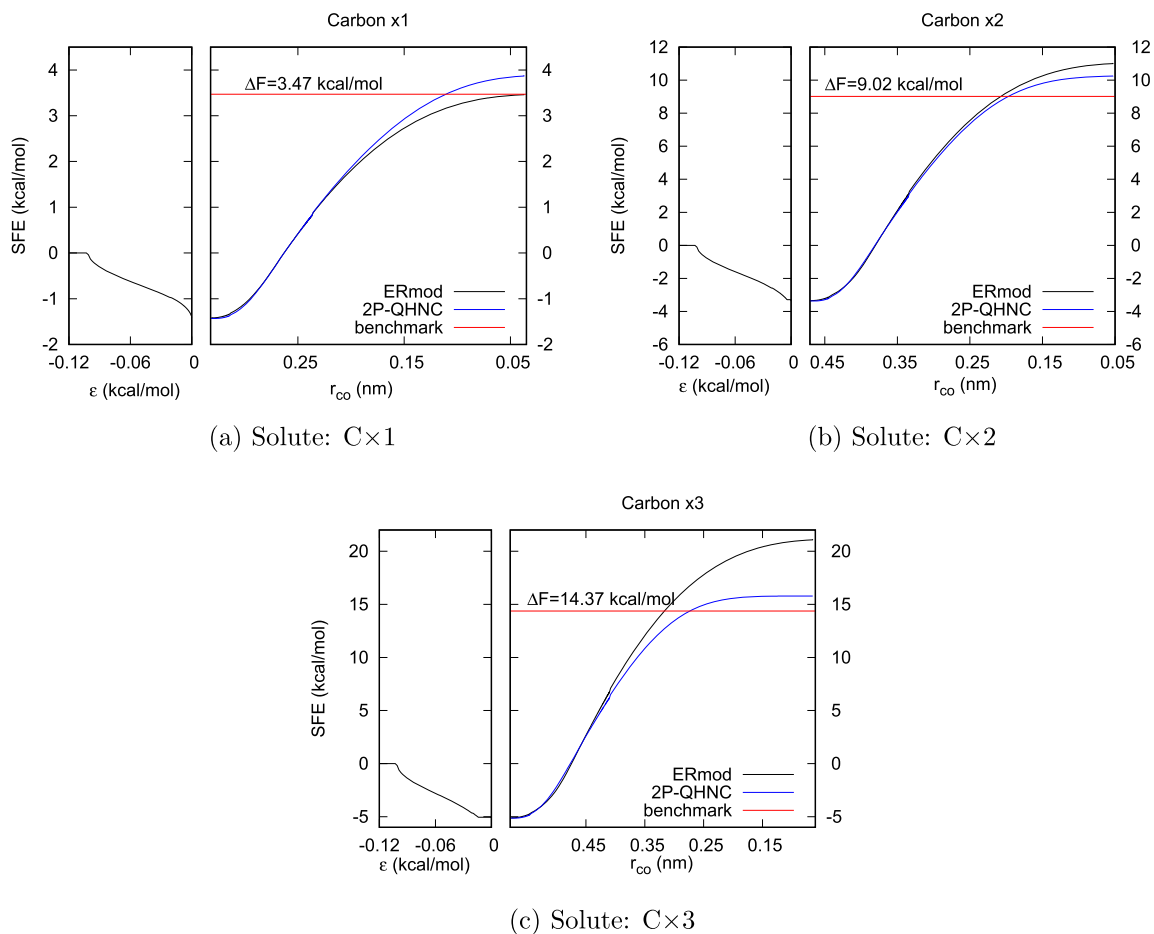


Figure 8. Comparison between the original ERmod and the (modified) ERmod using the 2P-QHNC approximation. Note that the cumulative results for these two do not include the Lennard-Jones long-range corrections. Therefore, the FEP values are further shifted from those in Table 2 by removing the long-range corrections.

solute–solvent interaction in the endpoints DFT scheme employed at present.³⁷ The contribution from that region to the solvation free energy will then be overestimated according to eq 49 because $\rho_0 > \rho$ there and the PY estimate of ω is more negative than the exact value.

In the repulsive-core region, only σ_0 has been used to approximate the integration over λ in eq 49 in the endpoints functional in ref 37. This means that the λ dependence of ω_λ is estimated as an extrapolation through eq 50 or 51. In the present work, we have determined ω (at $\lambda = 1$) and can therefore formulate an interpolation by using both σ_0 and ω . Because $\sigma_0 = (\partial\omega_\lambda/\partial\lambda)_{\lambda=0}$ and $\omega = \omega_1$, a simplest expression is quadratic and is given by

$$\omega_\lambda = \lambda(1 - \lambda)\sigma_0 + \lambda^2\omega \quad (53)$$

$$\int_0^1 d\lambda \omega_\lambda = \frac{1}{6}\sigma_0 + \frac{1}{3}\omega \quad (54)$$

Equation 53 enforces $\omega_1 = \omega$ at $\lambda = 1$, while it keeps the derivative at $\lambda = 0$ valid. The λ dependence of the indirect PMF ω_λ was linear in the HNC-type form of eq 50 and is extended in eq 53 to the quadratic form by using σ_0 and ω_1 determined at the two endpoints of $\lambda = 0$ and 1. We thus call eq 53 2-point quadratic HNC (2P-QHNC). We note that the term of quadratic HNC was originally adopted in ref 57 in an integral-equation study of multipolar hard spheres through the expansion of angle-dependent parts of correlation functions to the second order. Equation 53 is distinct from this and focuses on the dependence of the indirect PMF on the coupling parameter λ . With eq 54 adopted in eq 49, the approximate estimate of the solvation free energy is improved for the larger solutes as shown in Table 2.

Figure 8 shows the cumulative results of calculating the solvation free energy by using eq 49, which are running integrals obtained by setting the upper limit of the integration in eq 49 to ϵ . The full value of the solvation free energy is recovered at $\epsilon \rightarrow \infty$. Each picture in Figure 8 is separated into two panels. The left panel exhibits the cumulative result of the integral in eq 49 when the integral upper limit changes from -0.12 kcal/mol to 0; the right panel exhibits the cumulative result when the integral upper limit changes from 0 to $1.0\epsilon + 11$ kcal/mol. The horizontal variable in the right panel, the interaction energy between the tagged water molecule and the solute ϵ , is replaced by their distance r_{co} because of the one-to-one relation between ϵ and r_{co} when ϵ is positive. Since the first and the second terms are treated exactly in eq 49, the differences between the original ERmod and the (modified) ERmod using the 2P-QHNC approximation in the right panels of Figure 8 are contributed by the third (charging) term of the integrand in eq 49. For larger solutes, CX2 and CX3, the ERmod estimates using the 2P-QHNC approximation are less positive and the absolute differences compared to the benchmarks are smaller than the original ERmod estimates. Notice that the differences between the 2P-QHNC results and the benchmarks are similar for the CX2 and CX3 solutes, which suggests that the difference between the 2P-QHNC results and the “true” solvation free energies might approach a constant value (or increase very slowly) as we make the solute larger still (i.e., CX4, CX5, etc.), whereas the differences between the original ERmod results and the “true” results continue to increase.

Further improvements are possible by examining the λ dependence of ω_λ more closely, while it was observed that the

free-energy differences due to conformational changes are well-reproduced even with the currently employed functional (see section 5.4).

5.4. Possible Future Directions and Practical Applications of the Density Functionals. Table 2 shows that the performance of the endpoints approximation deteriorates with the solute size. When seen in terms of the solute volume, however, the deviation of the solvation free energy $\Delta\mu_u$ from its numerically exact value in Table 2 is larger than those in previous studies.^{13,58,59} The volumes of the CX2 and CX3 solutes are ~ 3 and ~ 5 times that of the CX1 (methane-like) solute (note that we adopted the geometric combination rule for LJ σ). Many of the amino acid analogues examined in refs 13 and 58 have more than 3 or 5 non-hydrogen atoms, while the difference between the approximate and exact values is less than 1 kcal/mol.^{13,58} Alanine dipeptide and β -cyclodextrin have even more non-hydrogen atoms (10 and 77, respectively), and the deviation from the exact value was still found to be within 2 and 10 kcal/mol, respectively.⁵⁹ This results from a cancellation of errors between the hydrophobic and polar contributions for realistic solutes. As noted in section 5.3, the contribution from the repulsive region of the solute–solvent interaction to $\Delta\mu_u$ is overestimated in the endpoints method. Table 2 shows that $\Delta\mu_u$ is well-reproduced for the CX1 (methane-like) solute, indicating that the attractive contribution cancels the overestimation. It should also be noted that realistic molecules are formed from such non-hydrogen atoms as carbon, oxygen, and nitrogen with overlapping excluded-volume regions. This is distinct from the present models in which the repulsive region of a single atomic core grows. The sizes of the carbon, oxygen, and nitrogen atoms are comparable among one another and comparable to that for the CX1 solute. The approximation for ω is worse at positions that are more distant from the solute surface, whereas the deep-core regions for those atoms cannot be as separated from the surface as the deep cores for CX2 and CX3. It is thus expected that the overestimation from the repulsive region is less severe than in the cases of the larger solutes treated in this work. Indeed, the cumulative results in Figure 8 from the original ERmod estimates deviate significantly from those from the 2P-QHNC approximation only in the deep-core part of the repulsive region of the solute–solvent interaction. The original and quadratic estimates agree with each other in the “skin” parts that are close to the solute surfaces, and the volumes of the deep-core parts for realistic molecules are not as large as those for the CX2 and CX3 solutes. It was seen from our previous analysis for alanine dipeptide and β -cyclodextrin, in addition, that the conformational dependence of the solvation free energy is reproduced by the approximate endpoints DFT method within $\sim k_B T$.⁵⁹ This reflects the observation that the density of the solvation free energy depends weakly on the conformation.¹¹ When the conformation of a protein changes, its excluded volume varies only because of the rearrangements of the “skin” parts. The deep-core parts keep their volumes, so that their contributions will be canceled for conformational changes.

When the core volume of the solute, V , is large enough, the solvation free energy, $\Delta\mu_w$, scales with $p_0 V$, with p_0 being the pressure of the system. This asymptotic property holds for eqs 17 and 49 because they are exact formulas. For the HNC approximation as well as for the information-theoretic approach, however, it was pointed out that the proportionality constant for the V dependence is $1/2\kappa_T$ or $(\rho_0 k_B T/2 + 1/2\kappa_T)$,

where ρ_0 is the bulk density of solvent and κ_T is the isothermal compressibility.^{21,60,61} The V dependence then has a proportionality constant of 10^4 atm for ambient water and is a source of overestimation of $\Delta\mu_u$ for a large solute. The PY approximation provides a milder constant of proportionality of $-\rho_0 k_B T \log(\rho_0 k_B T \kappa_T)$, which is still ~ 3500 atm.^{60,61} The too large volume dependence has been fixed with the volume-correction scheme.^{62–64} In this scheme, the V -dependent term is forced to be $p_0 V$ and the performance improves for the $\Delta\mu_u$ expression. $\Delta\mu_u$ is too positive for CX2 or CX3 in Table 2, and it was noted in section 5.3 that the contribution from the deep-core region of the solute is overestimated with too favorable ω in the HNC and PY approximations. Molecular simulations can provide numerically exact values of ω_λ , which serve as a guide to develop improved functionals that suppress the deep-core contribution toward possible improvement of the endpoints DFT method.

The 2P-QHNC approximation proposed in the current work leads to improved estimates of the solute chemical potential, but it relies on cavity particle simulations to estimate the indirect PMF ω at the $\lambda = 1$ solution state that are computationally costly and not easily performed for more complex solutes. To obtain the indirect PMF in this way, a free-energy method such as umbrella sampling needs to be employed. Recently we proposed an alternative method to estimate the indirect PMF ω at the solute–solvent interface using reweighting techniques to combine data from the $\lambda = 0$ and $\lambda = 1$ states derived from UWHAM.⁶⁵ It may be possible to use a similar approach to extrapolate values of $\omega(\mathbf{x})$ into the solute repulsive core. In this event the 2P-QHNC functional could be evaluated without requiring explicit cavity particle simulations, and it could then be used as a practical alternative to the linear endpoints DFT functional that does not use ω at $\lambda = 1$ to estimate the contribution to the DFT charging integral from the region within the solute core.

Our previous observations^{11,59} that errors in the endpoints density functional which are located deeper in the solute core region tend to cancel when calculating conformational free-energy profiles suggests that the use of reweighting techniques to estimate the indirect PMF ω in the solute “outer skin” region may be sufficient to construct accurate solute conformational free-energy profiles in solution using 2P-QHNC. It is precisely in this solute outer skin region where the use of UWHAM to estimate ω is expected to be most accurate. We plan to report in a future communication on the use of UWHAM and 2P-QHNC to estimate ω and solvation free energies using physical endpoint simulations without simulating cavity particles explicitly.

6. CONCLUSIONS

Endpoints DFT provides a density functional theory framework for estimating solvation thermodynamic properties from simulations of the two physical endpoint states, the pure liquid and the solution containing one or more solutes. A key role is played by the indirect part of the solute–solvent potential of mean force (indirect PMF), and in this work the equivalence with the standard formulation employing the direct correlation functions is established. The indirect PMF describes the collective effects of solvent on the solute–solvent correlation, and our formulation connects the solvent effects at the molecular level to the solvation free energy of the solute. Approximations were introduced with respect to the indirect PMF, and we examined the indirect PMF for spherical,

hydrophobic solutes with varied sizes in water. It was observed that the indirect PMF is favorable (negative) at distances not far from the solute surface for each solute examined and becomes unfavorable (positive) in the deep-core regions of larger solutes. This reflects the dependencies on the cavity–solute distance of the interaction energy between the cavity particle and solvent water and the solvent-reorganization contribution. Upon penetration of the cavity particle within the solute repulsive core, the former reduces to a lesser extent than the free-energy penalty from solvent reorganization when the cavity particle stays near solvent water molecules, while the opposite is true when the cavity particle penetrates more deeply into the core for the two larger hydrophobic solutes (CX2 and CX3). When the cavity particle stays near the solute surface, it maintains the hydrogen bonding with solvent water, and at the same time, the water molecules do not sacrifice the hydrogen bonds with the other water molecules. An approximation to the solvation free energy of the solute can be formulated using HNC-type and PY-type relationships in the endpoints DFT method, and numerical evaluation of the solvation free energy was conducted by adopting the one-dimensional, energy representation. The indirect PMF computed in a numerically exact scheme was actually found to be less negative than the HNC-type and PY-type estimates, and these approximations could not reproduce the sign changes of the exact, indirect PMF from negative to positive in the deep-core regions of larger-sized solutes. This leads to the overestimation of the solvation free energies of the two larger solutes from the approximate, endpoints method compared to the benchmarks without approximations. It was shown indeed that the approximate estimate improves when the correct indirect PMF at the $\lambda = 1$ endpoint was incorporated into the endpoints DFT functional using the 2-points quadratic HNC (2P-QHNC) approximation.

The endpoints DFT method has a firm theoretical basis and is advantageous in terms of computational speed and interpretability with a variety of applications.^{13,37} The method combines a statistical–mechanical DFT formalism with molecular simulations that are currently possible at nanometer and microsecond length and time scales to approach the free energetics of solvation efficiently and accurately with extended applicability to diverse problems in chemistry and biophysics, including protein allostery and ligand binding, micellar and lipid membrane chemistry, polymers, surface phenomena, and supercritical fluids.^{12,13} The key quantity is the indirect PMF, and the approximation adopted for it determines the accuracy of the endpoints method. Molecular simulations of cavity particles can approach the numerically exact, indirect PMF, which can be helpful for further improvement of the endpoints DFT scheme.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcb.0c02721>.

Density of the cavity particle; Lennard-Jones interactions; potential of mean force of the cavity particle for the CX3 solute; interaction between the cavity particle and the solvent for the CX3 solute; validation of the indirect PMF $\omega(r_{co} = 0)$ (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Nobuyuki Matubayasi – Division of Chemical Engineering, Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka 560-8531, Japan; orcid.org/0000-0001-7176-441X; Email: nobuyuki@cheng.es.osaka-u.ac.jp

Ronald M. Levy – Center for Biophysics and Computational Biology, and Department of Chemistry, Temple University, Philadelphia, Pennsylvania 19122, United States; orcid.org/0000-0001-8696-5177; Email: ronlevy@temple.edu

Author

Bin W. Zhang – Center for Biophysics and Computational Biology, and Department of Chemistry, Temple University, Philadelphia, Pennsylvania 19122, United States; orcid.org/0000-0003-3007-4900

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jpcb.0c02721>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by an NIH grant (R35GM132090), an NSF grant (1665032), and an NIH computer equipment grant (OD020095). This research includes calculations carried out on HPC resources (Owl's Nest of Temple University) supported in part by the National Science Foundation through major research instrumentation grant number 1625061 and by the U.S. Army Research Laboratory under contract number W911NF-16-2-0189. We are further grateful to the Grant-in-Aid for Scientific Research (No. JP19H04206) from the Japan Society for the Promotion of Science and to the Elements Strategy Initiative for Catalysts and Batteries (No. JPMXP0112101003) and the Fugaku Supercomputer Project from the Ministry of Education, Culture, Sports, Science, and Technology.

■ DEDICATION

It is a great pleasure for me (R.M.L.) to contribute to this Festschrift honoring the scientific contributions of Peter Rossky. I have known Peter since 1974 when we were very briefly officemates in Old Prince House at Harvard. We shared an addiction to nicotine and not much else, I thought, but I later found out we shared other interests, including repairing old cars as graduate students and more recently watching formula 1 cars fly around tracks at crazy speeds. I got to know Peter better when we had nearby offices in New Prince House. It was during the period in Prince House from 1976 to 1978 that I grew to have tremendous respect for Peter as a scientist, and during that time he helped inspire my long-term interest in the statistical mechanics of solutions. Those years were also during the heyday of the classical density functional theory of solutions, and in the 2000s these integral equation methods are getting a modern makeover by combining them with computer simulations in various ways; the “endpoints DFT” approach originally proposed by one of the authors (N.M.) is further developed in the present paper.

■ REFERENCES

- (1) Matubayasi, N.; Nakahara, M. Theory of solutions in the energetic representation. I. Formulation. *J. Chem. Phys.* **2000**, *113*, 6070.
- (2) Matubayasi, N.; Nakahara, M. Theory of solutions in the energy representation. II. Functional for the chemical potential. *J. Chem. Phys.* **2002**, *117*, 3605.
- (3) Matubayasi, N.; Nakahara, M. Theory of solutions in the energy representation. III. Treatment of the molecular flexibility. *J. Chem. Phys.* **2003**, *119*, 9686.
- (4) Cui, D.; Zhang, B. W.; Matubayasi, N.; Levy, R. M. The Role of Interfacial Water in Protein–Ligand Binding: Insights from the Indirect Solvent Mediated Potential of Mean Force. *J. Chem. Theory Comput.* **2018**, *14*, 512.
- (5) Llano-Restrepo, M.; Chapman, W. G. Bridge function and cavity correlation function for the Lennard-Jones fluid from simulation. *J. Chem. Phys.* **1992**, *97*, 2046–2054.
- (6) Llano-Restrepo, M.; Chapman, W. G. Bridge function and cavity correlation function for the soft sphere fluid from simulation: Implications on closure relations. *J. Chem. Phys.* **1994**, *100*, 5139–5148.
- (7) Vyalov, I.; Chuev, G.; Georgi, N. Solute-solvent Cavity and Bridge Functions. I. Varying Size of the Solute. *J. Chem. Phys.* **2014**, *141*, No. 074505.
- (8) Pratt, L. R.; Chandler, D. Theory of the hydrophobic effect. *J. Chem. Phys.* **1977**, *67*, 3683–3704.
- (9) Jeanmairet, G.; Levesque, M.; Vuilleumier, R.; Borgis, D. Molecular density functional theory of water. *J. Phys. Chem. Lett.* **2013**, *4*, 619–624.
- (10) Belloni, L. Exact Molecular Direct, Cavity, and Bridge Functions in Water System. *J. Chem. Phys.* **2017**, *147*, 164121.
- (11) Ishii, Y.; Yamamoto, N.; Matubayasi, N.; Zhang, B. W.; Cui, D.; Levy, R. M. Spatially-Decomposed Free Energy of Solvation Based on the Endpoint Density-Functional Method. *J. Chem. Theory Comput.* **2019**, *15*, 2896–2912.
- (12) Matubayasi, N. Free-Energy Analysis of Protein Solvation with All-Atom Molecular Dynamics Simulation Combined with a Theory of Solutions. *Curr. Opin. Struct. Biol.* **2017**, *43*, 45–54.
- (13) Matubayasi, N. Energy-Representation Theory of Solutions: Its Formulation and Application to Soft, Molecular Aggregates. *Bull. Chem. Soc. Jpn.* **2019**, *92*, 1910–1927.
- (14) Henderson, D., Ed. *Fundamentals of Inhomogeneous Fluids*; Marcel Dekker, Inc.: New York, 1992.
- (15) Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids: with Applications to Soft Matter*, 4th ed.; Academic Press: Amsterdam, 2013.
- (16) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Academic Press: London, U.K., 2002.
- (17) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*, 2nd ed.; Oxford University Press: Oxford, 2017.
- (18) Hirata, F.; Rossky, P. J. On the interpretation of solute induced solvent structure. *J. Chem. Phys.* **1981**, *74*, 5324–5326.
- (19) Lee, C.-Y.; McCammon, J. A.; Rossky, P. J. The structure of liquid water at an extended hydrophobic surface. *J. Chem. Phys.* **1984**, *80*, 4448–4455.
- (20) Lee, S. H.; Rossky, P. J. A comparison of the structure and dynamics of liquid water at hydrophobic and hydrophilic surfaces—a molecular dynamics simulation study. *J. Chem. Phys.* **1994**, *100*, 3334–3345.
- (21) Hummer, G.; Garde, S.; García, A. E.; Paulaitis, M. E.; Pratt, L. R. Hydrophobic Effects on a Molecular Scale. *J. Phys. Chem. B* **1998**, *102*, 10469.
- (22) Hummer, G.; Garde, S.; García, A. E.; Pratt, L. R. New perspectives on hydrophobic effects. *Chem. Phys.* **2000**, *258*, 349–370.
- (23) Pratt, L. R. Molecular theory of hydrophobic effects: “She is too mean to have her name repeated. *Annu. Rev. Phys. Chem.* **2002**, *53*, 409–436.

- (24) Chandler, D. Interfaces and the driving force of hydrophobic assembly. *Nature* **2005**, *437*, 640–647.
- (25) Imai, T.; Hirata, F. Hydrophobic effects on partial molar volume. *J. Chem. Phys.* **2005**, *122*, No. 094509.
- (26) Choudhury, N.; Pettitt, B. M. The Dewetting Transition and The Hydrophobic Effect. *J. Am. Chem. Soc.* **2007**, *129*, 4847–4852.
- (27) Giovambattista, N.; Debenedetti, P. G.; Rossky, P. J. Effect of Surface Polarity on Water Contact Angle and Interfacial Hydration Structure. *J. Phys. Chem. B* **2007**, *111*, 9581–9587.
- (28) Mittal, J.; Hummer, G. Static and dynamic correlations in water at hydrophobic interfaces. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 20130–20135.
- (29) Patel, A. J.; Varilly, P.; Chandler, D. Fluctuations of Water near Extended Hydrophobic and Hydrophilic Surfaces. *J. Phys. Chem. B* **2010**, *114*, 1632–1637.
- (30) Garde, S.; Patel, A. J. Unraveling the hydrophobic effect, one molecule at a time. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 16491–16492.
- (31) Patel, A. J.; Varilly, P.; Jamadagni, S. N.; Hagan, M. F.; Chandler, D.; Garde, S. Sitting at the Edge: How Biomolecules use Hydrophobicity to Tune Their Interactions and Function. *J. Phys. Chem. B* **2012**, *116*, 2498–2503.
- (32) Harris, R. C.; Pettitt, B. M. Effects of geometry and chemistry on hydrophobic solvation. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 14681–14686.
- (33) Xi, E.; Patel, A. J. The hydrophobic effect, and fluctuations: The long and the short of it. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 4549–4551.
- (34) Pratt, L. R.; Chaudhari, M. I.; Rempe, S. B. Statistical Analyses of Hydrophobic Interactions: A Mini-Review. *J. Phys. Chem. B* **2016**, *120*, 6455–6460.
- (35) Hirata, F.; Sugita, M.; Yoshida, M.; Akasaka, K. Perspective: Structural fluctuation of protein and Anfinsen's thermodynamic hypothesis. *J. Chem. Phys.* **2018**, *148*, No. 020901.
- (36) Torrie, G.; Patey, G. Monte Carlo calculation of $y(r)$ for the hard-sphere fluid. *Mol. Phys.* **1977**, *34*, 1623–1628.
- (37) Sakuraba, S.; Matubayasi, N. ERmod: Fast and versatile computation software for solvation free energy with approximate theory of solutions. *J. Comput. Chem.* **2014**, *35*, 1592–1608.
- (38) Levy, R. M.; Cui, D.; Zhang, B. W.; Matubayasi, N. Relationship between Solvation Thermodynamics from IST and DFT Perspectives. *J. Phys. Chem. B* **2017**, *121*, 3825–3841.
- (39) Beck, T. L.; Paulaitis, M. E.; Pratt, L. R. *The Potential Distribution Theorem and Models of Molecular Solutions*; Cambridge University Press: New York, 2006.
- (40) Kjellander, R.; Sarman, S. The chemical potential of simple fluids in a common class of integral equation closures. *J. Chem. Phys.* **1989**, *90*, 2768–2775.
- (41) Ramshaw, J. D. Functional derivative relations for a finite non-uniform molecular fluid in the canonical ensemble. *Mol. Phys.* **1980**, *41*, 219–227.
- (42) Hernando, J. A. Density functional theory in the canonical ensemble: I. General formalism. *J. Phys.: Condens. Matter* **2002**, *14*, 303–317.
- (43) White, J. A.; Velasco, S. The Ornstein-zernike Equation in the Canonical Ensemble. *Europhys. Lett.* **2001**, *54*, 475–481.
- (44) Krüger, P.; Schnell, S. K.; Bedeaux, D.; Kjelstrup, S.; Vlugt, T. J. H.; Simon, J.-M. Kirkwood–Buff Integrals for Finite Volumes. *J. Phys. Chem. Lett.* **2013**, *4*, 235–238.
- (45) Cortes-Huerto, R.; Kremer, K.; Potestio, R. Kirkwood–buff Integrals in the Thermodynamic Limit from Small-sized Molecular Dynamics Simulations. *J. Chem. Phys.* **2016**, *145*, 141103.
- (46) Rogers, D. M. Extension of Kirkwood–buff Theory to the Canonical Ensemble. *J. Chem. Phys.* **2018**, *148*, No. 054102.
- (47) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (48) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.
- (49) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- (50) Parrinello, M.; Rahman, A. Polymorphic Transitions In Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (51) Van Gunsteren, W. F.; Berendsen, H. J. C. A Leap-frog Algorithm for Stochastic Dynamics. *Mol. Simul.* **1988**, *1*, 173–185.
- (52) Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952–962.
- (53) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (54) Zhang, B. W.; Arasteh, S.; Levy, R. M. The UWHAM and SWHAM Software Package. *Sci. Rep.* **2019**, *9*, 2803.
- (55) Hummer, G.; Garde, S.; Garcia, A. E.; Pohorille, A.; Pratt, L. R. An information theory model of hydrophobic interactions. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 8951–8955.
- (56) Wallqvist, A.; Gallicchio, E.; Levy, R. M. A Model for Studying Drying at Hydrophobic Interfaces: Structural and Thermodynamic Properties†. *J. Phys. Chem. B* **2001**, *105*, 6745–6753.
- (57) Patey, G. Integral equation approximations for fluids of hard spheres with linear quadrupoles. *Mol. Phys.* **1978**, *35*, 1413–1428.
- (58) Karino, Y.; Fedorov, M. V.; Matubayasi, N. End-Point Calculation of Solvation Free Energy of Amino-Acid Analogs by Molecular Theories of Solution. *Chem. Phys. Lett.* **2010**, *496*, 351–355.
- (59) Harris, R. C.; Deng, N.; Levy, R. M.; Ishizuka, R.; Matubayasi, N. Computing Conformational Free Energy Differences in Explicit Solvent: An Efficient Thermodynamic Cycle Using an Auxiliary Potential and a Free Energy Functional Constructed from the End Points. *J. Comput. Chem.* **2017**, *38*, 1198–1208.
- (60) Matubayasi, N.; Kinoshita, M.; Nakahara, M. Hydration free energy of hard-sphere solute over a wide range of size studied by various types of solution theories. *Condens. Matter Phys.* **2007**, *10*, 471–480.
- (61) Fujita, T.; Yamamoto, T. Assessing the accuracy of integral equation theories for nano-sized hydrophobic solutes in water. *J. Chem. Phys.* **2017**, *147*, No. 014110.
- (62) Chuev, G. N.; Fedorov, M. V.; Crain, J. Improved Estimates for Hydration Free Energy Obtained by the Reference Interaction Site Model. *Chem. Phys. Lett.* **2007**, *448*, 198–202.
- (63) Ratkova, E. L.; Palmer, D. S.; Fedorov, M. V. Solvation Thermodynamics of Organic Molecules by the Molecular Integral Equation Theory: Approaching Chemical Accuracy. *Chem. Rev.* **2015**, *115*, 6312–6356.
- (64) Luukkonen, S.; Levesque, M.; Belloni, L.; Borgis, D. Hydration free energies and solvation structures with molecular density functional theory in the hypernetted chain approximation. *J. Chem. Phys.* **2020**, *152*, No. 064110.
- (65) Zhang, B. W.; Cui, D.; Matubayasi, N.; Levy, R. M. The Excess Chemical Potential of Water at the Interface with a Protein from End Point Simulations. *J. Phys. Chem. B* **2018**, *122*, 4700–4707.