THE JOURNAL OF PHYSICAL CHEMISTRY B

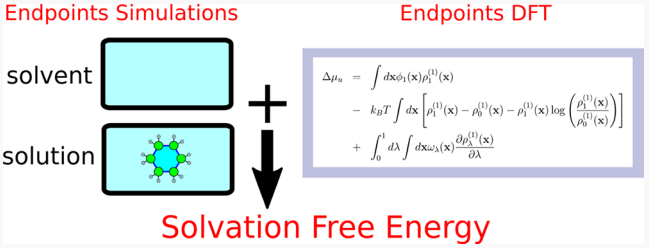
pubs.acs.org/JPCB Perspective

Solvation Thermodynamics from the Perspective of Endpoints DFT

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

Cite This: J. Phys. Chem. B 2020, 124, 11771−11782

ABSTRACT: Solvation thermodynamics is concerned with the evaluation and physical interpretation of solvation free energies. Endpoints DFT provides a framework for computing solvation free energies by combining molecular simulations with a version of the classical density-functional theory of solutions which focuses on ω, the indirect (solvent-mediated) part of the solute−solvent potential of mean force (indirect PMF). The simulations are performed at the endpoints of a hypothetical charging process which transforms the solvent density from the pure liquid state to that of the solution state. The endpoints DFT expression for solvation free energy can be shown to be equivalent to the standard expression for which the key quantity is the direct correlation function, but it has the advantage that the indirect term ω is more focused on the change in solvent−solvent correlations with respect to the pure liquid as the solute is inserted into the solution. In this Perspective, we review recent developments of endpoints DFT, highlighting a series of papers we have written together beginning in 2017. We emphasize the importance of dimensionality reduction as the key to the evaluation of endpoints DFT expressions and present a recently developed, spatially resolved version of the theory. The role of interfacial water at certain positions which stabilize or destabilize a solute in solution can be analyzed with the spatially resolved version, and it is of considerable interest to investigate how changes in solvation affect protein−ligand binding and conformational landscapes from an endpoints DFT perspective. Endpoints DFT can also be employed in materials science; an example involving the rational design strategy for polymer membrane separation is described. The endpoints DFT method is a scheme to evaluate the solvation free energy by introducing approximations to integrate the classical density functional over a charging parameter. We have further proposed a new functional which captures the correct dependence of the indirect PMF ω at both endpoints of the charging process, and we review how it might be employed in future work.



1. INTRODUCTION

Statistical thermodynamic theories of solvation have deep roots in chemical physics, with the two main branches being inhomogeneous solvation theory (IST)[1−5] and classical density-functional theory (DFT).[2,6,7] In the 1990s, we wrote a series of three papers from the IST perspective about hydration shell models of solvation, which attributes the thermodynamic response when a solute is inserted into a liquid to a locally perturbed region of solvent adjacent to the solute.[8−10] The central focus of the IST formalism is the expansion of the solvation energy and entropy[11] in a series of solute−solvent multipoint correlation functions; the numerical implementation requires a judicious choice of truncation scheme. Classical DFT leads through the variational principle to an integral equation theory of condensed phase fluids and solutions that dates back to the 1950s;[2,7] it has elements in common with the density-functional theory formalism employed in widely used quantum chemistry computational packages. The central focus of classical DFT is the direct correlation function, which formally corresponds to the functional derivative of the free energy with respect to a change in the density. The earliest applications of classical DFT were to simple atomic liquids.[2] The structure and thermodynamics of these liquids which have a high degree of symmetry can be determined by first solving a pair of integral equations for the total and direct correlation functions, given the intermolecular potential. For complex liquids like water and solutions containing solutes dissolved in water, it is much more difficult to extract solute excess chemical potentials by solving the DFT equations starting from knowledge only of the intermolecular potentials, and the integral-equation methods have been mainly developed in terms of site−site radial distribution functions and three-dimensional versions of the reference interaction site model (RISM);[2,12−20] a local molecular field integral equation theory of solvation has also been developed in recent years.[21−23] A feature of the DFT approach is that it does not use an alchemical path for the insertion of the solute into solution. This stands in contrast to

Received: October 2, 2020
Revised: November 19, 2020
Published: December 11, 2020

https://dx.doi.org/10.1021/acs.jpcb.0c08988
J. Phys. Chem. B 2020, 124, 11771−11782

the much more common computational approaches like free-energy perturbation which involve explicit simulations along an alchemical change of the state,[24−28] or approaches which involve reweighting the two endpoint simulations when the states have sufficient overlap.[29−31] Explicit (brute force) free-energy simulations over alchemical intermediate states can be much more readily carried out today than a decade or two ago, but they are still computationally demanding. Many thermodynamic processes of interest are not accessible by this brute force approach, and little physical insight is provided by the structural data generated at the alchemical intermediate states.

In the early 2000s, one of us (N.M.) constructed a DFT theory of solvation thermodynamics which departed in two important respects from other DFT-based approaches employed at the time.[32−34] First, structural data extracted from simulations of both physical endpoints of the solvation process, the pure solvent and the solution with fully coupled solute, were used to evaluate the DFT expression for the excess chemical potential. The name "endpoints DFT" makes the use of data from both physical endpoints explicit. Second, the DFT equations for the excess chemical potential were formulated in a one-dimensional energy representation, rather than the standard six-dimensional coordinate (position and orientation) representation; this greatly facilitated the ability to converge the density functionals which appear in the DFT expression for the excess chemical potential. Another important feature of "endpoints DFT" is its focus on evaluating ω, the indirect (solvent-mediated) contribution to the solute−solvent potential of mean force (called indirect PMF hereafter). ω plays the central role in endpoints DFT as the direct correlation function c does in classical DFT, and is more focused on the solvent−solvent effects on the solute−solvent correlations, as discussed below.

This Perspective reviews recent developments of endpoints DFT, highlighting a series of papers we have written together beginning in 2017.[35−40] In the following section, we provide a brief introduction to, and overview of, our recent collaborative work. In section 3, we review the theoretical framework and implementation of endpoints DFT, including the approximations used to evaluate the endpoints DFT equations, especially dimensionality reduction and our construction of a new endpoints density functional based on the two-points quadratic HNC (hypernetted-chain) approximation. In section 4, we review some recent applications of endpoints DFT to problems in biological and materials science.

2. RECENT DEVELOPMENTS IN THE ENDPOINTS DFT APPROACH TO SOLVATION THERMODYNAMICS

2.1. DFT Framework in Comparison to IST. In 2017, we published a paper which examined the relationship between the IST and DFT frameworks for analyzing solvation thermodynamics based on solute−solvent and solvent−solvent correlation functions.[35] When the expansion of the entropy in terms of the correlation functions is carried out to all orders in IST and the evaluation of the density-functional charging integral is performed exactly in DFT, both DFT and IST provide exact expressions for the solute chemical potential in solution. In practice, approximations must be made. Both IST and DFT contain an explicit term which when integrated over the solution gives the one-body solute−solvent contribution to the solvation free energy. We pointed out that, when the excess energy and excess entropy are integrated separately, there can be a nonlocal and ensemble-dependent contribution to both terms which indeed cancels when both terms are evaluated exactly but may not cancel when the two terms are each approximated and evaluated separately. The "standard" lowest-order approximation in IST is to include the second-order solvent−solvent energy change in the IST expression, while the "standard" approximation employed in DFT is to use the HNC and PY (Percus−Yevick) closure relations to evaluate the density functional. The relative accuracy of these different approximations can only be ascertained by comparison with exact free-energy simulations of solute insertion given a common force field. However, it is customary in the IST approaches to include the two-body solvent−solvent energy and omit the two-body and higher-order entropy terms. This approximation may be unbalanced because the energy and entropy will tend to compensate each other. Finally, we noted that the one-body free-energy density computed from the position-dependent energy and entropy of the solvent molecule does not correspond to the free energy of transfer of the solvent from the bulk to a position proximate to the solute. This observation has important implications for IST- and DFT-based analyses of the role of solvation in molecular association.

2.2. Endpoints DFT in Reduced Dimensions. The endpoints DFT formalism was introduced in the early 2000s.[32−34] As was emphasized in the first papers, the numerical implementation of the fundamental equations is not practical in the full coordinate representation because it requires the calculation of high-dimensional density functions involving the position, orientation, and intramolecular degrees of freedom of all of the solvent and solute molecules, but this problem can be overcome by projecting to lower dimensions. The original implementation of endpoints DFT and virtually all of the subsequent work has been carried out in a one-dimensional energy representation.[41] A widely distributed software package which uses this one-dimensional representation, ERmod, is available for endpoints DFT calculations.[42] The price paid for this reduction in dimensionality is the loss of spatial resolution, but there are problems for which a spatially resolved version of endpoints DFT is needed. In work published in 2019, we derived a new endpoints DFT expression in a mixed four-dimensional representation (three dimensions for position and one for energy).[39] This opens the way to identifying regions proximate to a solute which provide net stabilizing vs destabilizing contributions to the solvation free energy. We expect this mixed four-dimensional position-energy representation to be especially useful in future applications of endpoints DFT for analyzing the role that interfacial water plays in protein−ligand binding and many other molecular recognition processes as well as to mapping protein conformational free-energy landscapes.

2.3. The Cavity Particle and a New Endpoints DFT Functional. In our most recent joint publication, we showed the equivalence between the density-functional theory of solvation thermodynamics based on standard DFT expressions that employ the direct correlation function $c(\mathbf{x})$ and the endpoints DFT formalism which focuses on the indirect PMF $\omega(\mathbf{x})$.[40] $\omega(\mathbf{x})$ is closely related to the so-called cavity distribution function $y(\mathbf{x})$ which is the distribution function of a tagged solvent molecule which interacts normally with all of the other solvent molecules but does not interact with the solute. The relationship between $\omega(\mathbf{x})$ and $y(\mathbf{x})$ is given by $\omega(\mathbf{x}) = -k_BT \log y(\mathbf{x})$. A key idea underlying endpoints DFT is to use simulation data to determine the correct behavior of the functional at the physical endpoints, and the integral over the charging parameter is then approximated by a linear form using

https://dx.doi.org/10.1021/acs.jpcb.0c08988
J. Phys. Chem. B 2020, 124, 11771−11782

either HNC or PY approximation. Because the integration within the excluded volume region of a solute makes a substantial contribution to the solute chemical potential, it is desirable to estimate this contribution as accurately as possible. However, it is not generally possible to determine $\omega(\mathbf{x})$ within the excluded volume region of the solute molecule without employing special sampling methods that amount to simulating a cavity particle in the solution containing the solute and solvent. In order to assess the accuracy of the endpoints DFT approximations used within the solute excluded volume region, we performed explicit simulations of a cavity particle in solutions containing model hydrophobic solutes of different sizes.[40] We found that, for solutes whose sizes are greater than or equal to ~1.5× the diameter of a water molecule, $\omega(\mathbf{x})$ changes sign from negative to positive as the cavity particle penetrates deeper into the hydrophobic core, whereas the sign remains negative when the HNC or PY approximations are used to estimate $\omega(\mathbf{x})$. It was also observed that the water molecules in the first shell adapt well to the hydrogen bonds of the cavity particle located near the solute surface, which may be related to the enhanced fluctuations of water at hydrophobic interfaces.[43] Based on the results of the cavity simulations, we constructed a new density functional which gives the correct behavior of $\omega(\mathbf{x})$ inside the cores of larger solutes as well as having the correct gradient with respect to the variation of the solute−solvent distribution at the pure liquid state. The new functional varies quadratically with the density change and gives the exact result at the two endpoints, hence our name 2P-QHNC (two-points quadratic HNC; see ref 40 for the detailed naming convention concerning 2P-QHNC). We found that the new 2P-QHNC functional leads to improved estimates for the excess chemical potential of idealized hydrophobic solutes as the solute size grows, but its construction relies on cavity particle simulations which are costly. We have begun to explore alternative routes to the calculation of $\omega(\mathbf{x})$ in the solution system with fully coupled solute−solvent interaction in the interfacial solute−solvent region using reweighting techniques instead of explicitly simulating cavity particles.[40]

We now present a more complete description of the endpoints DFT framework for estimating solvation free energies from simulations of the two physical endpoint states, the pure liquid and the solute in solution.

3. THEORETICAL DEVELOPMENTS

3.1. DFT Approach to Solvation Free Energy. The focus of our developments is the solvation free energy $\Delta\mu$. It is the free-energy change for turning on the solute−solvent interaction and is to be expressed as a functional of solute−solvent distribution functions in the DFT approach. Our formulation starts from the Kirkwood charging formula.[35,40,41] We let $\mathbf{x}$ be the full coordinate of position and orientation of the solvent molecule relative to the solute (with the intramolecular degrees of freedom if the molecule is flexible), and the solute−solvent intermolecular interaction is assumed to be pairwise additive. The solute−solvent pair potential of interest is $v(\mathbf{x})$, and to formulate the charging formula, the coupling parameter λ ($0 \le \lambda \le 1$) is employed to introduce a set of solute−solvent interaction potentials $u_\lambda(\mathbf{x})$. When $\lambda = 0$, $u_0(\mathbf{x}) = 0$ and the statistical ensemble is generated without solute−solvent interactions. When $\lambda = 1$, $u_1(\mathbf{x}) = v(\mathbf{x})$ and the system is the solution system of interest. $\Delta\mu$ is then expressed as an integral over λ as

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

where $\rho_\lambda^f(\mathbf{x})$ is the one-body distribution function of solvent around the solute and the superscript f is attached to show that the distribution is represented over the full coordinate. It should be noted that $\rho_\lambda^f(\mathbf{x})$ is obtained in the presence of the solute−solvent interaction $u_\lambda(\mathbf{x})$.

The key quantity in our DFT treatment is the indirect (solvent-mediated) part of the potential of mean force $\omega_\lambda^f(\mathbf{x})$ (denoted as indirect PMF) introduced by

$$\rho_\lambda^f(\mathbf{x}) = \rho_0^f(\mathbf{x}) \exp(-\beta(u_\lambda(\mathbf{x}) + \omega_\lambda^f(\mathbf{x}))) \tag{2}$$

where $\rho_0^f(\mathbf{x})$ is the distribution function in the pure solvent (reference solvent) at $\lambda = 0$ and β is the inverse of k_BT with the Boltzmann constant k_B and the temperature T. $\omega_\lambda^f(\mathbf{x})$ is called the indirect part, since $-k_BT \log(\rho_\lambda^f(\mathbf{x})/\rho_0^f(\mathbf{x}))$ and $u_\lambda(\mathbf{x})$ are the potential of mean force (PMF) and direct interaction between the solute and solvent molecules, respectively. $\exp(-\beta\omega_\lambda^f(\mathbf{x}))$ is further termed the cavity distribution function.[2] This function corresponds to the distribution of a "cavity" particle, for which the molecular structure is the same as that for the solvent molecule and only the interaction with the solute is turned off. $\omega_\lambda^f(\mathbf{x})$ quantifies the effect on the solute−solvent distribution of the change in the solvent−solvent correlation due to the solute insertion. With eq 2, eq 1 is modified through the use of partial integration into[33,35,40,41]

$$\Delta\mu = \int d\mathbf{x} v(\mathbf{x}) \rho^f(\mathbf{x}) - k_BT \int d\mathbf{x} \left[(\rho^f(\mathbf{x}) - \rho_0^f(\mathbf{x})) - \rho^f(\mathbf{x}) \log\left(\frac{\rho^f(\mathbf{x})}{\rho_0^f(\mathbf{x})}\right)\right] + \int_0^1 d\lambda \int d\mathbf{x} \omega_\lambda^f(\mathbf{x}) \frac{\partial\rho_\lambda^f(\mathbf{x})}{\partial\lambda} \tag{3}$$

where $\rho^f(\mathbf{x})$ denotes $\rho_1^f(\mathbf{x})$ in the solution system ($\lambda = 1$) for notational brevity. Equation 3 is an exact expression for the solvation free energy $\Delta\mu$. Its first term is the average sum of the interaction energy between the solute and solvent in the solution system of interest ($\lambda = 1$), and the second term corresponds to the pair entropy (in the units of k_BT) to quantify the deviation of the solute−solvent distribution in the solution ($\lambda = 1$) from that in the pure solvent ($\lambda = 0$). The third term is written as an integral of $\omega_\lambda^f(\mathbf{x})$ over the coupling parameter λ and incorporates the many-body effects from the variations in solvent−solvent correlations with λ.

The standard DFT approach has been formulated by employing the intrinsic free energy and a series of direct correlation functions.[2,7,44] The one-body direct correlation function $c_\lambda^{(1)}(\mathbf{x})$ is introduced by

$$\rho_\lambda^{(1)}(\mathbf{x}) = z \exp(-\beta u_\lambda(\mathbf{x}) + c_\lambda^{(1)}(\mathbf{x})) \tag{4}$$

where $z = \exp(\beta\mu_v)/\Lambda^3$ with the chemical potential μ_v and the thermal de Broglie length Λ of the solvent. The solvation free energy is then expressed in the grand-canonical ensemble as

https://dx.doi.org/10.1021/acs.jpcb.0c08988
J. Phys. Chem. B 2020, 124, 11771−11782

$$\Delta\mu = \int d\mathbf{x} v(\mathbf{x})\rho^f(\mathbf{x}) + (F_1 - F_0) - \mu_v \int d\mathbf{x}(\rho^f(\mathbf{x}) - \rho_0^f(\mathbf{x})) \quad (5)$$

where F_λ is the intrinsic free energy at the coupling parameter λ. It was shown[40] by virtue of

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

that eq 3 is equivalent to eq 5. According to eq 6, $\omega_\lambda^f(\mathbf{x})$ is more focused on the solute−solvent effects than $c_\lambda^{(1)}(\mathbf{x})$. This is because the one-body direct correlation function $c_\lambda^{(1)}(\mathbf{x})$ corresponds to the free-energy change (divided by $-k_BT$) for transferring a cavity particle from vacuum to configuration $\mathbf{x}$. It is referenced to vacuum and is zero when the solvent is absent. On the other hand, $\omega_\lambda^f(\mathbf{x})$ vanishes beyond a certain correlation length and becomes zero in the bulk region (far from the solute) of the solution (when the thermodynamic limit is achieved).

Our formulation starts from the Kirkwood charging formula of eq 1 and is advantageous in that it is valid in any ensemble including the grand-canonical, canonical, and isothermal−isobaric. The intricacies due to the conservation of the number of solvent molecules in the canonical and isothermal−isobaric ensembles can be further resolved by setting the solute−solvent potentials to zero at far separations.[34,40−42] Therefore, eq 3 is readily amenable for evaluation of $\Delta\mu$ when combined with molecular simulations through the scheme of dimensionality reduction described in the next subsection.

An approximate expression for $\Delta\mu$ is called endpoints DFT when it is constructed from distribution functions obtained from molecular simulations at the two endpoint states of physical interest: the pure solvent ($\lambda = 0$) and the solution ($\lambda = 1$). Within the framework of eq 3, an endpoints expression can be obtained by introducing approximations to $\omega_\lambda^f(\mathbf{x})$. The Percus−Yevick (PY) and hypernetted-chain (HNC) approximations are formulated by adopting the linear dependencies on λ of the cavity distribution function $\exp(-\beta\omega_\lambda^f(\mathbf{x}))$ and the indirect PMF $\omega_\lambda^f(\mathbf{x})$, respectively, under the linear variation of the solute−solvent distribution through

$$\rho_\lambda^f(\mathbf{x}) = \lambda\rho^f(\mathbf{x}) + (1-\lambda)\rho_0^f(\mathbf{x}) \quad (7)$$

The integration over λ in the third term of eq 3 can then be performed analytically with eq 7, leading to an endpoints formula for $\Delta\mu$. The PY and HNC relationships correspond to the first-order expansions with respect to the variation of the solute−solvent distribution. A higher-order scheme can be proposed through a quadratic dependence of the indirect PMF on λ, as will be presented in the later subsection which describes an improved endpoints functional. It should be noted that the contrast between integral-equation approaches and the endpoints method lies in the schemes for obtaining the distribution functions and carrying out the integration over the coupling parameter λ to determine $\Delta\mu$. In the former,[2,6,7] a set of distribution functions are usually determined in self-consistent and iterative manners and care needs to be taken for the integration over λ for some choices of closure relationships (e.g., when using bridge functions).[45] In the latter, the integration is already performed over λ for the approximate functional of free energy and a set of distribution functions obtained from molecular simulations are directly substituted into the functional without using iterative procedures.

3.2. DFT Approach in Reduced Dimensions. When the molecules of interest are polyatomic, in fact, the numerical implementation of eq 3 is not practical. This is because the full set of coordinates of position and orientation (with intramolecular degrees of freedom for flexible species) is high-dimensional and a distribution function represented over a coordinate of high dimension is not straightforward to handle. The problem of dimensionality can be overcome by projection to a lower-dimensional coordinate. In the present subsection, we introduce the energy representation and provide a DFT formalism over the energy coordinate.

The energy representation is a scheme for dimensionality reduction in the density-functional method.[32−34,40−42] In this representation, the solvent distribution around the solute is expressed with respect to the solute−solvent pair interaction $v(\mathbf{x})$, where $\mathbf{x}$ is the full coordinate. The instantaneous distribution $\hat{\rho}^e(\epsilon)$ is then introduced as

$$\hat{\rho}^e(\epsilon) = \sum_i \delta(\epsilon - v(\mathbf{x}_i)) \quad (8)$$

where $v(\mathbf{x}_i)$ refers to the pair interaction between the solute and the ith solvent molecule and the superscript e signifies that the function is represented over the energy coordinate ϵ. With eq 8, the high-dimensional $\mathbf{x}$ is projected onto the one-dimensional ϵ and the information other than the pair-energy value is projected out.

In the energy representation, the density-functional method is formulated by restricting the solute−solvent interaction potentials u_λ in the charging formula of eq 1 to those with the composite form of $u_\lambda(v(\mathbf{x}))$, where $v(\mathbf{x})$ is the potential function in the solution system of interest ($\lambda = 1$). u_λ can then be expressed as $u_\lambda(\epsilon)$ when ϵ denotes the value of $v(\mathbf{x})$. At the endpoints, $u_0(\epsilon) = 0$ and $u_1(\epsilon) = \epsilon$, since $v(\mathbf{x})$ itself is the potential at $\lambda = 1$. Equation 1 then reduces to

$$\Delta\mu = \int_0^1 d\lambda \int d\epsilon \frac{\partial u_\lambda(\epsilon)}{\partial\lambda}\rho_\lambda^e(\epsilon) \quad (9)$$

where $\rho_\lambda^e(\epsilon)$ is the ensemble average of eq 8 in the presence of the solute−solvent interaction u_λ. This is the charging formula in the energy representation. It is an exact expression with the distribution function represented over the one-dimensional coordinate. The formulation in the energy representation can be developed in parallel to that from eqs 1−3 in the full-coordinate representation. The indirect PMF $\omega_\lambda^e(\epsilon)$ is introduced by

$$\rho_\lambda^e(\epsilon) = \rho_0^e(\epsilon)\exp(-\beta\{u_\lambda(\epsilon) + \omega_\lambda^e(\epsilon)\}) \quad (10)$$

and the solvation free energy $\Delta\mu$ is expressed as

$$\Delta\mu = \int d\epsilon\epsilon\rho^e(\epsilon) - k_BT\int d\epsilon\left[(\rho^e(\epsilon) - \rho_0^e(\epsilon)) - \rho^e(\epsilon)\log\left(\frac{\rho^e(\epsilon)}{\rho_0^e(\epsilon)}\right)\right] + \int_0^1 d\lambda\int d\epsilon\omega_\lambda^e(\epsilon)\frac{\partial\rho_\lambda^e(\epsilon)}{\partial\lambda} \quad (11)$$

In eqs 10 and 11, $\rho_0^e(\epsilon)$ refers to the pure solvent at $\lambda = 0$ and $\rho^e(\epsilon)$ denotes $\rho_1^e(\epsilon)$ for the brevity of notation. The parallelism is striking between eqs 9, 10, and 11 and eqs 1, 2, and 3. This is due to the use of the composite form of u_λ, and it is anticipated

https://dx.doi.org/10.1021/acs.jpcb.0c08988
J. Phys. Chem. B 2020, 124, 11771−11782

that many of theoretical features in the full-coordinate DFT that cannot be implemented in practice due to the high dimensionality of the full coordinate will be numerically feasible over the one-dimensional coordinate ϵ. Equation 11 is exact, and its structure is in correspondence to that of eq 3. The first term is the average sum of the solute−solvent interaction energy in the solution ($\lambda = 1$), and the second term refers to the pair entropy expressed over the one-dimensional coordinate ϵ. The variations in solvent−solvent correlations through the introduction of the solute−solvent interaction are incorporated in the third term through $\omega_\lambda^e(\epsilon)$, and an endpoint expression for $\Delta\mu$ is obtained by formulating the λ dependence of $\omega_\lambda^e(\epsilon)$.

An endpoints expression for $\Delta\mu$ is obtained in the energy representation when the λ dependence is approximated for the indirect PMF $\omega_\lambda^e(\epsilon)$ and the integration over λ is performed analytically in the third term of eq 11. The PY-type and HNC-type approximations in the energy representation can be introduced by taking $u_\lambda(\epsilon)$ so that

$$\rho_\lambda^e(\epsilon) = \lambda\rho^e(\epsilon) + (1-\lambda)\rho_0^e(\epsilon) \tag{12}$$

as in eq 7. These approximations are formulated as the linear dependencies on λ of the cavity distribution function $\exp(-\beta\omega_\lambda^e(\epsilon))$ and the indirect PMF $\omega_\lambda^e(\epsilon)$, respectively, over the energy coordinate ϵ. Since $\omega_\lambda^e(\epsilon) = 0$ at $\lambda = 0$ by definition of eq 10, the linear variation with respect to λ can be written in terms of $\omega^e(\epsilon)$ or $\sigma_0^e(\epsilon)$, where $\omega^e(\epsilon)$ means $\omega_\lambda^e(\epsilon)$ at $\lambda = 1$ for notational brevity and $\sigma_0^e(\epsilon)$ is defined as the derivative at $\lambda = 0$ through

$$\sigma_0^e(\epsilon) = \left.\frac{\partial\omega_\lambda^e(\epsilon)}{\partial\lambda}\right|_{\lambda=0} \tag{13}$$

The combined PY−HNC relationship is adopted to approximate the integral over λ in the third term of eq 11, and the endpoints expression for $\Delta\mu$ thus formulated has been implemented in a software package ERmod.[42] It is publicly available on the web at http://sourceforge.net/projects/ermod and can be used in combination with common molecular dynamics (MD) simulation packages.

The formulation in the energy representation is parallel to that in the full-coordinate representation. It is required in a density-functional treatment, in fact, that the coordinate used to construct the distribution functions be one that suffices to determine the value of the interaction potential. The full coordinate certainly satisfies this requirement, and the energy is the coordinate with minimum dimensionality that satisfies the requirement. The relationship between the interaction potential and the distribution function is further one-to-one when the potential is set to be zero at far separations between the solute and solvent in the canonical and isothermal−isobaric ensembles. The spatial information on the solvent distribution around the solute is lost in the energy representation, though. To restore the spatial resolution at the lowest possible dimensionality, we proposed a mixed representation over the position and energy, in which the energy serves as a proxy for all of the variables other than the position, including the orientation and the intramolecular flexibility.[39] The mixed representation over the position $\mathbf{r}$ and the pair-energy ϵ of a solvent molecule is formulated by defining the instantaneous distribution $\hat{\rho}^m(\mathbf{r}, \epsilon)$ as

$$\hat{\rho}^m(\mathbf{r}, \epsilon) = \sum_i \delta(\mathbf{r} - \mathbf{r}_i)\delta(\epsilon - v(\mathbf{x}_i)) \tag{14}$$

where $\mathbf{r}_i$ and $v(\mathbf{x}_i)$ are the position of the ith solvent molecule relative to the solute and the pair energy between them, respectively, and the superscript m stands for the mixed representation. $\mathbf{r}$ can be set to the center of mass of the solvent molecule or refer to the oxygen site in the case of water.

A DFT expression for $\Delta\mu$ in the mixed representation can be formulated in parallel to those in the full-coordinate and energy representations by adopting the set of solute−solvent interactions with the composite form of $u_\lambda(\mathbf{r}, \epsilon)$, where $u_0(\mathbf{r}, \epsilon) = 0$ and $u_1(\mathbf{r}, \epsilon) = \epsilon$. The Kirkwood charging formula is then written as

$$\Delta\mu = \int_0^1 d\lambda \int d\mathbf{r}d\epsilon \frac{\partial u_\lambda(\mathbf{r}, \epsilon)}{\partial\lambda}\rho_\lambda^m(\mathbf{r}, \epsilon) \tag{15}$$

where $\rho_\lambda^m(\mathbf{r}, \epsilon)$ is the ensemble average of eq 14 at the coupling parameter λ. $\rho_\lambda^m(\mathbf{r}, \epsilon)$ is obtained by marginalizing the full-coordinate distribution $\rho_\lambda^f(\mathbf{x})$, and the information other than the position and energy is projected out. It should be noted that eq 15 can be formulated in this way, since the energy ϵ is retained as a coordinate along with the position $\mathbf{r}$. In fact, the four-dimensional density $\rho_\lambda^m(\mathbf{r}, \epsilon)$ is a function determined uniquely through the one-to-one relationship by the solute−solvent interaction potential. When the coordinate set used to construct the solute−solvent density function is not enough to specify the value of the interaction potential, for example, it is the water oxygen coordinate $\mathbf{r}$ only, an expression like eq 15 cannot be formulated. The indirect PMF $\omega_\lambda^m(\mathbf{r}, \epsilon)$ is further introduced by

$$\rho_\lambda^m(\mathbf{r}, \epsilon) = \rho_0^m(\mathbf{r}, \epsilon)\exp(-\beta\{u_\lambda(\mathbf{r}, \epsilon) + \omega_\lambda^m(\mathbf{r}, \epsilon)\}) \tag{16}$$

and the spatially decomposed DFT expression is given by

$$\Delta\mu = \int d\mathbf{r}\Delta\mu(\mathbf{r}) \tag{17}$$

$$\Delta\mu(\mathbf{r}) = \int d\epsilon\epsilon\rho^m(\mathbf{r}, \epsilon) - k_BT\int d\epsilon\left[\rho^m(\mathbf{r}, \epsilon) - \rho_0^m(\mathbf{r}, \epsilon) - \rho^m(\mathbf{r}, \epsilon)\log\left(\frac{\rho^m(\mathbf{r}, \epsilon)}{\rho_0^m(\mathbf{r}, \epsilon)}\right)\right] + \int_0^1 d\lambda\int d\epsilon\omega_\lambda^m(\mathbf{r}, \epsilon)\frac{\partial\rho_\lambda^m(\mathbf{r}, \epsilon)}{\partial\lambda} \tag{18}$$

Equations 17 and 18 are an exact formula. $\Delta\mu(\mathbf{r})$ is interpreted as the free-energy density at $\mathbf{r}$, and each term in eq 18 has a similar physical meaning to the corresponding terms in eqs 3 and 11. Approximations to the third term of eq 18 can be introduced similarly to the case of (one-dimensional) energy representation, and the PY-type and HNC-type forms were adopted in ref 39 as in ref 42. In numerical implementations, the spatial position in eq 14 needs to be discretized. A grid can be employed, or the space is divided with respect to the solvation shells around chemically important groups of atoms.

3.3. Improved Endpoints Functional. The indirect PMF $\omega_\lambda^e(\epsilon)$ plays a key role in eq 11 and is approximated in the previous subsection as the linear dependencies on λ of $\exp(-\beta\omega_\lambda^e(\epsilon))$ and $\omega_\lambda^e(\epsilon)$. The linear approximations are formulated in terms of either $\sigma_0^e(\epsilon)$ or $\omega^e(\epsilon)$. An improved

https://dx.doi.org/10.1021/acs.jpcb.0c08988
J. Phys. Chem. B 2020, 124, 11771−11782

approximation may then be formulated when both of $\sigma_0^e(\epsilon)$ and $\omega^e(\epsilon)$ are employed to model the λ dependence of $\omega_\lambda^e(\epsilon)$ under eq 12.[40] Equation 13 at $\lambda = 0$ and $\omega_\lambda^e(\epsilon) = \omega^e(\epsilon)$ at $\lambda = 1$ can both be enforced with a quadratic form of

$$\omega_\lambda^e(\epsilon) = \lambda(1-\lambda)\sigma_0^e(\epsilon) + \lambda^2\omega^e(\epsilon) \tag{19}$$

and with this form, the third term of eq 11 leads with eq 12 to

$$\int_0^1 d\lambda \int d\epsilon \omega_\lambda^e(\epsilon)\frac{\partial \rho_\lambda^e(\epsilon)}{\partial \lambda} = \left(\frac{1}{6}\sigma_0^e(\epsilon) + \frac{1}{3}\omega^e(\epsilon)\right)(\rho^e(\epsilon) - \rho_0^e(\epsilon)) \tag{20}$$

The λ dependence of $\omega_\lambda^e(\epsilon)$ is linear in the HNC-type expression and is extended in eq 19 to the quadratic form by employing $\sigma_0^e(\epsilon)$ and $\omega^e(\epsilon)$ at the two endpoint states of $\lambda = 0$ and 1. Equation 19 was thus called two-points quadratic HNC (2P-QHNC) in ref 40, where the $\Delta\mu$ estimation was observed to improve significantly for large solutes.

In the endpoints functional introduced in the previous subsection, $\omega_\lambda^e(\epsilon)$ is approximately treated in terms of $\sigma_0^e(\epsilon)$ in the repulsive-core region of the solute due to the use of a weighting function.[42] $\sigma_0^e(\epsilon) < 0$ is usually observed in the solute core, and the PY-type and HNC-type approximations lead to $\omega_\lambda^e(\epsilon) < 0$ at $\epsilon \gg 0$. $\omega^e(\epsilon)$ ($=\omega_\lambda^e(\epsilon)$ at $\lambda = 1$) is equal to the negative of the solvation free energy of the solvent molecule in its bulk, on the other hand, when the solute is large enough and the cavity particle is located in the deep-core region of the solute. This is because $\omega_\lambda^e(\epsilon)$ is the transfer free energy of the cavity particle from the bulk to the location specified by the energy coordinate ϵ and the cavity particle does not interact with the other solvent molecules in the deep core of a large enough solute. $\omega_\lambda^e(\epsilon)$ is thus positive at $\epsilon \gg 0$ when the solvent is ambient water, for example, and the contribution from the deep-core region will be too positive for a large solute in the approximate calculation with the original functional in ref 42 compared to the exact treatment with the exact $\omega_\lambda^e(\epsilon)$. Equation 19 is a scheme to capture the correct behavior of $\omega_\lambda^e(\epsilon)$ in the deep-core region of a large solute, and the results obtained in ref 40 suggest that it is an improved functional for $\Delta\mu$.

Given that $\epsilon \gg k_BT$ and $\rho^e(\epsilon) \approx 0$ in the deep-core region, the evaluation of $\omega^e(\epsilon)$ from eq 10 at $\lambda = 1$ is usually not possible there when molecular simulations are conducted only at $\lambda = 0$ and 1 without employing an advanced strategy such as umbrella sampling.[40] It was actually observed, though, that the density of the solvation free energy in the excluded-volume region depends weakly on the structure (conformation) of the solute molecule,[39] and accordingly, the repulsive-interaction effect on the structural change is expected to reflect only the contributions from the "skin" parts that are close to the solute surface. A reweighting technique based on UWHAM (unbinned weighted histogram analysis method)[38,46] can then be used to construct ω^e in the "skin" parts efficiently, and the 2P-QHNC functional will be of use to study the solvent effects on the conformational changes of biomolecules.

4. APPLICATIONS

4.1. The Role of Interfacial Water in Protein−Ligand Binding and Conformational Free-Energy Differences. It is widely believed that the displacement of specific water molecules from the binding site at protein receptor surfaces plays a key role in determining protein−ligand binding affinities.[47−51] The simple idea, cast in the IST framework, is that interfacial water molecules located at positions for which the sum of the excess energy and entropy is unfavorable will make a favorable contribution to binding if they are replaced by a ligand functional group. This is the basis of the popular "Watermap" tool that has been used for lead optimization in drug design.[47−51] The idea can also be analyzed within the context of endpoints DFT, with the goal of identifying the thermodynamic signatures of interfacial water molecules, some of which are observed in crystal structures, which are predicted to provide the largest contribution to the binding affinity when displaced by the functional group of a ligand.[37] It is important to distinguish the contribution of interfacial water to the stability (excess chemical potential) of the protein from its possible role in protein ligand binding. To clarify this point, we rewrite eq 3 for a protein solute as a functional of the density of water $\rho(\mathbf{x})$ at the interface and the indirect PMF $\omega(\mathbf{x})$

$$\Delta\mu = -k_BT\int d\mathbf{x}[\rho(\mathbf{x}) - \rho_0(\mathbf{x})] + \int d\mathbf{x}\rho(\mathbf{x})[-\omega(\mathbf{x})] + \int_0^1 d\lambda \int d\mathbf{x}\omega_\lambda(\mathbf{x})\frac{\partial \rho_\lambda(\mathbf{x})}{\partial \lambda} \tag{21}$$

where the superscript f has been removed. It was shown in ref 10 that the behaviors of distribution functions at far separations between the solute and solvent give rise to the ensemble dependence of a thermodynamic quantity. In eq 21, the contributions from the far separations are actually canceled, and eq 21 is valid in any ensemble. For a more detailed discussion of this point, see refs 35 and 37. Equation 21 highlights the central role of $\omega(\mathbf{x})$. When a water molecule located at the interface $\mathbf{x}$ interacts more favorably with the other solvent molecules than one in the bulk, $\omega(\mathbf{x}) < 0$ and this water makes an unfavorable contribution to the stability of the protein in solution through the second term in eq 21. When the displaced water is subject to an unfavorable interaction with the other solvent molecules compared with bulk $\omega(\mathbf{x}) > 0$, water at this location makes a favorable contribution to $\Delta\mu$ through the second term in eq 21.

Furthermore, the total PMF $\mathrm{WT}(\mathbf{x})$ to transfer a water molecule from the pure solvent to location $\mathbf{x}$ at the interfacial region can be expressed as

$$\mathrm{WT}(\mathbf{x}) = -k_BT\ln\frac{\rho(\mathbf{x})}{\rho_0(\mathbf{x})} = Ts^{(1)}(\mathbf{x}) \tag{22}$$

where $s^{(1)}(\mathbf{x})$ is the one-body term of the space-resolved entropy at location $\mathbf{x}$ in solution from the IST expression.[37] Since the direct part of the PMF $v(\mathbf{x})$ is equivalent to the one-body term (solute−solvent term) of the space-resolved energy in the IST expression $e^{(1)}(\mathbf{x})$, the indirect PMF corresponds to the one-body term in the IST (energy + entropy) expansion for the solvent excess chemical potential:

$$\omega(\mathbf{x}) = \mathrm{WT}(\mathbf{x}) - v(\mathbf{x}) = -[e^{(1)}(\mathbf{x}) - Ts^{(1)}(\mathbf{x})] \tag{23}$$

When expressed in this way, it is now clear from eqs 21, 22, and 23 why water molecules at the interface with a repulsive indirect PMF make a favorable (stabilizing) contribution at the one-body level to the excess chemical potential of the solute, while water molecules with an attractive indirect PMF make an unfavorable (destabilizing) contribution at the one-body level. When the indirect PMF $\omega(\mathbf{x})$ at $\mathbf{x}$ is positive, the one-body

https://dx.doi.org/10.1021/acs.jpcb.0c08988
J. Phys. Chem. B 2020, 124, 11771−11782

contribution to the free energy to move a water to **x** from the bulk (or pure solvent) is attractive and vice versa.

In approximate implementations of IST formulas for the solute chemical potential and for the analysis of interfacial solvent effects on protein−ligand binding, the two-body energy replaces the third term of eq 21, and the first term is dropped. Two-body and higher-order terms in the IST entropy expansion are usually not included in the IST analysis of the thermodynamic signatures of interfacial waters. A key distinction between IST and DFT is that IST includes two-body energies in the analysis while DFT takes care of the many-body effects as the variation of the indirect PMF against λ (third term of eq 21). It is conjectured that the DFT formulas which are based on the analysis of the indirect solute−solvent PMF, and therefore include the IST one-body energy and one-body entropy terms but not the two-body energy, provide a well balanced approximation, especially for strongly associating liquids such as water for which cancelation between the second-order energy and the second- and higher-order entropy terms is more likely.

In order to further motivate why the analysis of the indirect PMF ω holds information about the role of interfacial water in protein−ligand binding, we have designed a thermodynamic cycle to show how the relative binding free energy $\Delta\Delta G_{\text{bind}}$ between a pair of congeneric ligands can be expressed in a way that includes the contribution ω from displaced water explicitly.[37] The relative binding free energy $\Delta\Delta G_{\text{bind}}$ between a large ligand L and a small ligand S can be approximately calculated by the following formula

$$\begin{aligned}\Delta\Delta G_{\text{bind}} &= \Delta G_{\text{bind-L}} - \Delta G_{\text{bind-S}} \\ &\approx k_{\text{B}}T \int^{V^{\text{disp}}} \mathrm{d}\mathbf{x}[\rho_{\text{PS}}(\mathbf{x}) - \rho_0] \\ &\quad + \int^{V^{\text{disp}}} \mathrm{d}\mathbf{x}\rho_{\text{PS}}(\mathbf{x})\omega_{\text{PS}}(\mathbf{x}) + \Delta U_{\text{pro-lig}} \\ &\quad - \left\{k_{\text{B}}T \int^{V^{\text{disp}}} \mathrm{d}\mathbf{x}[\rho_{\text{S}}(\mathbf{x}) - \rho_0] \right. \\ &\quad \left. + \int^{V^{\text{disp}}} \mathrm{d}\mathbf{x}\rho_{\text{S}}(\mathbf{x})\omega_{\text{S}}(\mathbf{x})\right\} \\ &\quad + [\text{change in charging term} \\ &\quad + \text{terms for the integration outside } V^{\text{disp}} \\ &\quad \text{arising from the changes in } \rho \text{ and } \omega] \end{aligned} \tag{24}$$

where the subscript PS denotes the case that the solute is the complex of the protein and the small ligand; the subscript S denotes the case that the solute is the small ligand located in bulk water; V^{disp} is the region in which water is found in the system containing the small ligand but not in the system containing the large ligand; and $\Delta U_{\text{pro-lig}}$ is the difference of the interaction between the protein and ligand as the ligand which binds changes from small (S) to large (L). The first three terms capture the thermodynamic effects of displacing interfacial water closest to the protein and the change in the protein−ligand interactions when a functional group is added to a ligand core of a congeneric pair.

According to eq 24, we can suggest a scheme to search for water molecules to displace by ligand functional groups in terms of the thermodynamic signatures of interfacial water in Table 1.[37] We suppose that the functional group added to the ligand is polar when the displaced water is at the high density hydrophilic or bulk density hydrophilic position and is nonpolar when the displaced water is at the high density hydrophobic or the low density dry water position. For a "high density hydrophilic water", the difference of the interaction energies between the protein and ligand $\Delta U_{\text{pro-lig}}$ needs to be strongly negative to increase the binding strength because both the first and second terms of eq 24 are positive ($\rho_{\text{PS}}(\mathbf{x}) > \rho_0$ and $\omega_{\text{PS}}(\mathbf{x}) \gg 0$). Thus, a polar functional group is added to replace the interfacial water that has this thermodynamic signature. A similar argument holds for "bulk density hydrophilic water", since $\rho_{\text{PS}}(\mathbf{x}) \approx \rho_0$ and $\omega_{\text{PS}}(\mathbf{x}) \approx -\nu(\mathbf{x}) \gg 0$. When the displaced water is at a high density hydrophobic or low density dry water position, $\Delta U_{\text{pro-lig}}$ is considered to be small based on our supposition and $\nu \approx 0$ for both cases. Under the assumption that ν is approximately equal to zero, it can be proved that the sum of the first two terms in eq 24 is negative.[37] Therefore, these two terms account for most of $\Delta\Delta G_{\text{bind}}$ and make a favorable contribution. We also show in ref 37 that the magnitude of the sum of these two terms increases with $|\omega_{\text{PS}}|$, which agrees with the following argument that those interfacial water molecules with indirect PMF that are large in magnitude are the prime candidates to be replaced by a ligand functional group during a search of a protein receptor for possible binding sites. Furthermore, for the "low density dry water" ($\omega_{\text{PS}} < 0$), the sum of the first and second terms is bounded by $-k_{\text{B}}T\rho_0V^{\text{disp}}$.[37] This also agrees with our summary in Table 1 that the "low density dry water" is not a good target for ligand design because the corresponding $\Delta\Delta G_{\text{bind}}$ is small in magnitude.

Table 1. Thermodynamic Signatures of Interfacial Water (the Unit of Energy Is kcal/mol)[a]

type	WT	ρ/ρ_0	ν	ω	one-body effect on $\Delta\mu$	design target
high density hydrophilic	<0	>1	≤ -4	$\gg 0$	stabilize	yes
bulk density hydrophilic	≈ 0	≈ 1	≤ -4	$\approx -\nu$	stabilize	yes
high density hydrophobic	<0	>1	≈ 0	<0	destabilize	yes
low density dry water	>0	<1	≈ 0	>0	stabilize (small)	?
bulk density water	≈ 0	≈ 1	≈ 0	≈ 0		no

[a]Reprinted (adapted) with permission from ref 37. Copyright 2018 American Chemical Society.

In ref 37, we examined the total PMF WT, the direct interaction energy ν, and the indirect PMF ω of water molecules at binding sites on the surface of three proteins which are drug design targets and for which congeneric binding data are available. We constructed a simple fitting function to illustrate the empirical correlation between the change of binding free energy $\Delta\Delta G_{\text{bind}}$ of congeneric ligand pairs and the magnitude of the indirect PMF ω from the extra displaced water molecules

$$\Delta\Delta G_{\text{bind}} = A\left|\sum_{i=1}^{n}\omega_i\right| + B \tag{25}$$

where A and B are two fitting parameters and ω_i is the indirect PMF from the ith water molecule. $\Delta\Delta G_{\text{bind}}$ values based on eq 25 versus $\Delta\Delta G_{\text{bind}}$ values using the double decoupling charging method over alchemical states for 12 congeneric ligand pairs are compared in Figure 1. As can be seen, they agree well with a correlation of $R^2 = 0.84$.

https://dx.doi.org/10.1021/acs.jpcb.0c08988
J. Phys. Chem. B 2020, 124, 11771−11782

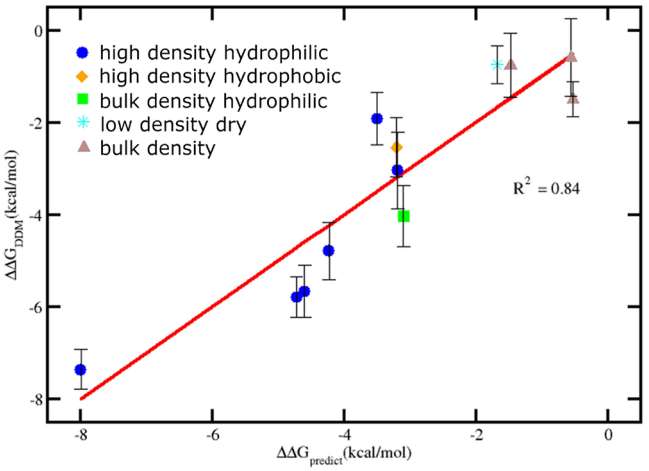


Figure 1. $\Delta\Delta G_{predict}$ based on eq 25 versus $\Delta\Delta G_{DDM}$ using the double decoupling method from 12 congeneric ligand pairs. The displaed water molecules are located at binding sites on the surface of coagulation factor Xa (FXa), streptavidin, and the mouse major urinary protein (MUP). Reprinted (adapted) with permission from ref 37. Copyright 2018 American Chemical Society.

There are two challenges to implementing a more systematic search for high affinity ligand binding sites based on the thermodynamic signatures of receptor bound waters. They both involve the problem of how to rapidly calculate the potential of mean force (corresponding to the excess chemical potential of a water molecule) WT for all of the interfacial locations. A straightforward approach to calculate WT is to measure the work of dragging a tagged water molecule from the bulk to the specific position on the interface,[37] but this method is impractical for searching binding sites throughout the entire solute−solvent interface. In ref 38, we proposed a novel approach to estimate the excess chemical potential of water molecules in solution by applying UWHAM[46] to the simulation data generated from the endpoints states. The excess chemical potential of the interfacial waters (not the solute) estimated in this way agrees with the benchmark within a 95% confidence interval for most interfacial locations. Second, because the analysis of a water "position" involves six variables (three for position and three for orientation), evaluation of the indirect PMF ω at many locations for the purposes of identifying waters to displace by the ligand is not straightforward. Here we can make use of the four-dimensional representation of endpoints DFT for aqueous solutions, consisting of the three-dimensional coordinates of the water oxygen and the interaction energy of the water with the solute.[39] Use of this reduced four-dimensional representation which we introduced in 2019 should facilitate the development of a DFT-based method of "Watermap"-type analysis. We will report on this in a future communication.

Conformational free-energy differences play an essential role in many areas of biophysics, including inhibitor specific binding, and allosteric effects. The reorganization of the hydration shell solvent can make a significant contribution to conformational free-energy differences of peptides and proteins. The spatially resolved four-dimensional version of endpoints DFT that we developed can be used to investigate how the reorganization of interfacial solvent contributes to the conformational preferences of peptides and proteins. As a first effort in this direction,[39] we used eq 11 for an analysis of how the solute−solvent interaction and solvent reorganization affect the conformational preferences of alanine dipeptide in water, a frequently used model system. We further defined the concept of the hydration free energy within nonoverlapping regional volumes by suitably discretizing eq 18. We found that the regional hydration free-energy contributions to the conformational preferences of the peptide could be used as an accurate proxy for the total contribution including the entire solvent volume as shown in Figure 2, and furthermore, when the intramolecular conformational energy of alanine dipeptide was added to the hydration free energy, the sum of the terms accurately reflected the conformational free energies obtained by explicit long MD free-energy simulations. Mapping conformational free-energy landscapes of proteins, and

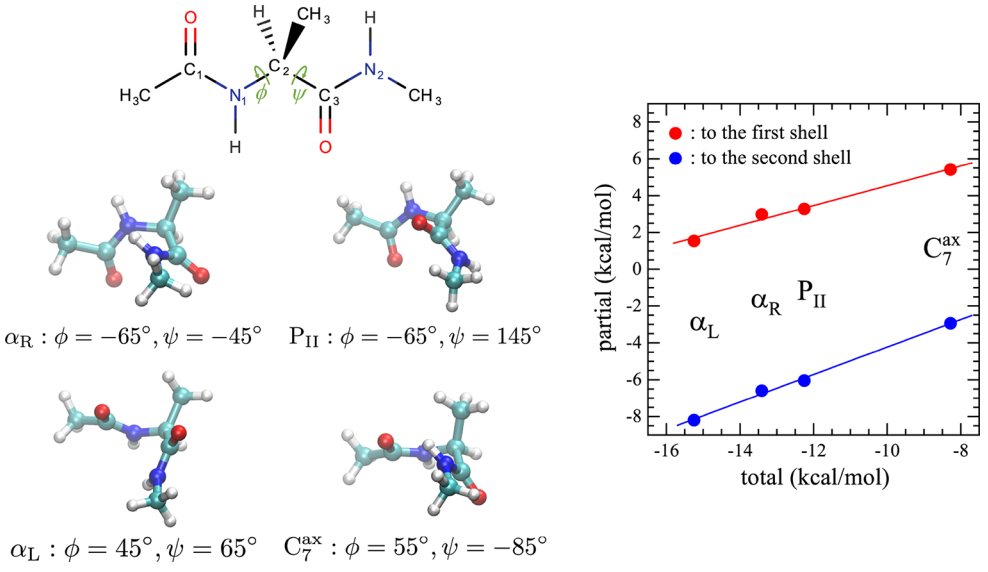


Figure 2. Correlations of the total free energy of hydration of alanine dipeptide against the partial contribution from the excluded-volume and first-shell regions of C_1, N_1, C_2, C_3, and N_2 atomic sites and that from the excluded-volume to second-shell regions. A single point in the right panel corresponds to a single conformation of alanine dipeptide shown in the left panel. With the linear regression against the total sum, the slope is 0.54 and 0.75 for the sums to the first and second shells, respectively, with correlation coefficients of 0.99 and 1.00.

https://dx.doi.org/10.1021/acs.jpcb.0c08988
J. Phys. Chem. B 2020, 124, 11771−11782

particularly how changing the protein sequence alters the landscape, is currently one of the most challenging problems in computational biophysics.[52] We are hopeful that the spatially resolved four-dimensional version of endpoints DFT can be further developed and will become a useful tool to better understand the central role that solvation plays in determining the conformational landscapes of proteins.

4.2. Water Dissolution into Polymers. Here we treat polymers. Polymer material is useful in membrane separation, and the performance as the separation medium is quantified by the permeability of the molecule to be separated. The standard framework to describe the permeability is the solubility–diffusion mechanism, in which the permeant dissolves into the polymer medium and diffuses across it. The permeability is then governed often by the extent of dissolution, that is determined in turn by the accompanying change in the free energy. The dissolution free energy reflects the intermolecular interaction between the permeant and polymer at atomic resolution. Atomistic schemes to analyze the dissolution are thus necessary in order to establish a theoretical and computational framework for the rational design of polymer membranes for separation.

MD simulation is useful for the atomistic analysis of polymer physical properties. In MD, the intra- and intermolecular interactions can be described at the atomic level and the structural flexibility of polymer species is incorporated naturally during the generation of the statistical ensemble. A key for the design of a polymer membrane is then to develop a computation method for the dissolution free energy. A polymer is usually a highly flexible molecule, and the free-energy method is required to be suitable for treating flexible species. The endpoints DFT method in the energy representation meets this requirement. It is applicable to flexible molecules, and the dissolution free energy can be obtained as a free energy of solvation by viewing the permeant as the solute and the polymer as the solvent.

The dissolution of water was examined[53] in nine polymers depicted in Figure 3. The polymer medium was treated at the

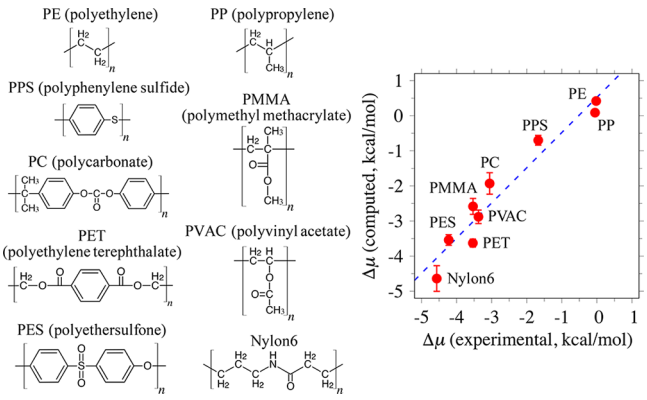


Figure 3. Polymers examined in ref 53 and the correlation plot between the computed and experimental free energies $\Delta\mu$ of water dissolution. The dashed line in the right panel stands for the least-squares fit of the computed $\Delta\mu$ to the experimental, and its slope is 1.0 with an intercept of 0.5 kcal/mol.

state of amorphous bulk and was a homogeneous "solvent" when the solvation free energy $\Delta\mu$ of a water molecule was calculated. Figure 3 shows $\Delta\mu$ obtained through the endpoints functional, with the experimental values. The mean absolute deviation is 0.5 kcal/mol between the computed and experimental sets of $\Delta\mu$, and their correlation coefficient is 0.97. The agreement is thus satisfactory given that, even for small solute molecules in water, the computational $\Delta\mu$ deviates from the experimental value by ~1 kcal/mol with well tuned force fields.[54] It is demonstrated with Figure 3 that the endpoints method is predictive for the free energy of water dissolution in a diverse class of polymers.

5. CONCLUSIONS AND FUTURE DIRECTIONS

Endpoints DFT is a framework for fast computation and physical interpretation of the excess chemical potential (solvation free energy) through the combination of molecular simulation and the density-functional theory of solutions. The simulation is to be performed only in the pure liquid and the solution system of interest, which are the endpoints of the hypothetical insertion process of the solute into the system. Our formulation starts from the Kirkwood charging formula and is valid in any ensembles including the canonical and isothermal–isobaric. The central role is played in the formulation by the indirect (solvent-mediated) part of the solute–solvent potential of mean force (indirect PMF), and we have shown that the expression for the solvation free energy is equivalent to the one in classical DFT in which the key quantity is the direct correlation function. As seen in eqs 3 and 21, our density functional can identify the stabilizing or destabilizing effect of a solvent molecule on the solute. The thermodynamic signatures of interfacial water can be categorized according to the density and indirect PMF, and through eq 24, a scheme can be devised to search for druggable targets on protein receptors.

A difficulty in the DFT approach is the high dimensionality when the coordinate for the distribution functions is the full set of positions and orientations of a solvent molecule relative to the solute. This difficulty can be circumvented by introducing the one-dimensional energy representation, in which the solute–solvent pair energy is adopted as a projected coordinate and the free-energy functional is formulated in terms of the energy distribution functions. The energy representation enables applications of the endpoints DFT method to a wide variety of systems, and we saw that the dissolution of a water molecule in a polymer is described at chemical accuracy. The spatial resolution of endpoints DFT can be restored in a numerically feasible manner by implementing the four-dimensional representation of mixed position and energy. When the hydration effect is analyzed for the conformational change of alanine dipeptide, it was found that the total effect of hydration is well correlated to the contribution up to the first shell with quantitatively non-negligible effects from farther distances.

With the endpoints DFT approach, all-atom analyses are possible for solvation effects on a protein and its complex in direct connection to the free energy, given that molecular simulations are now possible at nanometer and microsecond length and time scales. The role of interfacial water at a certain position to stabilize or destabilize a solute can be analyzed with the spatially resolved version, and it is of interest to specify which water molecules are responsible for conformational changes in such pharmaceutically important targets as kinase family proteins.[55,56] Another possibly is ligand design. The Watermap analysis has shown the importance of the energetics of water around a protein receptor, and with eq 24, the endpoints DFT approach provides a systematic route to estimating the free-energy change due to the displacement of solvent resulting from the variation of the ligand. Key properties are the one-body distribution function, the indirect PMF, and a gridding procedure based on the four-dimensional coordinates $(\mathbf{r}, \epsilon)$.

https://dx.doi.org/10.1021/acs.jpcb.0c08988
J. Phys. Chem. B 2020, 124, 11771–11782

With these computational tools, the DFT-based design of ligands can be anticipated.

The endpoints DFT method is an approximate scheme to determine the solvation free energy. Approximations are introduced to the integral of the indirect PMF over the coupling parameter. A systematic improvement of the free-energy functional can thus be possible by incorporating the correct dependence of the indirect PMF on the coupling parameter at the endpoints. We proposed 2P-QHNC and, indeed, observed improved performance of the functional for large, hydrophobic solutes. It is then necessary to identify which region in the energy or spatial coordinate is important and which is not for implementing the correct behavior of the indirect PMF, in view of its computational demand. The deep-core region of the solute seems less important in the analysis of conformational changes of a protein, and we are now developing a scheme to obtain the indirect PMF in the "outer skin" of the excluded volume.

AUTHOR INFORMATION

Corresponding Authors

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

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

Author

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

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.0c08988

Notes

The authors declare no competing financial interest.

Biographies

Ronald M. Levy received his Ph.D. in Biophysics from Harvard University in 1976 and was a Postdoctoral Fellow in the Chemistry Department at Harvard between 1976 and 1980. He started his independent career as Assistant Professor of Chemistry at Rutgers University in 1980 where he rose to the rank of Board of Governors Professor. In 2014, he moved to Temple University where he is Professor of Chemistry, Laura H. Carnell Professor, and Director of the Center for Biophysics and Computational Biology. His research interests are currently focused on surveying evolutionary fitness and conformational free-energy landscapes of proteins that determine the statistical thermodynamic basis for protein−ligand binding, molecular recognition, and allostery.

Nobuyuki Matubayasi received his Ph.D. degree in chemistry from Rutgers University in 1995, and after working at Institute for Chemical Research, Kyoto University, he has been a professor in Division of Chemical Engineering, Graduate School of Engineering Science, Osaka University, since 2014. Through the combination with state-of-the-art molecular simulation, he has been developing statistical−mechanical theories of solvation and transport properties, with molecular-level analysis of solvent effects on proteins, partitioning functions of such molecular aggregates as micelles, lipid membranes, and polymers, and electrical conductivity and diffusion in ionic liquids.

Bin W. Zhang received his Ph.D. degree in physics from the University of Pittsburgh in 2008. He joined the Center for Biophysics and Computational Biology at Temple University in 2013 and currently holds an Associate Research Professor position. His research interests are focused on advanced sampling algorithms, reweighting techniques, and protein−ligand binding.

ACKNOWLEDGMENTS

This work was supported by NIH grant (R35GM132090), NSF grant (1665032), and an NIH computer equipment grant (OD020095). This research includes calculations carried out on HPC resources (Owl's Nest of Temple University) supported in part by the National Science Foundation through major research instrumentation grant number 1625061 and by the US Army Research Laboratory under contract number W911NF-16-2-0189 and calculations carried out on the Extreme Science and Engineering Discovery Environment (XSEDE), which is

supported by National Science Foundation grant number ACI-1548562. We are further grateful to the Grant-in-Aid for Scientific Research (No. JP19H04206) from the Japan Society for the Promotion of Science and to the Elements Strategy Initiative for Catalysts and Batteries (No. JPMXP0112101003) and the Fugaku Supercomputer Project from the Ministry of Education, Culture, Sports, Science, and Technology.

REFERENCES

(1) Friedman, H. L. *A course in statistical mechanics*; Prentice-Hall: Englewood Cliffs, NJ, 1985.
(2) Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids: with Applications to Soft Matter*, 4th ed.; Academic Press: Amsterdam, The Netherlands, 2013.
(3) Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory. *J. Phys. Chem. B* **1998**, *102*, 3531−3541.
(4) Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 2. Applications to Simple Fluids. *J. Phys. Chem. B* **1998**, *102*, 3542−3550.
(5) Wallace, D. C. On the role of density fluctuations in the entropy of a fluid. *J. Chem. Phys.* **1987**, *87*, 2282.
(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., Ed. *Fundamentals of Inhomogeneous Fluids*; Marcel Dekker, Inc.: New York, 1992.
(8) 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.
(9) 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.
(10) 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.
(11) Lazaridis, T.; Paulaitis, M. E. Entropy of hydrophobic hydration: a new statistical mechanical formulation. *J. Phys. Chem.* **1992**, *96*, 3847−3855.
(12) Chandler, D.; Andersen, H. C. Optimized Cluster Expansions for Classical Fluids. II. Theory of Molecular Liquids. *J. Chem. Phys.* **1972**, *57*, 1930.
(13) Stell, G.; Patey, G. N.; Høye, J. S. *Advances in Chemical Physics*; John Wiley & Sons, Inc.: 1981; Vol. *48*, pp 183−328.
(14) Hirata, F.; Rossky, P. J. An extended RISM equation for molecular polar fluids. *Chem. Phys. Lett.* **1981**, *83*, 329−334.
(15) 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.
(16) Beglov, D.; Roux, B. An Integral Equation To Describe the Solvation of Polar Molecules in Liquid Water. *J. Phys. Chem. B* **1997**, *101*, 7821−7826.
(17) Kovalenko, A.; Hirata, F. Self-consistent description of a metal-water interface by the Kohn−Sham density functional theory and the three-dimensional reference interaction site model. *J. Chem. Phys.* **1999**, *110*, 10095−10112.
(18) Hirata, F. *Molecular Theory of Solvation*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2003.
(19) Ratkova, E. L.; Palmer, D. S.; Fedorov, M. V. Solvation Thermodynamics of Organic Molecules by the Molecular Integral Equation Theory: Approaching Chemical Accuracy. *Chem. Rev.* **2015**, *115*, 6312−6356.
(20) Sergiievskyi, V.; Jeanmairet, G.; Levesque, M.; Borgis, D. Solvation free-energy pressure corrections in the three dimensional reference interaction site model. *J. Chem. Phys.* **2015**, *143*, 184116.
(21) Weeks, J. D. Connecting Local Structure to Interface Formation: A Molecular Scale van der Waals Theory of Nonuniform Liquids. *Annu. Rev. Phys. Chem.* **2002**, *53*, 533−562.
(22) Remsing, R. C.; Liu, S.; Weeks, J. D. Long-ranged contributions to solvation free energies from theory and short-ranged models. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 2819−2826.
(23) Gao, A.; Remsing, R. C.; Weeks, J. D. Short solvent model for ion correlations and hydrophobic association. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 1293−1302.
(24) Jorgensen, W. L. Free energy calculations: a breakthrough for modeling organic chemistry in solution. *Acc. Chem. Res.* **1989**, *22*, 184−189.
(25) Gilson, M.; Given, J.; Bush, B.; McCammon, J. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.* **1997**, *72*, 1047−1069.
(26) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *J. Phys. Chem. B* **2003**, *107*, 9535−9551.
(27) Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W. Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *J. Chem. Theory Comput.* **2010**, *6*, 1509−1519.
(28) Procacci, P. Solvation free energies via alchemical simulations: let 's get honest about sampling, once more. *Phys. Chem. Chem. Phys.* **2019**, *21*, 13826−13834.
(29) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245−268.
(30) Hummer, G.; Szabo, A. Calculation of free-energy differences from computer simulations of initial and final states. *J. Chem. Phys.* **1996**, *105*, 2004−2010.
(31) Lu, N.; Singh, J. K.; Kofke, D. A. Appropriate methods to combine forward and reverse free-energy perturbation averages. *J. Chem. Phys.* **2003**, *118*, 2977.
(32) Matubayasi, N.; Nakahara, M. Theory of solutions in the energetic representation. I. Formulation. *J. Chem. Phys.* **2000**, *113*, 6070.
(33) Matubayasi, N.; Nakahara, M. Theory of solutions in the energy representation. II. Functional for the chemical potential. *J. Chem. Phys.* **2002**, *117*, 3605.
(34) Matubayasi, N.; Nakahara, M. Theory of solutions in the energy representation. III. Treatment of the molecular flexibility. *J. Chem. Phys.* **2003**, *119*, 9686.
(35) Levy, R. M.; Cui, D.; Zhang, B. W.; Matubayasi, N. Relationship between Solvation Thermodynamics from IST and DFT Perspectives. *J. Phys. Chem. B* **2017**, *121*, 3825−3841.
(36) Harris, R. C.; Deng, N.; Levy, R. M.; Ishizuka, R.; Matubayasi, N. Computing Conformational Free Energy Differences in Explicit Solvent: An Efficient Thermodynamic Cycle Using an Auxiliary Potential and a Free Energy Functional Constructed from the End Points. *J. Comput. Chem.* **2017**, *38*, 1198−1208.
(37) Cui, D.; Zhang, B. W.; Matubayasi, N.; Levy, R. M. The Role of Interfacial Water in Protein−Ligand Binding: Insights from the Indirect Solvent Mediated Potential of Mean Force. *J. Chem. Theory Comput.* **2018**, *14*, 512−526.
(38) Zhang, B. W.; Cui, D.; Matubayasi, N.; Levy, R. M. The Excess Chemical Potential of Water at the Interface with a Protein from End Point Simulations. *J. Phys. Chem. B* **2018**, *122*, 4700−4707.
(39) Ishii, Y.; Yamamoto, N.; Matubayasi, N.; Zhang, B. W.; Cui, D.; Levy, R. M. Spatially-Decomposed Free Energy of Solvation Based on the Endpoint Density-Functional Method. *J. Chem. Theory Comput.* **2019**, *15*, 2896−2912.
(40) Zhang, B. W.; Matubayasi, N.; Levy, R. M. Cavity Particle in Aqueous Solution with a Hydrophobic Solute: Structure, Energetics, and Functionals. *J. Phys. Chem. B* **2020**, *124*, 5220−5237.
(41) Matubayasi, N. Energy-Representation Theory of Solutions: Its Formulation and Application to Soft, Molecular Aggregates. *Bull. Chem. Soc. Jpn.* **2019**, *92*, 1910−1927.
(42) 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.

https://dx.doi.org/10.1021/acs.jpcb.0c08988
J. Phys. Chem. B 2020, 124, 11771−11782

(43) Xi, E.; Patel, A. J. The hydrophobic effect, and fluctuations: The long and the short of it. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 4549−4551.
(44) Beck, T. L.; Paulaitis, M. E.; Pratt, L. R. *The Potential Distribution Theorem and Models of Molecular Solutions*; Cambridge University Press: 2006.
(45) Kjellander, R.; Sarman, S. The chemical potential of simple fluids in a common class of integral equation closures. *J. Chem. Phys.* **1989**, *90*, 2768−2775.
(46) Zhang, B. W.; Arasteh, S.; Levy, R. M. The UWHAM and SWHAM Software Package. *Sci. Rep.* **2019**, *9*, 2803.
(47) 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.
(48) 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.
(49) 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.
(50) Huggins, D. J. Quantifying the Entropy of Binding for Water Molecules in Protein Cavities by Computing Correlations. *Biophys. J.* **2015**, *108*, 928−936.
(51) Ramsey, S.; Nguyen, C.; Salomon-Ferrer, R.; Walker, R. C.; Gilson, M. K.; Kurtzman, T. Solvation thermodynamic mapping of molecular surfaces in AmberTools: GIST. *J. Comput. Chem.* **2016**, *37*, 2029−2037.
(52) He, P.; Zhang, B. W.; Arasteh, S.; Wang, L.; Abel, R.; Levy, R. M. Conformational Free Energy Changes via an Alchemical Path without Reaction Coordinates. *J. Phys. Chem. Lett.* **2018**, *9*, 4428−4435.
(53) Kawakami, T.; Shigemoto, I.; Matubayasi, N. Free-Energy Analysis of Water Affinity in Polymer Studied by Atomistic Molecular Simulation Combined with the Theory of Solutions in the Energy Representation. *J. Chem. Phys.* **2012**, *137*, 234903.
(54) Shirts, M. R.; Pande, V. S. Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. *J. Chem. Phys.* **2005**, *122*, 134508.
(55) Haldane, A.; Flynn, W. F.; He, P.; Vijayan, R.; Levy, R. M. Structural propensities of kinase family proteins from a Potts model of residue co-variation. *Protein Sci.* **2016**, *25*, 1378−1384.
(56) Haldane, A.; Flynn, W. F.; He, P.; Levy, R. M. Coevolutionary Landscape of Kinase Family Proteins: Sequence Probabilities and Functional Motifs. *Biophys. J.* **2018**, *114*, 21−31.

https://dx.doi.org/10.1021/acs.jpcb.0c08988
J. Phys. Chem. B 2020, 124, 11771−11782