Published in final edited form as:

J Phys Chem B. 2016 August 25; 120(33): 8230–8237. doi:10.1021/acs.jpcb.6b01898.

Solute–Solvent Energetics Based on Proximal Distribution Functions

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

Abstract

We consider the hydration structure and thermodynamic energetics of solutes in aqueous solution. On the basis of the dominant local correlation between the solvent and the chemical nature of the solute atoms, proximal distribution functions (pDF) can be used to quantitatively estimate the hydration pattern of the macromolecules. We extended this technique to study the solute–solvent energetics including the van der Waals terms representing excluded volume and tested the method with butane and propanol. Our results indicate that the pDF-reconstruction algorithm can reproduce van der Waals solute–solvent interaction energies to useful kilocalorie per mole accuracy. We subsequently computed polyaniline–water interaction energies for a variety of conformers, which also showed agreement with the simulated values.

Graphical Abstract

*Corresponding Author: mpettitt@utmb.edu. Tel: (409)772-0723.

The authors declare no competing financial interest.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.6b01898.

Calculated pDFs of water hydrogen atoms; atomic sequence of ala₁₀ and energetic difference between simulation and pDF reconstruction; relationship between simulated and pDF reconstructed interaction energies and atomic solvent accessible surface area; error comparison between different grid resolutions; error reduction with interpolation during pDF reconstruction (PDF)

INTRODUCTION

Calculation of thermodynamic properties such as the average internal energy and enthalpy for molecules in solution is often straightforward but tedious. Simulation of a force field^{1,2} using molecular mechanics allows calculation of the correlation between molecules and solvent, making it a primary tool for such purposes.^{3,4} In general, there are basically two sources of approximation to be considered for this route: a functional form to represent the interactions and the sampling of the distribution of solvent.

The mathematical form of the nonbonded potential is typically expressed with the Coulomb electrostatic and van der Waals terms.^{5–9} However, the calculation of nonbonded potential is usually the most expensive process on account of the number of water molecules in the solvent environment and consequent pairs of interparticle interactions. Realistically, simulating $\sim 10^6$ atoms is roughly the limit using modern hardware. Hence, alternative approaches have been developed to reduce the demanding calculations in aqueous systems. These approaches can essentially be classified as two broad categories: (a) to model the solvent as an implicit continuum dielectric^{10–13} such as Poisson–Boltzmann and generalized Born models and (b) to create a cell of a computationally affordable amount of explicit solvent to approximate an infinite system. By modeling the solvent implicitly, the electrostatic solvation of biomolecules has been shown to be sufficiently accurate for some purposes, which contrasts with the situation for the nonpolar component because of the length scale and geometry dependence in hydrophobic solvation.^{14–18} On the other hand, the explicit treatment of solvent molecules offers the full details of intermolecular interactions, but in order to extract meaningful thermodynamic properties from biomolecular simulations, the configuration space must be explored, which increases the computational cost.

To quantify the hydration features, it is common to use probability-based pair correlation functions.^{19,20} However, due to the inherent chemical and topographical heterogeneity of macromolecules,^{21–26} some modifications are convenient for a quantitative description of the hydration structures.^{27–30} These methods select sets of solute surface atoms and calculate the corresponding perpendicular/proximal distribution functions (pDF, also referred to as pRDF elsewhere) from simulations^{28,31} or experiments³⁰ to describe the locally unique hydration patterns.

The pDFs are the nearest neighbor contribution and thus the dominant leading term of the total distribution function.^{27,32} More recently, pDFs have been shown to be a nearly universal descriptor based on the dominant local interaction between the solvent and the chemical nature of the solute atoms³³ and, therefore, transferable between different solute molecules. This universality allows us to predict the hydration of complex biological macromolecules from the pDFs of similar molecules or even small solute molecules with chemical similarity.³⁴ In view of the quality of the solution structure, it seems reasonable to consider the energetics between solute and solvent molecules through pDF predictions to avoid lengthy simulations. Progress regarding the intermolecular electrostatic potential energies and solvation free energetics has been reported in previous work.^{35,36} Our goal here is to evaluate the nonbonded van der Waals solute–solvent average energy using the pDF reconstruction technique.

This study is organized as follows. We begin with the calculation of pDF and the reconstruction of solvent density distribution, along with the simulation details. Our results are organized into two parts. We start by examining two simple molecules, butane and propanol, comparing the simulated and pDF-reconstructed solute–solvent van der Waals interaction energies. Next, we use the pDF reconstruction algorithm to estimate polypeptide–water interaction energies for some alanine peptides. We conclude our study with a summary of our findings and discussion for future work.

METHODS

Distribution and Potential Energy Functions

The proximal/perpendicular distributions (pDF), $g_{\perp}(r)$, are a conditional case of the quasi-component distributions^{27,37,38} that describes the solvent structure at a given distance r perpendicular to the solute surface. The set of $g_{\perp}(r)$ for all atoms of the solute defines a nearest neighbor approximation to total three-dimensional distribution of solvent around a solute, $g(x,y,z)$. This gives rise to the approximation that the effective solute–solvent interactions are largely local,³⁹ and consequently, the physical correlations can be represented by the nearest neighbor contributions.³³ In this manner, it is feasible to estimate the surrounding solvent environment of a given solute structure based on chemical similarities.^{39,40}

As in the literature cited above, we write $g_{\perp}(r)$ for the k_{th} solute atom as

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

where $\vec{r}_{ij}(t)$ represents the position vector from i_{th} solute atom to j_{th} solvent atom at time t . Here, m is the number of atoms in the solute molecule, with n the total amount of solvent molecules in the system, $\text{Inf}[|\vec{r}_{ij}(t)|]_{i=1,m}$ takes the minimum distance combination between solute and solvent atoms, and k is the assigned solute atom with the closest distance with j_{th} solvent atom. The solvent space is thereby partitioned into Voronoi polyhedra.⁴¹ $\delta\tau(\vec{r}_j(t), k)$ is the volume element around the j_{th} solvent molecule, which is defined by all $\vec{r}$ vectors, where $|\vec{r}_k - \vec{r}| < |\vec{r}_i - \vec{r}|$. In practice, a pDF can be computed on a three-dimensional grid. In this framework, each individual grid point along with its corresponding time averaged solvent density is assigned to the closest solute atom for $g_{\perp}(r)$ calculations. The detailed theory and computational procedure can be found in the literature.^{27,28}

Given a set of pDFs and the assumption of transferability, we can construct the solvation density around other solutes with a similar set of atom types. To reconstruct the solvent density distribution for a given solute configuration, we use a three-dimensional grid and assign solvent density to the grid point (u, v, w) with the precomputed $g_{\perp}^{Xk}(r')$

$$g(u, v, w) = g_{\perp}^{\chi_k}(r') \quad (2)$$

where $r' = |\vec{r}_i - (u, v, w)|$ is the minimum distance between the grid point and all solute atoms i , χ_k is the atom type of the closest solute atom, and u , v , and w denote the center of the grid volume along the x , y , and z dimensions, respectively. Consequently, the solvent density at grid point (u, v, w) is

$$\rho(u, v, w) = \rho g_{\perp}^{\chi_k}(r') \quad (3)$$

where ρ is the bulk solvent number density.

During the reconstruction process, we need to consider the solute/solvent molecule as a whole due to differing geometries and atomic radii. So, while a grid point is assigned to a solute atom χ_1 , this grid point could still be within the exclusion distance of another solute atom χ_2 , such that the solvent–solute interaction energy at this grid point with atom χ_2 becomes unrealistically repulsive. This happens frequently when the grid point is near or between two atoms of disparate size. To prevent these situations, we modify the procedure in eq 3 by applying an exclusion rule during the reconstructing process:

$$\rho(u, v, w) = \rho g_{\perp}^{\chi_k}(r') \left\{ \prod_{i \neq \chi_k}^m \delta_{\perp}^i(|(u, v, w) - \vec{r}_i|) \right\} \quad (4)$$

Again, m is the number of solute atoms. The exclusion factor $\delta_{\perp}^i$ is defined as

$$\delta_{\perp}^i = \begin{cases} 1, & \text{if } g_{\perp}^i(|(u, v, w) - \vec{r}_i|) > 0 \\ 0, & \text{if } g_{\perp}^i(|(u, v, w) - \vec{r}_i|) = 0 \end{cases} \quad (5)$$

Any solute atom may exclude solvent from any grid point. Once we determine the solvent density distribution, the average solute–solvent interaction energy can then be written as

$$\begin{aligned} \langle U_{\text{solute-solv}} \rangle &= \rho \int g(x, y, z) U(x, y, z) \, dx \, dy \, dz \\ &\approx \sum_{p=1}^s \sum_{i=1}^m \Delta v \rho(u, v, w) U_{ij}(|(u, v, w) - \vec{r}_i|) \end{aligned} \quad (6)$$

where $U(x, y, z)$ is the total solute–solvent interaction energy at x, y, z , v is the unit volume which depends on the spatial grid resolution, and s is the total number of grid points.

Next, we address two parameters used in conducting pDF calculations and solvent density reconstructions. The first is the grid spacing (Δs) on the three-dimensional target grid around the solute, and the second is the radial resolution (Δr) for collecting pDF's. While the reconstructed solvent densities are relatively insensitive to the choice of Δs for a given required reconstruction resolution (refer to the Supporting Information), the effect of different Δr can be significant. Unless otherwise specified, the Δr and Δs in the present work are set to 0.01 and 0.2 Å, respectively. We note that the resolution of 0.01 Å is somewhat fine for most aqueous distribution functions. More detail will be discussed in this context below.

In this study, we focus on the van der Waals contribution to the average interaction energy in solution. In the underlying molecular mechanics we use the Lennard-Jones 6–12 potential

$$U_{ij}^{\text{vdW}}(r) = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (7)$$

where ϵ is the well depth and σ is the contact distance.

Simulation Details

The solutes (butane, propanol, ala₁, ala₃, and ala₁₀) were chosen for this study. Molecular dynamics simulations were performed with NAMD 2.9^{42,43} and the CHARMM 36 force field parameters.⁴⁴ Each solute was solvated with TIP3P water⁴⁵ with at least 10 Å from each boundary to the solute atoms. The peptide terminals were capped with neutral acetate (ACE) and methyl amide (NME) residues (Figure 1). The initial structures of deca-alanines were from our previous studies.^{18,35} Three-dimensional periodic boundary conditions were applied. A rigid water geometry is enforced using SHAKE.⁴⁶ Particle Mesh Ewald (PME)⁴⁷ was used to treat electrostatic interactions using a grid of 1.0 Å resolution. Lennard-Jones (L-J) interactions were gradually switched off over the range 10 Å to 11 Å. Temperature was fixed at 300 K via a Langevin method with damping coefficient of 5 ps⁻¹. A time step of 1.0 fs was used to integrate the equations of motion.

Throughout the simulations, the solute molecules were fixed. The simulations were initially carried out in the NVT ensemble for 5 ns. The final configurations were then used as the starting structure for NPT simulations at 1 atm pressure, with the first 5 ns as equilibration. Sampling time for collecting the density distributions and the calculation of simulated intermolecular energetics for each system was between 20 and 50 ns. Trajectory snapshots were saved every 1 ps for analyses.

RESULTS AND DISCUSSION

Butane and Propanol

We begin by comparing the solute–solvent interaction energy of a single butane molecule between the direct value from simulation and the reconstructed value using the same solute

configuration. For any given grid point around the solute, the energy function and distances to the solute atoms are determined. Thus, the accuracy of estimated average energies depends on the reconstructed solvent probability densities only. In the current study, owing to the strong repulsion in the Lennard-Jones potential, even small deviations in densities in the near vicinity of solute atoms may cause dramatic discrepancies with respect to the energy contribution. In the previous work, we used the heavy atoms and polar hydrogen atoms as the pDF atom set.^{28,33,35,39} Essentially, the richer the set of atom types we include, the higher the accuracy in solvent densities we can expect during the reconstructions, especially when we consider the error within the solute molecule's first shell.³⁴

We consider the contribution of every single atom type of the solute molecules in our pDF set. As for butane, it consists four atom types: methylene carbon (CT2), methyl carbon (CT3), methylene hydrogen (HA2) and methyl hydrogen (HA3). The pDFs for water oxygen atoms are presented in Figure 2a. As expected, the characteristics of the same atom types are highly similar. Figure 2b shows the comparisons of individual solute atom–water oxygen interaction energies between simulated and reconstructed. CT3 shows the largest energetic difference (~ 0.2 kcal/mol). The pDFs of water hydrogen atoms (HT) are included in the Supporting Information. The butane–water interaction energies are listed in Table 1. The deviation of butane–OT is roughly 0.2 kcal/mol, while the deviation of butane–HT is even less than 0.02 kcal/mol. The electrostatic contribution is only off by 0.02 kcal/mol. Overall, these quantities exhibit good consistencies.

We next increased the complexity by replacing one methyl group with a hydroxyl group to form propanol. The pDF set is therefore {CT2, CT3, HA2, HA3, O1, HO1}, with O1 and HO1 representing the oxygen and hydrogen atoms of the hydroxyl group. The corresponding pDFs of water oxygen are shown in Figure 2c. We observed distinct characteristics between the hydroxyl hydrogen and hydrocarbon hydrogens. Due to the polarity of the O1–HO1 bond, on one hand, the size of the hydroxyl hydrogen is much smaller than the hydrocarbon hydrogens; on the other hand, the hydroxyl hydrogen carries a large positive charge. As a result, the water oxygen atoms (OT) are capable of approaching the hydroxyl hydrogen at a shorter distance. This also leads to a strong peak in the O1–OT pDF and a positive interaction energy accordingly. The discrepancies of simulated and reconstructed interaction energies were all within 0.2 kcal/mol, as illustrated in Figure 2d. The total propanol–water interaction energies are listed in Table 1. Compared with the case of butane, we observed a slightly larger deviation from the propanol–OT. However, the components are in good accord with simulation while the largest difference comes from the electrostatic contribution.

Estimation: Trialanine and Deca-alanine

We now consider estimating the energetics between larger solutes and solvent by our method of precomputed pDFs. We chose to compute the pDFs from a single alanine and assess the poly alanines. In our previous work, we adopted pDFs from decaalanine to capture the electrostatic features surrounding the deca-alanine conformers.³⁵ The pDF set of the alanine is {CT3, C, O, NH1, CT1, HA3, HN, HB1}, while 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. The calculated pDFs of water oxygen atoms are illustrated in Figure 3a,b.

The simulated and the reconstructed ala₃–water interaction energies are itemized in Table 2. The difference in vdW interactions ($\Delta U_{\text{solu-wat}}^{\text{vdW}}$) between simulated and reconstructed ala₃–OT and ala₃–HT increase to 0.96 and 0.22 kcal/mol, respectively. The difference in the electrostatic component is 1.04 kcal/mol, which is negligible considering $U_{\text{solu-wat}}^{\text{elec}}$ is roughly three times that of $U_{\text{solu-wat}}^{\text{vdW}}$. For convenience, we further subtracted the difference between each simulated and reconstructed solute atom–solvent energies in Figure 4. The difference between solute–OT is less than 0.2 kcal/mol except for the case of carbonyl oxygen atoms, while the reconstructed $U_{\text{O-OT}}^{\text{vdW}}$ is 0.5 kcal/mol more negative than simulated $U_{\text{O-OT}}^{\text{vdW}}$. All of the pDF reconstructed solute–HT interaction energies are slightly more negative than the simulated (note the difference of solute–HT is scaled by a factor of 20 in the figure), while the heavy C, O, and NH1 atom types show relatively larger but less significant deviations.

We examined ala₁₀–water interaction energies using four different deca-alanine conformers from previous work, referred as d, d1, d2, and d3.³⁵ The results of the simulations and the values from the pDF reconstructions are listed in Table 2. For $U_{\text{solu-OT}}^{\text{vdW}}$, the differences between simulations and reconstructions are 0.79, 1.09, 1.42, and 0.34 kcal/mol, respectively. Deviations for $U_{\text{solu-HT}}^{\text{vdW}}$ also increased to roughly 0.7 kcal/mol. The electrostatic energetics are off by 2 to 5 kcal/mol except for d1, which was also observed in our previous study³⁵ because of its extremely twisted configuration. By dissecting the total solute–OT interaction energies into individual atomic components for comparison of simulated and pDF reconstructed van der Waals Energy in the same manner as done above for ala₃ (included in the Supporting Information), similar behaviors are found. The majority of deviations arise from the carbonyl oxygen atom contributions. This trend can be quantified by breaking down the error components for each atom type on deca-alanine. The error (ε_{χ}) of atom type χ is defined as

$$\varepsilon_{\chi} = \sqrt{\sum (\Delta U_{i,\chi})^2} \quad (8)$$

The results for all atom types of deca-alanine are shown in Figure 5. The error for carbonyl oxygen is larger than that of any other atom types.

We might speculate that reconstructed quality may relate to the solvent accessible regions for solute atoms.³⁵ However, we observe little if any correlation between $\Delta U_{\text{O-OT}}^{\text{vdW}}$ of carbonyl oxygen atoms and corresponding solvent accessible surface areas (refer to the Supporting Information). As discussed in the previous section, the energetic estimations essentially depend on the accuracy of the reconstructed density, particularly in the region near each solute atom. Although the relative error within the neighboring hydration shell of the solutes may not be dominant in global solvent densities reconstructions,³⁴ the steep

repulsion in proximity to the solute atoms can numerically amplify these small errors. We demonstrate this by plotting the accumulated $U_{\text{O-O-T}}^{\text{vdW}}$ as a function of distance from the carbonyl oxygen (as shown in Figure 6a). The energetic deviations originated from the first few tenths of an angstrom contact of the solvent with the solute atoms since the divergence does not expand thereafter. We find that for $U_{\text{solute-HT}}^{\text{vdW}}$ deviations in O-HT are small (as shown in Figure 6b) since there is not a steep repulsive contact the difference is trivial.

There are essentially two numerical factors that are related to our reconstructed density quality. These are the resolution for collecting our pDF's (Δr) and the grid resolution (Δs) for solvent reconstructions. While further reducing Δs at a given Δr does not show any interesting distinctions, the effect from Δr is noticeable, as we examined the errors of different atom types as a function of Δr in Figure 7a. For carbonyl oxygen, the error drops significantly when we change the Δr resolution from 0.1 Å to 0.05 and 0.02 Å, while ϵ for all the other atom types hardly depends on Δr . If we adopt $\epsilon = 1$ as the acceptable value in our reconstruction algorithm, a finer resolution of $\Delta r = 0.01$ or 0.02 Å is probably necessary for the carbonyl oxygen. This requires careful sampling for the pDF calculations at short range. We include the accumulated $U_{\text{O-O-r}}^{\text{vdW}}$ using $\Delta r = 0.05$ Å as a comparison with the $\Delta r = 0.01$ Å case in order to elucidate how Δr affects our reconstructed density quality especially at small distances from the carbonyl oxygen atom. We can also perform interpolation with coarser Δr , as displayed in Figure 7b. Even simple linear interpolation during reconstruction significantly reduces the error of the carbonyl oxygen atoms.

SUMMARY

In this work, we have extended the application of using reconstructed densities to estimate the van der Waals solute-solvent energetics. Previous work had only considered the electrostatic component.³⁵ Based on the precomputed solute-solvent pDFs from small molecules, we are capable of reconstructing the solvent densities and estimating the interaction energies of larger solute with solvent using the same proximal criteria, without performing lengthy simulations on the large systems. In the current examples of poly alanines, the energetic reconstructions of carbonyl oxygen-solvent components show the larger deviations with simulations. The total errors in the van der Waals average potential energies calculated for the grids in this paper are within a couple of kilocalories per mole and often better. Combined with functional theories, pDF reconstruction can serve as a rapid and economical tool to estimate the thermodynamics of a variety of molecules.^{35,36} This warrants further ongoing investigations on how free energy techniques can be applied to predict the nonbonded van der Waals contributions.

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. MacKerell AD Jr. Empirical Force Fields for Biological Macromolecules: Overview and Issues. *J Comput Chem.* 2004; 25:1584–1604. [PubMed: 15264253]
2. Huang J, Lopes PEM, Roux B, MacKerell AD Jr. Recent Advances in Polarizable Force Fields for Macromolecules: Micro-second Simulations of Proteins Using the Classical Drude Oscillator Model. *J Phys Chem Lett.* 2014; 5:3144–3150. [PubMed: 25247054]
3. Karplus M, McCammon JA. Molecular Dynamics Simulations of Biomolecules. *Nat Struct Biol.* 2002; 9:646–652. [PubMed: 12198485]
4. Shea JE, Brooks CL III. From Folding Theories to Folding Proteins: A Review and Assessment of Simulation Studies of Protein Folding and Unfolding. *Annu Rev Phys Chem.* 2001; 52:499–535. [PubMed: 11326073]
5. Brooks BR, Brooks CL III, MacKerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, et al. CHARMM: The Biomolecular Simulation Program. *J Comput Chem.* 2009; 30:1545–1614. [PubMed: 19444816]
6. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Caldwell J, et al. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J Comput Chem.* 2003; 24:1999–2012. [PubMed: 14531054]
7. van Duin ACT, Dasgupta S, Lorant F, Goddard WA III. ReaxFF: A Reactive Force Field for Hydrocarbons. *J Phys Chem A.* 2001; 105:9396–9409.
8. Jorgensen WL, Tirado-Rives J. The OPLS [Optimized Potentials for Liquid Simulations] Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J Am Chem Soc.* 1988; 110:1657–1666. [PubMed: 27557051]
9. Ponder JW, Case DA. Force Fields for Protein Simulations. *Adv Protein Chem.* 2003; 66:27–85. [PubMed: 14631816]
10. Tomasi J, Persico M. Molecular Interactions in Solution: An Overview of Methods Based on Continuous Distributions of the Solvent. *Chem Rev.* 1994; 94:2027–2094.
11. Roux B, Simonson T. Implicit Solvent Models. *Biophys Chem.* 1999; 78:1–20. [PubMed: 17030302]
12. Feig M, Brooks CL III. Recent Advances in the Development and Application of Implicit Solvent Models in Biomolecule Simulations. *Curr Opin Struct Biol.* 2004; 14:217–224. [PubMed: 15093837]
13. Chen J, Brooks CL III, Khandogin J. Recent Advances in Implicit Solvent Based Methods for Biomolecular Simulations. *Curr Opin Struct Biol.* 2008; 18:140–148. [PubMed: 18304802]
14. 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]
15. Chandler D. Interfaces and the Driving Force of Hydrophobic Assembly. *Nature.* 2005; 437:640–647. [PubMed: 16193038]
16. Wagoner JA, Baker NA. Assessing Implicit Models for Nonpolar Mean Solvation Forces: the Importance of Dispersion and Volume Terms. *Proc Natl Acad Sci U S A.* 2006; 103:8331–8336. [PubMed: 16709675]
17. Chen J, Brooks CL III. Critical Importance of Length-Scale Dependence in Implicit Modeling of Hydrophobic Interactions. *J Am Chem Soc.* 2007; 129:2444–2445. [PubMed: 17288425]
18. 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]
19. Hansen, JP., McDonald, IR. *Theory of Simple Liquids.* Academic Press; San Diego, CA: 1976.
20. Chandler, D. *Introduction to Modern Statistical Mechanics.* Oxford University Press; New York: 1987.

21. Jamadagni SN, Godawat R, Garde S. Hydrophobicity of Proteins and Interfaces: Insights from Density Fluctuations. *Annu Rev Chem Biomol Eng.* 2011; 2:147–171. [PubMed: 22432614]
22. Israelachvili J, Wennerström H. Role of Hydration and Water Structure in Biological and Colloidal Interactions. *Nature.* 1996; 379:219–225. [PubMed: 8538786]
23. Svergun DI, Richard S, Koch M, Sayers Z, Kuprin S, Zaccai G. Protein Hydration in Solution: Experimental Observation by X-Ray and Neutron Scattering. *Proc Natl Acad Sci U S A.* 1998; 95:2267–2272. [PubMed: 9482874]
24. 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]
25. 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]
26. 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.
27. Mezei M, Beveridge DL. Structural Chemistry of Biomolecular Hydration via Computer Simulation: The Proximity Criterion. *Methods Enzymol.* 1986; 127:21–47. [PubMed: 3755494]
28. Lounnas V, Pettitt BM, Phillips G Jr. A Global Model of the Protein-Solvent Interface. *Biophys J.* 1994; 66:601–614. [PubMed: 8011893]
29. Phillips GN Jr, Pettitt BM. Structure and Dynamics of the Water around Myoglobin. *Protein Sci.* 1995; 4:149–158. [PubMed: 7757005]
30. 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]
31. 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]
32. Virtanen JJ, Makowski L, Sosnick TR, Freed KF. Modeling the Hydration Layer around Proteins: Applications to Small-Angle Wide-Angle X-Ray Scattering. *Biophys J.* 2011; 101:2061–2069. [PubMed: 22004761]
33. Lin B, Pettitt BM. Note: On the Universality of Proximal Radial Distribution Functions of Proteins. *J Chem Phys.* 2011; 134:106101. [PubMed: 21405193]
34. 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]
35. 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]
36. 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]
37. Ben-Naim, A. Water and Aqueous Solutions. Plenum Press; New York: 1992.
38. 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.
39. Makarov VA, Andrews BK, Pettitt BM. Reconstructing the Protein-Water Interface. *Biopolymers.* 1998; 45:469–478. [PubMed: 9577228]
40. Rudnicki WR, Pettitt BM. Modeling the DNA-Solvent Interface. *Biopolymers.* 1997; 41:107–119. [PubMed: 8986123]
41. 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]
42. 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]

43. 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.
44. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, et al. 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.* 2010; 31:671–690. [PubMed: 19575467]
45. 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.
46. 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.
47. 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.
48. Grossfield A, Zuckerman DM. Quantifying Uncertainty and Sampling Quality in Biomolecular Simulations. *Annu Rep Comput Chem.* 2009; 5:23–48. [PubMed: 20454547]
49. Humphrey W, Dalke A, Schulten K. VMD: Visual Molecular Dynamics. *J Mol Graphics.* 1996; 14:33–38.

Figure 1. Structures of alanine peptides used in this study: (a) trialanine (ala₃); (b–e) deca-alanine (ala₁₀). For ala₁₀, peptide backbones are represented with red ribbons. For convenience, we designate d, d1, d2, and d3 for the configuration of deca-alanines, respectively. Configurations are visualized using VMD.⁴⁹

Figure 2. pDF of solute atoms and reconstructed solute-solvent van der Waals interaction energy, compared with the values obtained via simulation: (a) pDF of butane atoms to water oxygen with (b) corresponding van der Waals interaction energies. (c) pDF of propanol atoms to water oxygen with (d) corresponding van der Waals interaction energies. CT2, CT3, HA2, HA3, O1, and HO1 are the atom types of butane and propanol. OT is the water oxygen atom label.

Figure 3.
Calculated pDFs for atoms from ala₁. CT3, C, O, NH1, CT1, HA3, HN, HB1 are the atom types of alanine. OT is the water oxygen atom type.

Figure 4. van der Waals solute solvent potential atomic component difference ΔU_1^{vdW} of ala₃ between simulation and the pDF reconstruction. Note the difference for the solute atom–water hydrogen atoms is scaled by a factor of 20 for clarity. The horizontal axis is the atom type i .

Figure 5.
Error (ϵ) for each atom type of deca-alanine for each conformer.

Figure 6.

(a) O–OT (b) O–HT van der Waals accumulated interaction energies as functions of distance from the carbonyl oxygen to water atoms.

Figure 7. Error of each atom type in deca-alanine (configuration d) using (a) different pDF collecting resolution and (b) linear interpolation.

Table 1

Comparison of Solute–Solvent Interaction Energies of Butane and Propanol Using Simulations and pDF Reconstruction^a

	simulated	reconstructed
butane		
$U_{\text{solu-O}^{\text{T}}}^{\text{vdW}}$	−8.75 (0.008)	−8.55
$U_{\text{solu-HT}}^{\text{vdW}}$	−0.71 (0.001)	−0.73
$U_{\text{solu-wat}}^{\text{elec}}$	−0.25 (0.004)	−0.27
$U_{\text{solu-wat}}$	−9.71 (0.009)	−9.55
propanol		
$U_{\text{solu-O}^{\text{T}}}^{\text{vdW}}$	−5.76 (0.011)	−5.42
$U_{\text{solu-HT}}^{\text{vdW}}$	−0.80 (0.001)	−0.80
$U_{\text{solu-wat}}^{\text{elec}}$	−13.68 (0.027)	−12.83
$U_{\text{solu-wat}}$	−20.24 (0.029)	−19.05

^aUncertainties obtained as the block standard errors⁴⁸ are denoted in parentheses.

Table 2

Comparison of Solute–solvent Interaction Energies of Tri-alanine and Deca-alanines Using Simulations and pDF Reconstruction^a

	simulated	reconstructed	simulated	reconstructed
ala ₃				
$U_{\text{solu}-\text{OT}}^{\text{vdW}}$	−20.78 (0.022)	−19.82		
$U_{\text{solu}-\text{HT}}^{\text{vdW}}$	−2.90 (0.002)	−3.12		
$U_{\text{solu}-\text{wat}}^{\text{elec}}$	−68.82 (0.048)	−67.78		
$U_{\text{solu}-\text{wat}}$	−92.50 (0.052)	−90.72		
ala ₁₀ (d)				
$U_{\text{solu}-\text{OT}}^{\text{vdW}}$	−48.56 (0.037)	−47.77	ala ₁₀ (d2) −48.48 (0.034)	−47.06
$U_{\text{solu}-\text{HT}}^{\text{vdW}}$	−6.89 (0.003)	−7.61	−6.56 (0.003)	−6.94
$U_{\text{solu}-\text{wat}}^{\text{elec}}$	−152.38 (0.076)	−148.51	−179.93 (0.101)	−177.99
$U_{\text{solu}-\text{wat}}$	−207.83 (0.085)	−203.89	−234.97 (0.107)	−231.99
ala ₁₀ (d1)				
$U_{\text{solu}-\text{OT}}^{\text{vdW}}$	−45.92 (0.037)	−44.83	ala ₁₀ (d3) −47.14 (0.035)	−46.80
$U_{\text{solu}-\text{HT}}^{\text{vdW}}$	−5.87 (0.004)	−6.26	−6.99 (0.003)	−7.62
$U_{\text{solu}-\text{wat}}^{\text{elec}}$	−135.50 (0.076)	−121.39	−173.75 (0.082)	−171.71
$U_{\text{solu}-\text{wat}}$	−187.29 (0.084)	−172.48	−227.88 (0.089)	−226.13

^aUncertainties obtained as the block standard errors⁴⁸ are denoted in parentheses.