

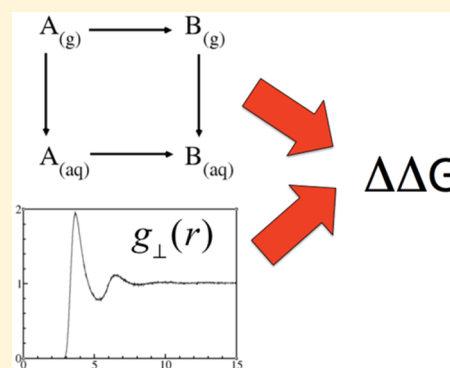
Free Energy Calculations Based on Coupling Proximal Distribution Functions and Thermodynamic Cycles

Shu-Ching Ou and B. Montgomery Pettitt*

Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, 301 University Boulevard, Galveston, Texas 77555-0304, United States

 Supporting Information

ABSTRACT: Techniques to calculate the free energy changes of a system are very useful in the study of biophysical and biochemical properties. In practice, free energy changes can be described with thermodynamic cycles, and the free energy change of an individual process can be computed by sufficiently sampling the corresponding configurations. However, this is still time-consuming especially for large biomolecular systems. Previously, we have shown that by utilizing precomputed solute–solvent correlations, so-called proximal distribution functions (pDF), we are capable of reconstructing the solvent environment near solute atoms, thus estimating the solute–solvent interactions and solvation free energies of molecules. In this contribution, we apply the technique of pDF-reconstructions to calculate chemical potentials and use this information in thermodynamic cycles. This illustrates how free energy changes of nontrivial chemical processes in aqueous solution systems can be rapidly estimated.



1. INTRODUCTION

The calculation of free energies has long been of interest for computational drug design as is the recognition of ligands by their receptors and interpreting the consequences of protein mutation and binding.¹ There is considerable literature developing methodologies for free energy calculations utilizing the property of state functions: the free energy difference between two state points in configuration space is independent of the selection of the pathway.² Thus, by using thermodynamic cycles we can calculate the free energy changes for convenient nonphysical processes to circumvent the complicated or computationally infeasible physical processes.³

The free energy difference associated with mutating a protein can be formulated as the difference in solvation free energy (ΔG_{solv}) of the initial and final states.^{1,4–7} Consider the pathway of creating a neutral/nonpolar cavity in a liquid solution prior to adding charges. This is an old problem in liquid state theory. We can decompose ΔG_{solv} into van der Waals (ΔG_{vdW})⁸ and electrostatic (ΔG_{elec}) components. We note here that such components of a free energy are defined by the process and are not state functions even though the sum of the components is. In practice, ΔG_{elec} has often been approximated using dielectric continuum theory based techniques^{8–11} such as the generalized Born model,^{12,13} or the Poisson–Boltzmann equation,^{14,15} while the more general Linear Response theory (LRT) can use more atomic detail.^{16–18}

The cavity or nonpolar ΔG_{vdW} component of the solvation free energy has a long history in the literature.^{19–21} An often used rough approximation is to assume this component is a linear function of the solvent accessible surface area (SASA).^{22–25} However, the justification of such simple models

to estimate ΔG_{vdW} remains a challenge due to multibody interactions between neighboring atoms,^{26–28} and the complex geometry of the solute–solvent interface.^{29–32} Thus, explicit solvent models which give rise to nontrivial correlations in packing structure must be coupled with techniques such as thermodynamic integration and free energy perturbation for accurate calculations of ΔG_{vdW} .^{28,33,34}

Despite the rather straightforward theory behind free energy calculations, the implementation of numerical calculations is not trivial. One requires a force field that correctly reproduces conformational properties, and solvation interactions including partitioning behavior between different environments. Classical force fields continue to advance sometimes at the expense of adding complexity.^{35–37} In addition, sampling a representative number of configurations within the given ensemble is also required for converged thermodynamic quantities, which leads to computational expense.^{38–40} Methods relying on precomputed quantities from explicit simulations of small solutes to estimate ΔG_{vdW} of large solutes^{41–46} are therefore being developed to circumvent expensive free energy simulations.

In this contribution, we explore the use of precomputed perpendicular/proximal distribution functions (pDFs) to rapidly calculate ΔG_{solv} and associated free energy differences with appropriate thermodynamic cycles. pDFs approximate solute–solvent pair correlation functions to describe the local solvent structure around a solute^{47–51} and thus can be used to predict the solvent structures around complex biological macromolecules that are composed of chemically similar

Received: November 14, 2018

Published: February 15, 2019



components.^{52–54} Average solute–solvent interaction energies, and subsequently solvation free energies, can then be calculated from these reconstructed solvent structures. We have shown that LRT electrostatic contributions to the free energy can be calculated with pDFs and compare well to explicit simulations with a substantial computational savings compared to not only simulation but Poisson–Boltzmann calculations.^{43,44} When coupled with the van der Waals components of the excess chemical potentials from pDF-reconstructions the total solvation free energy difference for small molecules is reasonably approximated (within kcal/mol accuracy).^{43,44,55,56} In this contribution, we extend the method to challenging thermodynamic chemical processes, including the length dependence of a polypeptide,⁵⁷ mutating a residue on a polypeptide, and finally, association between two alanine peptides.

2. THEORY

In this section, we present the thermodynamic cycles used in this study and the method of calculating free energy differences. Next, we briefly review the definition of pDFs and how they are used to reconstruct solvent environment around a given solute and solute–solvent (free) energetics. More details can be found from the literature references.^{44,47,49,55,56}

2.1. Thermodynamic Cycle. We present three studies in this contribution: extending a polypeptide ($\text{Gly}_2 \rightarrow \text{Gly}_3 \rightarrow \text{Gly}_4$), mutating a residue on tripeptide (alanine–tyrosine–alanine [AYA] $\rightarrow$ Ala₃ [AAA], or alanine–histidine–alanine [AHA] $\rightarrow$ Ala₃ [AAA]), and finally, association between two alanine peptides ($2 \times \text{Ala}_1 \rightarrow \text{Ala}_1\cdot\text{Ala}_1$). The thermodynamic cycles of these processes are illustrated in Figures 1 to 3. The

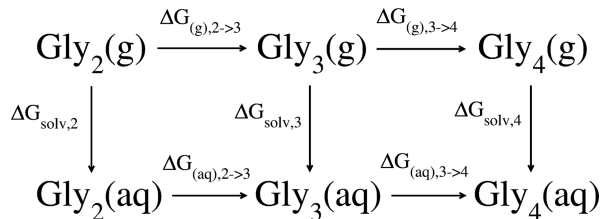


Figure 1. The thermodynamic cycle of growing polyglycine. The vertical terms represent the solvation free energies of solutes, while horizontal terms represent the free energies of growing an additional glycine in the (g) gas phase or (aq) aqueous phase.

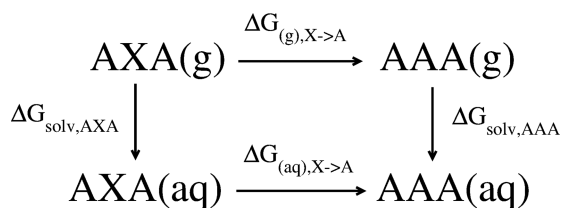


Figure 2. Thermodynamic cycle of mutating the central residue of a tripeptide. The vertical terms represent the solvation free energies of solutes, while horizontal terms represent the free energies of mutating the central residue into alanine, in the (g) gas phase or (aq) aqueous phase.

definition of each ΔG_i term is marked next to each process. For each thermodynamic cycle the sum of the free energies around any loop must be zero. Using the example in Figure 2 and given the sign of each contribution, we have

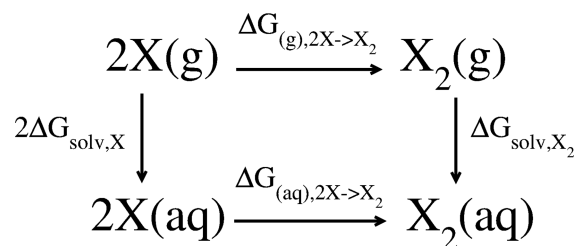


Figure 3. Thermodynamic cycle of the association between two Ala₁. X and X₂ represent the monomer and dimer states, respectively. The vertical terms represent the solvation free energies of solutes, while the horizontal terms represent the dimerization free energies in the (g) gas phase or (aq) aqueous phase.

$$\begin{aligned}\Delta\Delta G &= \Delta G_{(\text{aq}),X\rightarrow A} - \Delta G_{(\text{g}),X\rightarrow A} \\ &= \Delta G_{\text{solv},AAA} - \Delta G_{\text{solv},AXA}\end{aligned}\quad (1)$$

The horizontal terms can usually be approached by constructing intermediate states with single or dual topology methods.⁵⁸ The former method considers changes of atom types and atomic interactions during simulations, while the latter method includes dummy atoms to avoid the changes of atom types. The dual topology method requires more computing power particularly when simulations are run with Hamiltonian exchange or expanded ensembles. In practice, the path of calculating solvation free energy differences is more efficient. There are theories using implicit models to rapidly estimate the solvation free energies regardless the solvation details in the vicinity of the solute. Thus, we introduce the use of precomputed pDFs for the calculations of solvation free energies, providing the solvation detail via solvent density reconstructions.

2.2. Distribution Functions and Density Reconstruction. pDFs, or $g_{\perp}(r)$, are average density distributions calculated from solvent molecules to a nearest solute atom.^{47,59,60} The solvent probability distribution is most nearly perpendicular to the k th solute atom $g_{\perp}^k(r)$ and can be written as

$$g_{\perp}^k(r) = \sum_{t=0}^T \sum_{j=1}^n \frac{\delta(\text{Inf}[|\vec{r}_{ij}(t)|]_{i=1,m} - r)}{\delta\tau(\vec{r}_j(t), k)} \quad (2)$$

where $\vec{r}_{ij}(t)$ represents the position vector from the i th solute atom to the j th solvent atom at time t , m is the number of solute atoms, and n is the total number of solvent molecules in the system. $\text{Inf}[|\vec{r}_{ij}(t)|]_{i=1,m}$ returns the distance between a particular solvent atom and the nearest solute atom k ; that is, k is the assigned solute atom with the closest distance to the j th solvent atom. The underlying concept is to divide the solvent space into Voronoi polyhedra⁶¹ such that the resulting distribution functions are defined essentially perpendicularly to the exposed surface of the respective solute atoms.⁴³ $\delta\tau(\vec{r}_j(t), k)$ is the volume element around the j th solvent molecule, which is defined by all $\vec{r}$ vectors, where $|\vec{r}_k - \vec{r}| \leq |\vec{r}_i - \vec{r}|$. In practice, a pDF can be computed from a MD simulation mapped to a three-dimensional grid where each individual grid point along with its corresponding time-averaged solvent density is assigned to the closest solute atom for $g_{\perp}^k(r)$ calculations.

A variety of solute atom sets or groupings can be used to define the pDFs. Previous work shows that the pDFs classified by the chemical identity (carbon, nitrogen, etc.) of solute atoms are good enough for some purposes and using the force field

atom types for proteins and nucleic acids yields nearly universal functions which are thus transferable to chemically similar solute molecules.^{49,53,62} In this contribution, we construct pDFs for each force field atom type. We refer to a particular grouping or atom set by χ and refer to a particular pDF defined by this grouping with χ_k .

We can reconstruct the solvation density around a given solute configuration with a three-dimensional grid and assign solvent density to the grid point (x, y, z) with the precomputed $g_{\perp}^{\chi_k}(r')$:

$$g(x, y, z) = g_{\perp}^{\chi_k}(r') \quad (3)$$

where $r' = |\vec{r}_i - (x, y, z)|$ is the minimum distance between the grid point and all solute atoms i , χ_k is the assigned atom type set of the closest solute atom and (x, y, z) denotes the center of the grid volume. In this framework, the solvent density at grid point (x, y, z) is therefore

$$\rho(x, y, z) = \rho g_{\perp}^{\chi_k}(r') \quad (4)$$

where ρ is the bulk solvent number density. In our pDF calculations (both pDF-collections and solvent density reconstructions), we used 0.2 and 0.01 Å as the reconstruction grid space resolution and $g_{\perp}(r)$ collection resolution, respectively. The selection of $g_{\perp}(r)$ resolution is more important in the van der Waals reconstruction process and a resolution of 0.01 Å is suggested.⁵⁵

In practice, any solute atom may exclude solvent from any grid point due to overlap, or volume exclusion. A two-step-process was implemented to ensure grid points are assigned with the appropriate solvent density during reconstruction. Details are addressed in [Supporting Information](#) and the references.^{55,56}

2.3. Thermodynamic Integration for Solvation Free Energetics. Once the solvent density is reconstructed around the solute, the average solute–solvent interaction energy can then be written as

$$\begin{aligned} \langle U_{\text{solute-solv}} \rangle &= \rho \int g(x, y, z) U(x, y, z) dx dy dz \\ &\approx \sum_{i=1} \sum_{x,y,z}^m \Delta v \rho(x, y, z) U_{ij}(|(x, y, z) - \vec{r}_i|) \end{aligned} \quad (5)$$

where $U(x, y, z)$ is the total solute–solvent interaction energy at (x, y, z) , Δv is the unit volume which depends on the spatial grid resolution and the outer sum is taken over all grid points. Depending on the potential energy function, $U(x, y, z)$ may be decomposed into the sum of electrostatic and van der Waals interaction energies. The average electrostatic and van der Waals solute–solvent interaction energies can then be calculated using eq 5.

We have previously shown that the solute–solvent electrostatic interaction energy and electrostatic solvation free energy (ΔG_{elec}) can be reasonably approximated from the pDF-reconstructed solvent densities with Linear Response Theory.^{44,55} To calculate the free energy of cavity formation (ΔG_{vdW}), we applied a soft-core van der Waals potential and thermodynamic integration (TI) approach. The pathway between the initial gas phase and final solvated state by means of a coupling parameter, λ , which varies between 0 and 1 such that when $\lambda = 0$ the interaction energy (U_{ij}) between solute atom i and solvent atom j is zero and when $\lambda = 1$, $U_{ij} = U_{ij}^{\text{vdW}}$, or the typical van der Waals potential energy function. To avoid singularities and numerical instabilities at the $\lambda = 0$ end-point, a

soft-core potential is commonly used to scale U_{ij}^{vdW} along the pathway^{63,64} with a radius-shifting coefficient δ . The total solute–solvent van der Waals interaction energy (U^{vdW}) is the sum of U_{ij}^{vdW} over all solute–solvent atom pairs. The van der Waals solvation free energy is calculated as the integral of the average derivative of U^{vdW} with respect to λ . One can calculate ΔG_{vdW} using approximate numerical integration methods with the ensemble averages of $\partial U^{\text{vdW}} / \partial \lambda$ at various values of λ .

In our pDF approach, we first construct $g_{\perp}^{\chi_k}$ for the atom type set of a small representative chemical group (e.g., a peptide substituent) from MD simulations performed at regularly spaced values of λ . For each λ , we obtain unique $g_{\perp}^{\chi_k}$ values. Then using these precomputed pDFs, we reconstruct the average, λ -dependent solvent density distributions on a 3D grid around a given solute configuration (e.g., a polypeptide) for each λ value. We directly estimate $\langle \partial U^{\text{vdW}} / \partial \lambda \rangle_{\lambda}$ using a form like eq 5 but with U_{ij} replaced by $\partial U_{ij}^{\text{vdW}} / \partial \lambda$. Finally, numerically integrating $\langle \partial U^{\text{vdW}} / \partial \lambda \rangle_{\lambda}$ along λ yields ΔG_{vdW} . Note that the λ -dependent pDFs only need to be constructed once but may be used to predict ΔG_{vdW} of more complex solutes that are composed of similar atom type sets for which the pDFs were generated.

3. METHOD

3.1. Simulations for Calculating Solvation Free Energies. The solvation free energies of Gly₂, Gly₃, Gly₄, Ala₁, Ala₃ (AAA), alanine–tyrosine–alanine (AYA), alanine–histidine–alanine (AHA, with H representing neutral histidine with a protonated δ nitrogen), and the doubly blocked Ala₁ dimer (or Ala₁–Ala₁) were computed in this study. Molecular dynamics simulations were performed with NAMD 2.10^{65,66} with the CHARMM 36 force field parameters⁶⁷ to generate pDFs and calculate free energy benchmarks. Each solute was solvated with TIP3P water⁶⁸ in a volume with at least 10 Å from each boundary to the solute atoms. Peptide terminals were capped with neutral acetyl (ACE) and *N*-methyl amide (NME) chemical groups. Initial structures of these peptides are presented in the [Supporting Information](#). Three-dimensional periodic boundary conditions were applied. A rigid water geometry is enforced using SHAKE.⁶⁹ Particle Mesh Ewald (PME)⁷⁰ was used to treat electrostatic interactions using a grid of 1.0 Å resolution. The Lennard-Jones (L-J) interactions were gradually switched off over the range 10 Å to 11 Å. The temperature was fixed at 300 K via a Langevin thermostat with damping coefficient of 5 ps^{−1}. A time step of 2.0 fs was used to integrate the equations of motion.

Throughout the simulations, the solute molecules were fixed. The simulations were initially equilibrated in the NVT ensemble for 1 ns. The final configuration was then used as the starting condition for NPT production simulations at 1 atm pressure, with the first 1 ns excluded as further equilibration.

A soft-core van der Waals potential function introduced from refs 63 and 64 with a radius-shifting coefficient $\delta = 5.0$ was used. The simulated $\langle \partial U^{\text{vdW}} / \partial \lambda \rangle$ were calculated from more than 10 ns production simulations at each λ . The convergence of $\langle \partial U^{\text{vdW}} / \partial \lambda \rangle$ at each λ window was discussed in our previous work.⁵⁶ Trajectory snapshots were saved every 0.2 ps for analyses. For each λ window, the uncertainty of $\langle \partial U^{\text{vdW}} / \partial \lambda \rangle$ is estimated as the block standard error.⁷¹ The final uncertainty of ΔG_{vdW} is calculated by the propagation of errors across all λ values.

The transferable pDFs in this study were from our previous work^{55,56} and examples are presented in [Supporting Informa-](#)

tion. For each solvation free energy calculation, there are 11 windows with a λ -spacing of 0.1 from 0 to 1. Solvent reconstruction and calculations of ΔG_{vdW} were performed with these same λ values. pDFs are used to reproduce simulated $\langle \partial U^{\text{vdW}} / \partial \lambda \rangle$ at all λ -windows, which are subsequently integrated to estimate ΔG_{vdW} . One may wish to sample a more fine grid in certain λ ranges depending on the chosen soft-core potential function (near end points in our case) to better capture the repulsive van der Waals forces when inserting the solute into the system.⁴ The ΔG_{elec} term is approximated as half of solute–solvent electrostatic interaction energy based on Linear Response theory, which can be estimated via simulations with full electrostatic interactions or via solvent reconstruction with pDFs.⁴⁴ For the electrostatic components the simulation details are the same as the above free energy simulations except for having charges on solute atoms and a total sampling time for each system of 100 ns with a 2 ps save frequency.

Throughout this manuscript, we applied Simpson's rule for numerical integrations.

3.2. Potential of Mean Force for the Dimerization of Ala₁. A stringent test of the pDF theory is the calculation of a coordinate dependent free energy surface or a potential of mean force (PMF). The dimerization/association between two Ala₁ was studied via the thermodynamic cycle in Figure 3 and calculations of the solvent-induced PMF, with X and X₂ representing the monomer and dimer. The solvation contribution to the free energy of dimerization can be written as⁷²

$$\begin{aligned} \Delta \Delta G &= \Delta G_{(\text{aq}), 2X \rightarrow X_2} - \Delta G_{(\text{g}), 2X \rightarrow X_2} \\ &= \Delta G_{\text{solv}, X_2} - 2\Delta G_{\text{solv}, X} \end{aligned} \quad (6)$$

Figure 4 shows the geometry of the system, which is adopted from the previous study.⁷² Two Ala₁ monomers were placed in

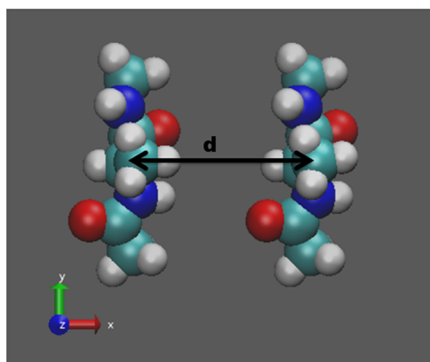


Figure 4. Geometry of the dimerization of Ala₁.

parallel, with the C_α–C_α vector along the x -direction. The reaction coordinate for obtaining the PMF is defined as the C_α–C_α separation distance (d). In this study we considered 45 separations/windows between 6.3 (associated/dimerized state) and 15.1 Å (dissociated state), with the positions of all solute atoms being fixed at each separation window d_i throughout the simulations and pDF calculations. We calculate the reversible work at separation d_i using the free energy perturbation (FEP) formula:

$$\Delta G_{i+1,i} = G(d_i + \delta d) - G(d_i) = -kT \ln \langle \exp(-\beta \Delta U_i) \rangle_i \quad (7)$$

where $\Delta U_i = U(d_i + \delta d) - U(d_i)$ is the variation of potential energy due to changing of separation distance from d_i to $d_i + \delta d$.

$\langle \cdots \rangle_i$ indicates an ensemble average in state i . The solute–solute interactions are fixed at each simulation window. Thus, we focus only on the solute–solvent potential energy for eq 7. The solute–solute PMF is presented in the Supporting Information. If we define $i = 0$ as the associated state and $i = N$ as the dissociated state, the free energy of dimerization is then obtained as

$$\Delta \Delta G = -kT \sum_{i=0}^{N-1} \ln \langle \exp(-\beta \Delta U_i) \rangle_i \quad (8)$$

The control simulation is prepared by placing the dimer ($d = 6.3$ Å) at the center of a box of dimensions $60 \times 60 \times 60$ Å with 6814 TIP3P water molecules. The simulation setup remains the same for each window with a 10 ns production run with a 2 ps save frequency. Throughout the simulation the dimer atoms are fixed. The final configuration was then used to generate the next simulation window by translating one monomer 0.2 Å along the x -axis. This process was repeated until $d = 15.1$ Å. We calculate $\Delta \Delta G$ along the reaction coordinate d . Since the process is reversible, in the results section for the dimerization process, we use the dissociated state as the reference point (zero) for discussion and better visualization.

In the pDF-reconstruction algorithm, the van der Waals and electrostatic interactions were treated separately. Thus, we have one set of full solute–solvent interactions simulations, while the other set of simulations removed all solute atomic charges.

4. RESULTS

4.1. Growth of Polyglycine. The simulated and pDF-reconstructed ΔG values are listed in Table 1. The uncertainties

Table 1. Comparison of Solvation Free Energies of Gly₂, Gly₃, and Gly₄ Using Thermodynamic Integration with Simulations and pDF-Reconstructions^a

ΔG	simulated T.I.	pDF T.I.
Gly ₂		
ΔG_{vdW}	−0.33 (0.39)	−0.75
ΔG_{elec}	−23.96 (0.05)	−23.14
ΔG_{solv}	−24.29 (0.39)	−23.89
Gly ₃		
ΔG_{vdW}	−1.01 (0.41)	−1.27
ΔG_{elec}	−29.96 (0.06)	−28.65
ΔG_{solv}	−30.97 (0.41)	−29.92
Gly ₄		
ΔG_{vdW}	−1.75 (0.46)	−1.91
ΔG_{elec}	−36.61 (0.07)	−35.16
ΔG_{solv}	−38.36 (0.47)	−37.07

^a ΔG_{vdW} is obtained using thermodynamic integration of $\langle \partial U^{\text{vdW}} / \partial \lambda \rangle_i$. ΔG_{elec} is approximated as half of electrostatic solute–solvent interaction energy using linear response theory. Uncertainties obtained as the block standard errors⁷¹ are denoted in parentheses. Uncertainties for ΔG_{vdW} and ΔG_{solv} are calculated using error propagation. The units are kcal/mol.

of ΔG_{elec} are calculated as the block standard errors⁷¹ of solute–solvent electrostatic interaction energies. The uncertainties of ΔG_{vdW} are calculated from the error propagation of the uncertainties of $\langle \partial U^{\text{vdW}} / \partial \lambda \rangle_i$, which are usually larger at small λ . Thus, uncertainties of simulated ΔG_{vdW} in general are larger than uncertainties of ΔG_{elec} . The differences between simulated and pDF-reconstructed ΔG_{vdW} are 0.42, 0.26, and 0.16 kcal/mol

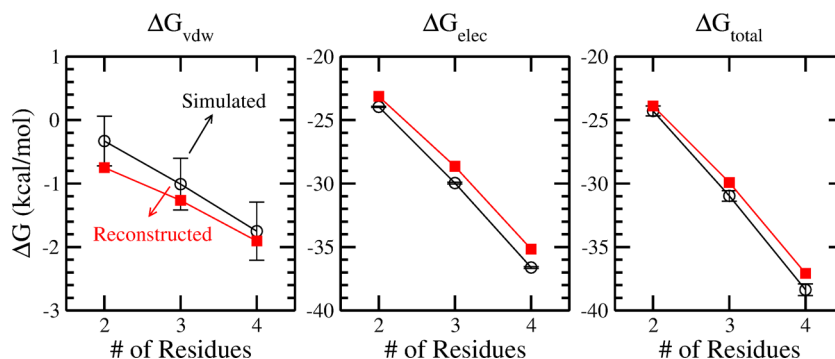


Figure 5. Solvation free energies of polyglycine.

for Gly₂, Gly₃, and Gly₄, which are generally within the uncertainties of ΔG_{vdw} (~ 0.4 kcal/mol). Differences between simulated and pDF-reconstructed ΔG_{elec} are 0.82, 1.31, 1.45 kcal/mol. Although these differences are larger than the uncertainties of the simulated ΔG_{elec} (~ 0.06 kcal/mol), they are still within $\pm 4\%$ of the simulated values. The largest difference between simulated and pDF-reconstructed ΔG_{solv} is 1.29 kcal/mol.

Figure 5 shows the dependence of ΔG_{vdw} , ΔG_{elec} , and ΔG_{solv} on peptide chain length. The simulated and pDF-reconstructed yield a similar near linear dependence between all ΔG terms versus chain length. On the basis of the thermodynamic cycle in Figure 1, we have

$$\Delta\Delta G_{2\rightarrow 3} = \Delta G_{\text{solv},3} - \Delta G_{\text{solv},2} \quad (9)$$

$$\Delta\Delta G_{3\rightarrow 4} = \Delta G_{\text{solv},4} - \Delta G_{\text{solv},3} \quad (10)$$

The simulated and pDF-reconstructed $\Delta\Delta G_{2\rightarrow 3}$ are -6.68 (± 0.57 using error propagation) and -6.03 kcal/mol, while the simulated and pDF-reconstructed $\Delta\Delta G_{3\rightarrow 4}$ are -7.39 (± 0.62) and -7.15 kcal/mol, respectively. These values are comparable with other similar studies (ca. -5.5 kcal/mol^{57,73,74}). The particular values are affected by the λ resolution for the FEP, the force field, the numerical method for doing integrals, and the linear response approximation we used for ΔG_{elec} . As a control, the consistency between simulated and pDF-reconstructed $\Delta\Delta G$ under the same conditions demonstrates that densities reconstructed via pDFs are not only sufficient for solute–solvent interaction energies ($\langle U \rangle$) or solvation free energies ($\int \langle \partial U / \partial \lambda \rangle_{\lambda} d\lambda$), but also sufficient to recapitulate the features of free energy differences ($\Delta\Delta G$).

4.2. Mutation of a Tripeptide. We now present results which can be thought of as a computational alanine scan in a tripeptide. The simulated and pDF-reconstructed ΔG values are listed in Table 2. The differences between simulated control and pDF-reconstructed ΔG_{vdw} are -0.37 , -1.11 , and -1.04 kcal/mol for AAA, AHA, and AYA, while the differences for ΔG_{elec} are 0.6, 1.12, 0.54. Thus, overall the ΔG_{solv} are consistent with previous quality measures. With the thermodynamic cycle in Figure 2, the free energy differences are

$$\Delta\Delta G_{\text{AYA} \rightarrow \text{AAA}} = \Delta G_{\text{solv},\text{AAA}} - \Delta G_{\text{solv},\text{AYA}} \quad (11)$$

$$\Delta\Delta G_{\text{AHA} \rightarrow \text{AAA}} = \Delta G_{\text{solv},\text{AAA}} - \Delta G_{\text{solv},\text{AHA}} \quad (12)$$

and from Table 2 we can derive $\Delta\Delta G$ values listed in Table 3 along with the corresponding hydration free energy differences between amino acid side chains using the same force fields from ref 75 and side chain analogues from ref 76–78.

Table 2. Comparison of Solvation Free Energies of Tripeptides (Ala-Ala-Ala, Ala-His-Ala, Ala-Tyr-Ala) Using Thermodynamic Integration with Simulations and pDF-Reconstructions^a

ΔG	simulated T.I.	pDF T.I.
AAA		
ΔG_{vdw}	0.33 (0.23)	0.70
ΔG_{elec}	−29.12 (0.03)	−29.72
ΔG_{solv}	−28.79 (0.23)	−29.02
AHA		
ΔG_{vdw}	0.67 (0.25)	1.78
ΔG_{elec}	−41.38 (0.04)	−42.50
ΔG_{solv}	−40.71 (0.25)	−40.72
AYA		
ΔG_{vdw}	0.77 (0.26)	1.81
ΔG_{elec}	−34.53 (0.03)	−35.07
ΔG_{solv}	−33.76 (0.27)	−33.26

^aUncertainties obtained as the block standard errors are denoted in parentheses. The units are kcal/mol.

In general, all of our simulated and pDF-reconstructed $\Delta\Delta G$ terms are within 1 kcal/mol difference from the reference values. We note that all of the calculated $\Delta\Delta G_{\text{solv}}$ are smaller than the experimental values. Although the experimental ΔG_{solv} of these tripeptides are not reported in this study, there seems to be common over estimations of ΔG_{solv} for nonpolarizable force fields.^{75,79,80} These deviations between computations and experiments are usually larger than relative solvation free energies (i.e., $\Delta\Delta G_{\text{solv}}$).⁸¹ We have shown that by using solvent density reconstructions via pDFs, the free energy differences of residue mutations are well-reproduced for a given force field.

4.3. Dimerization of Ala₁. The solvent contribution to Ala₁ dimerization is given via eq 6. The solvation free energies of Ala₁ monomer and dimer from simulations and pDF-reconstructions are listed in Table 4. For the monomer, both ΔG_{vdw} and ΔG_{elec} terms are given by pDF-reconstructions to within a difference of 0.4 kcal/mol. For the dimer, there is a slightly larger difference between the simulated and pDF-reconstructed ΔG_{elec} . Using eq 6 with the shift to allow the dissociated state as the reference (zero), we get $\Delta\Delta G_{\text{vdw}} = 2.48 \pm 0.62$, $\Delta\Delta G_{\text{elec}} = 1.77 \pm 0.10$, and $\Delta\Delta G_{\text{solv}} = 4.25 \pm 0.63$ kcal/mol from the solvation free energy simulations. From the pDF-reconstructions we obtain $\Delta\Delta G_{\text{vdw}} = 1.57$, $\Delta\Delta G_{\text{elec}} = 2.15$, and $\Delta\Delta G_{\text{solv}} = 3.72$ kcal/mol. From the solvation free energy difference calculations (i.e., vertical contributions from the thermodynamic cycle in Figure 3), the simulated and pDF-reconstructed results of solvent contributions to dimerization are consistent.

Table 3. Comparison of $\Delta\Delta G$ Values for AXA $\rightarrow$ AAA Mutations^a

$\Delta\Delta G$	simulated T.I.	pDF T.I.	CHARMM 22 ^b	CHARMM 27 ^c	exp ^d
AYA $\rightarrow$ AAA					
$\Delta\Delta G_{\text{vdW}}$	−0.44 (0.35)	−1.11	−0.92		
$\Delta\Delta G_{\text{elec}}$	5.41 (0.04)	5.35	6.04		
$\Delta\Delta G_{\text{solv}}$	4.97 (0.35)	4.24	5.12	4.9	8.05
AHA $\rightarrow$ AAA					
$\Delta\Delta G_{\text{vdW}}$	−0.34 (0.34)	−1.08	−0.63		
$\Delta\Delta G_{\text{elec}}$	12.26 (0.05)	12.78	11.45		
$\Delta\Delta G_{\text{solv}}$	11.92 (0.34)	11.70	10.82	9.1	12.21

^aAll units are in kcal/mol. Values in parentheses show the corresponding uncertainties calculated from error propagations. ^bValues are taken from ref 75 using CHARMM 22 force field. ^cValues are taken from ref 80 using CHARMM 27 force field with pseudoglycine method. ^d $\Delta\Delta G$ is calculated as the ΔG difference between experimental solvation free energies of methane, *p*-Cresol, methylimidazole (with protonated δ nitrogen), which are the side chain analogues of Ala, Tyr, and His. These values are taken from refs 76–78.

Table 4. Comparison of Solvation Free Energies of Using Thermodynamic Integration with Simulations and pDF-Reconstructions

ΔG	simulated T.I.	pDF T.I.
Ala ₁ (monomer)		
ΔG_{vdW}	1.08 (0.36)	1.48
ΔG_{elec}	−16.11 (0.03)	−15.75
ΔG_{solv}	−15.03 (0.36)	−14.27
Ala ₁ -Ala ₁ (dimer)		
ΔG_{vdW}	4.64 (0.51)	4.53
ΔG_{elec}	−30.45 (0.10)	−29.35
ΔG_{solv}	−25.81 (0.52)	−24.82

We next calculate the solvent contribution to the PMF along a simple separation path coordinate. The solvent-induced PMF can be decomposed into van der Waals and electrostatic components. We first consider the van der Waals component from the set of simulations without charges on solute atoms, as shown in Figure 6a. The simulated PMF (black solid line) shows the expected barrier between the contact and solvent separated states as expected from studies of nonpolar hydrophobic assemblies.^{82–84} The shape of the pDF-reconstructed PMF (red dashed line) shows a weaker feature, and the barrier is not as structured compared with simulated PMF. However, the barrier from simulation is less than 0.5 kcal/mol, which is challenging since the accuracy of pDF-reconstructions is ~ 1 kcal/mol. We find that the effective interaction near contact is underestimated by the pDF procedure. This is to be expected given the near neighbor approximation used. We would expect a substantial correction from next near neighbor terms at close distance between solutes. Between $d = 6.3$ and 15.1 Å the

$\Delta\Delta G_{\text{vdW}}$ is 2.32 kcal/mol from simulated PMF and 1.64 kcal/mol from the pDF-reconstructed PMF, as listed in Table 5.

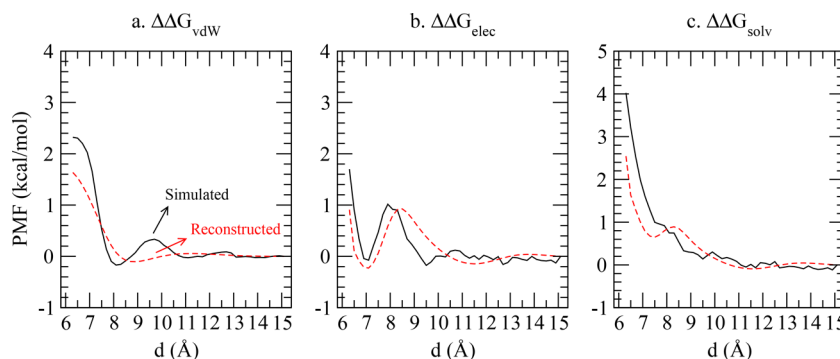
Table 5. Comparison of $\Delta\Delta G$ Terms Using Different Calculation Paths

$\Delta\Delta G$	simulated	pDF
Via Differences of ΔG_{solv}		
$\Delta\Delta G_{\text{vdW}}$	2.48(0.62)	1.57
$\Delta\Delta G_{\text{elec}}$	1.77(0.10)	2.15
$\Delta\Delta G_{\text{solv}}$	4.25(0.63)	3.72
Via Sum of Intermediate States		
$\Delta\Delta G_{\text{vdW}}$	2.32	1.64
$\Delta\Delta G_{\text{elec}}$	1.70	0.91
$\Delta\Delta G_{\text{solv}}$	4.02	2.55

These results are close to the results using solvation free energy differences between associated/dissociated states, both from control simulations and pDF-reconstructions individually.

The electrostatic component of the solvent-induced PMF from simulation and pDF-reconstruction is shown in Figure 6b. The minimum at short distance is well reproduced in depth and position between the simulation and reconstruction. The first barrier of the simulated PMF is qualitatively given via the pDF-reconstructions with a barrier ~ 1 kcal/mol but shifted a few tenths of an Å compared to simulation. Again the corrections to this feature at close proximity would be expected to come from the next near neighbor contributions. The $\Delta\Delta G_{\text{elec}}$ between $d = 6.3$ and 15.1 Å is 1.70 kcal/mol from simulation and 0.91 kcal/mol from the pDF-reconstructed PMF, as itemized in Table 5.

The sum of van der Waals and electrostatic components yields the total $\Delta\Delta G_{\text{solv}}$ shown in Figure 6c. The shape of the PMF is

Figure 6. Potential of mean force for Ala₁ dimerization, from simulations and pDF-reconstructions.

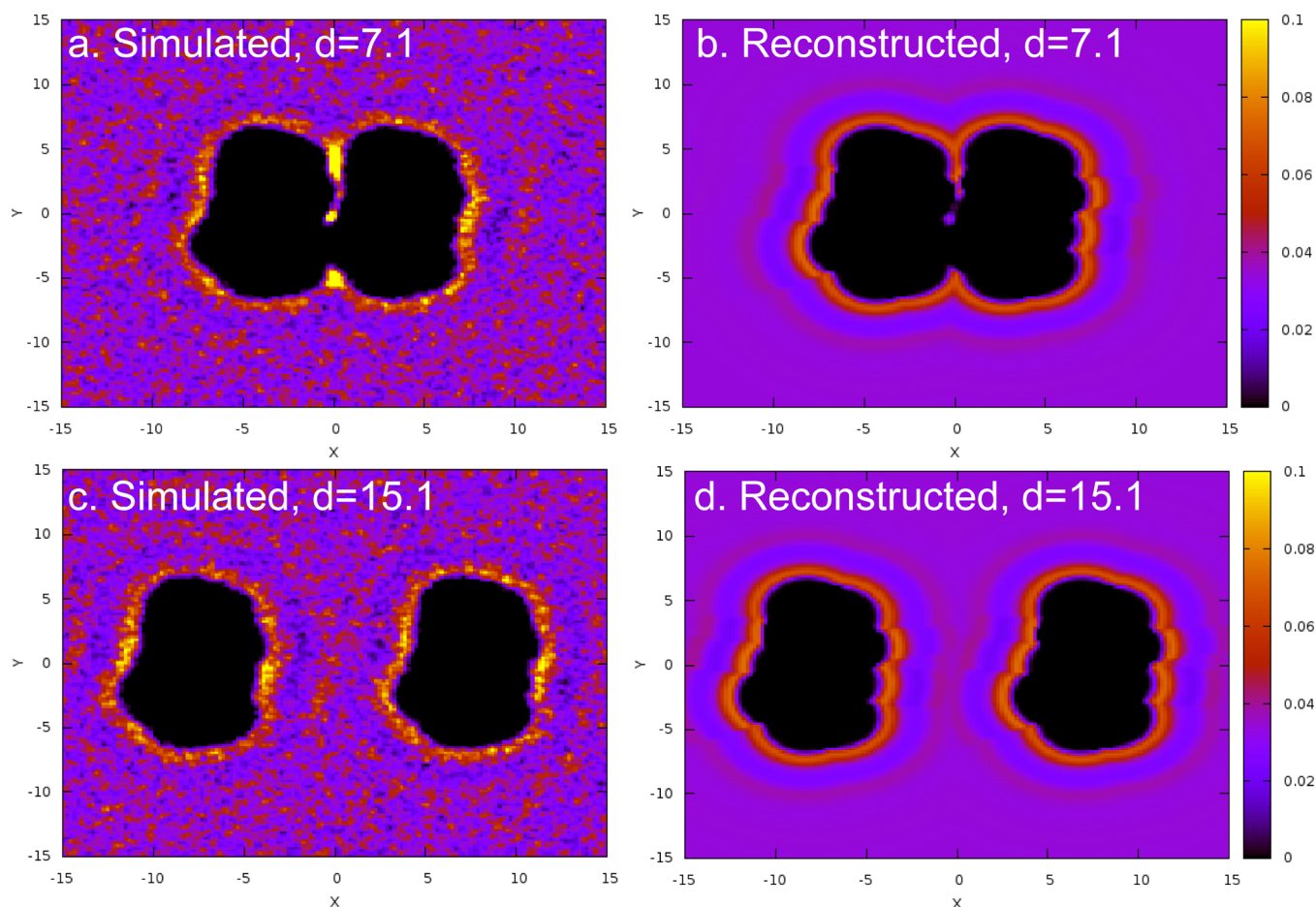


Figure 7. (a) Simulated (b) pDF-reconstructed 2-D water density maps on $z = 0$ plane for the Ala_1 – Ala_1 at $d = 7.1$ Å. (c) Simulated and (d) pDF-reconstructed 2-D water density maps on $z = 0$ plane for Ala_1 – Ala_1 at $d = 15.1$ Å.

reasonable at larger distances and shows the expected flaw near contact. The final $\Delta\Delta G_{\text{solv}}$ values from simulated and pDF-reconstructed PMF differ by about 1.5 kcal/mol.

We use the average solvent density maps (as presented in Figure 7) to consider consequences of the near neighbor approximation we use in pDF-reconstructions. We select the separation distance of 7.1 Å and calculate the average solvent densities from simulations and the pDF-reconstructed solvent densities. Figure 7 panels a and b show the average solvent densities on the x – y plane at $z = 0$. Some higher-density features were found between the solutes via simulation, but they are not evident from reconstructions. Solvent molecules within intersolute regions are usually adjacent to both solutes and thus multibody correlations play a more significant role in determining the distributions of solvent. In the pDF-reconstruction, solvent densities were determined by the nearest solute atoms on one of the Ala_1 peptides, which means the effects from the next-nearest solute atom (which might be on the other Ala_1) are ignored. These differences in densities from simulations and pDF-reconstructions, contribute to larger differences in energetics and free energetics near contact in all components of the PMF. When the solutes are sufficiently separated, as shown in the case of Figure 7 panels c and d, the solvation structures around each solute are not as affected.

5. CONCLUSION

In this contribution, we used precomputed solvent density distributions with free energy methods including thermodynamic integration and Linear Response Theory to calculate van der Waals and electrostatic solvation free energies for a variety of solutes. With appropriate free energy cycles, these solvation free energies can be used for free energy differences in a variety of chemical processes. Using this framework, we have shown by comparing with free energy simulations on the same potential surface that $\Delta\Delta G$ can be reproduced via pDF-reconstructions within useful accuracy. The peptide case studies in this contribution are relatively simple model compounds. The trends explored should have implications for many applications including protein engineering and drug design, etc. While the properties of single molecule surfaces showed considerable promise, we found details of the more stringent PMF calculations near contact somewhat less accurate. In the future, we will further extend the pDF-reconstruction algorithm to terms including next near neighbor and other explicit three body effects. Other obvious applications include protein pK_a calculations and catalytic site predictions, which are of great interest in a variety of applications.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.8b01157.

Details of cavity and exclusion factors, initial structures of solutes used in this study, selected pPDFs at various λ , solute–solute potential of mean force (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: mpettitt@utmb.edu. Phone: (409)772-0723.

ORCID

B. Montgomery Pettitt: 0000-0003-4902-3046

Funding

We gratefully acknowledge the Robert A. Welch Foundation (H-0037), the National Science Foundation (CHE-1152876), and the National Institutes of Health (GM-037657) for partial support of this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank the scientific computing staff at the Sealy Center for Structural Biology and Molecular Biophysics for computing support.

REFERENCES

- (1) Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.
- (2) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*; John Wiley & Sons: Chichester, UK, 2013.
- (3) Straatsma, T. P.; McCammon, J. A. Computational Alchemy. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407–435.
- (4) Jorgensen, W. L.; Ravimohan, C. Monte Carlo Simulation of Differences in Free Energies of Hydration. *J. Chem. Phys.* **1985**, *83*, 3050–3054.
- (5) Caldwell, J. W.; Agard, D. A.; Kollman, P. A. Free Energy Calculations on Binding and Catalysis by Alpha-Lytic Protease: the Role of Substrate Size in the P1 Pocket. *Proteins: Struct., Funct., Genet.* **1991**, *10*, 140–148.
- (6) Prevost, M.; Wodak, S. J.; Tidor, B.; Karplus, M. Contribution of the Hydrophobic Effect to Protein Stability: Analysis Based on Simulations of the Ile96-Ala Mutation in Barnase. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*, 10880–10884.
- (7) Dang, L. X.; Merz, K. M., Jr.; Kollman, P. A. Free Energy Calculations on Protein Stability: Thr-157—Val-157 Mutation of T4 Lysozyme. *J. Am. Chem. Soc.* **1989**, *111*, 8505–8508.
- (8) Cramer, C. J.; Truhlar, D. G. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.* **1999**, *99*, 2161–2200.
- (9) Schaefer, M.; van Vlijmen, H. W. T.; Karplus, M. Electrostatic Contributions to Molecular Free Energies in Solution. *Adv. Protein Chem.* **1998**, *51*, 1–58.
- (10) Roux, B.; Simonson, T. Implicit Solvent Models. *Biophys. Chem.* **1999**, *78*, 1–20.
- (11) Simonson, T.; Roux, B. Concepts and Protocols for Electrostatic Free Energies. *Mol. Simul.* **2016**, *42*, 1090–1101.
- (12) Roux, B.; Yu, H. A.; Karplus, M. Molecular Basis for the Born Model of Solvation. *J. Phys. Chem.* **1990**, *94*, 4683–4688.
- (13) Bashford, D.; Case, D. A. Generalized Born Models of Macromolecular Solvation Effects. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129–152.
- (14) Netz, R. R.; Orland, H. Beyond Poisson-Boltzmann: Fluctuation Effects and Correlation Functions. *Eur. Phys. J. E: Soft Matter Biol. Phys.* **2000**, *1*, 203–214.
- (15) Grochowski, P.; Trylska, J. Continuum Molecular Electrostatics, Salt Effects, and Counterion Binding - A Review of the Poisson-Boltzmann Theory and Its Modifications. *Biopolymers* **2008**, *89*, 93–113.
- (16) Chandler, D. *Introduction to Modern Statistical Mechanics*; Oxford University Press: New York, NY, 1987.
- (17) Hansen, J. P.; McDonald, I. R. *Theory of Simple Liquids*; Academic Press: San Diego, CA, 1976.
- (18) Aqvist, J.; Luzhkov, V. B.; Brandsdal, B. O. Ligand Binding Affinities from MD Simulations. *Acc. Chem. Res.* **2002**, *35*, 358–365.
- (19) Barker, J. A.; Henderson, D. Perturbation Theory and Equation of State for Fluids. II. A Successful Theory of Liquids. *J. Chem. Phys.* **1967**, *47*, 4714.
- (20) Weeks, J. D.; Chandler, D.; Andersen, H. C. Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids. *J. Chem. Phys.* **1971**, *54*, 5237.
- (21) Ben-Amotz, D.; Stell, G. Reformulation of Weeks-Chandler-Andersen Perturbation Theory Directly in Terms of a Hard-Sphere Reference System. *J. Phys. Chem. B* **2004**, *108*, 6877–6882.
- (22) Lee, B.; Richards, F. M. The Interpretation of Protein Structures: Estimation of Static Accessibility. *J. Mol. Biol.* **1971**, *55*, 379–400.
- (23) Sharp, K. A.; Nicholls, A.; Fine, R. F.; Honig, B. Reconciling the Magnitude of the Microscopic and Macroscopic Hydrophobic Effects. *Science* **1991**, *252*, 106–109.
- (24) Cramer, C. J.; Truhlar, D. G. An SCF Solvation Model for the Hydrophobic Effect and Absolute Free Energies of Aqueous Solvation. *Science* **1992**, *256*, 213–217.
- (25) Raschke, T. M.; Tsai, J.; Levitt, M. Quantification of the Hydrophobic Interaction by Simulations of the Aggregation of Small Hydrophobic Solutes in Water. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 5965–5969.
- (26) Czaplewski, C.; Ripoll, D. R.; Liwo, A.; Rodziewicz-Motowidlo, S.; Wawak, J.; Scheraga, H. A. Can Cooperativity in Hydrophobic Association be Reproduced Correctly by Implicit Solvation Models? *Int. J. Quantum Chem.* **2002**, *88*, 41–45.
- (27) Shimizu, S.; Chan, H. S. Anti-Cooperativity and Cooperativity in Hydrophobic Interactions: Three-Body Free Energy Landscapes and Comparison with Implicit-Solvent Potential Functions for Proteins. *Proteins: Struct., Funct., Genet.* **2002**, *48*, 15–30.
- (28) Harris, R. C.; Pettitt, B. M. Effects of Geometry and Chemistry on Hydrophobic Solvation. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 14681–14686.
- (29) Jamadagni, S. N.; Godawat, R.; Garde, S. Hydrophobicity of Proteins and Interfaces: Insights from Density Fluctuations. *Annu. Rev. Chem. Biomol. Eng.* **2011**, *2*, 147–171.
- (30) Patel, A. J.; Varilly, P.; Jamadagni, S. N.; Acharya, H.; Garde, S.; Chandler, D. Extended Surfaces Modulate Hydrophobic Interactions of Neighboring Solutes. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 17678–17683.
- (31) Cui, D.; Ou, S.; Patel, S. Free Energetics of Rigid Body Association of Ubiquitin Binding Domains: A Biochemical Model for Binding Mediated by Hydrophobic Interaction. *Proteins: Struct., Funct., Genet.* **2014**, *82*, 1453–1468.
- (32) Giovambattista, N.; Debenedetti, P.; Rossky, P. Hydration Behavior under Confinement by Nanoscale Surfaces with Patterned Hydrophobicity and Hydrophilicity. *J. Phys. Chem. C* **2007**, *111*, 1323–1332.
- (33) Gallicchio, E.; Kubo, M. M.; Levy, R. M. Enthalpy-Entropy and Cavity Decomposition of Alkane Hydration Free Energies: Numerical Results and Implications for Theories of Hydrophobic Solvation. *J. Phys. Chem. B* **2000**, *104*, 6271–6285.
- (34) Tan, C.; Tan, Y. H.; Luo, R. Implicit Nonpolar Solvent Models. *J. Phys. Chem. B* **2007**, *111*, 12263–12274.
- (35) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
- (36) Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D., Jr. An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chem. Rev.* **2016**, *116*, 4983–5013.

- (37) Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *J. Chem. Theory Comput.* **2015**, *11*, 3499–3509.
- (38) Mazar, M.; Pettitt, B. M. Convergence of the Chemical Potential in Aqueous Simulations. *Mol. Simul.* **1991**, *6*, 1–4.
- (39) Mitchell, M. J.; McCammon, J. A. Free Energy Difference Calculations by Thermodynamic Integration: Difficulties in Obtaining a Precise Value. *J. Comput. Chem.* **1991**, *12*, 271–275.
- (40) Chipot, C.; Pohorille, A. *Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Springer: Berlin, Germany, 2007.
- (41) Fennell, C. J.; Kehoe, C.; Dill, K. A. Oil/Water Transfer Is Partly Driven by Molecular Shape, Not Just Size. *J. Am. Chem. Soc.* **2010**, *132*, 234–240.
- (42) Fennell, C. J.; Kehoe, C. W.; Dill, K. A. Modeling Aqueous Solvation with Semi-Explicit Assembly. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 3234–3239.
- (43) Lin, B.; Wong, K. Y.; Hu, C.; Kokubo, H.; Pettitt, B. M. Fast Calculations of Electrostatic Solvation Free Energy from Reconstructed Solvent Density using Proximal Radial Distribution Functions. *J. Phys. Chem. Lett.* **2011**, *2*, 1626–1632.
- (44) Lin, B.; Pettitt, B. M. Electrostatic Solvation Free Energy of Amino Acid Side Chain Analogs: Implications for the Validity of Electrostatic Linear Response in Water. *J. Comput. Chem.* **2011**, *32*, 878–885.
- (45) Li, L.; Fennell, C. J.; Dill, K. A. Field-SEA: A Model for Computing the Solvation Free Energies of Nonpolar, Polar, and Charged Solutes in Water. *J. Phys. Chem. B* **2014**, *118*, 6431–6437.
- (46) Lake, P. T.; McCullagh, M. Implicit Solvation Using the Superposition Approximation (IS-SPA): An Implicit Treatment of the Nonpolar Component to Solvation for Simulating Molecular Aggregation. *J. Chem. Theory Comput.* **2017**, *13*, 5911–5924.
- (47) Mezei, M.; Beveridge, D. L. Structural Chemistry of Biomolecular Hydration via Computer Simulation: The Proximity Criterion. *Methods Enzymol.* **1986**, *127*, 22–47.
- (48) Subramanian, P. S.; Ravishanker, G.; Beveridge, D. L. Theoretical Considerations on the 'Spine of Hydration' in the Minor Groove of d(CGCGAATTCGCG).d(GCGCTTAAGCGC): Monte Carlo Computer Simulation. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 1836–1840.
- (49) Lounnas, V.; Pettitt, B. M.; Phillips, G., Jr. A Global Model of the Protein-Solvent Interface. *Biophys. J.* **1994**, *66*, 601–614.
- (50) Phillips, G. N., Jr.; Pettitt, B. M. Structure and Dynamics of the Water around Myoglobin. *Protein Sci.* **1995**, *4*, 149–158.
- (51) Burling, F. T.; Weis, W. I.; Flaherty, K. M.; Brünger, A. T. Direct Observation of Protein Solvation and Discrete Disorder with Experimental Crystallographic Phases. *Science* **1996**, *271*, 72–77.
- (52) Rudnicki, W. R.; Pettitt, B. M. Modeling the DNA-Solvent Interface. *Biopolymers* **1997**, *41*, 107–119.
- (53) Makarov, V. A.; Andrews, B. K.; Pettitt, B. M. Reconstructing the Protein-Water Interface. *Biopolymers* **1998**, *45*, 469–478.
- (54) Nguyen, B. L.; Pettitt, B. M. Effects of Acids, Bases, and Heteroatoms on Proximal Radial Distribution Functions for Proteins. *J. Chem. Theory Comput.* **2015**, *11*, 1399–1409.
- (55) Ou, S. C.; Pettitt, B. M. Solute-Solvent Energetics Based on Proximal Distribution Functions. *J. Phys. Chem. B* **2016**, *120*, 8230–8237.
- (56) Ou, S. C.; Drake, J. A.; Pettitt, B. M. Nonpolar Solvation Free Energy from Proximal Distribution Functions. *J. Phys. Chem. B* **2017**, *121*, 3555–3564.
- (57) Drake, J. A.; Harris, R. C.; Pettitt, B. M. Solvation Thermodynamics of Oligoglycine with Respect to Chain Length and Flexibility. *Biophys. J.* **2016**, *111*, 756–767.
- (58) Baron, R. *Computational Drug Discovery and Design*; Humana: New York, NY, 2012.
- (59) Ben-Naim, A. *Water and Aqueous Solutions*; Plenum Press: New York, NY, 1992.
- (60) Mehrotra, P. K.; Beveridge, D. L. Structural Analysis of Molecular Solutions Based on Quasi-Component Distribution Functions. Application to [H₂CO]_{aq} at 25.Degree.C. *J. Am. Chem. Soc.* **1980**, *102*, 4287–4294.
- (61) Gerstein, M.; Tsai, J.; Levitt, M. The Volume of Atoms on the Protein Surface: Calculated from Simulation, using Voronoi Polyhedra. *J. Mol. Biol.* **1995**, *249*, 955–966.
- (62) Lin, B.; Pettitt, B. M. Note: On the Universality of Proximal Radial Distribution Functions of Proteins. *J. Chem. Phys.* **2011**, *134*, 106101.
- (63) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. Avoiding Singularities and Numerical Instabilities in Free Energy Calculations Based on Molecular Simulations. *Chem. Phys. Lett.* **1994**, *222*, 529–539.
- (64) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. Separation-Shifted Scaling, A New Scaling Method for Lennard-Jones Interactions in Thermodynamic Integration. *J. Chem. Phys.* **1994**, *100*, 9025.
- (65) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (66) Kale, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. NAMD2: Greater Scalability for Parallel Molecular Dynamics. *J. Comput. Phys.* **1999**, *151*, 283–312.
- (67) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell, A. D., Jr. CHARMM General Force Field (CGenFF): A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2009**, *31*, 671–690.
- (68) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (69) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of *n*-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (70) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (71) Grossfield, A.; Zuckerman, D. M. Quantifying Uncertainty and Sampling Quality in Biomolecular Simulations. *Annu. Rev. Comput. Chem.* **2009**, *5*, 23–48.
- (72) Su, Y.; Gallicchio, E. The Non-Polar Solvent Potential of Mean Force for the Dimerization of Alanine Dipeptide: the Role of Solute-Solvent van der Waals Interactions. *Biophys. Chem.* **2004**, *109*, 251–260.
- (73) Tomar, D. S.; Asthagiri, D.; Weber, V. Solvation Free Energy of the Peptide Group: Its Model Dependence and Implications for the Additive-Transfer Free-Energy Model of Protein Stability. *Biophys. J.* **2013**, *105*, 1482–1490.
- (74) Hu, C. Y.; Kokubo, H.; Lynch, G. C.; Bolen, D. W.; Pettitt, B. M. Backbone Additivity in the Transfer Model of Protein Solvation. *Protein Sci.* **2010**, *2010*, 1011–1022.
- (75) Deng, Y.; Roux, B. Hydration of Amino Acid Side Chains: Nonpolar and Electrostatic Contributions Calculated from Staged Molecular Dynamics Free Energy Simulations with Explicit Water Molecules. *J. Phys. Chem. B* **2004**, *108*, 16567–16576.
- (76) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **1994**, *98*, 1978–1988.
- (77) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Model for Aqueous Solvation Based on Class IV Atomic Charges and First Solvation Shell Effects. *J. Phys. Chem.* **1996**, *100*, 16385–16398.
- (78) Wolfenden, R.; Andersson, L.; Cullis, P. M.; Southgate, C. C. B. Affinities of Amino Acid Side Chains for Solvent Water. *Biochemistry* **1981**, *20*, 849–855.
- (79) Chang, J.; Lenhoff, A. M.; Sandler, S. I. Solvation Free Energy of Amino Acids and Side-Chain Analogues. *J. Phys. Chem. B* **2007**, *111*, 2098–2106.

- (80) König, G.; Bruckner, S.; Boresch, S. Absolute Hydration Free Energies of Blocked Amino Acids: Implications for Protein Solvation and Stability. *Biophys. J.* **2013**, *104*, 453–462.
- (81) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. Extremely Precise Free Energy Calculations of Amino Acid Side Chain Analogs: Comparison of Common Molecular Mechanics Force Fields for Proteins. *J. Chem. Phys.* **2003**, *119*, 5740.
- (82) Hua, L.; Zangi, R.; Berne, B. J. Hydrophobic Interaction and Dewetting between Plates with Hydrophobic and Hydrophilic Domains. *J. Phys. Chem. C* **2009**, *113*, 5244–5253.
- (83) Choudhury, N.; Pettitt, B. M. Dynamics of Water Trapped between Hydrophobic Solutes. *J. Phys. Chem. B* **2005**, *109*, 6422–6429.
- (84) Bauer, B. A.; Ou, S.; Patel, S.; Siva, K. Dynamics and Energetics of Hydrophobically Confined Water. *Phys. Rev. E* **2012**, *85*, 051506.