# Relationship between Solvation Thermodynamics from IST and DFT Perspectives

Ronald M. Levy,*,[†] Di Cui,[†] Bin W. Zhang,[†] and Nobuyuki Matubayasi*,[‡]

[†]Center for Biophysics and Computational Biology, Department of Chemistry, and Institute for Computational Molecular Science, Temple University, Philadelphia, Pennsylvania 19122, United States

[‡]Division of Chemical Engineering, Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka 560-8531, Japan and Elements Strategy Initiative for Catalysts and Batteries, Kyoto University, Katsura, Kyoto 615-8520, Japan

ABSTRACT: Inhomogeneous solvation theory (IST) and classical density functional theory (DFT) each provide a framework for relating distribution functions of solutions to their thermodynamic properties. As reviewed in this work, both IST and DFT can be formulated in a way that use two "end point" simulations, one of the pure solvent and the other of the solution, to determine the solute chemical potential and other thermodynamic properties of the solution and of subvolumes in regions local to the solute containing hydrating waters. In contrast to IST, where expressions for the excess energy and entropy of solution are the object of analysis, in the DFT end point formulation of the problem, the solute−solvent potential of mean force (PMF) plays a central role. The indirect part of the PMF corresponds to the lowest order (1-body) truncation of the IST expression. Because the PMF is a free energy function, powerful numerical methods can be used to estimate it. We show that the DFT expressions for the solute excess chemical potential can be written in a form which is local, involving integrals only over regions proximate to the solute. The DFT end point route to estimating solvation free energies provides an alternative path to that of IST for analyzing solvation effects on molecular recognition and conformational changes in solution, which can lead to new insights. In order to illustrate the kind of information that is contained in the solute−solvent PMF, we have carried out simulations of β-cyclodextrin in water. This solute is a well studied "host" molecule to which "guest" molecules bind; host−guest systems serve as models for molecular recognition. We illustrate the range of values the direct and indirect parts of the solute−solvent PMF can have as a water molecule is brought to the interface of β-cyclodextrin from the bulk; we discuss the "competition" between these two terms, and the role it plays in molecular recognition.
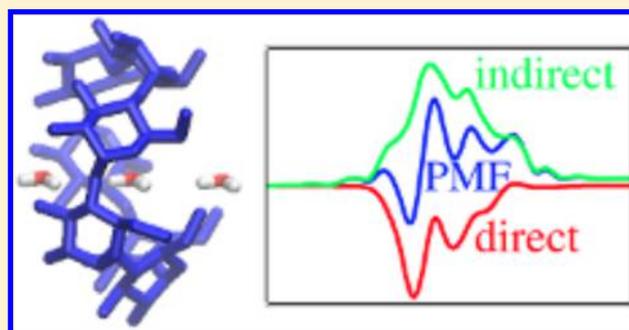
INTRODUCTION

Statistical thermodynamic theories of solutions have a storied history. One line of research, now called inhomogeneous solvation theory (IST) by some researchers in the field,[1−3] can be traced back to the cluster expansion techniques introduced by Mayer and Montroll[4] to classify and enumerate molecular distribution functions. In IST, the entropy change upon insertion of a solute into solution can be expressed as a series expansion in terms of the solvent multipoint correlation functions.[3,5] The solution thermodynamic process can be described in an alternate way as the turning on of an external potential that corresponds to the solute interacting with the solvent. This description of the solution process, together with the concept of the solvent excess chemical potential, provides the starting point for a classical density functional theory (DFT) description of solution thermodynamics.[6−8] In both IST and DFT theories, the solute excess chemical potential can be expressed as an integral over the volume of the solution, and the corresponding integrands therefore have the units of a free energy density. However, as we show later in the Correspondence between IST and DFT Section, the local

IST and DFT free energy density expressions (the integrands) are not in general equivalent even though both theories provide formulas for the excess chemical potential of the solute subject to the approximations needed to evaluate the integrals.

In the 1990s we wrote a series of three papers[9−11] about the statistical thermodynamic basis for hydration shell models of solvation; we investigated the theoretical basis for assigning the thermodynamic changes associated with inserting a solute into solution to a locally perturbed solvent region in the vicinity of the solute. Our viewpoint was that of IST; wherein the excess chemical potential of the solute is written as the sum of the excess energy ΔE and the excess entropy contribution $T\Delta S$ to the free energy change. A central result of our analysis was that although the individual terms ΔE and $T\Delta S$ are nonlocal and depend on the ensemble, it is possible to isolate ensemble

dependent contributions to ΔE and $T\Delta S$ which cancel when combined.

Taking the IST perspective, Lazaridis developed a set of practical formulas that could be used to evaluate the local excess solvent reorganization energy and entropy of water at a solute interface[1,2] and this framework became the basis for analyzing the thermodynamic properties of water at bio-molecular binding interfaces.[12−31] The strategy, as laid out by Friesner, Berne, and co-workers is to exploit the thermodynamic signatures of waters at protein receptors to identify waters which may be more or less readily displaced by ligands, and thereby inform the process of designing more potent ligands with higher binding affinity. There have been notable successes,[14−16] but it has proved difficult to apply the analysis to more complicated binding environments that are frequently encountered. The IST framework is also being further developed through a grid based implementation of the IST equations.[32]

Classical DFT can also be used as a conceptual framework to try to identify solvent patterns at protein receptors that may serve as motifs for molecular recognition. One of us (N.M.) has used DFT to construct a theory of solvation thermodynamics for which the approximate density functionals are constructed from the solute−solvent energy distributions computed from end point molecular dynamics simulations;[33−36] similar in some sense to the observation that data derived from end point simulations is used as the basis of IST formulas. The way the end point data is manipulated however is very different in the two approaches. In contrast to IST, where local approximate expressions for the excess energy and excess entropy are the object of analysis, in the DFT end point formulation of the problem, the solute−solvent potential of mean force (PMF) $WT(\mathbf{x})$ to move a solvent molecule from the bulk to the interface plays a central role, where $\mathbf{x}$ denotes the configuration (position and orientation) of the solvent molecule with respect to the solute. By analyzing the solute−solvent PMF $WT(\mathbf{x})$ and decomposing it into direct $u(\mathbf{x})$ and indirect parts $\omega(\mathbf{x})$, further insights can be obtained about the role of water at protein−ligand binding sites beyond those afforded by the IST perspective.

In this article we review and contrast the IST and DFT frameworks for analyzing the structure and thermodynamics of water at a solute interface, and highlight the possibilities that the DFT perspective can offer. In order to fully implement a DFT analysis based on the results of end point molecular dynamics simulations of the kind that are currently used for IST analysis, further work will be required to build in the spatial resolution that is needed. We describe what new computational tools will be needed to modify the formulation of the approximate density functional in the energy representation[33,34] currently used to calculate the excess chemical potential of the solute with only the solute−solvent energy distributions from simulations at the end points, in order to provide spatially resolved local expressions for the solvation thermodynamics similar to the kind that have been developed based on IST. In the Comparison of the IST and DFT Framework for Solvation Section we review the IST and DFT formulations of the problem, compare them, and describe what is needed in order to implement a DFT-based analysis of solvation thermodynamics which retains the spatial resolution inherent in IST. In the Model PDT Calculations for the Solvent Excess Chemical Potential at a Solute Interface Section we present some results for a model system concerning the direct and indirect

contributions to the excess chemical potential of the solvent in solution. Our conclusions are presented in the Conclusions and Future Directions Section.

■ COMPARISON OF THE IST AND DFT FRAMEWORK FOR SOLVATION THERMODYNAMICS

In this section, we review how the chemical potential of a solute can be expressed as an integral over the whole space of certain correlation functions in the solution containing 1 solute and N solvent molecules. This is possible in both IST and DFT. The integrand can then be viewed as a free-energy density at each space point, but expressions that are to be integrated depend upon the theoretical framework within which the solute chemical potential is to be computed. In the IST Expression for the Excess Chemical Potential of a Solute in Solution Section we first provide a free-energy expression within IST, and in the DFT Expression for the Excess Chemical Potential of a Solute in Solution Section, we discuss the DFT expression. Their connection will be discussed in the Correspondence between IST and DFT Section. In the equations that follow, we use the Ben-Naim standard state wherein a solute is inserted at a fixed origin with a fixed orientation, and denote the free energy to insert the solute into the liquid, the solute excess chemical potential, by $\Delta\mu$. It is the free-energy change for turning on the solute−solvent interaction while keeping the solute position and orientation fixed, and it corresponds to the standard change in the free energy when the solubility is expressed in units of molar concentration (Ostwald coefficient in the case of gas dissolution).

IST Expression for the Excess Chemical Potential of a Solute in Solution. One route to the free-energy change $\Delta\mu$ for inserting a solute into solution is based on inhomogeneous solvation theory. It starts by writing $\Delta\mu$ as a sum of two/three terms in the canonical/isobaric ensemble; one corresponds to the excess energy, the other to the excess entropy of solution; in the isobaric ensemble there is an additional $P\Delta V$ term.

$$\Delta\mu = \Delta E_{\text{const}-P} - T\Delta S_{\text{const}-P} + P\Delta V$$
$$= \Delta E_{\text{const}-V} - T\Delta S_{\text{const}-V} \tag{1}$$

The equation above emphasizes the fact that the excess chemical potential is independent of the ensemble, although the individual terms do depend on the ensemble. In our 1998 paper we analyzed the ensemble dependence of the excess energy and entropy and showed that the ensemble dependence is entirely contained in the nonlocal components of these thermodynamic quantities;[11] the nonlocal component refers to the integration over the whole space of the $O(1/V)$ deviation of the correlation functions at distances far from the solute from those in pure solvent, where V is the system volume. We use the notation $O(1/V)$ to indicate a quantity whose volume integral does not vanish in the thermodynamic limit, while $o(1/V)$ is used to indicate a quantity whose volume integral goes to zero in the thermodynamic limit.

The next step is to write the excess energy and entropy in terms of correlation functions of the initial and final states of the process of solute insertion, so that $\Delta\mu$ can be expressed as an end point formula, with terms that can be evaluated by two simulations; one of the pure solvent and the other of the solute in solution. In the following developments, the solute is taken as fixed at the origin with fixed orientation. When the intermolecular interaction is pairwise additive, the excess energy ΔE is then expressed as

$$e(\mathbf{x}) = U_{u\bar{v}}(\mathbf{x}) + \frac{1}{2}\int d\mathbf{x}_1 \frac{\rho^{(2)}(\mathbf{x}, \mathbf{x}_1)}{\rho(\mathbf{x})} U_{w\bar{v}}(\mathbf{x}, \mathbf{x}_1) \tag{2}$$

$$\Delta E = \int d\mathbf{x}\rho(\mathbf{x})[e(\mathbf{x}) - e_0] \tag{3}$$

where $U_{u\bar{v}}$ and $U_{w\bar{v}}$ are the solute−solvent (water) and solvent−solvent (water−water) interaction potentials of a tagged solvent molecule, respectively, $\mathbf{x}$ refers to the position and orientation of the tagged solvent molecule relative to the solute, ρ is the 1-body distribution of solvent molecules with respect to the solute fixed at the origin (it is a 2-body distribution, but it is called 1-body since the solute is fixed), $\rho^{(2)}$ is the 2-body distribution of solvent with respect to the fixed solute (i.e., it is the solute−solvent−solvent distribution), $e(\mathbf{x})$ is the average interaction energy of a tagged solvent molecule located at $\mathbf{x}$ and e_0 is its value in pure solvent. Let $e(\infty)$ be the value of $e(\mathbf{x})$ when the solvent is separated far from the solute. In the canonical ensemble, $e(\infty)$ deviates by $O(1/V)$ from e_0 and this $O(1/V)$ deviation is integrated over the whole space to provide an observable, $O(1)$ effect of the ensemble dependence of ΔE. In the canonical ensemble the excess energy can be separated into local and nonlocal terms. The local and nonlocal terms are listed in eqs 4 and 5, respectively.

$$\Delta E_{\text{local}} = \int d\mathbf{x}\rho(\mathbf{x})[e(\mathbf{x}) - e(\infty)] \tag{4}$$

$$\Delta E_{\text{nonlocal}} = \int d\mathbf{x}\rho(\mathbf{x})[e(\infty) - e_0] \rightarrow \Delta V(T\alpha/\kappa - P) \tag{5}$$

In eq 5, ΔV is the excess partial molar volume of the solution in the Ben Naim convention with the ideal term ($\kappa k_{\text{B}}T$) removed, α is the isobaric thermal expansion coefficient, and κ is the thermal compressibility, see ref 9 for further details. We emphasize that eq 5 is the nonideal part of the nonlocal contribution to the excess energy, as the ideal contribution is excluded from the outset when using the Ben Naim standard state. The difference between $e(\infty)$ and e_0 is of $o(1/V)$ in the NPT ensemble, however, meaning that it decreases faster than $1/V$ in the limit of large systems size. The nonlocal component is absent in the NPT ensemble. It should be noted in eq 3 that the limit of large system size (thermodynamic limit) should be taken only after the integration is done over the whole space. The thermodynamic limit and integration process are not interchangeable in general, they are so only for the local component. Thus, the local component is introduced to identify that part of the thermodynamic quantity for which the thermodynamic limit can be taken freely without regard to whether or not the spatial integration is carried out before taking the thermodynamic limit.

The corresponding expression for the excess entropy ΔS is based on the Green-Wallace expansion.[3] In this expansion, the entropy is expressed in terms of the correlation functions $\delta g^{(i)}$ which are defined by a hierarchy of distributions functions:

$$g^{(1)}(\mathbf{x}) = \frac{\rho(\mathbf{x})}{\rho_0} \equiv g_{\text{sw}}(\mathbf{x})$$

$$g^{(i)}(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_i) = \frac{\rho^{(i)}(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_i)}{\rho(\mathbf{x}_1)\rho(\mathbf{x}_2)\cdots\rho(\mathbf{x}_i)}$$

$$\delta g^{(2)}(\mathbf{x}_1, \mathbf{x}_2) = g^{(2)}(\mathbf{x}_1, \mathbf{x}_2) \equiv g_{\text{sww}}(\mathbf{x}_1, \mathbf{x}_2)$$

$$\delta g^{(3)}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = \frac{g^{(3)}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)}{g^{(2)}(\mathbf{x}_1, \mathbf{x}_2)g^{(2)}(\mathbf{x}_2, \mathbf{x}_3)g^{(2)}(\mathbf{x}_3, \mathbf{x}_1)}$$

$$\cdots \tag{6}$$

where ρ_0 is the density in pure solvent. Following our previous work[11] the entropy density of the solution can be expressed in terms of:

$$s(\mathbf{x}) = -k_{\text{B}}\sum_{i=1}^{\infty}\frac{1}{i!}\int d\mathbf{x}_1\cdots d\mathbf{x}_{i-1}\frac{\rho^{(i)}(\mathbf{x}, \mathbf{x}_1, \cdots, \mathbf{x}_{i-1})}{\rho(\mathbf{x})}\ln \delta g^{(i)}(\mathbf{x}, \mathbf{x}_1, \cdots, \mathbf{x}_{i-1}) \tag{7}$$

An expression can also be obtained for the entropy density of the pure solvent, where s_0 corresponds to $s_0(\mathbf{x})$ and is independent of $\mathbf{x}$ due to the homogeneity of the pure solvent. The excess entropy of the solution is

$$\Delta S = \int d\mathbf{x}\rho(\mathbf{x})[s(\mathbf{x}) - s_0] \tag{8}$$

Several groups[12−16,29,32,37−40] have used the inhomogeneous solvation theory equations summarized in this section in an attempt to identify those water molecules at a protein receptor site that are most advantageous to replace by a ligand. The terms that are typically evaluated in practice are the 1- and 2-body contributions to the excess energy (eqs 2 and 3 above), and the 1-body contribution to the excess entropy. This term can be written as

$$\Delta S_{\text{sw}}^{(1)} = -k_{\text{B}}\rho_0\int d\mathbf{x}g_{\text{sw}}(\mathbf{x})\ln g_{\text{sw}}(\mathbf{x}) \tag{9}$$

where g_{sw} is the solute−solvent pair correlation function. As in the discussion of the excess energy, this expression also has a nonlocal component in the canonical ensemble since g_{sw} deviates from 1 by order $1/V$ at large distances from the solute. As shown in ref 11, far from the solute g_{sw} is equal to $V/(V - \Delta V)$, where ΔV is the excess volume. Eq 9 can then be written as

$$\Delta S_{\text{sw}}^{(1)} = -k_{\text{B}}\rho_0\int d\mathbf{x}g_{\text{sw}}(\mathbf{x})\ln\frac{g_{\text{sw}}(\mathbf{x})}{g_{\text{sw}}(\infty)} - k_{\text{B}}\rho_0\Delta V \tag{10}$$

in the canonical ensemble. The first term is local and the limit of large system size (thermodynamic limit) can then be taken before the integration over the whole volume, so that $g_{\text{sw}}(\infty)$ can be set to unity and a cutoff can be introduced when the integral is evaluated. Eq 9 then becomes the commonly known expression in the thermodynamic limit given by

$$\Delta S_{\text{sw}}^{(1)} = -k_{\text{B}}\rho_0\int d\mathbf{x}g_{\text{sw}}^{\text{TL}}(\mathbf{x})\ln g_{\text{sw}}^{\text{TL}}(\mathbf{x})$$
$$+ k_{\text{B}}\rho_0\int d\mathbf{x}(g_{\text{sw}}^{\text{TL}}(\mathbf{x}) - 1) \tag{11}$$

where $g_{\text{sw}}^{\text{TL}}$ denotes the g_{sw} in the thermodynamic limit. See Appendix A for a detailed derivation of eqs 9−11. In ref 9, eq 7 and eq 9 were provided without separation into the local and nonlocal components. As discussed for ΔE, the correlation

C

Table 1. Local and Nonlocal Components in the Canonical Ensemble with the Local Estimator[a]

quantity	local estimator	local component	nonlocal component
excess energy	$e(\mathbf{x})$ (eq 2)	$\int d\mathbf{x}\rho(\mathbf{x})[e(\mathbf{x}) - e(\infty)]$	$\Delta V(T\alpha/\kappa - P)$
excess entropy	$s(\mathbf{x})$ (eq 7)	$\int d\mathbf{x}\rho(\mathbf{x})[s(\mathbf{x}) - s(\infty)]$	$\dfrac{\alpha}{\kappa}\Delta V$
excess entropy (1-body)	$-k_B \ln g_{uw}(\mathbf{x})$	$-k_B\rho_0 \int d\mathbf{x} g_{sw}(\mathbf{x})\ln \dfrac{g_{sw}(\mathbf{x})}{g_{sw}(\infty)}$	$-k_B\rho_0\Delta V$
excess chemical potential	$-k_B T\rho(\mathbf{x}) - \omega(\mathbf{x}) + \dfrac{1}{\rho(\mathbf{x})}\int_0^1 d\lambda\omega_\lambda(\mathbf{x})\dfrac{\partial\rho_\lambda(\mathbf{x})}{\partial\lambda}$	$-k_B T\int d\mathbf{x}(\rho(\mathbf{x}) - \rho(\infty)) - \int d\mathbf{x}\rho(\mathbf{x})(\omega(\mathbf{x}) - \omega(\infty))$ $+ \int_0^1 d\lambda \int d\mathbf{x}\omega_\lambda(\mathbf{x})\dfrac{\partial\rho_\lambda(\mathbf{x})}{\partial\lambda}$	none

[a]The nonlocal components are absent in the *NPT* ensemble using the local estimators listed in the table.

functions evaluated in the thermodynamic limit can be employed only when the limit is interchangeable with the integration over the whole space. Although eq 9 is correct if the integration is to be done before the thermodynamic limit, eq 10 is more useful since the order of limit and integration does not matter there. It should be emphasized again that g_{sw} appearing in eq 9 is the one before the thermodynamic limit, while g_{sw}^{TL} in eq 11 is the one at the thermodynamic limit which approaches 1 far from the solute, so that a cutoff can be introduced when the integral is evaluated. Eq 11 is the one used to evaluate the 1-body IST entropy.

The 2-body contribution to the excess entropy, which is usually ignored but has recently been considered,[41−43] can be expressed as

$$\Delta S_{sww}^{(2)} = -\frac{1}{2}k_B(\rho_0)^2 \int d\mathbf{x}_1 d\mathbf{x}_2[g_{sw}(\mathbf{x}_1)g_{sw}(\mathbf{x}_2)]$$
$$[g_{sww}(\mathbf{x}_1, \mathbf{x}_2)\ln g_{sww}(\mathbf{x}_1, \mathbf{x}_2)] \quad (12)$$

We note that while the nonlocal contributions to the excess entropy in the canonical ensemble cancel the corresponding contributions to the excess energy when the integration of the correlation functions is taken over the entire volume, when the entropy expansion is truncated after the 1-body or 2-body terms, the cancellation of nonlocal effects between the energy and entropy terms present in the canonical ensemble simulations is no longer assured. The summary of the local and nonlocal terms is given in Table 1 for the excess energy and entropy.

By combining eq 3 for the excess energy of solution written as an integral over the interaction energy density times the solvent density, with eq 8 for the excess entropy also written as an integral over the solvent density, we have an expression for the excess free energy of the solute in solution written as an integral over the solvent density:

$$\Delta\mu = \int d\mathbf{x}\rho(\mathbf{x})[(e(\mathbf{x}) - Ts(\mathbf{x})) - (e_0 - Ts_0)] \quad (13)$$

This expression is valid in the canonical ensemble when the integrals are carried out over the entire simulation box (nonlocal integrals), or they can be treated as local integrals by introducing a cutoff as they usually are in IST calculations. If so, a $P\Delta V$ term must be added when comparing IST calculations with FEP simulations of the solute chemical potential, (although the $P\Delta V$ is usually small). The $P\Delta V$ term appears because this is the part of the nonlocal contribution to the excess energy in the canonical ensemble which is not canceled by the nonlocal part of the excess entropy. The integrand of eq 13 can formally be considered a free energy

density, but the term in [] brackets in eq 13 does not correspond to the work required to move a solvent molecule from the bulk to the location at $\mathbf{x}$, or pure liquid. Further consideration of this point and the distinction between the bulk and pure liquid is provided in the DFT Expression for the Excess Chemical Potential of a Solute in Solution Section and Correspondence between IST and DFT Section.

DFT Expression for the Excess Chemical Potential of a Solute in Solution. Another route to the excess chemical potential of the solute is provided by classical density-functional theory (DFT). Its formulation is closely connected to Percus' idea of viewing the solute as a source of an external field for the solvent.[8] DFT is particularly useful in treating inhomogeneous systems and has been recently applied to molecular fluids, especially water.[6,44−47] In DFT, the chemical potential of the solute is expressed as an (approximate) functional of the solvent density around the solute. A self-consistent set of integral equations is then obtained by combining it with the variational principle[48−52] and a closure relation which introduces an approximation for the direct correlation function. It is also possible to derive a form of classical DFT which uses the input from molecular dynamics simulations at the two end points of the solute insertion process to estimate the solute insertion free energy, as one of us has done.[33−35]

In this subsection, we review a derivation of the DFT equations related to the potential distribution theorem (PDT). The derivation follows closely that in ref 34. We introduce the density-functional through partial integration of the Kirkwood charging formula and emphasize the role of the solvent-mediated (indirect) part of the solute−solvent potential of mean force (PMF).

Let the solute−solvent interaction $u_\lambda(\mathbf{x})$ be gradually turned on with the coupling parameter λ ($0 \leq \lambda \leq 1$), where $\mathbf{x}$ is the solvent coordinate relative to the solute. When $\lambda = 0$, the solute−solvent interaction is absent and $u_\lambda(\mathbf{x}) = 0$. When $\lambda = 1$, the solute interacts with the solvent at full coupling and $u_\lambda(\mathbf{x}) = u(\mathbf{x})$, where $u(\mathbf{x})$ is the solute−solvent interaction potential of interest. (Note: $u(\mathbf{x})$ is equivalent to $U_{uv}(\mathbf{x})$ in the IST expression.) When $\rho_\lambda(\mathbf{x})$ denotes the (1-body) solvent distribution around the solute at the coupling parameter λ, the Kirkwood charging formula provides the excess chemical potential $\Delta\mu$ through

$$\Delta\mu = \int_0^1 d\lambda \int d\mathbf{x}\frac{\partial u_\lambda(\mathbf{x})}{\partial\lambda}\rho_\lambda(\mathbf{x}) \quad (14)$$

This is then integrated partially to obtain

D

$$\Delta\mu = \int d\mathbf{x}\, u(\mathbf{x})\rho(\mathbf{x}) - \int_0^1 d\lambda \int d\mathbf{x} \frac{\partial \rho_\lambda(\mathbf{x})}{\partial \lambda} u_\lambda(\mathbf{x}) \quad (15)$$

where $\rho(\mathbf{x})$ stands for $\rho_\lambda(\mathbf{x})$ at $\lambda = 1$ and is the solvent density in the solution system of interest. The second term of eq 15 is approximated by a free energy density functional (see Appendix D).

In the IST Expression for the Excess Chemical Potential of a Solute in Solution Section, the IST expressions were separated into a term that depends explicitly on the density weighted solute−solvent interaction potential and correlation (1-body energy and 1-body entropy) and remaining terms which depend on solvent−solvent interactions and correlations. The DFT expression (eq 15) can also be separated into the solute−solvent terms (1-body energy and 1- body entropy) plus the solvent−solvent term. To do so, we define the solvent-mediated (indirect) part of the solute−solvent potential of mean force (PMF) $\omega_\lambda(\mathbf{x})$ by

$$\rho_\lambda(\mathbf{x}) = \exp\{-\beta[u_\lambda(\mathbf{x}) + \omega_\lambda(\mathbf{x})]\}\rho_0(\mathbf{x}) \quad (16)$$

where $\rho_0(\mathbf{x})$ denotes $\rho_\lambda(\mathbf{x})$ at $\lambda = 0$. When the system is homogeneous, $\rho_0(\mathbf{x})$ is independent of $\mathbf{x}$ and is equal to the density in pure solvent. Our formulation is valid also when the system is inhomogeneous and ρ_0 depends on $\mathbf{x}$; we thus leave the argument $\mathbf{x}$ of ρ_0 for the sake of generality. It should be further noted that even when the pure solvent is homogeneous, ρ_λ ($\lambda \neq 0$) at distances far from the solute is different from ρ_0 by $O(1/N)$ in the canonical ensemble, where N is the system size. In other words, the density in the bulk (the region far beyond the solute−solvent correlation length) differs from that in pure solvent by $O(1/N)$ in the canonical ensemble. In the grand canonical and NPT ensembles, on the other hand, the difference between ρ_λ ($\lambda \neq 0$) in the bulk and ρ_0 in the pure solvent is of $o(1/N)$ and it is not necessary to be taken into account when the integration is taken over the whole space. Thus, $\omega_\lambda(\mathbf{x})$ depends on the ensemble by $O(1/N)$, while the following developments are valid in any ensemble even when the system is of finite size and the thermodynamic limit is not taken. Accordingly, when the value of $\omega_\lambda(\mathbf{x})$ in the bulk is written as $\omega_\lambda(\infty)$, it depends on the ensemble and its values are discussed below with respect to eq 20. In any case, $\omega_\lambda(\mathbf{x})$-$\omega_\lambda(\infty)$ is that part of the free-energy change for bringing a tagged solvent molecule from the bulk, far from the solute, to the specified configuration $\mathbf{x}$ after the direct interaction with the solute is subtracted. The tagged solvent molecule interacts only with the other solvent molecules upon being dragged to $\mathbf{x}$. Note that eq 16 is equivalent to the potential distribution theorem, which is also commonly called "Widom particle insertion".[53−55] A key for the following developments is the separation of the direct solute−solvent interaction and the solvent-mediated term within the exponential.

Eq 15 can be modified, by virtue of eq 16, resulting in the following expression:

$$\Delta\mu = \int d\mathbf{x}\, u(\mathbf{x})\rho(\mathbf{x}) - k_{\mathrm{B}}T \int d\mathbf{x}\Bigg[(\rho(\mathbf{x}) - \rho_0(\mathbf{x}))$$
$$- \rho(\mathbf{x})\log\left(\frac{\rho(\mathbf{x})}{\rho_0(\mathbf{x})}\right)\Bigg] + \int_0^1 d\lambda \int d\mathbf{x}\,\omega_\lambda(\mathbf{x}) \frac{\partial \rho_\lambda(\mathbf{x})}{\partial \lambda} \quad (17)$$

See Appendix B for the derivation. This is an exact expression. The first term is the 1-body energy corresponding to the

interaction of the solute with the solvent. The second term corresponds to the one-body entropy. It is always non-negative and is a measure of the "deviation" of the structure of the solvent in the solution containing the solute from that of the pure solvent system. The third term expresses (the change in) the solvent−solvent correlations upon gradual insertion of the solute with the coupling parameter λ. The solvent reorganization energy or two body energy (i.e the change in the solvent−solvent interaction energy due to the insertion of the solute) is part of the third term of eq 17. The sum of the second plus third terms corresponds to the solvent reorganization free energy. The free energy to insert the solute into solution is equal to the reversible work required to reorganize the solvent from the pure solvent structure to the solution structure (the sum of the second plus third terms of eq 17) which is always non-negative, plus the interaction energy of the solute with the solvent after the solvent structure has been reorganized. An approximate free energy functional may be introduced to evaluate the third term, by adopting one that is analogous to the Percus−Yevick or hypernetted-chain relationship as described in ref 34 (also see Appendix D).

Eq 17 is exact and has the form of a coupling-parameter integration in the last term. One of us (N.M.) formulated an approximate free energy functional to evaluate eq 17 in the one-dimensional (1-d) energy representation. In this formulation, the distribution functions are introduced over the solute−solvent interaction potential function $u(\mathbf{x})$ and an equation similar to eq 17 is derived for the 1-dimensional coordinate ε which represents the possible values of $u(\mathbf{x})$, and corresponding distribution functions $\rho(\varepsilon)$ and $\rho_0(\varepsilon)$ by analogy with eq 17 (see Appendix D). The Percus−Yevick-type and hypernetted-chain-type approximations are then adopted for the coupling-parameter integration, and $\Delta\mu$ can be computed from distribution functions in the "end point" states ($\lambda = 0$ and 1 in the coupling-parameter integration). This means that a DFT formulation can be used to evaluate the solute excess chemical potential using data extracted from end point simulations, just as IST can be used to evaluate the solute excess chemical potential from end point simulations.

Correspondence between IST and DFT. The solute−solvent 1-body energy and 1- body entropy terms are common between eq 17 and the corresponding IST expressions for $\Delta\mu$ in the IST Expression for the Excess Chemical Potential of a Solute in Solution Section. The change in solvent−solvent interaction energy (the 2-body energy) which is treated explicitly in IST is included in the charging term in DFT. The higher-order entropy terms which are expressed as a series in IST, are included in DFT in the integral over the coupling parameter in eq 17. When the series in IST is treated exactly and the λ integral in eq 17 is evaluated exactly, IST and DFT are in principle equivalent to each other. An approximate treatment needs to be carried out in practice, however, both for IST and for DFT, and physical insights and numerical accuracy determine the usefulness of the approximate IST or DFT expressions.

Eq 17 shows the key role played by the solvent-mediated (indirect) part of the solute−solvent potential of mean force (PMF) $\omega_\lambda(\mathbf{x})$. When the $\lambda = 1$ expression for the solvent density, eq 16, is inserted into eq 17, the solute chemical potential can be expressed in terms of the indirect part of the solute−solvent PMF:[34]

$$\Delta\mu = -k_BT \int d\mathbf{x}(\rho(\mathbf{x}) - \rho_0(\mathbf{x})) + \int d\mathbf{x}\rho(\mathbf{x})(-\omega(\mathbf{x}))$$
$$+ \int_0^1 d\lambda \int d\mathbf{x}\omega_\lambda(\mathbf{x})\frac{\partial\rho_\lambda(\mathbf{x})}{\partial\lambda} \tag{18}$$

where $\omega(\mathbf{x})$ stands for $\omega_\lambda(\mathbf{x})$ at $\lambda = 1$ and is the solvent-mediated part of solute−solvent PMF in the solution when the solute is fully coupled to the solvent. The first term is related to the change in the number of solvent molecules upon insertion of the solute. It is zero in the canonical ensemble but nonzero in the grand canonical ensemble, and the sum of the first and second terms is independent of the ensemble (the condition of solute insertion), as shown below.

The expression for the solute chemical potential, eq 18, has a simple physical interpretation in the canonical ensemble (for which the first term in eq 18 is zero). According to the definition of $\omega(\mathbf{x})$ given by eq 16, $-\omega(\mathbf{x})$ is the indirect (solvent-mediated) part of the "potential of mean force" to transfer a solvent molecule from the position at $\mathbf{x}$ in the solution to the pure liquid. Equivalently, $-\omega(\mathbf{x})$ is the difference between two potential fields acting on the one particle solvent density field: $-\omega(\mathbf{x}) = -[w(\mathbf{x}) - w_0]$ where $w(\mathbf{x})$ is the indirect solvent−solvent contribution to the solvent excess chemical potential at $\mathbf{x}$ in solution, and w_0 is the excess chemical potential of the solvent in the pure liquid. eq 18 can then be written in the canonical ensemble as

$$\Delta\mu = -\int d\mathbf{x}\rho(\mathbf{x})[w(\mathbf{x}) - w_0] + \int_0^1 d\lambda \int d\mathbf{x}\omega_\lambda(\mathbf{x})\frac{\partial\rho_\lambda(\mathbf{x})}{\partial\lambda} \tag{19}$$

Therefore, we see that in the DFT view of the solvation process, the effect of turning on an external potential which corresponds to the solute interacting with the solvent, is to change an effective one-body potential field acting on the one-body density field $\rho(\mathbf{x})$ from $-w_0$ to $-w(\mathbf{x})$. The first term accounts for the difference in the free energy between the initial and final potential fields acting on the one particle solvent density field. The second term in eq 18 accounts for the reversible work required to change the effective potential field from its initial value $-w_0$ to final value $-w(\mathbf{x})$. It is interesting to note that in this formulation of the DFT expression for the solvation process, the direct solute−solvent interaction does not appear explicitly.

We now show that the DFT expression for the solute excess chemical potential is independent of the ensemble (canonical vs grand canonical). As noted with respect to eq 16, $\omega(\infty)$ depends on the ensemble, it corresponds to the value of $\omega(\mathbf{x})$ in the bulk (the region far beyond the correlation length between the solute and solvent). To see the ensemble dependence of the individual terms of eq 18, we exploit this property and rewrite it as

$$\Delta\mu = -k_BT \int d\mathbf{x}(\rho(\mathbf{x}) - \rho_0(\mathbf{x})) - N\omega(\infty)$$
$$- \int d\mathbf{x}\rho(\mathbf{x})(\omega(\mathbf{x}) - \omega(\infty))$$
$$+ \int_0^1 d\lambda \int d\mathbf{x}\omega_\lambda(\mathbf{x})\frac{\partial\rho_\lambda(\mathbf{x})}{d\lambda}$$
$$= -k_BT \int d\mathbf{x}(\rho(\mathbf{x}) - \rho(\infty))$$
$$- \int d\mathbf{x}\rho(\mathbf{x})(\omega(\mathbf{x}) - \omega(\infty))$$
$$+ \int_0^1 d\lambda \int d\mathbf{x}\omega_\lambda(\mathbf{x})\frac{\partial\rho_\lambda(\mathbf{x})}{\partial\lambda} \tag{20}$$

In the canonical ensemble, $\omega(\infty)$ is equal to $-k_BT\Delta V/V$ since $\rho(\infty)$ in the bulk is $N/(V - \Delta V)$ and ρ_0 is N/V, while in the grand canonical and NPT ensembles, $\omega(\infty)$ vanishes as $o(1/N)$ since the difference between $\rho(\infty)$ and ρ_0 is $o(1/N)$. The sum of the first and second terms in the first line of eq 20 thus reduces to $k_BT\rho_0\Delta V$ because the first term is zero in the canonical ensemble. Given the value of $\rho(\infty)$ described above, in turn, the first and second terms in the first line is equal to the first term in the second line. Thus, although the first and second terms in the first line are dependent on the ensemble when treated separately, their sum is independent of the ensemble. The second line of eq 20 is a local expression in the sense that the integrand vanishes as $o(1/N)$ when the solvent is far separated from the solute. As shown in ref 11, a local expression is independent of the ensemble and the thermodynamic limit for the correlations functions and the integration over the whole space can be interchanged. The solvation free energy $\Delta\mu$ should be independent of the ensemble, and the second line of eq 20 shows that $\Delta\mu$ is indeed so in the DFT framework (see Table 1).

The second term of eq 20 is a sum of $-(\omega(\mathbf{x})\text{-}\omega(\infty))$ with a weight of $\rho(\mathbf{x})$. According to the definition of $\omega(\mathbf{x})$, $-(\omega(\mathbf{x})\text{-}\omega(\infty))$ is the indirect solvent mediated part of the free-energy change of displacing a tagged molecule from the configuration $\mathbf{x}$ to the bulk region; it is emphasized that the tagged solvent molecule interacts only with the other solvent molecules and not with the solute. The second term of eq 20 shows that when a tagged solvent molecule at $\mathbf{x}$ interacts more favorably with the other solvent molecules than in the bulk ($\omega(\mathbf{x}) < \omega(\infty)$), it contributes unfavorably to the excess chemical potential $\Delta\mu$ and conversely when the tagged molecule is subject to an unfavorable interaction with the other solvent molecules ($\omega(\mathbf{x}) > \omega(\infty)$), it makes a favorable contribution to $\Delta\mu$.

The central role of $\omega(\mathbf{x})$ played in eq 18 and eq 20 provides a contrast between IST and DFT. As $\omega(\mathbf{x})$ is a free energy, a wealth of numerical methods may be used to estimate it, including FEP charging methods, end point methods which employ WHAM, or by the use of liquid state theory, such as the HNC or PY approximations. The IST equations however are formulated in terms of the excess energy and entropy of solvation separately. As shown next, the integrand $[e(\mathbf{x})\text{-}Ts(\mathbf{x}) - (e_0 - Ts_0)]$ in eq 13 is not the free-energy change for a process of displacing a solvent molecule from the bulk to $\mathbf{x}$ and thus cannot be evaluated with a free-energy method that can be used for such a process.

Finally in this section, we comment on the relationship between the IST free energy density and the free energy (PMF) to displace a tagged solvent molecule from the bulk to the solute interface. It is sometimes stated that the integrand in the

IST expression eq 13 (the term in brackets) corresponds to the free energy to transfer a solvent molecule from the bulk to the interface at $\mathbf{x}$ (or pure liquid). To see that this is not the case when the ensemble is canonical, we add and subtract the term $\rho(\mathbf{x})\,u(\mathbf{x})$ to $\omega(\mathbf{x})$ in eq 18:

$$\Delta\mu^{\mathrm{DFT}} = -\int d\mathbf{x}\rho(\mathbf{x})\omega(\mathbf{x}) + \int_0^1 d\lambda \int d\mathbf{x}\omega_\lambda(\mathbf{x})\frac{\partial\rho_\lambda(\mathbf{x})}{\partial\lambda}$$

$$= -\int d\mathbf{x}\rho(\mathbf{x})WT(\mathbf{x}) + \left[\int d\mathbf{x}\rho(\mathbf{x})u(\mathbf{x}) + \int_0^1 d\lambda \int d\mathbf{x}\omega_\lambda(\mathbf{x})\frac{\partial\rho_\lambda(\mathbf{x})}{\partial\lambda}\right]$$

$$\Delta\mu^{\mathrm{IST}} = \int d\mathbf{x}\rho(\mathbf{x})[(e(\mathbf{x}) - Ts(\mathbf{x})) - (e_0 - Ts_0)] \tag{21}$$

where $WT(\mathbf{x})$ is the total PMF between the solute and solvent and is the sum of $u(\mathbf{x})$ and $\omega(\mathbf{x})$. By comparing the DFT and IST expressions for the solute excess chemical potential, it is apparent that the IST free energy density is not equal to the free energy to transfer a solvent molecule from the bulk (or pure liquid) to the interface at $\mathbf{x}$:

$$WT(\mathbf{x}) \neq [(e(\mathbf{x}) - Ts(\mathbf{x})) - (e_0 - Ts_0)] \tag{22}$$

Instead, it should be noted that

$$-WT(\mathbf{x}) = k_{\mathrm{B}}T\log\left(\frac{\rho(\mathbf{x})}{\rho_0}\right) = -T(s^{(1)}(\mathbf{x}) - s_0^{(1)}) \tag{23}$$

holds, where $s^{(1)}$ is the one-body term (the first term) of the space-resolved entropy of eqs 6 and 7. This lowest-order term in the entropy component of the IST expression actually corresponds to the free energy (PMF) for displacing a solvent molecule from the interface at $\mathbf{x}$ to the bulk. And with addition of the one-body term of the space-resolved energy of eq 2, we have

$$-\omega(\mathbf{x}) = -WT(\mathbf{x}) + u(\mathbf{x})$$
$$= (e^{(1)}(\mathbf{x}) - Ts^{(1)}(\mathbf{x})) - (e_0^{(1)} - Ts_0^{(1)}) \tag{24}$$

where $e_0^{(1)}$ is the solute–solvent energy in the bulk and is actually zero. Accordingly, the term:

$$-\int d\mathbf{x}\rho(\mathbf{x})\omega(\mathbf{x}) \tag{25}$$

in the DFT expression is the lowest-order (1-body) truncation of the IST expression. The IST route to the excess chemical potential is often implemented with the 1-body energy, 1-body entropy, and 2-body energy terms. This is an approximate procedure, and given that the DFT and IST expressions are both exact when the charging integral is performed in the former and the entropy is evaluated to full order in the later, the common approximation in IST corresponds to equating the charging integral in DFT to the 2-body energy.

Orientational Averaging in IST and Dimensionality Reduction in DFT. The IST and DFT equations for the solute chemical potential presented in the preceding sections, depend on performing integrations over the full six dimensional set of position and orientational coordinates $\mathbf{x}$ of the solvent (water) molecules and corresponding correlation functions.

In order to develop a computationally tractable route to evaluate the excess chemical potential of the solute within either the IST or the DFT framework, it becomes necessary to perform orientational averaging (within IST) or dimensionality reduction (within DFT). In this section we make some comparisons between the orientationally averaged 1-body (energy + entropy) and 2-body (energy) terms in IST and

the 1-body (energy + entropy) terms in DFT. To separately treat the position and orientation, we denote them by $\mathbf{r}$ and $\alpha\beta\gamma$, respectively, in the following; note that $\mathbf{x}$ in preceding sections corresponds to $(\mathbf{r}, \alpha\beta\gamma)$. Based on the analysis of Lazaridis and Karplus,[1,56] Lazaridis separated the local part of the 1 body expression for the IST entropy, eq 9 above, into a pure translational term that depends on $\mathbf{r}$, and an orientationally averaged entropic term which is conditional on $\mathbf{r}$:

$$\langle\Delta S_{1-\mathrm{body}}^{\mathrm{IST}}\rangle_{\alpha\beta\gamma} = \Delta S^{\mathrm{IST}}(\mathbf{r})$$
$$= -k_{\mathrm{B}}\int P(\alpha\beta\gamma|\mathbf{r})\ln P(\alpha\beta\gamma|\mathbf{r})d\alpha d\beta d\gamma \tag{26}$$

where $P(\alpha\beta\gamma|\mathbf{r})$ is the probability of finding the orientation of a solvent molecule at $\alpha\beta\gamma$ given its position is $\mathbf{r}$.

$$\Delta S_{1\text{-body}}^{\mathrm{IST}} = \int\rho(\mathbf{r})\left(\Delta S^{\mathrm{IST}}(\mathbf{r}) - k_{\mathrm{B}}\log\left(\frac{\rho(\mathbf{r})}{\rho_0}\right)\right)d\mathbf{r} \tag{27}$$

The 1-body +2-body IST expressions for the excess energy can also be orientationally averaged:

$$\langle\Delta E_{1\text{-body}}^{\mathrm{IST}} + \Delta E_{2\text{-body}}^{\mathrm{IST}}\rangle_{\alpha\beta\gamma,\lambda=1}$$
$$\equiv \langle U_{u\bar{v}}(\mathbf{r}, \alpha\beta\gamma) + U_{v\bar{v}}(\mathbf{r}, \alpha\beta\gamma) - U_{v\bar{v}}^0\rangle_{\alpha\beta\gamma,\lambda=1}$$
$$\equiv \Delta E^{\mathrm{IST}}(\mathbf{r}) \tag{28}$$

We can combine the IST expressions for the orientationally averaged 1-body +2-body excess energy with the solvent oxygen position fixed at $\mathbf{r}$, with the 1-body expression for the excess entropy, and define the quantity $WT^{\mathrm{IST}}(\mathbf{r})$ by the following expression:

$$WT^{\mathrm{IST}}(\mathbf{r}) \equiv \Delta E^{\mathrm{IST}}(\mathbf{r}) - T\Delta S^{\mathrm{IST}}(\mathbf{r}) \tag{29}$$

We emphasize that eq 29 is not an expression for the potential of mean force to move a solvent water molecule from the bulk to the location with oxygen fixed at $\mathbf{r}$; but it can be related to a PDT/DFT expression for the PMF to move a solvent water molecule from the bulk to a position with the oxygen centered at $\mathbf{r}$, and which is otherwise orientationally unrestrained. The PMF to move the water from the bulk to the position $\mathbf{r}$ is

$$WT^{\mathrm{DFT/PDT}}(\mathbf{r}) = -k_{\mathrm{B}}T\ln\left(\frac{1}{8\pi^2}\right)\int g_{u\bar{v}}(\mathbf{r}, \alpha\beta\gamma)d\alpha d\beta d\gamma$$
$$= -k_{\mathrm{B}}T\ln\left(\frac{1}{8\pi^2}\right)\int \exp\{-\beta WT(\mathbf{r}, \alpha\beta\gamma)\}d\alpha d\beta d\gamma$$
$$= -k_{\mathrm{B}}T\ln Z_{\alpha\beta\gamma}(\mathbf{r}) \tag{30}$$

where $WT(\mathbf{r}, \alpha\beta\gamma)$ is the potential of mean force to move a tagged solvent molecule from the bulk to the position $\mathbf{x}= (\mathbf{r},\alpha\beta\gamma)$. We now differentiate eq 30 with respect to the temperature and if we assume that $WT^{\mathrm{DFT/PDT}}(\mathbf{r},\alpha\beta\gamma)$ is independent of the temperature, we can write the effective energy term as

$$U_{\text{eff}}(\mathbf{r}) = k_B T^2 \left(\frac{\partial \ln Z_{\alpha\beta\gamma}(\mathbf{r})}{\partial T} \right)$$

$$= \langle WT^{\text{DFT/PDT}}(\mathbf{r}, \alpha\beta\gamma) \rangle_{\alpha\beta\gamma, \lambda=1}$$

$$- T \left\langle \left(\frac{\partial WT^{\text{DFT/PDT}}(\mathbf{r}, \alpha\beta\gamma)}{\partial T} \right)_{V,N} \right\rangle_{\alpha\beta\gamma, \lambda=1}$$

$$= \langle WT^{\text{DFT/PDT}}(\mathbf{r}, \alpha\beta\gamma) \rangle_{\alpha\beta\gamma, \lambda=1} \qquad (31)$$

This leads to the result:

$$WT^{\text{DFT/PDT}}(\mathbf{r}) = U_{\text{eff}}(\mathbf{r}) - T\Delta S^{\text{DFT/PDT}}(\mathbf{r})$$

$$= \langle WT^{\text{DFT/PDT}}(\mathbf{r}, \alpha\beta\gamma) \rangle_{\alpha\beta\gamma, \lambda=1} - T\Delta S^{\text{DFT/PDT}}(\mathbf{r})$$

$$(32)$$

where eq 32 provides the definition of $\Delta S^{\text{DFT/PDT}}(\mathbf{r})$:

$$\Delta S^{\text{DFT/PDT}}(\mathbf{r}) = \frac{\langle WT^{\text{DFT/PDT}}(\mathbf{r}, \alpha\beta\gamma) \rangle_{\alpha\beta\gamma, \lambda=1}}{T}$$

$$- \frac{WT^{\text{DFT/PDT}}(\mathbf{r})}{T}$$

$$= \frac{1}{T} \int_{\alpha\beta\gamma} P(\alpha\beta\gamma|\mathbf{r}) WT(\mathbf{r}, \alpha\beta\gamma) d\alpha d\beta d\gamma$$

$$+ k_B \ln Z_{\alpha\beta\gamma}(\mathbf{r}) \int_{\alpha\beta\gamma} P(\alpha\beta\gamma|\mathbf{r}) d\alpha d\beta d\gamma$$

$$= -k_B \int_{\alpha\beta\gamma} P(\alpha\beta\gamma|\mathbf{r}) \ln \frac{\exp\{-\beta WT(\mathbf{r}, \alpha\beta\gamma)\}}{Z_{\alpha\beta\gamma}(\mathbf{r})} d\alpha d\beta d\gamma$$

$$= -k_B \int_{\alpha\beta\gamma} P(\alpha\beta\gamma|\mathbf{r}) \ln P(\alpha\beta\gamma|\mathbf{r}) d\alpha d\beta d\gamma \qquad (33)$$

From this analysis it is clear that while orientational averaging of the 1-body IST excess entropy is necessary in order to make the IST calculations tractable, and this term can also be derived from a DFT analysis if the full six dimensional DFT expression for $WT(\mathbf{r},\alpha\beta\gamma)$ is assumed to be independent of temperature; in general the corresponding orientationally averaged DFT and IST energy terms (eq 33) are not equivalent.

$$\Delta S^{\text{DFT/PDT}}(\mathbf{r}) = \Delta S^{\text{IST}}(\mathbf{r})$$

$$\langle WT^{\text{DFT/PDT}}(\mathbf{r}, \alpha\beta\gamma) \rangle_{\alpha\beta\gamma, \lambda=1} \neq \langle \Delta E^{\text{IST}}_{\text{1-body}} + \Delta E^{\text{IST}}_{\text{2-body}} \rangle_{\alpha\beta\gamma, \lambda=1}$$

$$(34)$$

This highlights the fact that while the terms that are typically evaluated in the IST framework (1-body +2-body energies +1-body entropy) can be orientationally averaged and combined to yield a quantity which may be considered to be a free energy, they are not a true potential of mean force or a free-energy change of displacing a solvent molecule and cannot be evaluated using standard free energy methods.

Finally, in this section we note that the DFT equations for the solute chemical potential starting from the Kirkwood charging formula, eq 14 through eq 20, are formulated in the six dimensional (position and orientation) space of a solvent water molecule in the coordinate frame of the solute. However, unlike the IST equations, it is not necessary to formulate the exact DFT equations in 6 dimensional space; this is potentially an advantage with respect to numerical evaluation of the DFT expressions from end-point formulas as it is difficult to construct densities in 6 dimensional space. When considering alternate formulations of DFT, the requirement for the

construction of the density functional starting from the Kirkwood charging formula is that the direct solute−solvent interaction and the indirect solvent-mediated term be separated as in the PDF expression of eq 16 and that the density be uniquely determined by the solute−solvent interaction potential energy. Matubayasi and co-workers showed how to construct a 1-dimensional version of the DFT equations for the solute chemical potential, where the central focus is the 1-dimensional solute−solvent interaction energy density $\rho(\varepsilon)$.[33,34] While reducing the dimensionality of the DFT equations to 1-d is advantageous from the standpoint of the construction of the required density functionals from end-point simulations because of the substantial amount of statistical averaging that occurs when $\rho(\varepsilon)$ is used to construct the DFT functional, the spatial resolution of the 6 dimensional functional inherent in eqs 14−20 is lost. The challenge is to find a reduced set of coordinates lower than six to construct the density functional but which retains the spatial resolution implied by eq 14 through eq 20. The 3-dimensional coordinate $\mathbf{r}$ of the solvent water oxygen cannot be used to construct the DFT equations, because the density $\rho(\mathbf{r})$ does not correspond uniquely to the solute−solvent interaction energy $u(\mathbf{x})$. However, the mixed four dimensional coordinate-energy representation $(\mathbf{r}, \varepsilon)$ is a valid representation for constructing a free energy functional from end-point simulations. It retains the desired spatial resolution inherent in the full six dimensional representation (eq 14 to eq 20), and numerically it is easier to evaluate the required functionals than in the full $(\mathbf{r}, \alpha\beta\gamma)$ coordinate space. The development of a set of DFT equations for the solute excess chemical potential in this mixed coordinate-energy representation and methods to solve these equations will be the subject of a future communication (see also Appendix D).

■ MODEL PDT CALCULATIONS FOR THE SOLVENT EXCESS CHEMICAL POTENTIONAL AT A SOLUTE INTERFACE

In this section, calculations of PMF to move a tagged water molecule from the bulk to the interface of a solute are considered in order to illustrate the range of values the direct solute−solvent and indirect solvent−solvent PMF terms can have. We have chosen β-cyclodextrin (β-CD) to be the solute; it is a classic "host" molecule used for modeling molecular recognition[57−62] which we have studied previously.[63−65] Structurally, β-CD is made up of 7 D-glucose molecules linked in a 7-membered ring, which surrounds a central opening to which "guest" molecules can bind. β-CD is frustum shaped with a wider opening at one end lined by 14 secondary alcohols, and a narrower opening at the other end lined by 7 primary alcohols. In aqueous solution, β-CD displays two distinct conformations as shown in Figure 1. Viewed from the top with the wider opening above the plane, all the cyclic glucose rings are almost perpendicular to the paper plane in Figure 1A; while in Figure 1B one of the cyclic glucose rings is flipped over by almost 90°, so that it is nearly parallel to the paper plane and partially occludes the opening. We refer to Figure 1A as the open conformation and Figure 1 B as the closed conformation. The simulation details can be found in Appendix C.

In the previous section, we decomposed the total solute−solvent PMF into two parts: the direct and indirect terms. The direct term can be straightforwardly obtained from the solute−solvent pair interaction; the indirect contribution is derived from the difference between the total PMF and the direct
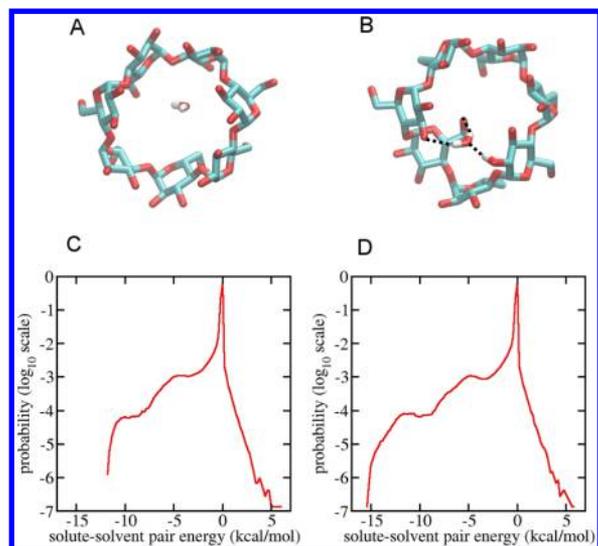
Figure 1. (A) The open conformation of β-CD and a water located in the central cavity of the β-CD solute. (B) The closed conformation of β-CD and the location of the water with the strongest direct interaction. (C) The probability distribution of solute−solvent pair interaction energy for open configuration of β-CD. (D) Probability distribution of the solute−solvent pair interaction energy for the closed conformation of β-CD.

contribution. The total PMF is the reversible work to bring a solvent molecule from the bulk (or pure liquid, the distinction is only important when a volume integral is performed) to a location ($\mathbf{x}$) at the interface. It can be evaluated by computer simulations using one of two approaches: (a) by integrating the mean force exerted on a tagged solvent molecule by the other molecules from a position in the bulk to a position in the vicinity of the solute, or (b) by the Widom particle insertion method. Most of the results reported in the section were obtained by the first method, i.e., by integrating the mean force in the following form:[66]

$$\nabla_\xi WT(\xi) = -\langle F_\xi \rangle_\xi \tag{35}$$

where ξ is the reaction coordinate taken as the location of the tagged water molecule along the y-axis. $\langle F_\xi \rangle_\xi$ denotes the average force exerted on the center of mass of the tagged water molecule along the reaction coordinate. We have performed two sets of simulations; one for which the angular orientation of the tagged solvent molecule is fixed, and another where the tagged solvent is allowed to reorient about the oxygen nucleus which is fixed. We denote the associated total PMF where the angular orientation is fixed as $WT(\mathbf{x})$ and the direct term as $u(\mathbf{x})$ and indirect term as $\omega(\mathbf{x})$; and $WT(\mathbf{r})$, $u(\mathbf{r})$, and $\omega(\mathbf{r})$ for the corresponding calculation where the water oxygen atom is fixed at $\mathbf{r}$ and the angular degrees of freedom of the tagged solvent molecule are not fixed. $u(\mathbf{r})$ is computed as the orientational average of $u(\mathbf{x})$ and $\omega(\mathbf{r}) = WT(\mathbf{r})$-$u(\mathbf{r})$. As noted

above, the solute chemical potential cannot be determined from knowledge of the orientationally averaged PMF $WT(\mathbf{r})$; we include calculations for $WT(\mathbf{r})$ to illustrate differences between $WT(\mathbf{x})$ and $WT(\mathbf{r})$. Using eq 35 above, we calculate the total PMF by "dragging" the tagged solvent molecule from the bulk to the interface with the solute, and integrating the mean force along the path. An alternative approach to calculating the solute−solvent PMF $WT(\mathbf{x})$ is by applying the potential distribution theorem (e.g., Widom particle insertion);[67] the total PMF is equal to the difference in free energy between inserting the tagged solvent molecule in the bulk and inserting it at ($\mathbf{x}$) (or $\mathbf{r}$ if the angular variables are not fixed). The formula for $WT(\mathbf{x})$ based on the potential distribution theorem is

$$
\begin{aligned}
WT(\mathbf{x}) = &-k_B T \ln \frac{\int \Omega(\mathbf{x}, U_{v\overline{v}}) \exp\{-\beta(U_{v\overline{v}} + U_{u\overline{v}})\} dU_{v\overline{v}}}{\int \Omega(\mathbf{x}, U_{v\overline{v}}) dU_{v\overline{v}}} \\
&+ k_B T \ln \frac{\int \Omega(\mathbf{x}_\infty, U_{v\overline{v}}) \exp\{-\beta U_{v\overline{v}}\} dU_{v\overline{v}}}{\int \Omega(\mathbf{x}_\infty, U_{v\overline{v}}) dU_{v\overline{v}}} \\
= &-k_B T \ln \frac{\int \Omega(\mathbf{x}, U_{v\overline{v}}) \exp\{-\beta(U_{u\overline{v}} + U_{v\overline{v}})\} dU_{v\overline{v}}}{\int \Omega(\mathbf{x}_\infty, U_{v\overline{v}}) \exp\{-\beta U_{v\overline{v}}\} dU_{v\overline{v}}}
\end{aligned}
\tag{36}
$$

where $\Omega(\mathbf{x}, U_{v\overline{v}})$ is the normalized density of states for the interaction energy of the tagged molecule in the solution at $\mathbf{x}$ with the untagged solvent molecules and $\Omega(\mathbf{x}_\infty, U_{v\overline{v}})$ is the corresponding normalized density of states when the tagged molecule is in the bulk, far from the solute. We have verified that both the "drag" and "Widom insertion" methods give the same PMF, $WT(\mathbf{x})$.

Figure 1C,D, shows the probability distributions of the solute−solvent pair interaction energies for solutions containing β-CD in the open Figure 1C and closed Figure 1D conformations, respectively. The solute−solvent direct interaction energies span a range of values between −16 kcal/mol and +5 kcal/mol. The most favorable solvent interaction with the host is a water molecule interacting with β-CD in the closed conformation in which three strong solute−solvent hydrogen bonds are formed, two with the secondary alcohols at the wider entrance to the host and a third with the flipped glucose ring that occludes the opening (see Figure 1B). The solute−solvent direct interaction energy distribution has a very large peak at zero as most of the water molecules solvating β-CD are distant from the solute and have interaction energy very close to zero.

Table 2 shows the total PMF and the direct and indirect terms for five selected water molecules hydrating β-CD in the open and closed conformations; results for both $WT(\mathbf{x})$ and $WT(\mathbf{r})$ are shown. The corresponding PMF profiles are shown in Figure 2. Further, in Figure 3 we show the exact positions of these five water molecules with respect to the solute. We consider first the most favorably bound solvent molecules to β-CD in both the closed and open conformations (Waters #1 and

Table 2. Free Energy Values To Transfer a Tagged Water Molecule from the Bulk to the Interface with β-CD

number	$\rho(\mathbf{x})/\rho_0$	$WT(\mathbf{x})$ (kcal/mol)	$u(\mathbf{x})$ (kcal/mol)	$\omega(\mathbf{x})$ (kcal/mol)	$WT(\mathbf{r})$ (kcal/mol)	$u(\mathbf{r})$ (kcal/mol)	$\omega(\mathbf{r})$ (kcal/mol)
1	4.8×10^3	−5.06(0.05)	−15.09	10.03	−3.01(0.23)	−13.12(0.03)	10.11
2	3.3×10^4	−6.21(0.06)	−11.82	5.61	−3.41(0.05)	−11.16(0.01)	7.75
3	2.1×10^{-3}	3.69(0.03)	6.09	−2.40	0.80(0.06)	−4.28(0.22)	5.08
4	1.7×10^1	−1.69(0.02)	−5.62	3.93	−0.55(0.07)	−6.23(0.01)	5.68
5	2.1×10^1	−1.83(0.10)	−2.45	0.62	−0.52(0.09)	−2.54(0.06)	2.02
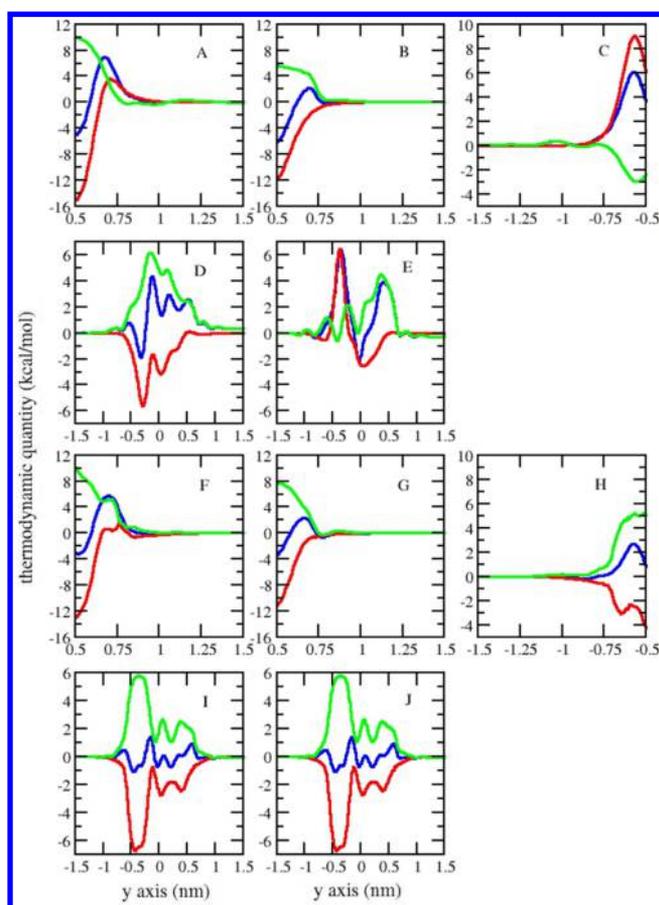
Figure 2. (A)−(E) $WT(\mathbf{x})$, $u(\mathbf{x})$ and $\omega(\mathbf{x})$ for transferring water molecules No.1−No.5 from bulk to interface. Among them, $WT(\mathbf{x})$ is in blue; $u(\mathbf{x})$ is in red and $\omega(\mathbf{x})$ is in green. (F)−(J) $WT(\mathbf{r})$, $u(\mathbf{r})$, and $\omega(\mathbf{r})$ for transferring water molecules No.1−No.5 from bulk to interface. $WT(\mathbf{r})$ is in blue; $u(\mathbf{r})$ is in red; and $\omega(\mathbf{r})$ is in green.
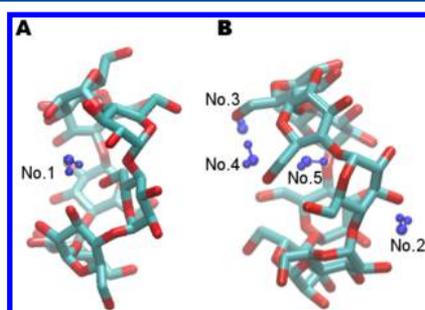
Figure 3. (A) The position of water No.1 with most favorable direct interaction with β-CD in closed conformation. (B) The position of water No.2 with most favorable direct interaction with β-CD in open conformation; the position of water No.3 with most unfavorable direct interaction with β-CD in open conformation; the position of water No.4 with favorable direct interaction around primary hydroxyl group of β-CD in open conformation; the position of water No.5 with favorable direct interaction inside core region of β-CD in open conformation.

#2). In the closed conformation, the very favorable solute−solvent interaction ($\sim$ −15 kcal/mol) of water #1 with the solute occurs because it is possible to form three hydrogen bonds as mentioned above. The direct interaction is opposed by a large indirect solvent−solvent free energy penalty of $\sim$ +10 kcal/mol (Figure 2A). Orientational averaging leads to a $\sim$ 2

kcal/mol loss of favorable solute−solvent energy, while the indirect part of the PMF remains $\sim$ +10 kcal/mol (see Table 2, and Figure 2F). The most favorably bound water molecule in the open conformation of β-CD (water #2, Table 2), forms a hydrogen bond to each of two secondary alcohols at the larger opening of the host cavity. The direct $u(\mathbf{x})$ and indirect $\omega(\mathbf{x})$ terms are −11.82 and +5.61 kcal/mol respectively, and the total is −6.21 kcal/mol (Figure 2B). Interestingly, when orientational averaging is performed, the total PMF $WT(\mathbf{r})$ becomes less attractive (−3.41 kcal/mol) relative to the fixed orientation, mostly due to an increase in the repulsive contribution of the indirect term by 2 kcal/mol ($\omega(\mathbf{r})$ = +7.75 kcal/mol) (Figure 2G). Water #3 exhibits the most repulsive direct interaction ($u(\mathbf{x})$ = +6.09 kcal/mol) (Figure 2C) due to a repulsive electrostatic interaction with the secondary hydroxyls in the open β-CD conformation, the indirect contribution to the PMF is attractive (−3.41 kcal/mol). This is a highly improbable conformation however; the orientationally averaged PMF for a water with oxygen at the same position is very different, with the average solute−solvent interaction contributing favorably (−4.28 kcal/mol); the indirect term $\omega(\mathbf{r})$ makes a large unfavorable contribution (+5.08 kcal/mol) (see Table 2, and Figure 2H).

Solvent molecules #4 and #5 (Table 2 and Figure 2D,E, and I,J) illustrate the potential of mean force to pull a solvent molecule through the center of the opening of the β-CD host. Both waters have the same oxygen position, they are located approximately on the central axis (x = 0, z = 0) of the host; their angular orientations differ. Water #4 is oriented so that its direct interaction with primary alcohols lining the smaller opening of the host ($y \sim$ −0.4 nm) is favorable to make a hydrogen bond with a primary alcohol ($u(\mathbf{x})$ = −5.62 kcal/mol) (Figure 2D). Water #5 has the most favorable interaction within the host cavity, $u(\mathbf{x})$ = −2.45 kcal/mol but it is not able to form hydrogen bonds with hydroxyl groups lining either of the openings as it is too distant from them. Water #5 has an angular orientation that creates a large barrier to exit, passing through either the narrower opening (the location of the primary hydroxyls at $y \sim$ −0.4) because of the repulsive solute−solvent direct interaction ($\sim$ +6 kcal/mol), or the wider opening (secondary hydroxyls) because of the repulsive solvent−solvent indirect interaction ($\sim$ +4 kcal/mol). Both waters #4 and #5 have the same orientationally average PMF as their oxygen atoms are located along the same axis (Figure 2I and J). The orientationally averaged PMF for pulling water molecules #4 and #5 through the β-CD opening are very different from the orientationally restricted results; they display much smaller barriers. Considering the orientationally averaged PMFs (Figure 2I and J), the total PMF as a function of displacement along the y-axis is only slightly repulsive. At the entrance to the host cavity where the primary hydroxyls are located, however, there is a large desolvation penalty ($\sim$6 kcal/mol) which is compensated by attractive solute−solvent interactions of the same magnitude.

In this section we have presented the results of calculations which show the range of values that the direct and indirect parts of the solute−solvent PMF can be expected to have, as a solvent molecule is moved from the bulk to the interface with β-CD, a solute that is used as a model system to illustrate features of solvation which are associated with molecular recognition events that occur when a ligand binds to its protein receptor target. Given the ability to calculate $WT(\mathbf{x})$ and $\omega(\mathbf{x})$ systematically throughout space, the chemical potential of the

solute can in principle be determined using either eq 17 or 18. In order to use these equations, spatially resolved approximate density functionals are needed to evaluate the charging integrals in eq 17 or 18 (see Appendix D for a further discussion of this point).

A second application that becomes possible using information of the kind shown in Table 2, is a "WaterMap"[14,15] type analysis of "druggable" sites at a receptor interface from the perspective of identifying water molecules that are more or less readily displaced by a ligand. It is clear from Table 2 that the direct and indirect components of the solute−solvent PMF are "competing" with each other. While there is no rigorous way to introduce these terms directly into a thermodynamic cycle that corresponds to the binding of a ligand to a receptor, it is possible to use knowledge of the individual PMF components (direct and indirect) to provide insights into ligand design and the binding process. One goal is to design ligands that capture as much of the favorable interactions of the tagged solvent molecule with the solute as possible while avoiding the competing solvent−solvent indirect term penalty. With a complete map of the solute−solvent PMF $WT(\mathbf{x})$ for a tagged solvent molecule in hand, solvent molecule locations at a protein receptor site targeted for displacement which exhibit both large attractive direct contributions and large repulsive indirect contributions to the solute−solvent PMF are prime targets for displacement when considering ligand design. Water molecule #1 in Table 2 is the example with the largest competing direct and indirect contributions to the PMF. It is interesting to note that water molecule #2 in Table 2, has a higher density relative to the bulk, although both the attractive and repulsive contributions to the PMF are smaller than those of molecule #1; for water #1 the indirect repulsion cancels more of the attractive interactions with the solute. Thus, solvent water molecule positions at the interface which may be good targets for displacement by adding a functional group to a ligand core are not necessarily located at positions where the solvent density is much higher than the bulk, and may not always be identified crystallographically.

■ CONCLUSIONS AND FUTURE DIRECTIONS

In this article we have reviewed the key equations that constitute inhomogeneous solvation theory (IST) and density functional theory (DFT) approaches to calculating the excess chemical potential of a solute in solution. The IST equations are expressly formulated in terms of distribution functions that can be calculated from the "end points" ($\lambda = 0$ and 1) of a virtual charging process, only simulations of the neat liquid and the fully coupled solution are actually needed. The DFT equations for the excess chemical potential can also be formulated in a way that only requires data from the "end points" of a virtual charging process.[33,34] An advantage of the DFT formulation over the IST formulation is the focus on free energies in DFT (i.e., the solute−solvent PMF), whereas in IST the focus is on calculating excess energies and entropies separately. Powerful methods to estimate the solute−solvent PMF $WT(\mathbf{x})$ and the indirect part $\omega(\mathbf{x})$ based on liquid state theory using HNC/PY approximations and WHAM can be used to evaluate the solute chemical potential, $\Delta\mu$ using end point simulation data.[68−71] Unlike IST for which the two-body and higher order terms in the entropy expansion become unwieldy to calculate from end point simulations, it is possible to calculate the corresponding charging integral in the DFT formulation.[33,34]

We have two goals in mind in our planned development of a spatially resolved DFT approach to solvation thermodynamics based on end point simulations of the pure liquid and solute in solution. The first concerns the calculation of the excess chemical potential of the solute as a local, spatially resolved integral within a region in proximity to the solute, the hydration shell volume, where the solvent structure is perturbed from that of the bulk solvent. One route to accomplish this is based on a modification of the DFT formulas in the 1-d energy representation. This leads to spatially resolved DFT formulas analogous to eq 20 in a 4-d mixed (coordinate, energy) representation where the coordinate is that of the water oxygen, and the energy refers to the solute−solvent interaction energy (see Appendix D). Using the 4-d representation of the DFT formulas for the solute chemical potential, it becomes possible to analyze the effects of specific waters on the solvation free energy of the solute and thereby identify locations where the solvating waters are stabilizing or destabilizing.

In the Model PDT Calculations for the Solvent Excess Chemical Potential at a Solute Interface Section, we presented the results of illustrative calculations of the solute−solvent PMF of a tagged solvent molecule in a solution containing β-CD in water. They are presented to illustrate the magnitudes and variance of the direct $u(\mathbf{x})$ and indirect $\omega(\mathbf{x})$ contributions to the PMF that are likely to be encountered in DFT calculations of the kind we are pursuing. They also serve as a reminder that while the chemical potential of the solvent is constant throughout the solution, the excess chemical potential is not; the solute−solvent PMF serves as a local probe that connects the solvation structure with the thermodynamics.

Our second goal is to use the DFT perspective to analyze the thermodynamic signatures of water molecules at the interface of protein receptors with a view toward the design of tight binding ligands. Previous efforts along this line have been based on the IST framework. A problem with the IST route for this analysis is that the free energy obtained by integrating the standard IST energy and entropy densities over a small volume at the solute surface, does not correspond to the physical process of displacing a water molecule from the bulk to the interface which can be evaluated by calculating a potential of mean force. Although the IST-based approach has yielded productive insights,[14−16,29−32,37−40] we think that an alternative DFT approach based on analyzing the competition between the direct and indirect contributions to the solute−solvent PMF can provide new insights into the thermodynamics of interfacial waters that will be useful in clarifying the role of water−water interactions in the thermodynamics of protein−ligand binding. One design strategy might involve adding a new functional group to a ligand core at a location which can mimic the strong attractive interaction of the solvating water the functional group is replacing at that position, but which disrupts the water structure to a smaller extent. Additional insights may be possible by analyzing how the direct and indirect terms of the solute−solvent PMF correlate individually and together with the charging term (the last terms of eqs 18, 20, and A23). Neither the IST nor the DFT equations for the solute excess chemical potential provide a direct route to ligand design, but the formulations of the interfacial thermodynamics based on end point formulas, as illustrated in Appendix D, suggest different approaches to the construction of scoring functions which may be helpful in the ligand design process. We are analyzing DFT based scoring functions and will report on this in a future communication. Our interest in this problem grows

out of our much earlier work together on IST,[10,11] and the development by one of us (N.M.) of a DFT based approach in the 1-d energy representation for the calculation of solvation thermodynamics using end point simulations.[33−36]

■ APPENDIX A: SEPARATION OF THE LOCAL AND NONLOCAL COMPONENTS IN EQ 9

The purpose of this Appendix is to resolve the intricacies involved in eqs 9−11. The discussion is based on the asymptotic analysis of the distribution functions at distances far from the solute[11] and is actually parallel to that described in Appendix A of Lazaridis's treatment of IST.[1]

Eq 9 involves an integral that is to be taken over the whole space (or the whole unit cell of simulation). When the system is in the canonical ensemble, the spatial integral needs to be done before the limit of large system size (thermodynamic limit) for the distribution function g_{sw}. This is because the asymptotic value of g_{sw} at distances far from the solute is expressed as

$$g_{sw}(\mathbf{x}) \rightarrow \frac{V}{V - \Delta V} + o\left(\frac{1}{V}\right) \equiv g_{sw}(\infty) \tag{A1}$$

where ΔV is the excess volume and $o(1/V)$ is a variable which vanishes faster than $1/V$ at large V, where V denotes the system size. Eq A1 shows that $g_{sw}(\infty)$ differs from 1 by order $1/V$, and the $O(1/V)$ effect is seen clearly by rewriting eq 9 as

$$\begin{aligned} \Delta S_{sw}^{(1)} &= -k_B \rho_0 \int d\mathbf{x} g_{sw}(\mathbf{x}) \ln \frac{g_{sw}(\mathbf{x})}{g_{sw}(\infty)} \\ &\quad - k_B \rho_0 \int d\mathbf{x} g_{sw}(\mathbf{x}) \ln g_{sw}(\infty) \\ &= -k_B \rho_0 \int d\mathbf{x} g_{sw}(\mathbf{x}) \ln \frac{g_{sw}(\mathbf{x})}{g_{sw}(\infty)} \\ &\quad - k_B N \ln \frac{1}{1 - \Delta V/V} \\ &= -k_B \rho_0 \int d\mathbf{x} g_{sw}(\mathbf{x}) \ln \frac{g_{sw}(\mathbf{x})}{g_{sw}(\infty)} - k_B \frac{N \Delta V}{V} + o(1) \end{aligned} \tag{A2}$$

where N is the number of solvent molecules and $o(1)$ refers to a variable which vanishes in the thermodynamic limit. The orders of the integral over the whole space and the thermodynamic limit can then be interchanged in the first term of the last line. When the thermodynamic limit is taken first, the first term of eq 11 is obtained; $g_{sw}(\infty) = 1$ in the thermodynamic limit. The ΔV in the second term is further expressed by virtue of the Kirkwood-Buff theory through

$$\Delta V = -\int d\mathbf{x}(g_{sw}^{TL}(\mathbf{x}) - 1) \tag{A3}$$

where g_{sw}^{TL} in eq A3 is the one in the thermodynamic limit when the system is treated in the canonical ensemble.

■ APPENDIX B: DERIVATION OF EQ 17

The key equation of the DFT Expression for the Excess Chemical Potential of a Solute in Solution Section is eq 17 and is obtained by combining eqs 15 and 16. In this Appendix, we describe the derivation in detail. The derivation is actually parallel to that in the energy representation shown in ref 34, and exploits eq 16 which is a PDT (potential distribution

theorem) expression and defines the solvent-mediated part of the solute−solvent potential of mean force (PMF) $\omega_\lambda(\mathbf{x})$.

Eq 15 is obtained simply by partial integration of the Kirkwood charging formula, eq 14. The second term of eq 15 is the density-functional and the λ integral in it is rewritten as

$$\begin{aligned} \int_0^1 d\lambda \frac{\partial \rho_\lambda(\mathbf{x})}{\partial \lambda} u_\lambda(\mathbf{x}) \\ = \int_0^1 d\lambda \frac{\partial \rho_\lambda(\mathbf{x})}{\partial \lambda} \left[-k_B T \log\left(\frac{\rho_\lambda(\mathbf{x})}{\rho_0(\mathbf{x})}\right) - \omega_\lambda(\mathbf{x}) \right] \end{aligned} \tag{A4}$$

Eq 17 then follows by noting that

$$\begin{aligned} \int_0^1 d\lambda \int d\mathbf{x} \frac{\partial \rho_\lambda(\mathbf{x})}{\partial \lambda} \log\left(\frac{\rho_\lambda(\mathbf{x})}{\rho_0(\mathbf{x})}\right) \\ = \int_0^1 d\lambda \int d\mathbf{x} \frac{\partial}{\partial \lambda}\left\{\rho_\lambda(\mathbf{x})\log\left(\frac{\rho_\lambda(\mathbf{x})}{\rho_0(\mathbf{x})}\right) - \rho_\lambda(\mathbf{x})\right\} \\ = \int d\mathbf{x}\left[\rho(\mathbf{x})\log\left(\frac{\rho(\mathbf{x})}{\rho_0(\mathbf{x})}\right) - (\rho(\mathbf{x}) - \rho_0(\mathbf{x}))\right] \end{aligned} \tag{A5}$$

where $\rho(\mathbf{x})$ means $\rho_\lambda(\mathbf{x})$ at $\lambda = 1$. Although the above derivation is a straightforward consequence of eq 16, it is possible because the direct, solute−solvent interaction term and the solvent-mediated term are separated within the exponential of eq 16.

Eq 16 provides the definition of $\omega_\lambda(\mathbf{x})$ and the thermodynamic limit is not taken there. The integration of $(\rho(\mathbf{x})-\rho_0(\mathbf{x}))$ in eq A5 then vanishes when the ensemble is canonical without the thermodynamic limit. The remaining contribution has the form of $\rho(\mathbf{x})\log(\rho(\mathbf{x})/\rho_0(\mathbf{x}))$ integration and corresponds to eq 9. This term can be further separated into the local and nonlocal components, and reduces to the expression corresponding to eq 11 through the arguments in Appendix A. eq A5 is thus valid both with and without the thermodynamic limit.

■ APPENDIX C: SIMULATION METHOD

The MD simulations were performed in GROMACS package version 4.6.3.[72] Simulations were carried out in NVT ensemble. Temperature was maintained at 300 K using Berendsen coupling scheme with time constant of $\tau_t = 1.0$ ps.[73] The simulation cell was cubic with all three dimensions as 3.5852 nm in the open β-CD system and 3.57866 nm in the closed β-CD system. In both systems, the slab was consisted of 1451 nonpolarizable TIP3P model[74] water molecules and one β-CD molecule fixed in the center of the box described with CHARMM General Force Field (CGenFF) 2b7.[75,76] In the first set of simulations to evaluate the distributions of solute−solvent pair energy in the open and closed β-CD systems, all the water molecules were considered as untagged. In the second set of simulations, a tagged water molecule was labeled with a special name different from other untagged water molecules in the topology file so that this water can be fixed in each simulation "windows". Extracting the forces on the tagged water and integration of the mean forces along the reaction coordinate in each window would obtain the PMF of transferring the tagged water from bulk to the specific interfacial region around open and closed β-CD.

The long-range electrostatic interactions were treated using smooth particle-mesh Ewald approach[77] with a real-space cutoff of 0.9 nm, a spline order of 4, a reciprocal-space mesh size of 32 for each of the x, y, and z directions. The Lennard-Jones interaction was handled by the twin range cut-offs with both neighbor list cutoff and VdW cutoff as 0.9 nm. Dynamics were propagated using a leapfrog stochastic dynamics integrator with a 2.0 fs time step. The systems were equilibrated for 1 ns, following by 10 ns production sampling.

■ APPENDIX D: DENSITY FUNCTIONALS FOR THE SOLUTE EXCESS CHEMICAL POTENTIAL IN 1-D AND 4-D

In the DFT Expression for the Excess Chemical Potential of a Solute in Solution Section, an expression for the excess chemical potential of a solute in solution was presented which can be evaluated using density functional theory. By this we mean the construction of approximate density functionals to evaluate the Kirkwood charging integral that appears in the DFT expressions in the DFT Expression for the Excess Chemical Potential of a Solute in Solution Section. The expression (eqs 17 and 18) was formulated over the full coordinate set of the solvent relative to the solute. The solvent coordinates were denoted as **x** and they are 6-dimensional to represent the position and orientation of a solvent (water) molecule with respect to a solute-fixed coordinate frame. The computation of the density (solute−solvent distribution) over the 6-dimentional coordinates is numerically difficult, however. To make the DFT route to the solute excess chemical potential more computationally practical, the problem can be transformed to one which requires the evaluation of distribution functions over the solvent coordinates of reduced dimension.

The method of energy representation was formulated as a DFT approach over a 1-dimensional coordinate.[33−36] In this representation, the distribution of the value ε of the solute−solvent pair interaction $u(\mathbf{x})$ plays the key role, where $u(\mathbf{x})$ is the solute−solvent interaction potential in the solution system of interest. The distribution functions are constructed from the instantaneous distribution introduced as

$$\hat{\rho}(\varepsilon) = \sum_i \delta(u(\mathbf{x}_i) - \varepsilon) \tag{A6}$$

where the sum is taken over the solvent molecules. A DFT formalism can then be implemented by restricting the set of solute−solvent interaction potentials $u_\lambda(\mathbf{x})$ to those which are constant over an equi-energy surface of $u(\mathbf{x})$. With this setup, the intermediate states can be written as $u_\lambda(\varepsilon)$ when the value of $u(\mathbf{x})$ is denoted as ε. At the end points ($\lambda = 0$ and 1), $u_0(\varepsilon) = 0$ and $u_1(\varepsilon) = \varepsilon$ since $u(\mathbf{x})$ itself is the potential function in the solution system of interest.

It is then possible to show that the Kirkwood charging formula is given by

$$\Delta\mu = \int_0^1 d\lambda \int d\varepsilon \frac{\partial u_\lambda(\varepsilon)}{\partial \lambda} \rho_\lambda(\varepsilon) \tag{A7}$$

where $\rho_\lambda(\varepsilon)$ is the ensemble average of eq A6 in the presence of the solute−solvent interaction u_λ. When $\rho(\varepsilon)$ denotes $\rho_\lambda(\varepsilon)$ at $\lambda = 1$ and is the distribution function of ε in the solution system of interest, the partial integration of eq A7 leads to

$$\Delta\mu = \int d\varepsilon \varepsilon \rho(\varepsilon) - \int_0^1 d\lambda \int d\varepsilon \frac{\partial \rho_\lambda(\varepsilon)}{\partial \lambda} u_\lambda(\varepsilon) \tag{A8}$$

and the indirect part ω of the potential of mean force in the energy representation can be introduced by rewriting **x** of eq 16 with ε as

$$\rho_\lambda(\varepsilon) = \exp(-\beta(u_\lambda(\varepsilon) + \omega_\lambda(\varepsilon)))\rho_0(\varepsilon) \tag{A9}$$

where $\rho_0(\varepsilon)$ is $\rho_\lambda(\varepsilon)$ at $\lambda = 0$. It is the (number) density of pure solvent multiplied by the density of states of the solute−solvent pair potential. The first term of eq A8 is the average sum of the solute−solvent interaction energy and the second term is the density-functional in the energy representation. $\omega_\lambda(\varepsilon)$ is defined by eq A9. The free energy change $WT(\varepsilon)$ in the energy representation is given by $-k_BT\ln[\rho_\lambda(\varepsilon)/\rho_0(\varepsilon)]$. It corresponds to $-k_BT$ times the log of the ratio of the number of configurations of the system for which a tagged solvent molecule has interaction energy ε with the solute in solution relative to the corresponding number in the pure solvent where the solute acts as a virtual particle (i.e., test particle insertion of the solute). $\omega_\lambda(\varepsilon)$ is the indirect solvent−solvent part of this free energy which is obtained by subtracting $u_\lambda(\varepsilon)$ from $WT(\varepsilon)$. Eq A9 is the potential distribution theorem in the 1-d energy representation. As was the case in the DFT Expression for the Excess Chemical Potential of a Solute in Solution Section, a key for the following developments is the separation of the direct solute−solvent interaction and the solvent-mediated term within the exponential.

Eq A8 can be modified, by virtue of eq A9, into

$$\Delta\mu = \int d\varepsilon \varepsilon \rho(\varepsilon) - k_BT \int d\varepsilon \left[(\rho(\varepsilon) - \rho_0(\varepsilon)) - \rho(\varepsilon)\log\left(\frac{\rho(\varepsilon)}{\rho_0(\varepsilon)}\right) \right] + \int_0^1 d\lambda \int d\varepsilon \omega_\lambda(\varepsilon) \frac{\partial \rho_\lambda(\varepsilon)}{\partial \lambda} \tag{A10}$$

This expression is exact and has a similar structure to eq 17. The second term corresponds to the one-body entropy in the energy representation. It is always non-negative and measures the "deviation" of the solution system of interest from the pure solvent. The third term expresses (the change in) the solvent−solvent correlation upon gradual insertion of the solute with the coupling parameter λ. The solvent reorganization energy (change in the solvent−solvent interaction due to the insertion of solute) is part of the third term of eq A10; the third term captures the energy and entropy of the solvent reorganization all together.

In the currently used version of the energy-representation method,[33−35] the third term of eq A10 is approximately expressed in terms of distribution functions at the end points ($\lambda = 0$ and 1). $\Delta\mu$ is evaluated with $\rho(\varepsilon)$ ($\rho_\lambda(\varepsilon)$ at $\lambda = 1$), $\rho_0(\varepsilon)$ ($\rho_\lambda(\varepsilon)$ at $\lambda = 0$), and $\chi_0(\varepsilon,\eta)$ defined as

$$\chi_0(\varepsilon, \eta) = \langle \hat{\rho}(\varepsilon)\hat{\rho}(\eta) \rangle_0 - \langle \hat{\rho}(\varepsilon) \rangle_0 \langle \hat{\rho}(\eta) \rangle_0 \tag{A11}$$

where $\langle \cdots \rangle_0$ denotes the ensemble average in the pure solvent ($\lambda = 0$). At $\lambda = 0$, $\hat{\rho}(\varepsilon)$ of eq A6 is constructed by placing the solute molecule in the pure solvent system as a test particle and χ_0 describes the two-body correlation of solvent molecules over the energy coordinate in the (physical) absence of the solute. An approximation to the coupling-parameter integration in the third term of eq A10 is introduced by adopting the Percus−Yevick-type or hypernetted-chain-type relationship and is given by a set of definitions and equations listed as

$$\sigma_0(\varepsilon) = -k_BT \int d\eta \left(\frac{\delta(\varepsilon - \eta)}{\rho_0(\varepsilon)} - (\chi_0)^{-1}(\varepsilon, \eta) \right)$$

$$(\rho(\eta) - \rho_0(\eta)) \tag{A12}$$

$$\int_0^1 d\lambda \int d\varepsilon \omega_\lambda(\varepsilon) \frac{\partial \rho_\lambda(\varepsilon)}{\partial \lambda}$$

$$= k_BT \int d\varepsilon (\rho^e(\varepsilon) - \rho_0^e(\varepsilon))[\alpha(\varepsilon)F(\varepsilon) + (1 - \alpha(\varepsilon))F_0(\varepsilon)] \tag{A13}$$

$$F(\varepsilon) = \begin{cases} \beta\omega(\varepsilon) + 1 + \dfrac{\beta\omega(\varepsilon)}{\exp(-\beta\omega(\varepsilon)) - 1} \\ \quad (\text{when } \omega(\varepsilon) \leq 0) \\ \dfrac{1}{2}\beta\omega(\varepsilon) \ (\text{when } \omega(\varepsilon) \geq 0) \end{cases} \tag{A14}$$

$$F_0(\varepsilon) = \begin{cases} -\log(1 - \beta\sigma_0(\varepsilon)) + 1 + \dfrac{\log(1 - \beta\sigma_0(\varepsilon))}{\beta\sigma_0(\varepsilon)} \\ \quad (\text{when } \sigma_0(\varepsilon) \leq 0) \\ \dfrac{1}{2}\beta\sigma_0(\varepsilon) \ (\text{when } \sigma_0(\varepsilon) \geq 0) \end{cases} \tag{A15}$$

$$\alpha(\varepsilon) = \begin{cases} 1 \ (\text{when } \rho(\varepsilon) \geq \rho_0(\varepsilon)) \\ 1 - \left(\dfrac{\rho(\varepsilon) - \rho_0(\varepsilon)}{\rho(\varepsilon) + \rho_0(\varepsilon)}\right)^2 \ (\text{when } \rho(\varepsilon) \leq \rho_0(\varepsilon)) \end{cases} \tag{A16}$$

where $\omega(\varepsilon) = \omega_1(\varepsilon)$ for notational brevity. $\omega(\varepsilon)$ introduced by eq A9 is the solvent-mediated part of the solute−solvent potential of mean force in the energy representation, and $\sigma_0(\varepsilon)$ is the solvent-mediated part of the response function of the solute−solvent distribution to the solute−solvent interaction in the pure solvent ($\lambda = 0$). Eqs A14 and A15 represent the PY-type and HNC-type approximations in their first and second lines, respectively, and are mixed with the weighting function α defined as eq A16.

Although an end point calculation is possible for $\Delta\mu$ using the DFT-type functional in the energy representation, the spatial resolution is then lost. To restore the spatial resolution and keep the $\Delta\mu$ calculation feasible, we can introduce a mixed space-energy representation. In this representation, the solute−solvent pair energy is used as a proxy for the solvent orientation relative to the solute and the instantaneous density is introduced as

$$\hat{\rho}(\mathbf{r}, \varepsilon) = \sum_i \delta(\mathbf{r}_i - \mathbf{r})\delta(u(\mathbf{x}_i) - \varepsilon) \tag{A17}$$

With the first delta function, the translational position (but not the orientation) of the solvent (water) molecule relative to the solute is specified, and the second delta function identifies the solute−solvent pair energy. A corresponding DFT expression can be formulated by considering the solute−solvent interaction potentials u_λ ($\mathbf{r}, \alpha\beta\gamma$) with the corresponding form of u_λ ($\mathbf{r},\varepsilon$). It is simply 0 at the initial state without solute ($\lambda = 0$) and is given by $u_1(\mathbf{r},\varepsilon) = \varepsilon$ at the final, solution state of interest ($\lambda = 1$); the explicit dependence on the position $\mathbf{r}$ may appear only at the intermediate states ($0 < \lambda < 1$). When the

ensemble average of eq A17 is $\rho_\lambda(\mathbf{r},\varepsilon)$ at the coupling parameter λ, the Kirkwood charging formula is expressed as

$$\Delta\mu = \int_0^1 d\lambda \int d\mathbf{r} d\varepsilon \frac{\partial u_\lambda(\mathbf{r}, \varepsilon)}{\partial \lambda}\rho_\lambda(\mathbf{r}, \varepsilon) \tag{A18}$$

We then let $\rho(\mathbf{r},\varepsilon)$ stand for $\rho_\lambda(\mathbf{r},\varepsilon)$ at $\lambda = 1$ (solution system of interest) and rewrite eq A18 through partial integration as

$$\Delta\mu = \int d\mathbf{r} d\varepsilon \varepsilon \rho(\mathbf{r}, \varepsilon) - \int_0^1 d\lambda \int d\mathbf{r} d\varepsilon \frac{\partial \rho_\lambda(\mathbf{r}, \varepsilon)}{\partial \lambda}u_\lambda(\mathbf{r}, \varepsilon) \tag{A19}$$

The first term is the average sum of the solute−solvent interaction energy and the second term is the density-functional. We can similarly introduce the indirect part of the solute−solvent potential of mean force (PMF) $\omega_\lambda(\mathbf{r},\varepsilon)$ by

$$\rho_\lambda(\mathbf{r}, \varepsilon) = \exp(-\beta(u_\lambda(\mathbf{r}, \varepsilon) + \omega_\lambda(\mathbf{r}, \varepsilon)))\rho_0(\mathbf{r}, \varepsilon) \tag{A20}$$

where $\rho_0(\mathbf{r}, \varepsilon)$ is $\rho_\lambda(\mathbf{r}, \varepsilon)$ at $\lambda = 0$. This is the potential distribution theorem in the mixed representation, and eq A19 can be modified into

$$\Delta\mu = \int d\mathbf{r} d\varepsilon \varepsilon \rho(\mathbf{r}, \varepsilon) - k_BT \int d\mathbf{r} d\varepsilon \left[(\rho(\mathbf{r}, \varepsilon) - \rho_0(\mathbf{r}, \varepsilon)) \right.$$
$$\left. - \rho(\mathbf{r}, \varepsilon)\log\left(\frac{\rho(\mathbf{r}, \varepsilon)}{\rho_0(\mathbf{r}, \varepsilon)}\right) \right]$$
$$+ \int_0^1 d\lambda \int d\mathbf{r} d\varepsilon \omega_\lambda(\mathbf{r}, \varepsilon)\frac{\partial \rho_\lambda(\mathbf{r}, \varepsilon)}{\partial \lambda} \tag{A21}$$

This is an exact expression and can be a basis of spatial decomposition expressed as

$$\Delta\mu = \int d\mathbf{r} \Delta\mu(\mathbf{r})$$
$$\Delta\mu(\mathbf{r}) = \int d\varepsilon \varepsilon \rho(\mathbf{r}, \varepsilon) - k_BT \int d\varepsilon \left[(\rho(\mathbf{r}, \varepsilon) - \rho_0(\mathbf{r}, \varepsilon)) \right.$$
$$\left. - \rho(\mathbf{r}, \varepsilon)\log\left(\frac{\rho(\mathbf{r},\varepsilon)}{\rho_0(\mathbf{r},\varepsilon)}\right) \right]$$
$$+ \int_0^1 d\lambda \int d\varepsilon \omega_\lambda(\mathbf{r}, \varepsilon)\frac{\partial \rho_\lambda(\mathbf{r},\varepsilon)}{\partial \lambda} \tag{A22}$$

The first term of the second line is the average sum of the interaction energy with the solute of the solvent molecules located at the translational position $\mathbf{r}$ but not orientationally restrained. The second line can be interpreted as the free-energy density at $\mathbf{r}$. It is expressed with 4-dimensional distribution functions. $\Delta\mu(\mathbf{r})$ is not necessarily zero even when $\rho(\mathbf{r})$ is almost zero. In the excluded-volume domain, $\rho(\mathbf{r})$ is almost zero while $\Delta\mu(\mathbf{r})$ will be positive due to the excluded-volume effect.

As done from eq 17 to 18, the second line of eq A22 is rewritten by virtue of eq A20 as

$$\Delta\mu(\mathbf{r}) = -k_BT \int d\varepsilon [(\rho(\mathbf{r}, \varepsilon)) - (\rho_0(\mathbf{r}, \varepsilon))]$$
$$- \int d\varepsilon \omega(\mathbf{r}, \varepsilon)\rho(\mathbf{r}, \varepsilon)$$
$$+ \int_0^1 d\lambda \int d\varepsilon \omega_\lambda(\mathbf{r}, \varepsilon)\frac{\partial \rho_\lambda(\mathbf{r}, \varepsilon)}{\partial \lambda} \tag{A23}$$

where $\omega_\lambda(\mathbf{r}, \varepsilon)$ at $\lambda = 1$ is denoted as $\omega(\mathbf{r}, \varepsilon)$ and the definition that $u_1(\mathbf{r}, \varepsilon) = \varepsilon$ is used. The second term of eq A23 is the weighted sum of the solvent-mediated (indirect) part of the solute−solvent PMF at fixed $\mathbf{r}$. When this is more positive, there arises a more favorable contribution to $\Delta\mu(\mathbf{r})$ due to the dragging of the orientationally unrestrained tagged solvent molecule at $\mathbf{r}$ to pure solvent with the direct solute−solvent interaction turned off, as is reminiscent of the idea of WaterMap. The indirect PMF can be a decisive component for analyzing the solvation thermodynamics if it is correlated to the space-resolved charging term (the second term of eqs 18 or A23) over the variation of the position $\mathbf{r}$. If this correlation holds and is combined with the WaterMap idea, furthermore, the displacement of a water molecule whose indirect solvent− solvent interactions are very unfavorable relative to the bulk contributes to stronger binding of a ligand. There is an advantage to using the second term in eq A23 for a WaterMap type analysis of the competition between the direct and indirect terms in 4-d as compared with 6-d, as eq A23 provides a framework for analyzing the effect on ligand binding of displacing water molecules at protein receptor sites where the target water(s) have some orientational freedom which must be taken into account in the analysis, and this is not convenient in the 6-d representation.

■ AUTHOR INFORMATION

Corresponding Authors
*E-mail: ronlevy@temple.edu; Phone: 1-215-204-1927.
*E-mail: nobuyuki@cheng.es.osaka-u.ac.jp; Phone: 81-6-6850-6565.

ORCID
Ronald M. Levy: 0000-0001-8696-5177

Notes
The authors declare no competing financial interest.

■ REFERENCES

(1) Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory. *J. Phys. Chem. B* **1998**, *102*, 3531−3541.
(2) Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 2. Applications to Simple Fluids. *J. Phys. Chem. B* **1998**, *102*, 3542−3550.
(3) Wallace, D. C. On the Role of Density Fluctuations in the Entropy of a Fluid. *J. Chem. Phys.* **1987**, *87*, 2282−2284.
(4) Mayer, J. E.; Montroll, E. Molecular Distribution. *J. Chem. Phys.* **1941**, *9*, 2−16.
(5) Baranyai, A.; Evans, D. J. Direct Entropy Calculation from Computer Simulation of Liquids. *Phys. Rev. A: At., Mol., Opt. Phys.* **1989**, *40*, 3817−3822.
(6) Evans, R. The Nature of the Liquid-Vapour Interface and Other Topics in the Statistical Mechanics of Non-Uniform, Classical Fluids. *Adv. Phys.* **1979**, *28*, 143−200.
(7) Henderson, D. *Fundamentals of Inhomogeneous Fluids*; Marcel Dekker, Inc.: New York, 1992.
(8) Hansen, J. P.; McDonald, I. R. *Theory of Simple Liquids*; Elsevier/ Academic Press: Amsterdam, 2013.
(9) Matubayasi, N.; Reed, L. H.; Levy, R. M. Thermodynamics of the Hydration Shell. 1. Excess Energy of a Hydrophobic Solute. *J. Phys. Chem.* **1994**, *98*, 10640−10649.
(10) Matubayasi, N.; Levy, R. M. Thermodynamics of the Hydration Shell. 2. Excess Volume and Compressibility of a Hydrophobic Solute. *J. Phys. Chem.* **1996**, *100*, 2681−2688.
(11) Matubayasi, N.; Gallicchio, E.; Levy, R. M. On the Local and Nonlocal Components of Solvation Thermodynamics and Their Relation to Solvation Shell Models. *J. Chem. Phys.* **1998**, *109*, 4864− 4872.
(12) Li, Z.; Lazaridis, T. Thermodynamics of Buried Water Clusters at a Protein-Ligand Binding Interface. *J. Phys. Chem. B* **2006**, *110*, 1464−1475.
(13) Li, Z.; Lazaridis, T. Water at Biomolecular Binding Interfaces. *Phys. Chem. Chem. Phys.* **2007**, *9*, 573−581.
(14) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding. *J. Am. Chem. Soc.* **2008**, *130*, 2817−2831.
(15) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. Motifs for Molecular Recognition Exploiting Hydrophobic Enclosure in Protein-Ligand Binding. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 808− 813.
(16) Wang, L.; Berne, B. J.; Friesner, R. A. Ligand Binding to Protein-Binding Pockets with Wet and Dry Regions. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 1326−1330.
(17) Beuming, T.; Farid, R.; Sherman, W. High-Energy Water Sites Determine Peptide Binding Affinity and Specificity of PDZ domains. *Protein Sci.* **2009**, *18*, 1609−1619.
(18) Pearlstein, R. A.; Hu, Q. Y.; Zhou, J.; Yowe, D.; Levell, J.; Dale, B.; Kaushik, V. K.; Daniels, D.; Hanrahan, S.; Sherman, W.; et al. New Hypotheses about the Structure-Function of Proprotein Convertase Subtilisin/Kexin Type 9: Analysis of the Epidermal Growth Factor-Like Repeat a Docking Site Using WaterMap. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 2571−2586.
(19) Higgs, C.; Beuming, T.; Sherman, W. Hydration Site Thermodynamics Explain SARs for Triazolylpurines Analogues Binding to the A2A Receptor. *ACS Med. Chem. Lett.* **2010**, *1*, 160− 164.
(20) Robinson, D. D.; Sherman, W.; Farid, R. Understanding Kinase Selectivity through Energetic Analysis of Binding Site Waters. *ChemMedChem* **2010**, *5*, 618−627.
(21) Abel, R.; Salam, N. K.; Shelley, J.; Farid, R.; Friesner, R. A.; Sherman, W. Contribution of Explicit Solvent Effects to the Binding Affinity of Small-Molecule Inhibitors in Blood Coagulation Factor Serine Proteases. *ChemMedChem* **2011**, *6*, 1049−1066.
(22) Snyder, P. W.; Mecinovic, J.; Moustakas, D. T.; Thomas, S. W., III; Harder, M.; Mack, E. T.; Lockett, M. R.; Heroux, A.; Sherman, W.; Whitesides, G. M. Mechanism of the Hydrophobic Effect in the Biomolecular Recognition of Arylsulfonamides by Carbonic Anhy-drase. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 17889−17894.
(23) Beuming, T.; Che, Y.; Abel, R.; Kim, B.; Shanmugasundaram, V.; Sherman, W. Thermodynamic Analysis of Water Molecules at the Surface of Proteins and Applications to Binding Site Prediction and Characterization. *Proteins: Struct., Funct., Genet.* **2012**, *80*, 871−883.
(24) Shah, F.; Gut, J.; Legac, J.; Shivakumar, D.; Sherman, W.; Rosenthal, P. J.; Avery, M. A. Computer-Aided Drug Design of Falcipain Inhibitors: Virtual Screening, Structure−Activity Relation-

ships, Hydration Site Thermodynamics, and Reactivity Analysis. *J. Chem. Inf. Model.* **2012**, *52*, 696−710.

(25) Tran, Q. T.; Williams, S.; Farid, R.; Erdemli, G.; Pearlstein, R. The Translocation Kinetics of Antibiotics through Porin OmpC: Insights from Structure-Based Solvation Mapping Using WaterMap. *Proteins: Struct., Funct., Genet.* **2013**, *81*, 291−299.

(26) Pearlstein, R. A.; Sherman, W.; Abel, R. Contributions of Water Transfer Energy to Protein-Ligand Association and Dissociation Barriers: WaterMap Analysis of a Series of p38α MAP Kinase Inhibitors. *Proteins: Struct., Funct., Genet.* **2013**, *81*, 1509−1526.

(27) Breiten, B.; Lockett, M. R.; Sherman, W.; Fujita, S.; Al-Sayah, M.; Lange, H.; Bowers, C. M.; Heroux, A.; Krilov, G.; Whitesides, G. M. Water Networks Contribute to Enthalpy/Entropy Compensation in Protein-Ligand Binding. *J. Am. Chem. Soc.* **2013**, *135*, 15579−15584.

(28) Fox, J. M.; Kang, K.; Sherman, W.; Heroux, A.; Sastry, G. M.; Baghbanzadeh, M.; Lockett, M. R.; Whitesides, G. M. Interactions between Hofmeister Anions and the Binding Pocket of a Protein. *J. Am. Chem. Soc.* **2015**, *137*, 3859−3866.

(29) Nguyen, C. N.; Cruz, A.; Gilson, M. K.; Kurtzman, T. Thermodynamics of Water in an Enzyme Active Site: Grid-Based Hydration Analysis of Coagulation Factor Xa. *J. Chem. Theory Comput.* **2014**, *10*, 2769−2780.

(30) Li, Z.; Lazaridis, T. Thermodynamic Contributions of the Ordered Water Molecule in HIV-1 Protease. *J. Am. Chem. Soc.* **2003**, *125*, 6636−6637.

(31) Li, Z.; Lazaridis, T. The Effect of Water Displacement on Binding Thermodynamics: Concanavalin A. *J. Phys. Chem. B* **2005**, *109*, 662−670.

(32) Nguyen, C. N.; Young, T. K.; Gilson, M. K. Grid Inhomogeneous Solvation Theory: Hydration Structure and Thermodynamics of the Miniature Receptor Cucurbit[7]uril. *J. Chem. Phys.* **2012**, *137*, 044101.

(33) Matubayasi, N.; Nakahara, M. Theory of Solutions in the Energetic Representation. I. Formulation. *J. Chem. Phys.* **2000**, *113*, 6070−6081.

(34) Matubayasi, N.; Nakahara, M. Theory of Solutions in the Energy Representation. II. Functional for the Chemical Potential. *J. Chem. Phys.* **2002**, *117*, 3605−3616; Erratum. *J. Chem. Phys.* **2003**, *118*, 2446.

(35) Matubayasi, N.; Nakahara, M. Theory of Solutions in the Energy Representation. III. Treatment of the Molecular Flexibility. *J. Chem. Phys.* **2003**, *119*, 9686−9702.

(36) 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.

(37) Vukovic, S.; Brennan, P. E.; Huggins, D. J. Exploring the Role of Water in Molecular Recognition: Predicting Protein Ligandability Using a Combinatorial Search of Surface Hydration Sites. *J. Phys.: Condens. Matter* **2016**, *28*, 344007.

(38) Haider, K.; Huggins, D. J. Combining Solvent Thermodynamic Profiles with Functionality Maps of the Hsp90 Binding Site to Predict the Displacement of Water Molecules. *J. Chem. Inf. Model.* **2013**, *53*, 2571−2586.

(39) Huggins, D. J.; Marsh, M.; Payne, M. C. Thermodynamic Properties of Water Molecules at a Protein-Protein Interaction Surface. *J. Chem. Theory Comput.* **2011**, *7*, 3514−3522.

(40) Huggins, D. J. Application of Inhomogeneous Fluid Solvation Theory to Model the Distribution and Thermodynamics of Water Molecules around Biomolecules. *Phys. Chem. Chem. Phys.* **2012**, *14*, 15106−15117.

(41) Huggins, D. J. Quantifying the Entropy of Binding for Water Molecules in Protein Cavities by Computing Correlations. *Biophys. J.* **2015**, *108*, 928−936.

(42) Huggins, D. J.; Payne, M. C. Assessing the Accuracy of Inhomogeneous Fluid Solvation Theory in Predicting Hydration Free Energies of Simple Solutes. *J. Phys. Chem. B* **2013**, *117*, 8232−8244.

(43) Nguyen, C. N.; Kurtzman, T.; Gilson, M. K. Spatial Decomposition of Translational Water-Water Correlation Entropy in Binding Pockets. *J. Chem. Theory Comput.* **2016**, *12*, 414−429.

(44) Jeanmairet, G.; Levesque, M.; Vuilleumier, R.; Borgis, D. Molecular Density Functional Theory of Water. *J. Phys. Chem. Lett.* **2013**, *4*, 619−624.

(45) Gendre, L.; Ramirez, R.; Borgis, D. Classical Density Functional Theory of Solvation in Molecular Solvents: Angular Grid Implementation. *Chem. Phys. Lett.* **2009**, *474*, 366−370.

(46) Zhao, S.; Ramirez, R.; Vuilleumier, R.; Borgis, D. Molecular Density Functional Theory of Solvation: From Polar Solvents to Water. *J. Chem. Phys.* **2011**, *134*, 194102.

(47) Borgis, D.; Gendre, L.; Ramirez, R. Molecular Density Functional Theory: Application to Solvation and Electron-Transfer Thermodynamics in Polar Solvents. *J. Phys. Chem. B* **2012**, *116*, 2504−2512.

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

(49) Hirata, F.; Rossky, P. J. An Extended RISM Equation for Molecular Polar Fluids. *Chem. Phys. Lett.* **1981**, *83*, 329−334.

(50) Hirata, F.; Pettitt, B. M.; Rossky, P. J. Application of an Extended RISM Equation to Dipolar and Quadrupolar Fluids. *J. Chem. Phys.* **1982**, *77*, 509−520.

(51) Hirata, F.; Redfern, P.; Levy, R. M. Viewing the Born Model for Ion Hydration through a Microscope. *Int. J. Quantum Chem.* **1988**, *34*, 179−190.

(52) Hirata, F. *Molecular theory of solvation*; Springer Science & Business Media: Dordrecht, 2006.

(53) Widom, B. Some Topics in the Theory of Fluids. *J. Chem. Phys.* **1963**, *39*, 2808−2812.

(54) Widom, B. Potential-Distribution Theory and the Statistical Mechanics of Fluids. *J. Phys. Chem.* **1982**, *86*, 869−872.

(55) Corti, D. S. Alternative Derivation of Widom's Test Particle Insertion Method Using the Small System Grand Canonical Ensemble. *Mol. Phys.* **1998**, *93*, 417−420.

(56) Lazaridis, T.; Karplus, M. Orientational Correlations and Entropy in Liquid Water. *J. Chem. Phys.* **1996**, *105*, 4294−4316.

(57) Chang, C.-E.; Gilson, M. K. Free Energy, Entropy, and Induced Fit in Host-Guest Recognition: Calculations with the Second-Generation Mining Minima Algorithm. *J. Am. Chem. Soc.* **2004**, *126*, 13156−13164.

(58) Rekharsky, M. V.; Inoue, Y. Complexation Thermodynamics of Cyclodextrins. *Chem. Rev.* **1998**, *98*, 1875−1918.

(59) Rekharsky, M.; Inoue, Y. Chiral Recognition Thermodynamics of β-Cyclodextrin: The Thermodynamic Origin of Enantioselectivity and the Enthalpy-Entropy Compensation Effect. *J. Am. Chem. Soc.* **2000**, *122*, 4418−4435.

(60) Rekharsky, M. V.; Inoue, Y. Solvent and Guest Isotope Effects on Complexation Thermodynamics of α-, β-, and 6-Amino-6-deoxy-β-cyclodextrins. *J. Am. Chem. Soc.* **2002**, *124*, 12361−12371.

(61) Szejtli, J. Introduction and General Overview of Cyclodextrin Chemistry. *Chem. Rev.* **1998**, *98*, 1743−1753.

(62) Del Valle, E. M. M. Cyclodextrins and Their Uses: A Review. *Process Biochem.* **2004**, *39*, 1033−1046.

(63) Wickstrom, L.; Deng, N.; He, P.; Mentes, A.; Nguyen, C.; Gilson, M. K.; Kurtzman, T.; Gallicchio, E.; Levy, R. M. Parameterization of an Effective Potential for Protein-Ligand Binding from Host-Guest Affinity Data. *J. Mol. Recognit.* **2016**, *29*, 10−21.

(64) Wickstrom, L.; He, P.; Gallicchio, E.; Levy, R. M. Large Scale Affinity Calculations of Cyclodextrin Host-Guest Complexes: Understanding the Role of Reorganization in the Molecular Recognition Process. *J. Chem. Theory Comput.* **2013**, *9*, 3136−3150.

(65) Harris, R.; 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.* **2016**, DOI: 10.1002/jcc.24668.

(66) Cui, D.; Ou, S.; Peters, E.; Patel, S. Ion-Specific Induced Fluctuations and Free Energetics of Aqueous Protein Hydrophobic Interfaces: Toward Connecting to Specific-Ion Behaviors at Aqueous Liquid-Vapor Interfaces. *J. Phys. Chem. B* **2014**, *118*, 4490−4504.

(67) Beck, T. L.; Paulaitis, M. E.; Pratt, L. R. *The Potential Distribution Theorem and Models of Molecular Solutions*; Cambridge University Press: Cambridge, 2012.

(68) Gallicchio, E.; Andrec, M.; Felts, A. K.; Levy, R. M. Temperature Weighted Histogram Analysis Method, Replica Exchange, and Transition Paths. *J. Phys. Chem. B* **2005**, *109*, 6722−6731.

(69) Tan, Z.; Gallicchio, E.; Lapelosa, M.; Levy, R. M. Theory of Binless Multi-State Free Energy Estimation with Applications to Protein-Ligand Binding. *J. Chem. Phys.* **2012**, *136*, 144102.

(70) Tan, Z.; Xia, J.; Zhang, B. W.; Levy, R. M. Locally Weighted Histogram Analysis and Stochastic Solution for Large-Scale Multi-State Free Energy Estimation. *J. Chem. Phys.* **2016**, *144*, 034107.

(71) Zhang, B. W.; Xia, J.; Tan, Z.; Levy, R. M. A Stochastic Solution to the Unbinned WHAM Equations. *J. Phys. Chem. Lett.* **2015**, *6*, 3834−3840.

(72) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91*, 43−56.

(73) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(74) Jorgensen, W. L.; Jenson, C. Temperature Dependence of TIP3P, SPC, and TIP4P Water from NPT Monte Carlo Simulations: Seeking Temperatures of Maximum Density. *J. Comput. Chem.* **1998**, *19*, 1179−1186.

(75) Vanommeslaeghe, K.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **2012**, *52*, 3144−3154.

(76) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **2012**, *52*, 3155−3168.

(77) 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.