Published in final edited form as:

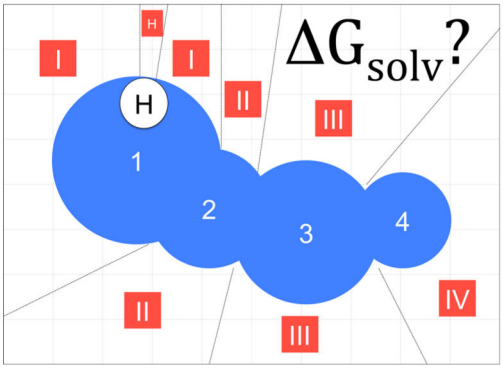
J Phys Chem B. 2017 April 20; 121(15): 3555–3564. doi:10.1021/acs.jpcb.6b09528.

Nonpolar Solvation Free Energy from Proximal Distribution Functions

Shu-Ching Ou, **Justin A. Drake**, and **B. Montgomery Pettitt***

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

Abstract



Using precomputed near neighbor or proximal distribution functions (pDFs) that approximate solvent density about atoms in a chemically bonded context one can estimate the solvation structures around complex solutes and the corresponding solute–solvent energetics. In this contribution, we extend this technique to calculate the solvation free energies (ΔG) of a variety of solutes. In particular we use pDFs computed for small peptide molecules to estimate ΔG for larger peptide systems. We separately compute the non polar (ΔG_{vdW}) and electrostatic (ΔG_{elec}) components of the underlying potential model. Here we show how the former can be estimated by thermodynamic integration using pDF-reconstructed solute–solvent interaction energy. The electrostatic component can be approximated with Linear Response theory as half of the electrostatic solute–solvent interaction energy. We test the method by calculating the solvation free energies of butane, propanol, polyalanine, and polyglycine and by comparing with traditional free energy simulations. Results indicate that the pDF-reconstruction algorithm approximately reproduces ΔG_{vdW} calculated by benchmark free energy simulations to within ~ kcal/mol

***Corresponding Author**, mpettitt@utmb.edu; Phone: (409)772-0723. .

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.6b09528. Graphical illustration of calculating pDFs, test of cavity and exclusion factors for pDF-reconstructions, fixed solute configurations, convergence of $\langle \partial U^{\mathrm{vdW}}/\partial\lambda \rangle$, pDFs for solute atoms with water hydrogen atoms, pDF-reconstructed $\langle \partial U^{\mathrm{vdW}}/\partial\lambda \rangle$ for all four conformers of deca-alanine, comparison of using different amount of λ windows for calculating ΔG_{vdW} of polyalanines (PDF)

ORCID

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

Notes

The authors declare no competing financial interest.

accuracy. The use of transferable pDFs for each solute atom allows for a rapid estimation of ΔG for arbitrary molecular systems.

INTRODUCTION

The accurate calculation of solvation free energies is critical to computational drug design and structure optimization as well as understanding important biophysical processes like aggregation and protein folding. Two central issues in computational free energy calculations include ensuring sufficient sampling of phase space and the accuracy of the force field used to model molecular interactions. The sampling issue remains a bottleneck in applications of free energy approaches to ever more complex systems; that is, in order to investigate thermodynamic behavior of a system, sampling of a representative number of configurations within the given ensemble is required, which often requires expensive simulations.[1–3]Force fields continue advancing in accuracy often at the expense of added complexity.[4–6]

Due to the transferable nature of atomic level force fields for a given chemical context, molecular mechanics force fields provide an extensible architecture to study the thermodynamic properties of a wide range of solutes.[7] Free energies of solvation (ΔG) depend on the solute–solvent potential,[8] and the distribution of solvent around a solute. Here, we seek a set of transferable functions to approximate solvent density distributions which will allow for rapid estimation of ΔG and its differences due to chemical processes in solution.

In assessing free energy differences it is common to consider a thermodynamic cycle for a given reaction process. For example, the relative change in free energy associated with the change or mutation of molecule A to B is given by the difference in solvation free energy ($\Delta\Delta G$) of the two molecules.[9–14] In computational free energy methods, the solvation process in which a solute is transferred from the gas phase to the aqueous phase is often decomposed into two stages or contributions. First an uncharged, solute cavity is created in the solution and the free energy required to insert this cavity is defined as ΔG_{vdW}. This cavity is then gradually charged and the change in free energy associated with this process is defined as ΔG_{elec}.[15] This process, though nonphysical, does allow us to calculate each component of the potential function with explicit simulations[16] or individual functional theories with less computational costs.[17,18] While ΔG is a state function and path independent, the components ΔG_{elec} and ΔG_{vdW} are well-known to be defined only by a process and therefore path-dependent.[19]

Many approximate techniques have been developed to rapidly estimate ΔG. ΔG_{elec} is often estimated using dielectric continuum theory based techniques,[15,20] such as the generalized Born model,[21,22] Poisson–Boltzmann equation,[23,24] or Linear Response theory (LRT).[25–27] The nonpolar cavity contribution, ΔG_{vdW}, has been approximated as a linear function of the solvent accessible surface area (SASA).[28–30] However, the correlation between ΔG_{vdW} calculated by SASA and explicit solvent models using rigorous thermodynamic integration or free energy perturbation is far from exact for organic molecules, especially at the level of small nonpolar molecules.[16,31–33] Further decomposition of ΔG_{vdW} into the solute–solvent

repulsive and attractive interactions has been proposed with an additional estimator accounting for solvent accessible volume.[18] The use of implicit solvation models to estimate ΔG_{vdW} remains a challenge due to nonadditive, multibody correlations with neighboring atoms,[33–35] and the complex geometry of the solute–solvent interface.[36–39] An approach that yields estimates of solvation free energies rapidly but retains the accuracy associated with explicit solvent models would be highly desirable. Methods relying on precomputed quantities from explicit simulations of small solutes to estimate ΔG_{vdW} of large solutes are therefore being developed. For example, the Semi-Explicit Assembly (SEA) approach[40–42] precomputes the modified nonbonded van der Waals interactions between a series of nonpolar solute spheres and water, then assembles the total interactions based on the given configurations of large solutes.

Here, using precomputed near neighbor or proximal distribution functions (pDFs) of solvation about atoms in chemically bonded (small molecule) context we estimate the solvation structures around larger and more complex solutes. We then use these approximately transferable distributions to calculate ΔG. pDFs approximate solute–solvent pair correlation functions at the near neighbor level to describe the local solvent structure around a solute.[43–47] They have been shown to be a near universal and transferable descriptor between different globular proteins.[48] That is, pDFs constructed for small molecules can be used to predict the solvent structures around complex biological macromolecules that are composed of chemically similar components.[49–51] Average solute–solvent interaction energies, and subsequently solvation free energies, can then be calculated from these reconstructed solvent structures.

Electrostatic interaction energies estimated using pDFs in combination with LRT have been shown to approximate ΔG_{elec} well for a variety of solutes compared to simulation.[52–54] We have recently shown that van der Waals solute–solvent interaction energies can be approximated with good accuracy using precomputed pDFs, but requires a finer spatial resolution than what is used to estimate the more smoothly varying electrostatic interactions.[55] Here, we extend the pDF-approach to calculate ΔG_{vdW} of a variety of solutes and assess the accuracy of this approach by a comparison with the more expensive explicit solvent thermodynamic integration.

Below we review the generation of pDFs from small solutes and how they can be used to reconstruct the distribution of solvent density around an arbitrary solute to estimate average solute–solvent interaction energies. Then we show how to calculate ΔG_{vdW} by constructing a set of pDFs as a function of a coupling parameter that gradually scales the van der Waals interactions between a small solute and solvent (i.e., we define a pathway along which to calculate ΔG_{vdW}). To benchmark this approach, we calculate ΔG_{vdW} of butane, and propanol and compare with more rigorous estimates obtained from alternative approaches. We then generate transferable pDFs from monomers of alanine and glycine and use those to calculate ΔG_{vdW} for various triglycine and deca-alanine systems. We study the effects of flexibility, or disorder, on ΔG_{vdW} by comparing ΔG_{vdW} of a fixed, extended conformation of triglycine to that obtained for a large ensemble of triglycine conformers generated from a molecular dynamics (MD) simulation. For completeness, we include linear response estimates of ΔG_{elec} and thus construct an approximate total solvation free energy.

THEORY

In this section, we briefly review nearest neighbor proximal distribution functions (pDFs) and how they are used to reconstruct solute–solvent density distributions and subsequently estimate solute–solvent energetics. Then we discuss how the pDF formalism can be modified to estimate ΔG_{vdW} using a thermodynamic integration scheme.

Distribution Functions and Solute–Solvent Energetics

Proximal distribution functions (pDFs, or $g_{\perp}(r)$) are average density distributions calculated from solvent molecules to their nearest solute atom.[43,56,57] The solvent probability distribution most nearly perpendicular to the k_{th} solute atom $g_{\perp}^{k}(r)$ can be written as

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

where $\overrightarrow{r}_{\mathrm{ij}}(t)$ 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. $\mathrm{Inf}[|\overrightarrow{r}_{\mathrm{ij}}(t)|]_{i=1,m}$ returns the distance between a particular solvent atom and the nearest solute atom k, i.e., k is the assigned solute atom with the closest distance to the j_{th} solvent atom. In this way, the solvent space is effectively divided into Voronoi polyhedra[58] and the resulting distribution functions are defined essentially perpendicularly to the exposed surface of the respective solute atoms[52] (illustrated in Supporting Information). $\delta\tau(\overrightarrow{r}_{j}(t),k)$ is the volume element around the j_{th} solvent molecule, which is defined by all $\overrightarrow{r}$ vectors, where $|\overrightarrow{r}_{k}-\overrightarrow{r}|\leq|\overrightarrow{r}_{i}-\overrightarrow{r}|$. A graphical illustration for this process can be found in the Supporting Information. 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. The detailed theory and computational procedure can be found elsewhere.[43,45]

A variety of solute atom sets or groupings can be used to define the pDFs. For example, at one extreme one could take each atom of a molecule as having a unique $g_{\perp}(r)$. Alternatively, pDFs can be constructed for each force field atom type or, using an even coarser grouping, by chemical identity or similarity.[49] In general, we refer to a particular grouping or atom set by χ and refer to a particular pDF defined by this grouping with χ_k. Previous work shows that the pDFs classified using force field atom types for proteins and nucleic acids yields nearly universal functions which are thus transferable to chemically similar solute molecules.[45,48,50]

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 $g_{\perp}^{\chi_k}(r')$:

$$g(x,y,z) = g_{\perp}^{\chi_k}(r') \quad (2)$$

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 set type 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

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

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 based on previous work.[55]

Any solute atom may exclude solvent from any grid point when there is overlap or volume exclusion. In solvent density reconstruction, a grid point is assigned zero solvent density if any $g_{\perp}^{\chi_k}(r') \approx 0$. As an example we consider a grid point density assignment near a hydrogen covalently bonded to a heavy atom whose van der Waals radius effectively eclipses that of the hydrogen's. While the hydrogen atom is the nearest grid point and $g_{\perp}^{\mathrm{H}}(r') > 0$, the pDF of the bonded heavy atom (i.e., next nearest neighbor) could dictate that solvent should be excluded at the grid point with the given distance, r'. This becomes increasingly important at small distances when using a soft-core van der Waals potential as is common in free energy calculations (addressed in next section, Supporting Information, and ref 55).

We use a two step process to ensure that grid points are assigned the appropriate solvent density during reconstruction. First we eliminate grid points at distances in which the pDFs of heavy atoms in the set χ predict essentially zero solvent occupation (referred to as the cavity rule). Conceptually, solvent space is divided into Vornoi polyhedra with heavy atoms at this step. Then, we exclude or zero out any grid points in a similar manner using $g_{\perp}^{\chi}$ of the remaining atoms (referred to as the exclusion rule). The whole reconstruction process including the cavity and exclusion rules can therefore be formulated as

$$\rho(x,y,z) = \{\prod_{i \neq \chi_k}^{m_{\mathrm{H}}} \delta_{\mathrm{cav}}^{i}(|(x,y,z) - \vec{r}_i|)\} \cdot \rho g_{\perp}^{\chi_k}(r') \cdot \{\prod_{j \neq \chi_k}^{m} \delta_{\mathrm{excl}}^{j}(|(x,y,z) - \vec{r}_j|)\} \quad (4)$$

Here m_{H} is the number of solute heavy atoms, and m is the number of all solute atoms. The cavity factor $\delta_{\mathrm{cav}}^{i}$ is defined as

$$\delta^i_{\text{cav}} = \begin{cases} 1, & \text{if } g^{\chi_i}_{\perp}(|(x,y,z) - \overrightarrow{r}_i|) > 0 \\ 0, & \text{if } g^{\chi_i}_{\perp}(|(x,y,z) - \overrightarrow{r}_i|) = 0 \end{cases} \quad (5)$$

and notice that *i* loops over solute heavy atoms only. Similarly, the exclusion factor δ^j_{excl} is written as

$$\delta^j_{\text{excl}} = \begin{cases} 1, & \text{if } g^{\chi_j}_{\perp}(|(x,y,z) - \overrightarrow{r}_j|) > 0 \\ 0, & \text{if } g^{\chi_j}_{\perp}(|(x,y,z) - \overrightarrow{r}_j|) = 0 \end{cases} \quad (6)$$

where *j* loops over the remaining atoms in the set χ.

Once the solvent density is reconstructed around the solute, the average solute–solvent interaction energy can then be written as

$$\langle U_{\text{solu-solv}} \rangle \approx \sum^{x,y,z} \sum_{i=1}^{m} \Delta\nu\rho(x,y,z) U(|(x,y,z) - \overrightarrow{r}_i|) \quad (7)$$

where $U(|(x,y,z) - \overrightarrow{r}_i|)$ is the solute–solvent interaction energy for solute atom *i*, Δv is the unit volume which depends on the spatial resolution and the outer sum is taken over all grid points. Depending on the potential energy function, *U* 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 7.

We have previously shown that the solute–solvent electrostatic interaction energy and electrostatic solvation free energy can be reasonably approximated from the pDF-reconstructed solvent density. Here, we extend the pDF approach to estimate the change in free energy (ΔG_{vdW}) upon inserting uncharged solutes in solution using a soft-core van der Waals potential and a thermodynamic integration scheme.

Thermodynamic Integration with Soft-Core Potentials

To calculate the solvation free energy of cavity formation (ΔG_{vdW}) we use a thermodynamic integration (TI) approach. We construct a 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 $\lambda = 1$, $U_{ij} = U^{\text{vdW}}_{ij}$, 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^{vdW}_{ij} along the pathway:[59,60]

$$U_{ij}^{\mathrm{vdW}}(r;\lambda)=4\epsilon_{ij}\lambda\left[\left(\frac{\sigma_{ij}^2}{r_{ij}^2+\delta(1-\lambda)}\right)^6-\left(\frac{\sigma_{ij}^2}{r_{ij}^2+\delta(1-\lambda)}\right)^3\right] \quad (8)$$

where r_{ij} is the distance between solute and solvent atoms, ϵ is the van der Waals well depth, and σ is the contact distance. δ is a radius-shifting coefficient of λ and here is taken to be 5.0 throughout the simulations and pDF-reconstructions. Trivially at λ = 1 this is the usual Lennard-Jones form. We note that the effective σ_{ij} is reduced as λ decreases, which leads to water residing at closer distances to the solute atoms.

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:

$$\Delta G_{\mathrm{vdW}}=\int_0^1\left\langle\frac{\partial U^{\mathrm{vdW}}(\lambda)}{\partial\lambda}\right\rangle_\lambda d\lambda \quad (9)$$

One can approximately calculate ΔG_{vdW} using numerical integration methods with the ensemble averages of $\partial U^{\mathrm{vdW}}/\partial\lambda$ at various values of λ. The derivative of U^{vdW} with respect to λ is readily obtained from the analytical expression:

$$\begin{aligned}\frac{\partial U_{ij}^{\mathrm{vdW}}(r;\lambda)}{\partial\lambda}=\;& 4\epsilon_{ij}\left\{\left[\frac{\sigma_{ij}^2}{r_{ij}^2+\delta(1-\lambda)}\right]^6\left[1+\frac{6\lambda\sigma_{ij}}{r_{ij}^2+\delta(1-\lambda)}\right]\right.\\ &-\left.\left[\frac{\sigma_{ij}^2}{r_{ij}^2+\delta(1-\lambda)}\right]^3\left[1+\frac{3\lambda\sigma_{ij}}{r_{ij}^2+\delta(1-\lambda)}\right]\right\}\\ \frac{\partial U^{\mathrm{vdW}}}{\partial\lambda}=&\sum_{i,j}\frac{\partial U_{ij}^{\mathrm{vdW}}}{\partial\lambda}\end{aligned} \quad (10)$$

In our pDF approach, we first construct $\mathrm{g}_\perp^{\chi_k}$ for the atom type set of a small representative chemical group (e.g., an alkane or peptide substituent) from MD simulations performed at regularly spaced values of λ. For each λ, we obtain unique $\mathrm{g}_\perp^{\chi_k}$s. 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 long alkane or polypeptide) for each λ value. We directly estimate $\langle\partial U^{\mathrm{vdW}}/\partial\lambda\rangle_\lambda$ using a form like eq 7 but with U_{ij} replaced with $\partial U_{ij}^{\mathrm{vdW}}/\partial\lambda$. Finally, numerically integrating $\langle\partial U^{\mathrm{vdW}}/\partial\lambda\rangle_\lambda$ 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.

METHODS

Simulation

The solutes butane, propanol, ala_1, ala_{10}, gly_1, gly_3 were chosen for this study. Molecular dynamics simulations were performed with NAMD 2.9 and 2.10[61,62] with the CHARMM36 force field parameters[63] to generate the pDFs and calculate free energy benchmarks. pDFs were generated for butane and propanol as a control to compare with more rigorous free energy calculations. The pDFs generated for ala_1 and gly_1 were used in the analysis of longer peptides. Each solute was solvated with TIP3P water[64] in a volume with at least 10 Å from each boundary to the solute atoms. Peptides were capped with neutral acetyl (ACE) and *N*-methyl amide (NME) chemical groups. The initial conformations of decaalanines and triglycine peptides are from previous studies[33,52,55,65] and presented in the Supporting Information. Three-dimensional periodic boundary conditions were applied. A rigid water geometry is enforced using SHAKE.[66] Particle Mesh Ewald (PME)[67] 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.

The pDFs for this free energy study were generated in 11 windows with a λ-spacing of 0.1 between 0 and 1. Solvent reconstruction and calculation of ΔG_{vdW} were performed with these λ values. In general, it is necessary to sample more extensively at certain λ-windows depending on the chosen soft-core potential function (near end points in our case) to accurately depict the repulsive van der Waals forces when inserting the solute into the system.[11] More detail on the analysis can be found in the Results Section below. In this article we show how to use pDFs to reproduce $\langle \partial U^{vdW}/\partial\lambda \rangle$ at all λ-windows with our pDF-reconstruction algorithm, which is subsequently integrated to estimate ΔG_{vdW}.

Throughout the simulations, the solute molecules were rigid except as noted below for gly_3. The pDF sampling simulations were initially equilibrated in the NVT ensemble for 1 ns. The simulations were then switched to NPT at 1 atm pressure, with the first 1 ns excluded as equilibration. The pDFs and $\langle \partial U^{vdW}/\partial\lambda \rangle$ were calculated from production simulations that ranged between 4 and 10 ns at each λ. Trajectory snapshots were saved every 0.2 ps for analyses. The convergence of $\langle \partial U^{vdW}/\partial\lambda \rangle$ at each λ window is discussed in the Supporting Information. For each λ window, the uncertainty of $\langle \partial U^{vdW}/\partial\lambda \rangle$ is estimated as the block standard error.[68] The final uncertainty of ΔG_{vdW} is calculated by the propagation of errors across all λ.

In addition to the fixed solute conformations, we also examined ΔG_{vdW} of gly_3 without positional constraints (i.e., allowed to be completely flexible during simulations). The simulation parameters and λ-windows are identical to those of simulations with gly_3 fixed in an extended conformation. We used 10000 solute configurations from a typical 50 ns MD simulation ($\lambda = 1$) of flexible gly_3 for calculating ΔG_{vdW} from pDF-reconstructions.

In the following sections, we also provide estimates of ΔG_{elec} based on LRT. That is, ΔG_{elec} can be approximated as half of the solute–solvent electrostatic interaction energy instead of

integrating the average derivative of U^{elec} with respect to a coupling parameter that gradually scales solute–solvent electrostatic interactions. Thus, a set of simulations with charged solutes is employed to achieve full electrostatic solute–solvent interactions. In our previous work, we have reported the averaged solute–solvent electrostatic interaction energies for configurations of butane, propanol, and ala_{10},[55] along with the corresponding pDFs used for solvent density reconstruction. In this contribution, with a goal of calculating total solvation free energy, we further include estimates of ΔG_{elec} for fixed ala_1, gly_1, and flexible gly_3 from free energy simulations to compare with pDF-reconstructions. The simulation details are the same as the above free energy simulations except for having charges on solute atoms and the total sampling time for each system is 100 ns with a 2 ps save frequency.

RESULTS AND DISCUSSIONS

Butane and Propanol

We compare the $\langle \partial U^{vdW}/\partial \lambda \rangle$ of a single butane/propanol molecule obtained from TI simulations and pDF-reconstruction using the same solute configuration and same number of λ points. In general, the more extensive the set of atom types we include, the higher the accuracy in solvent structure and solute–solvent energetics we can expect during the reconstructions,[51,55] especially for the nonpolar contributions. Here, we consider all six force field atom types (i.e., the set χ) for butane and propanol: methylene carbon (CT2), methyl carbon (CT3), methylene hydrogen (HA2), methyl hydrogen (HA3), the hydroxyl oxygen (O1), and the bonded polar hydrogen (HO1). The pDFs of each atom type with respect to water oxygen atom (OT) at selected λ windows are shown in Figure 1. The pDFs of each atom type with water hydrogen atoms (HT) are included in the Supporting Information.

Ideally, the pDFs at $\lambda = 0$ should be 1 (i.e., bulk solvent density) at all interatomic distances. Therefore, we set the pDFs at $\lambda = 0$ to 1 for all distances. Essentially, all atom types are uncharged and considered as pure covalently bonded L-J spheres, and thus show similar behaviors; as λ increases, the position of the peaks are shifted to larger distances with stronger magnitudes in the first peak.

The simulated and pDF-reconstructed estimates $\langle \partial U^{vdW}/\partial \lambda \rangle$ for each λ are presented in Figure 2. Differences between $\langle \partial U^{vdW}/\partial \lambda \rangle$ calculated by TI simulations and pDF reconstruction at each λ are shown in the insets. Overall $\langle \partial U^{vdW}/\partial \lambda \rangle$ calculated by pDF reconstruction for both solutes are consistent with the simulations with deviations for most windows less than 1 kcal/mol. The largest deviation (~2 kcal/mol) of $\langle \partial U^{vdW}/\partial \lambda \rangle$ between both methods occurs at $\lambda = 0.1$. The deviations are diluted upon integration when computing ΔG_{vdW}. ΔG_{vdW} calculated for both solutes and methods are listed in Table 1. Differences in simulated and pDF-reconstructed ΔG_{vdW} for butane and propanol are −0.33 and 0.18 kcal/mol, respectively. The uncertainty for each λ window is estimated using block standard errors,[68] and the final uncertainty of ΔG_{vdW} is computed using error propagation. We include the estimated ΔG_{elec} for butane and propanol from TI simulations and pDF-reconstructions in Table 1. These values are obtained as half of solute-water electrostatic interaction energies based on standard LRT[48,52] and are adapted from our previous work.[55]

Essentially all free energy components are in good accord with those obtained from TI simulations.

To demonstrate the utility of using this pDF approach to estimate solvation free energies, we consider a free energy cycle of mutating the methyl group in butane to a hydroxyl group to form propanol (illustrated in Figure 2). Since free energy is a state function, $\Delta\Delta G = \Delta G_4 - \Delta G_3 = \Delta G_2 - \Delta G_1$. ΔG_3 and ΔG_4 are the free energies of mutating butane to propanol in vacuum and solution, respectively, and their difference can be calculated by the difference in solvation free energies of butane and propanol. Using the values reported in Table 1, we obtain the pDF-reconstructed $\Delta\Delta G = -6.56$ kcal/mol, which is within 1 kcal/mol of the simulated $\Delta\Delta G = -7.50$ kcal/mol.

Deca-alanine

We next estimate ΔG_{vdW} from precomputed pDFs for a deca-peptide. pDFs were constructed from a single alanine monomer (ala_1) and used to compute ΔG_{vdW} of four different conformers of deca-alanine, ala_{10} (referred to as d, d1, d2, d3 in Supporting Information). The set of atoms chosen for pDF generation and solvent density reconstruction include {CT3, C, O, NH1, CT1, HA3, HN, HB1}, where C and O represent the carbon and oxygen atoms of the carbonyl group, NH1 is the peptide nitrogen, HN is the polar hydrogen bonded to the nitrogen, CT1 is the backbone carbon, and HB1 is the backbone hydrogen. We use the pDFs for CT3 and HA3 from our analysis of butane and propanol (Figure 1). The pDFs for the rest of the atom types at several values of λ are presented in Figure 3.

We start with comparing the simulated and pDF-reconstructed estimates of $\langle \partial U^{\mathrm{vdW}}/\partial\lambda \rangle$ for ala_1, as shown in Figure 4a, using the same solute configuration. From the inset we observe that the deviations of $\langle \partial U^{\mathrm{vdW}}/\partial\lambda \rangle$ between both methods are slightly larger than those observed for butane and propanol for all λ windows, but still within 2 kcal/mol. The simulated and the pDF-reconstructed ΔG_{vdW} and ΔG_{elec} are listed in Table 2. The difference in ΔG_{vdW} between simulation and pDF-reconstruction is 0.60 kcal/mol, while the difference in ΔG_{elec} is −0.51 kcal/mol. These differences cancel and yield a pDF reconstructed ΔG within the uncertainty of ΔG obtained by TI simulations. We note ΔG_{elec} is generally the dominating term in ΔG. Estimated uncertainties of ΔG_{elec} are smaller than the uncertainties of ΔG_{vdW} in most of our studies.

We use the pDFs calculated from ala_1 to estimate the ΔG_{vdW} of ala_{10}. The results of simulated and pDF-reconstructed $\langle \partial U^{\mathrm{vdW}}/\partial\lambda \rangle$ at each λ for conformer d are shown in Figure 4b. The results for the other three conformers are similar and included in the Supporting Information. The inset shows the deviations of $\langle \partial U^{\mathrm{vdW}}/\partial\lambda \rangle$ between simulations and pDF-reconstructions at each λ for all four conformers. While the deviations are large (10–15 kcal/mol/window) and increase with system size, they are suppressed upon integration when calculating ΔG_{vdw} in a manner similar to what we observed for butane and propanol. The pDF reconstructed estimates of ΔG_{vdW} of all ala_{10} conformers are consistent with those obtained from TI simulations (Table 2). The differences between simulated and reconstructed ΔG_{vdW} for the different conformers in order are 0.17, 2.88, 0.84, and −0.05 kcal/mol. Differences of ΔG_{elec} between methods (adapted from our previous work[55]) are larger than those observed for ΔG_{vdW}. Owing to the extremely twisted configuration of

conformer d1, we find the largest deviation between simulation and reconstruction for both ΔG_{vdW} and ΔG_{elec}. We note that Poisson–Boltzmann calculations fared worse in reproducing the ΔG_{elec} of d1.[52] This highly twisted conformation is a low probability structure with peculiar internal interatomic correlations. Thus, we consider this conformer to be a most challenging case for this theory based on pairwise near neighbor solvent correlations for the estimation of ΔG.

Flexible and Fixed Triglycine

Finally, we extend our approach to the calculation of ΔG_{vdW} for a large, diverse ensemble of gly_3 conformations and assess the effects of flexibility, or disorder, on these calculations by comparing to a single fixed/extended conformation of gly_3. We chose to calculate the pDFs from a single fixed glycine (gly_1) to evaluate the ΔG_{vdW} of fixed and flexible triglycines (gly_3). The configuration of fixed gly_3 is in the Supporting Information. The atom set used to generate the pDFs used here include {CT3, C, O, NH1, CT2, HA3, HN, HB2}. pDFs for these atom types have been presented in the previous sections except for HB2, which represents the aliphatic backbone hydrogen atoms bonded to the CT2 atoms. Results of simulated and pDF-reconstructed ΔG_{vdW} and ΔG_{elec} for fixed gly_1 are presented in Table 3. Differences of both components between methods are 0.06 kcal/mol. We next use these precomputed pDFs to estimate ΔG_{vdW} of fixed and flexible gly_3.

Consider the ΔG_{vdW} term. The comparisons of simulated and pDF-reconstructed $\langle \partial U^{vdW}/\partial \lambda \rangle$, along with the deviations (in the insets) at each λ for flexible and fixed gly_3 are presented in Figure 5a,b. The deviations are all within 2 kcal/mol. The values of ΔG_{vdW} are listed in Table 3. To estimate ΔG_{vdW} of flexible gly_3, we extracted 10000 solute configurations from a molecular dynamics simulation and, for each one, calculated $\langle \partial U^{vdW}/\partial \lambda \rangle$ at all λ values. The $\langle \partial U^{vdW}/\partial \lambda \rangle$ values are then configurationally averaged. We show the distribution over all conformations of $\langle \partial U^{vdW}/\partial \lambda \rangle$ at $\lambda = 0$ from simulation in Figure 5c. This shows the deviation among the solute configurations. We expect the distribution of $\langle \partial U^{vdW}/\partial \lambda \rangle$ to be different depending on the force field used given that we have previously shown that CHARMM36 and Amber ff12SB generate different ensembles of structures.[69] It would be an interesting comparison to test the distributions of ΔG_{vdW} using different force fields as they can display different distributions of solute configurations and flexibility.

The differences between simulated and pDF-reconstructed ΔG_{vdW} for flexible and fixed gly_3 are within 0.2 kcal/mol. In addition to the 11 λ windows used for thermodynamic integration, we include the results using the free energy perturbation (FEP) method with 50 equally spaced λ windows between 0 and 1 from recently published work.[65] Due to the nature of the soft-core potential used in this contribution, the integrand is changing more rapidly (as shown in all $\langle \partial U^{vdW}/\partial \lambda \rangle$ figures) at small λ. This is a likely source of error in the ΔG_{vdW} reported here. The comparison of using different numbers of λ windows for calculating the ΔG_{vdW} of the same solute can be found in Supporting Information. We observe that $\Delta\Delta G_{vdW} = (\Delta G^{flex}_{vdW} - \Delta G^{fix}_{vdW})$ calculated from the current free energy simulations, pDF-reconstructions, and FEP are 0.88, 0.77, and 0.81, respectively. The fact that the ΔG_{vdW} for flexible gly_3 is less favorable (or more repulsive) than the fixed gly_3 is not surprising. With solute flexibility, gly_3 can form configurations that prevent water from

accessing parts of the solute, consequently increasing ΔG_{vdW}. The striking part is that all three methods eventually lead to a consistent $\Delta\Delta G_{\mathrm{vdW}}$, which is not trivial. More extensive tests of our approximate $\Delta\Delta G_{\mathrm{vdW}}$ are certainly warranted.

To complete the estimation of the total solvation free energy, results for ΔG_{elec} are presented in Table 3. Here we do not include the ΔG_{elec} of flexible gly_3 calculated by TI simulations since the noise of ΔG_{elec} computed from the ensemble of solute and solvent configurations instead of from only the ensemble of solute configurations (like in continuum models or the current concept using pDF-reconstruction) is much larger.[70–72] For the fixed gly_3, the differences of ΔG_{elec} between these three methods are within 4 kcal/mol. For flexible gly_3, the difference of ΔG_{elec} between pDF-reconstruction and FEP increases slightly and is within 5 kcal/mol considering a 0.5 kcal/mol uncertainty for FEP reported ΔG_{elec}. Figure 5 d shows the distribution of pDF-reconstructed ΔG_{elec} using 10000 solute configurations generated from simulation. This distribution has a standard deviation of 7.42 and is larger than the distribution from pDF-reconstructed ΔG_{vdW}. $\Delta\Delta G_{\mathrm{elec}}=(\Delta G_{\mathrm{elec}}^{\mathrm{flex}}-\Delta G_{\mathrm{elec}}^{\mathrm{fix}})$ calculated from pDF-reconstruction and FEP are 4.62 and 1.93 kcal/mol, respectively.

DISCUSSION AND CONCLUSION

In this contribution, we used precomputed solvent density distributions at multiple alchemical windows with free energy methods to estimate the van der Waals and electrostatic solvation free energies for a variety of solutes. By computing solute–solvent pDFs from small molecules, we can reconstruct the solvent densities around arbitrary solutes with similar chemical composition and, subsequently, estimate solvation free energies without performing lengthy simulations. Using this framework, we have shown that ΔG_{vdW} can be reproduced to within useful accuracy. In the future, we will extend our work to the calculation of solvation free energies of complex macromolecules.

The examples of calculating $\Delta\Delta G$ encourage applications to different biological systems. Combinations of solvation free energies with appropriate free energy cycles may allow rapid estimations for molecular associations,[10] alanine scanning,[73] and the predictions of $\mathrm{p}K_{\mathrm{a}}$ values of proteins and amino acid side chains.[74]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

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

REFERENCES

(1). Mazor M, Pettitt BM. Convergence of the Chemical Potential in Aqueous Simulations. Mol. Simul. 1991; 6:1–4.

(2). Mitchell MJ, McCammon JA. Free Energy Difference Calculations by Thermodynamic Integration: Difficulties in Obtaining a Precise Value. J. Comput. Chem. 1991; 12:271–275.

(3). Chipot, C., Pohorille, A. Free Energy Calculations: Theory and Applications in Chemistry and Biology. Springer; Berlin, Germany: 2007.

(4). Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J. Chem. Theory Comput. 2015; 11:3696–3713. [PubMed: 26574453]

(5). Lemkul JA, Huang J, Roux B, MacKerell AD Jr. An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. Chem. Rev. 2016; 116:4983–5013. [PubMed: 26815602]

(6). Robertson MJ, Tirado-Rives J, Jorgensen WL. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. J. Chem. Theory Comput. 2015; 11:3499–3509. [PubMed: 26190950]

(7). Jorgensen W. Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers. Application to Liquid Water. J. Am. Chem. Soc. 1981; 103:335–340.

(8). Yu HA, Karplus M. A Thermodynamic Analysis of Solvation. J. Chem. Phys. 1988; 89:2366–2379.

(9). Tembre B, McCammon J. Ligand-Receptor Interactions. Comput. Chem. 1984; 8:281–283.

(10). Kollman P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. Chem. Rev. 1993; 93:2395–2417.

(11). Jorgensen WL, Ravimohan C. Monte Carlo Simulation of Differences in Free Energies of Hydration. J. Chem. Phys. 1985; 83:3050–3054.

(12). Caldwell JW, Agard DA, Kollman PA. 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. [PubMed: 1896427]

(13). Prevost M, Wodak SJ, 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. [PubMed: 1961758]

(14). Dang LX, Merz KM Jr. Kollman PA. Free Energy Calculations on Protein Stability: Thr-157—-Val-157 Mutation of T4 Lysozyme. J. Am. Chem. Soc. 1989; 111:8505–8508.

(15). Cramer CJ, Truhlar DG. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. Chem. Rev. 1999; 99:2161–2200. [PubMed: 11849023]

(16). Mobley DL, Bayly CI, Cooper MD, Shirts MR, Dill KA. Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations. J. Chem. Theory Comput. 2009; 5:350–358. [PubMed: 20150953]

(17). Simonson T, Roux B. Concepts and Protocols for Electrostatic Free Energies. Mol. Simul. 2016; 42:1090–1101.

(18). Levy RM, Zhang LY, Gallicchio E, Felts AK. On the Nonpolar Hydration Free Energy of Proteins: Surface Area and Continuum Solvent Models for the Solute-Solvent Interaction Energy. J. Am. Chem. Soc. 2003; 125:9523–9530. [PubMed: 12889983]

(19). Smith PE, van Gunsteren WF. When Are Free Energy Components Meaningful? J. Phys. Chem. 1994; 98:13735–13740.

(20). Schaefer M, van Vlijmen HWT, Karplus M. Electrostatic Contributions to Molecular Free Energies in Solution. Adv. Protein Chem. 1998; 51:1–58. [PubMed: 9615168]

(21). Roux B, Yu HA, Karplus M. Molecular Basis for the Born Model of Solvation. J. Phys. Chem. 1990; 94:4683–4688.

(22). Bashford D, Case DA. Generalized Born Models of Macromolecular Solvation Effects. Annu. Rev. Phys. Chem. 2000; 51:129–152. [PubMed: 11031278]

(23). Netz RR, Orland H. Beyond Poisson-Boltzmann: Fluctuation Effects and Correlation Functions. Eur. Phys. J. E: Soft Matter Biol. Phys. 2000; 1:203–214.

(24). 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. [PubMed: 17969016]

(25). Hansen, JP., McDonald, IR. Theory of Simple Liquids. Academic Press; San Diego, CA: 1976.

(26). Chandler, D. Introduction to Modern Statistical Mechanics. Oxford University Press; New York, NY: 1987.

(27). Aqvist J, Luzhkov VB, Brandsdal BO. Ligand Binding Affinities from MD Simulations. Acc. Chem. Res. 2002; 35:358–365. [PubMed: 12069620]

(28). Sharp KA, Nicholls A, Fine RF, Honig B. Reconciling the Magnitude of the Microscopic and Macroscopic Hydrophobic Effects. Science. 1991; 252:106–109. [PubMed: 2011744]

(29). Cramer CJ, Truhlar DG. An SCF Solvation Model for the Hydrophobic Effect and Absolute Free Energies of Aqueous Solvation. Science. 1992; 256:213–217. [PubMed: 17744720]

(30). Raschke TM, 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. [PubMed: 11353861]

(31). Gallicchio E, Kubo MM, Levy RM. 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.

(32). Tan C, Tan YH, Luo R. Implicit Nonpolar Solvent Models. J. Phys. Chem. B. 2007; 111:12263–12274. [PubMed: 17918880]

(33). Harris RC, Pettitt BM. Effects of Geometry and Chemistry on Hydrophobic Solvation. Proc. Natl. Acad. Sci. U. S. A. 2014; 111:14681–14686. [PubMed: 25258413]

(34). Czaplewski C, Ripoll DR, Liwo A, Rodziewicz-Motowidlo S, Wawak J, Scheraga HA. Can Cooperativity in Hydrophobic Association be Reproduced Correctly by Implicit Solvation Models? Int. J. Quantum Chem. 2002; 88:41–45.

(35). Shimizu S, Chan HS. 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. [PubMed: 12012334]

(36). Jamadagni SN, Godawat S, Garde S. Hydrophobicity of Proteins and Interfaces: Insights from Density Fluctuations. Annu. Rev. Chem. Biomol. Eng. 2011; 2:147–171. [PubMed: 22432614]

(37). Patel AJ, Varilly P, Jamadagni SN, 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. [PubMed: 21987795]

(38). 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. [PubMed: 24425498]

(39). 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.

(40). Fennell CJ, Kehoe C, Dill KA. Oil/Water Transfer Is Partly Driven by Molecular Shape, Not Just Size. J. Am. Chem. Soc. 2010; 132:234–240. [PubMed: 19961159]

(41). Fennell CJ, Kehoe CW, Dill KA. Modeling Aqueous Solvation with Semi-Explicit Assembly. Proc. Natl. Acad. Sci. U. S. A. 2011; 108:3234–3239. [PubMed: 21300905]

(42). Li L, Fennell CJ, Dill KA. 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. [PubMed: 24299013]

(43). Mezei M, Beveridge DL. Structural Chemistry of Biomolecular Hydration via Computer Simulation: The Proximity Criterion. Methods Enzymol. 1986; 127:21–47. [PubMed: 3755494]

(44). Subramanian PS, Ravishanker G, Beveridge DL. 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. [PubMed: 3162310]

(45). Lounnas V, Pettitt BM, Phillips G Jr. A Global Model of the Protein-Solvent Interface. Biophys. J. 1994; 66:601–614. [PubMed: 8011893]

(46). Phillips GN Jr. Pettitt BM. Structure and Dynamics of the Water around Myoglobin. Protein Sci. 1995; 4:149–158. [PubMed: 7757005]

(47). Burling FT, Weis WI, Flaherty KM, Brünger AT. Direct Observation of Protein Solvation and Discrete Disorder with Experimental Crystallographic Phases. Science. 1996; 271:72–77. [PubMed: 8539602]

(48). Lin B, Pettitt BM. Note: On the Universality of Proximal Radial Distribution Functions of Proteins. J. Chem. Phys. 2011; 134:106101. [PubMed: 21405193]

(49). Rudnicki WR, Pettitt BM. Modeling the DNA-Solvent Interface. Biopolymers. 1997; 41:107–119. [PubMed: 8986123]

(50). Makarov VA, Andrews BK, Pettitt BM. Reconstructing the Protein-Water Interface. Biopolymers. 1998; 45:469–478. [PubMed: 9577228]

(51). Nguyen BL, Pettitt BM. Effects of Acids, Bases, and Heteroatoms on Proximal Radial Distribution Functions for Proteins. J. Chem. Theory Comput. 2015; 11:1399–1409. [PubMed: 26388706]

(52). Lin B, Wong KY, Hu C, Kokubo H, Pettitt BM. Fast Calculations of Electrostatic Solvation Free Energy from Reconstructed Solvent Density using Proximal Radial Distribution Functions. J. Phys. Chem. Lett. 2011; 2:1626–1632. [PubMed: 21765968]

(53). Lin B, Pettitt BM. 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. [PubMed: 20941733]

(54). Virtanen JJ, Makowski L, Sosnick TR, Freed KF. Modeling the Hydration Layer around Proteins: Applications to Small-and Wide-Angle X-Ray Scattering. Biophys. J. 2011; 101:2061–2069. [PubMed: 22004761]

(55). Ou SC, Pettitt BM. Solute-Solvent Energetics Based on Proximal Distribution Functions. J. Phys. Chem. B. 2016; 120:8230–8237. [PubMed: 27095487]

(56). Ben-Naim, A. Water and Aqueous Solutions. Plenum Press; New York, NY: 1992.

(57). Mehrotra PK, Beveridge DL. 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.

(58). 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. [PubMed: 7540695]

(59). Beutler TC, Mark AE, van Schaik RC, Gerber PR, van Gunsteren WF. Avoiding Singularities and Numerical Instabilities in Free Energy Calculations Based on Molecular Simulations. Chem. Phys. Lett. 1994; 222:529–539.

(60). Zacharias M, Straatsma TP, McCammon JA. Separation-Shifted Scaling, A New Scaling Method for Lennard-Jones Interactions in Thermodynamic Integration. J. Chem. Phys. 1994; 100:9025.

(61). Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable Molecular Dynamics with NAMD. J. Comput. Chem. 2005; 26:1781–1802. [PubMed: 16222654]

(62). 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.

(63). Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, Mackerell AD, 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.

(64). Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of Simple Potential Functions for Simulating Liquid Water. J. Chem. Phys. 1983; 79:926–935.

(65). Drake JA, Harris RC, Pettitt BM. Solvation Thermodynamics of Oligoglycine with Respect to Chain Length and Flexibility. Biophys. J. 2016; 111:756–767. [PubMed: 27558719]

(66). Ryckaert JP, Ciccotti G, Berendsen HJC. 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.

(67). 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.

(68). Grossfield A, Zuckerman DM. Quantifying Uncertainty and Sampling Quality in Biomolecular Simulations. Annu. Rep. Comput. Chem. 2009; 5:23–48. [PubMed: 20454547]

(69). Drake JA, Pettitt BM. Force Field-Dependent Solution Properties of Glycine Oligomers. J. Comput. Chem. 2015; 36:1275–1285. [PubMed: 25952623]

(70). Morreale A, de la Cruz X, Meyer T, Gelpí JL, Luque FJ, Orozco M. Linear Response Theory: An Alternative to PB and GB Methods for the Analysis of Molecular Dynamics Trajectories? Proteins: Struct., Funct., Genet. 2004; 57:458–467. [PubMed: 15382247]

(71). Carlson HA, Jorgensen WL. An Extended Linear Response Method for Determining Free Energies of Hydration. J. Phys. Chem. 1995; 99:10667–10673.

(72). Carlson HA, Nguyen TB, Orozco M, Jorgensen WL. Accuracy of Free Energies of Hydration for Organic Molecules from 6-31g*-Derived Partial Charges. J. Comput. Chem. 1993; 14:1240–1249.

(73). Pomès R, Willson RC, McCammon JA. Free Energy Simulations of the HyHEL-10/HEL Antibody-Antigen Complex. Protein Eng., Des. Sel. 1995; 8:663–675.

(74). Li G, Cui Q. pKa Calculations with QM/MM Free Energy Perturbations. J. Phys. Chem. B. 2003; 107:14521–14528.

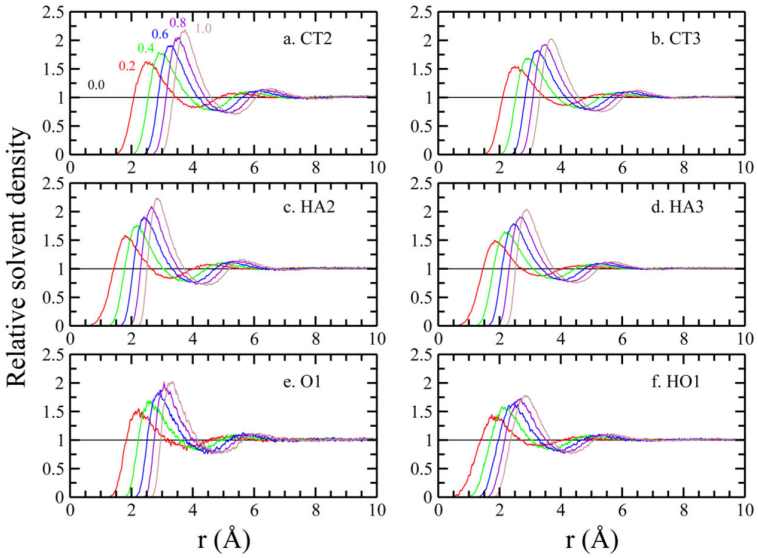


Figure 1.

pDF of the butane/propanol atoms to water oxygen atoms at different λ. Numbers above each pDF indicate the corresponding λ of this atom type.

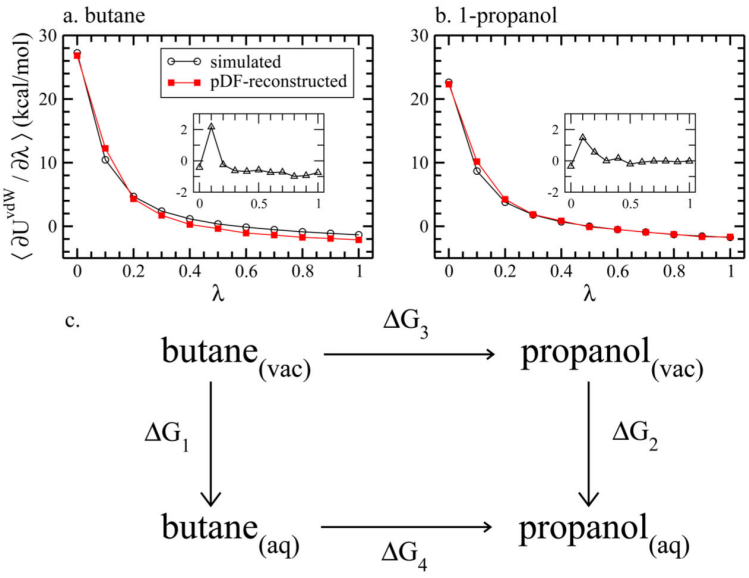


Figure 2.

Simulated and pDF-reconstructed $\langle \partial U^{vdW}/\partial \lambda \rangle$ at each λ window for (a) butane (b) propanol. The inset of each panel shows the difference between the simulated and pDF-reconstructed $\langle \partial U^{vdW}/\partial \lambda \rangle$. Panel (c) illustrates the thermodynamic cycle of mutating a methyl group of a butane into hydroxyl group to form a propanol.

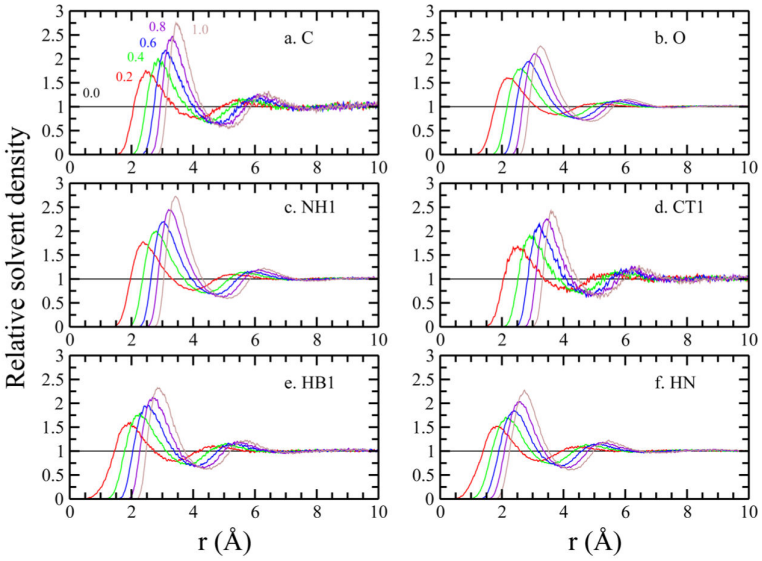


Figure 3.

pDF of alanine/glycine atoms to water oxygen atoms at different λ. Numbers above each pDF indicate the corresponding λ.

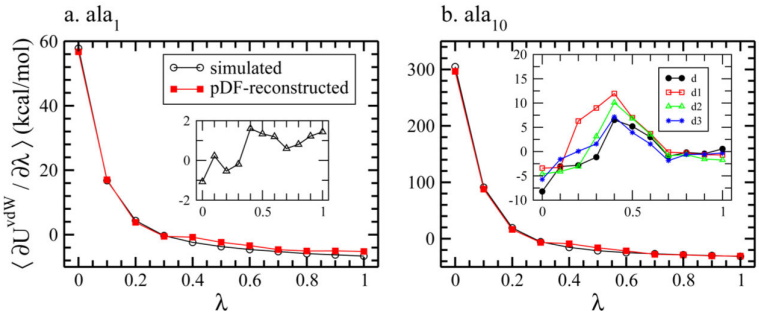


Figure 4.

Simulated and pDF-reconstructed $\langle \partial U^{vdW}/\partial \lambda \rangle$ at each λ window for (a) ala_1 (b) ala_{10} (the configuration referred to as d in Supporting Information). The inset of each panel shows the difference between the simulated and pDF-reconstructed $\langle \partial U^{vdW}/\partial \lambda \rangle$.

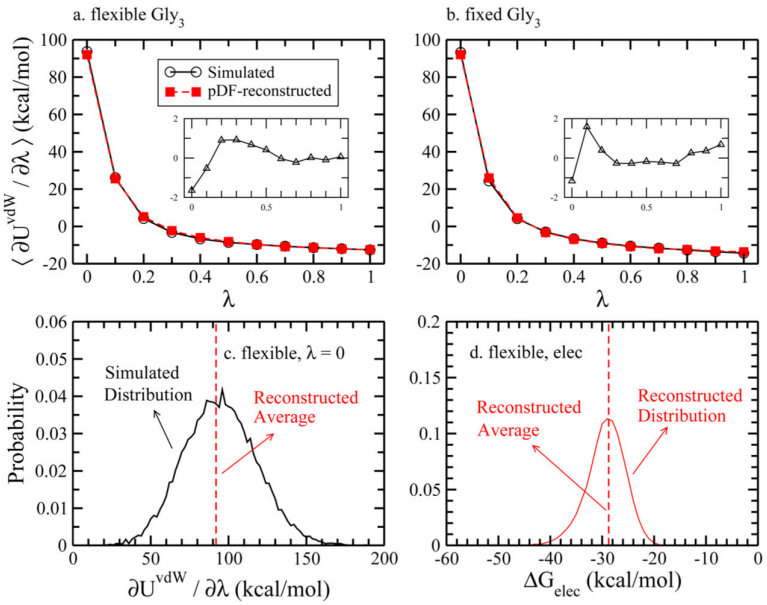


Figure 5.

Simulated and pDF-reconstructed $\langle \partial U^{vdW}/\partial \lambda \rangle$ at each λ window for (a) flexible gly_3 and (b) fixed gly_3. The inset of each panel shows the difference between the simulated and pDF-reconstruced $\langle \partial U^{vdW}/\partial \lambda \rangle$. Panel (c) shows the distribution of $\langle \partial U^{vdW}/\partial \lambda \rangle$ for flexible gly_3 at $\lambda = 0$. The average of pDF-reconstructed $\langle \partial U^{vdW}/\partial \lambda \rangle$ corresponds closely with the average of the simulated $\langle \partial U^{vdW}/\partial \lambda \rangle$. Panel (d) shows the distribution of pDF-reconstructed ΔG_{elec} using the solute configurations from flexible gly_3 simulations, with the red dashed line representing the ensemble average.

Table 1

Comparison of Solvation Free Energies of Butane and Propanol Using Thermodynamic Integration with Simulations and pDF-Reconstructions[a]

Δ *G*	**simulated T.I.**	**pDF T.I.**
	butane	
ΔG_{vdW}	2.67 (0.14)	2.34
ΔG_{elec}	−0.13 (0.01)	−0.14
ΔG_{total}	2.54 (0.14)	2.20
	propanol	
ΔG_{vdW}	1.88 (0.28)	2.06
ΔG_{elec}	−6.84 (0.03)	−6.42
ΔG_{total}	−4.96 (0.28)	−4.36

[a] ΔG_{vdW} is obtained using eq 9. ΔG_{elec} is approximated as half of electrostatic solute–solvent interaction energy using LRT and is adapted from ref 55. Uncertainties obtained as the block standard errors[68] are denoted in parentheses. Uncertainties for ΔG_{vdW} and ΔG_{total} are calculated using error propagation. The units are all in kcal/mol.

Table 2

Comparison of Solvation Free Energies of Single Alanine and Different Conformers of Deca-alanines Using Thermodynamic Integration with Simulations and pDF-Reconstructions[a]

Δ *G*	simulated T.I.	pDF T.I.
	ala_1	
ΔG_{vdW}	1.08 (0.36)	1.78
ΔG_{elec}	−20.21 (0.02)	−20.72
ΔG_{total}	−19.13 (0.36)	−18.94
	d	
ΔG_{vdW}	5.75 (1.77)	5.92
ΔG_{elec}	−78.19 (0.08)	−74.26
ΔG_{total}	−72.44 (1.77)	−68.34
	d1	
ΔG_{vdW}	5.83 (2.20)	8.70
ΔG_{elec}	−67.75 (0.08)	−60.70
ΔG_{total}	−61.92 (2.20)	−52.00
	d2	
ΔG_{vdW}	5.42 (1.81)	6.27
ΔG_{elec}	−89.97 (0.10)	−89.00
ΔG_{total}	−84.55 (1.81)	−82.73
	d3	
ΔG_{vdW}	5.72 (1.69)	5.67
ΔG_{elec}	−86.88 (0.08)	−85.86
ΔG_{total}	−81.16 (1.69)	−80.19

[a] ΔG_{vdW} is obtained using eq 9. ΔG_{elec} is approximated as half of electrostatic solute–solvent interaction energy using LRT and is adapted from ref 55. The units are all in kcal/mol.

Table 3

Comparison of Solvation Free Energies of Glycine and Tri-glycine Using Thermodynamic Integration with Simulations, pDF-Reconstructions, and Benchmark Free Energy Perturbation Calculations[a]

Δ *G*	simulated TI	pDF TI	FEP[b]
	gly_1		
ΔG_{vdW}	0.39 (0.34)	0.45	
ΔG_{elec}	−16.31 (0.12)	−16.25	
ΔG_{total}	−15.92 (0.36)	−15.80	
	gly_3, flexible		
ΔG^{flex}_{vdW}	−0.41 (0.32)	−0.33	−0.03 (0.05)
ΔG^{flex}_{elec}		−28.75	−24.01 (0.50)
ΔG^{flex}_{total}		−29.08	−24.04 (0.50)
	gly_3, fixed		
ΔG^{fix}_{vdW}	−1.29 (0.38)	−1.10	−0.84 (0.04)
ΔG^{fix}_{elec}	−25.62 (0.12)	−24.13	−22.08 (0.03)
ΔG^{fix}_{total}	−26.91 (0.40)	−25.23	−22.16 (0.05)

[a] ΔG_{vdW} is obtained using eq 9. For fixed configurations, ΔG_{elec} are approximated as half of electrostatic solute–solvent interaction energies using LRT. An average is taken over configurations for flexible solutes. The units are all in kcal/mol.

[b] Free energy perturbation results are adapted from ref 65.