THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Assessing the Accuracy of Inhomogeneous Fluid Solvation Theory in Predicting Hydration Free Energies of Simple Solutes

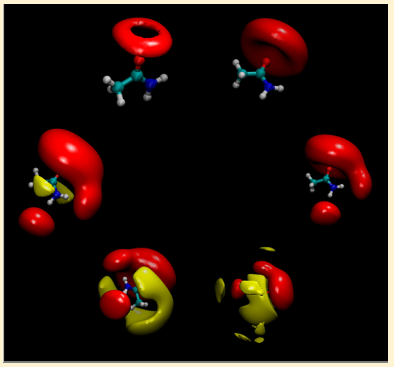
David J. Huggins*[,†,‡,§] and Mike C. Payne[†]

[†]Theory of Condensed Matter Group, Cavendish Laboratory, University of Cambridge, 19 J J Thomson Avenue, Cambridge CB3 0HE, United Kingdom

[‡]Cambridge Molecular Therapeutics Programme, Hutchison/MRC Research Centre, University of Cambridge, Hills Road, Cambridge, CB2 0XZ, United Kingdom

[§]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, UK CB2 1EW, United Kingdom

ABSTRACT: Accurate prediction of hydration free energies is a key objective of any free energy method that is applied to modeling and understanding interactions in the aqueous phase. Inhomogeneous fluid solvation theory (IFST) is a statistical mechanical method for calculating solvation free energies by quantifying the effect of a solute acting as a perturbation to bulk water. IFST has found wide application in understanding hydration phenomena in biological systems, but quantitative applications have not been comprehensively assessed. In this study, we report the hydration free energies of six simple solutes calculated using IFST and independently using free energy perturbation (FEP). This facilitates a validation of IFST that is independent of the accuracy of the force field. The results demonstrate that IFST shows good agreement with FEP, with an R^2 coefficient of determination of 0.99 and a mean unsigned difference of 0.7 kcal/mol. However, sampling is a major issue that plagues IFST calculations and the results suggest that a histogram method may require prohibitively long simulations to achieve convergence of the entropies, for bin sizes which effectively capture the underlying probability distributions. Results also highlight the sensitivity of IFST to the reference interaction energy of a water molecule in bulk, with a difference of 0.01 kcal/mol changing the predicted hydration free energies by approximately 2.4 kcal/mol for the systems studied here. One of the major advantages of IFST over perturbation methods such as FEP is that the systems are spatially decomposed to consider the contribution of specific regions to the total solvation free energies. Visualizing these contributions can yield detailed insights into solvation thermodynamics. An insight from this work is the identification and explanation of regions with unfavorable free energy density relative to bulk water. These regions contribute unfavorably to the hydration free energy. Further work is necessary before IFST can be extended to yield accurate predictions of binding free energies, but the work presented here demonstrates its potential.



INTRODUCTION

Inhomogeneous fluid solvation theory[1] (IFST) is a statistical mechanical framework for calculating the effect of a solute on the free energy of the surrounding solvent relative to its bulk state.[2] The solute can be a protein,[3] peptide,[4] or small molecule[5] and the solvent is commonly water.[6] IFST has found particular use in the pharmaceutical industry with the advent of Schrödinger's WaterMap software.[7] It has also been developed in the solvation theory of ordered water (STOW) package,[8] and recent work using grid inhomogeneous solvation theory (GIST) has explored the results of performing IFST calculations on a Cartesian grid.[9] One of the useful features of IFST is that the free energy changes are calculated for small subvolumes surrounding the solute and this allows the contribution of different regions of space to be calculated and visualized. This has been used to understand the determinants of binding affinity[10] and design new inhibitors in the hit-to-lead and lead optimization stages of drug development.[11] Work in this lab has focused on the importance of modeling solvation at protein surfaces[12] and the data requirements for convergence of the thermodynamic quantities computed by IFST.[13] One important issue that has not been addressed in detail is the quantitative accuracy of IFST. Initial work suggested a reasonable comparison with the experimental solvation free energy of methane[5] and recent work notes solvation free energies as a key benchmark for the method.[9] While a direct comparison with experiment is interesting, the results rely on the force field and particularly the water model that is used.[14] Thus, a more useful comparison is with an equivalent computational technique. This decouples testing of the method from testing of the parameters. In this work, we compute solvation free energies for six simple molecules using IFST and compare the results with solvation free energies calculated using free energy perturbation (FEP).[15]

SIMULATION DETAILS

Hydration free energies for six simple molecule were calculated using two statistical mechanical computational methods, FEP

Received: April 29, 2013
Revised: June 12, 2013
Published: June 13, 2013

   dx.doi.org/10.1021/jp4042233 | J. Phys. Chem. B 2013, 117, 8232−8244

and IFST. FEP calculates the hydration free energies directly, whereas IFST calculates the hydration enthalpies and entropies and these are combined to yield the predicted hydration free energy. All molecular dynamics (MD) simulations in this work were performed with the TIP4P-2005 water model.[16]

Systems Setup. The six solutes studied were acetamide, benzene, isobutane, methane, methanol, and *N*-methylacetamide. The parameters for acetamide, benzene, isobutane, methanol, and *N*-methylacetamide were taken from the CHARMM36 force field.[17] The parameters for methane were not present in the CHARMM36 force field and were thus adapted from ethane. The methane carbon atom was assigned an atom type of CG331 and a partial charge of −0.36. The methane hydrogen atoms were assigned an atom type of HGA3 and a partial charge of +0.09. The bond lengths, bond angles, and dihedral angles were set to their force field equilibrium values for all molecules.

Water Setup. To generate a reasonable initial water density, a water shell of radius 50.0 Å was generated around each biomolecule with the SOLVATE program version 1.0 from the Max Planck Institute.[18] The resulting water globules were then cut to rhombic dodecahedral unit cells with side lengths of 25.0 Å. To standardize the geometries of the water molecules, every hydrogen atom was deleted and all the necessary hydrogen atoms and lone pairs were built using the appropriate geometry for TIP4P-2005 water. The boxes contained 384, 371, 376, 394, 376, and 374 water molecules for the molecules acetamide, benzene, isobutane, methane, methanol, and *N*-methylacetamide, respectively. No ions were included in the systems.

IFST Equilibration. Equilibration was performed for 1.0 ns in an NPT ensemble at 300 K and 1 atm using Langevin temperature control and Nosé−Hoover[19] Langevin piston pressure control.[20] All systems were brought to equilibrium before continuing, by verifying that the energy fluctuations were stable. MD simulations were performed using an MD time step of 2.0 fs. Electrostatic interactions were modeled with a uniform dielectric and a dielectric constant of 1.0 throughout the equilibration and production runs. van der Waals interactions were truncated at 11.0 Å with switching from 9.0 Å. Electrostatics were modeled using the particle mesh Ewald method,[21] and the systems were treated using rhombic dodecahedral periodic boundary conditions. All solute atoms were fixed for the entirety of the equilibration and production simulations.

IFST Simulation. 100.0 ns of production simulation in an NPT ensemble were performed at 300 K and 1 atm for each system. System snapshots were saved every 10.0 fs, yielding 10 000 000 snapshots in total for each system. MD simulations for IFST were performed using NAMD[22] version 2.8 compiled for use with CUDA-accelerated GPUs.

IFST Implementation. An important decision in implementing IFST is the choice of subvolumes over which to perform the calculations. In protein binding sites, water molecules commonly cluster in distinct locations and the concept of a hydration site is useful to calculate contributions to the free energy from spherical regions. In the context of small molecule solvation, the water molecules do not cluster in distinct locations and the concept of a hydration site is not useful. Two approaches have been used to account for this in previous work. In the Lazaridis work on the solvation of methane,[5] the region surrounding the solute was split into subshells at different distances from the origin. For more complex molecules, spherical polar coordinates could usefully be used to further split these subshells. In the recent work on GIST, the region surrounding the solute was split into cubic voxels on a Cartesian grid.[9] This has the advantage that each subvolume has the same volume. There are a number of other possibilities for partitioning a volume into subvolumes and these different methods of honeycombing space are likely to affect the performance of IFST. In this work, we have used a Cartesian grid to honeycomb the volume and we use the term voxel to specify one cubic unit in this space. We consider all voxels within 12.0 Å of the origin in the calculation, where the centroid of each solute lies at the origin. In this study, we have not taken advantage of the symmetry of the solute molecules, which would provide more data, as we are interested in applying IFST to systems that do not have such symmetry in the future.

IFST Calculations. IFST calculates the difference in interaction energy (ΔE_{IFST}) and correlation entropy (ΔS_{IFST}) between each subvolume (v) and the equivalent number of bulk water molecules (n). These can be termed the local quantities. The contribution of each subvolume to ΔE_{IFST} was calculated from the mean solute−water interaction energy (E_{sw}), the mean water−water interaction energy (E_{ww}), and the mean interaction energy of a bulk water molecule (E_{bulk}) as follows:

$$\begin{aligned}\Delta E_{\mathrm{IFST}} &= \sum\nolimits_{v} [E_{\mathrm{sw}} + E_{\mathrm{ww}} - nE_{\mathrm{bulk}}] \\ &= \sum\nolimits_{v} [E_{\mathrm{sw}} + \Delta E_{\mathrm{ww}}]\end{aligned} \tag{1}$$

For the TIP4P-2005 water model, E_{bulk} is calculated from a 100 ns NPT simulation of bulk water at 300 K and 1 atm as −11.5813 kcal/mol. In this work, E_{bulk} and E_{ww} are defined as half the interaction energy of one water molecule with all other water molecules. Previous work have defined E_{bulk} and E_{ww} as the total interaction energy of one water molecule with all other water molecules and included factors of one-half in eq 1.[13,23] The two approaches are identical. The difference in correlation entropy is calculated from solute−water entropy (S_{sw}), the water−water entropy (S_{ww}), and the entropy of a bulk water molecule (S_{bulk}) as follows:

$$\begin{aligned}\Delta S_{\mathrm{IFST}} &= \sum\nolimits_{v} [S_{\mathrm{sw}} + S_{\mathrm{ww}} - nS_{\mathrm{bulk}}] \\ &= \sum\nolimits_{v} [S_{\mathrm{sw}} + \Delta S_{\mathrm{ww}}]\end{aligned} \tag{2}$$

All entropies calculated in this work exclude vibrational entropy changes. For each subvolume, S_{sw} is calculated from the translational and orientational contributions using the translational and orientational correlation functions, $g_{\mathrm{sw}}(r)$ and $g_{\mathrm{sw}}(\omega|r)$.

$$S_{\mathrm{sw}}^{\mathrm{trans}} = -k\rho^{\circ} \int g_{\mathrm{sw}}(r) \ln g_{\mathrm{sw}}(r)\, \mathrm{d}r \tag{3}$$

$$S_{\mathrm{sw}}^{\mathrm{orient}} = -k\frac{\rho^{\circ}}{\Omega} \int g_{\mathrm{sw}}(r) g_{\mathrm{sw}}(\omega|r) \ln g_{\mathrm{sw}}(\omega|r)\, \mathrm{d}r\, \mathrm{d}\omega \tag{4}$$

k is Boltzmann's constant, ρ° is the number density of bulk water, and Ω is the integral over the angles. For the TIP4P-2005 water model, ρ° is calculated from a 100 ns NPT simulation of bulk water at 300 K and 1 atm as 0.0331 707 molecules/Å^3. The solute−water orientational correlation functions were calculated using Euler angles by computing α, cos β, and γ in a fixed reference frame. The limits of integration for the water molecule are between 0 and 2π for α, between −1 and 1 for cos β, and between 0 and π for γ, due to water's C_{2v}

dx.doi.org/10.1021/jp4042233 | J. Phys. Chem. B 2013, 117, 8232−8244

symmetry. The orientational correlation functions were assumed to be independent of the position within the hydration sites. The default grid resolution was 0.5 Å and the default angular bin size was 45°. This leads to 8, 4, and 4 angular bins for α, cos β and γ and thus 128 angular bins in total. The ΔS_{ww} term can also be calculated from translational and orientational contributions. A rigorous treatment would use the inhomogeneous water−water pair-correlation functions $g_{ww}(r,r')$ and $g_{ww}(\omega,\omega'|r,r')$ from the system and the homogeneous water−water pair-correlation functions $g_{ww}(R)$ and $g_{ww}(\omega^{rel}|R)$ from bulk water. R and ω^{rel} are the distance and relative orientation between two water molecules in bulk water.

$$\Delta S_{ww}^{trans} = -\frac{1}{2}k_b(\rho^\circ)^2 \int g_{sw}(r)[\{g_{sw}(r')g_{ww}(r, r') \ln g_{ww}(r, r') - g_{ww}(r, r') + 1\} - \{g_{ww}^{bulk}(R) \ln g_{ww}^{bulk}(R) - g_{ww}^{bulk}(R) + 1\}]\, dr\, dr' \quad (5)$$

$$\Delta S_{ww}^{orient} = -\frac{1}{2}k_b\frac{(\rho^\circ)^2}{\Omega^2} \int g_{sw}(r)[g_{sw}(r')g_{ww}(r, r')g_{sw}(\omega|r) g_{sw}(\omega'|r')g_{ww}(\omega, \omega'|r, r') \ln g_{ww}(\omega, \omega'|r, r') - g_{ww}^{bulk}(R)g_{ww}^{bulk}(R)(\omega^{rel}|R) \ln g_{ww}^{bulk}(\omega^{rel}|R)]\, dr\, dr'\, d\omega\, d\omega' \quad (6)$$

However, converging calculations of the inhomogeneous water−water pair-correlation functions requires a very large amount of data and thus the Kirkwood superposition approximation (KSA) is used.[24] The KSA provides an approximation of a third-order correlation function in terms of lower order correlation functions. In this context, the KSA is applied to the homogeneous triplet solute−water−water correlation function. The inhomogeneous water−water correlation function is then assumed to be independent of the solute−water correlation functions and thus equal to the water−water correlation function in bulk solvent. The KSA is used for both $g_{ww}(r,r')$ and $g_{ww}(\omega,\omega'|r,r')$:

$$g_{ww}(r, r') = g_{ww}^{bulk}(R) \quad (7)$$

$$g_{ww}(\omega, \omega'|r, r') = g_{ww}^{bulk}(\omega^{rel}|R) \quad (8)$$

These correlation functions can be converged in bulk water due the increased data available from it being a homogeneous and isotropic system. Previous work on the solvation of methane suggests that these KSAs lead to reasonable results due to the cancellation of errors between more and less structured regions.[5] In this work, S_{ww} was only calculated for pairs of voxels within 3.6 Å. This is likely a good assumption for the translational contributions, which in bulk water are derived almost exclusively from the excluded volume and the first solvation shell.[25] The orientational contributions to S_{ww} (and S_{bulk}) are expected to be underestimated, as only approximately 75% of the orientational entropy in bulk water is derived from the first solvation shell.[25,26] Calculations for $g_{ww}(R)$ and $g_{ww}(\omega^{rel}|R)$ were performed on pairs of voxels on the Cartesian grid from the 100 ns NPT simulation of bulk water at 300 K and 1 atm. The difference in free energy (ΔG_{IFST}) for each voxel is then calculated from ΔE and ΔS.

$$\Delta G_{IFST} = \Delta E_{IFST} - T\Delta S_{IFST} \quad (9)$$

FEP Protocol. The equilibrated systems generated for the IFST simulations were used as the start points for the FEP systems. These systems consist of the solute in water and correspond to the λ = 0.0 states. FEP calculations were performed in both forward and backward directions to yield corresponding predictions for annihilation (λ = 0.0 to λ = 1.0) and creation (λ = 1.0 to λ = 0.0) of the solutes. Each annihilation and creation was split into 24 steps to yield 48 λ windows per system. The lambda schedules are reported in Table 1. Two measures were adopted to avoid the numerical instabilities termed "end-point catastrophes" that occur when λ approaches 0.0 or 1.0. First, a soft-core potential was employed with a van der Waals radius-shifting coefficient of 5.0.[27,28] Second, van der Waals interactions were scaled down to zero between λ = 0.0 and λ = 1.0 whereas electrostatic interactions were scaled down to zero between λ = 0.0 and λ = 0.4.[29] MD simulations for FEP were performed using NAMD version 2.8.[22]

Table 1. Lambda Schedules for FEP Annihilation and Creation[a]

	annihilation		creation	
window	$\lambda_{initial}$	λ_{final}	$\lambda_{initial}$	λ_{final}
1	0.000	0.015	1.000	0.975
2	0.015	0.030	0.975	0.925
3	0.030	0.045	0.925	0.850
4	0.045	0.060	0.850	0.600
5	0.060	0.075	0.600	0.450
6	0.075	0.090	0.450	0.350
7	0.090	0.105	0.350	0.300
8	0.105	0.120	0.300	0.275
9	0.120	0.135	0.275	0.250
10	0.135	0.150	0.250	0.230
11	0.150	0.170	0.230	0.210
12	0.170	0.190	0.210	0.190
13	0.190	0.210	0.190	0.170
14	0.210	0.230	0.170	0.150
15	0.230	0.250	0.150	0.135
16	0.250	0.275	0.135	0.120
17	0.275	0.300	0.120	0.105
18	0.300	0.350	0.105	0.090
19	0.350	0.450	0.090	0.075
20	0.450	0.600	0.075	0.060
21	0.600	0.850	0.060	0.045
22	0.850	0.925	0.045	0.030
23	0.925	0.975	0.030	0.015
24	0.975	1.000	0.015	0.000

[a]The 24 FEP windows used in the annihilation and creation simulations of the six solutes. The initial and final values of λ are given for each window.

FEP Equilibration. Starting with the equilibrated systems generated for the IFST simulations, further equilibration was performed for 800 ps in an NPT ensemble for each lambda window.

FEP Simulation. 7.0 ns of production simulation in an NPT ensemble were performed at 300 K and 1 atm for each lambda window.

FEP Calculations. For FEP, the free energy of a change (ΔG_{FEP}) is calculated as the sum of free energy changes for a series of N small steps between intermediate states a and b.[15]

dx.doi.org/10.1021/jp4042233 | J. Phys. Chem. B 2013, 117, 8232−8244

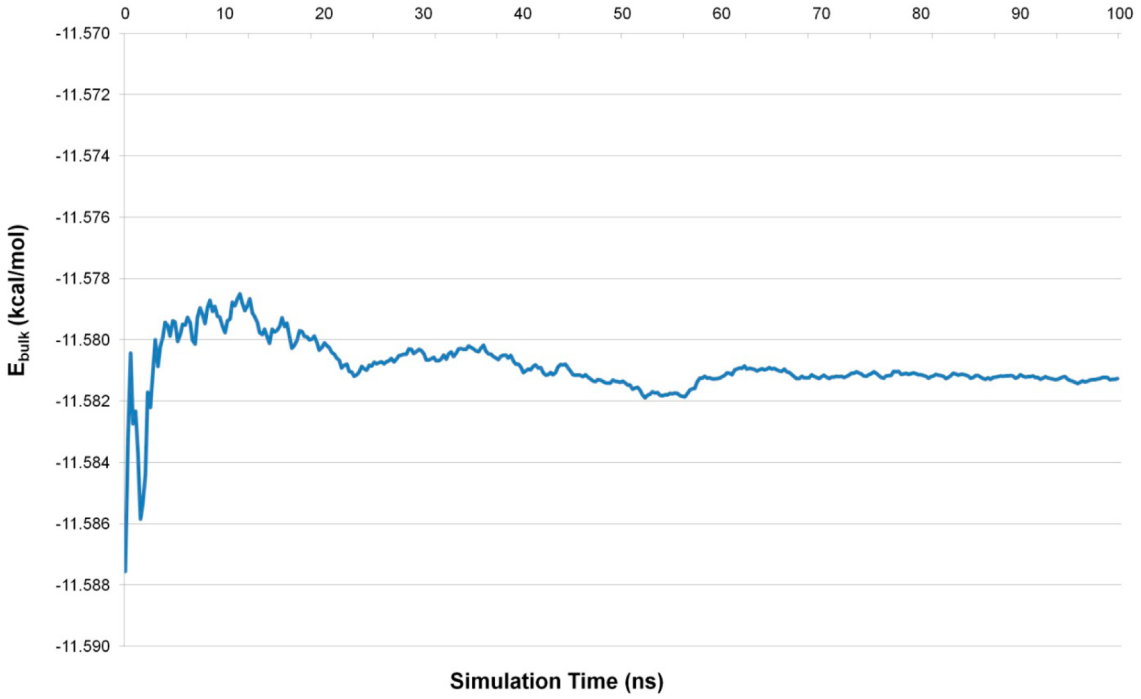


Figure 1. Convergence of E_{bulk} for a 100 ns simulation of bulk TIP4P-2005 water: the running average of E_{bulk} for a 100 ns NPT simulation of bulk TIP4P-2005 water at 300 K and 1 atm. The running average is calculated for 400 blocks of 250 ps. Each block is the average potential energy per water molecule taken from 2500 samples with one every 100 fs.

$$\Delta G_{FEP} = \sum_{a=1,b=a+1}^{N} \Delta G_{a\to b} \tag{10}$$

The free energy change for each small step ($\Delta G_{a\to b}$) is calculated using the partition functions (Q) for the two states which are calculated using the Hamiltonians (H) for the two states.

$$\Delta G_{a\to b} = G_b - G_a = -kT \ln\left(\frac{Q_b}{Q_a}\right) = -kT \ln(\langle \exp(-(H_b - H_a)/kT)\rangle_a) \tag{11}$$

We employed the Bennett acceptance ratio (BAR) method,[30] a technique that combines forward and backward FEP simulations to provide efficient estimates of free energy changes.[29] BAR was implemented using the ParseFEP Plugin from VMD. The ParseFEP Plugin is described in "A Toolkit for the Analysis of Free-Energy Perturbation Calculations" and the free energy and the statistical error are computed using eqs 11 and 14 from that paper.[31] The estimated statistical error in the FEP free energy predictions using BAR was less than 0.1 kcal/mol in all cases.

Additional Considerations. IFST calculations equate the change in enthalpy (ΔH) with the change in potential energy (ΔE). For calculations in an NPT ensemble, there is an additional contribution to the free energy of $P\Delta V$, where P is the pressure and ΔV is the change in volume. However, for solvation of small molecules in water under ambient conditions, $P\Delta V$is significantly smaller than ΔE and can be safely neglected.[32] In addition, the FEP and IFST calculations in this study both use a nonpolarizable force field. The parameters for the solutes are for prepolarized molecules and thus the free energy change associated with polarization of the solutes is ignored.[14] The calculated values are thus expected to differ from the experimental values. As noted previously, such calculations also ignore the difference in polarization of the water molecules between bulk water and the solution.[14] An additional consideration when comparing IFST calculations to experiment are terms that arise from the thermal motion of the solute and the change in ideal solvent entropy upon solute insertion. The additional terms are commonly termed the liberation contributions (ΔE_{lib} and ΔS_{lib}) and the volume entropy contribution (S_{ve}).[33] Collectively, they are termed the nonlocal contributions and are distinct from the correlation terms in eqs 1 and 2.[1] The local contributions ΔE_{IFST}, ΔS_{IFST}, and ΔG_{IFST} represent Ben-Naim's standard energy, entropy, and free energy of solvation.[34] The liberation enthalpy ΔE_{lib} is calculated from the solvent's thermal expansion coefficient α and isothermal compressibility κ:

$$\Delta E_{lib} = kT(\alpha T - \kappa P) \tag{12}$$

For solvation in water under standard ambient conditions, ΔE_{lib} is approximately +0.05 kcal/mol. IFST is formulated in such a way that portions of ΔS_{lib} and S_{ve} cancel with some of the correlation terms.[1] The entropy from the nonlocal contributions $\Delta S_{nonlocal}$ is calculated from the remaining terms:

$$\Delta S_{nonlocal} = k(T\alpha - 1) \tag{13}$$

For solvation in water under standard ambient conditions $\Delta S_{nonlocal}$ makes a contribution of +0.55 kcal/mol to the free energy. In this work, we compare the predictions of FEP and IFST with Ben-Naim's experimentally determined standard energy, entropy, and free energy of solvation. These do not include the nonlocal contributions, which are thus excluded from our calculations.

dx.doi.org/10.1021/jp4042233 | J. Phys. Chem. B 2013, 117, 8232−8244

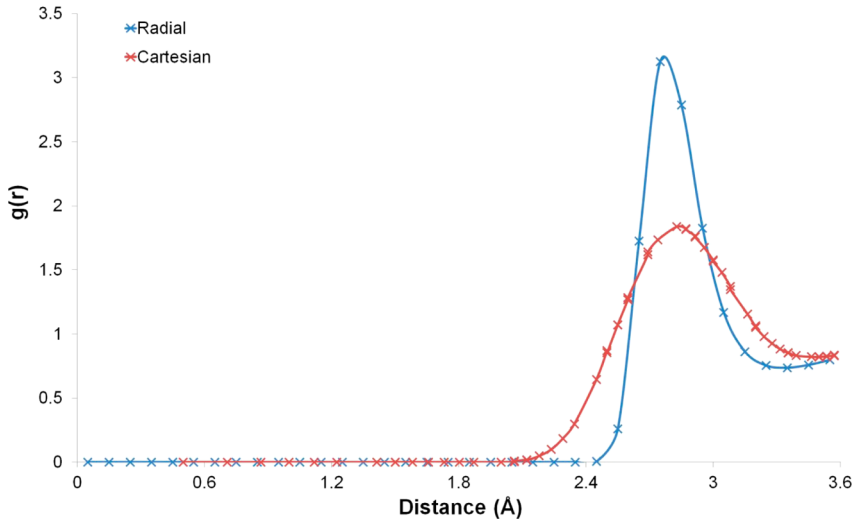


Figure 2. Translational correlation function for TIP4P-2005 between 0.0 and 3.6 Å. The RDF between 0.0 and 3.6 Å from the bulk water simulation of TIP4P-2005 (blue line) and an "effective RDF" computed for voxels at the given distance from the origin (red line). The effective RDF is calculated by first computing the distance between the origin and the center of each voxel. Voxels at the same distance from the origin are then grouped together. The value of $g(r)$ at each distance is calculated from the average number of water molecules in the group of voxels at that distance and their summed volume.

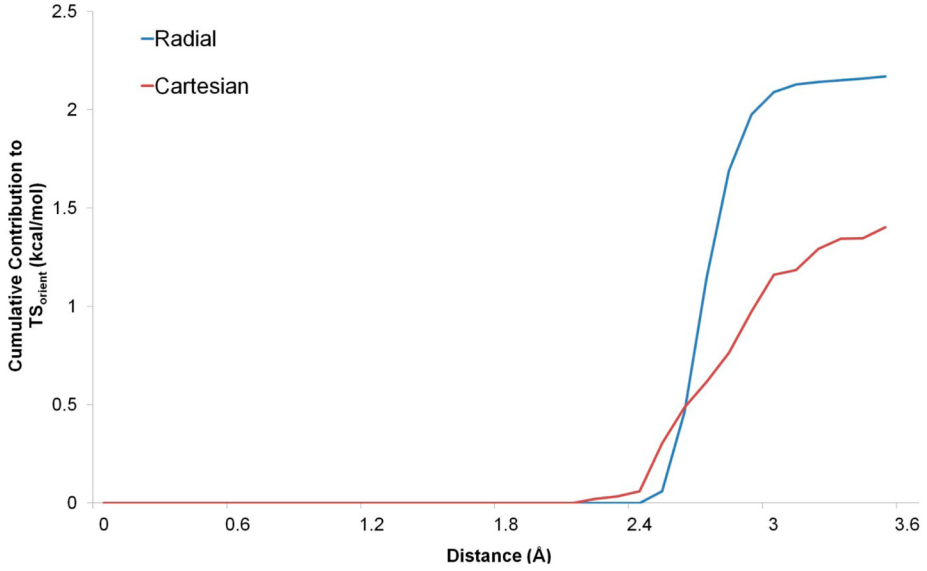


Figure 3. Cumulative orientational contribution to TS_{bulk} for bulk TIP4P-2005 water: the cumulative contribution to TS_{bulk} between 0.0 and 3.6 Å from the bulk water simulation of TIP4P-2005 using the radial coordinate system (blue line) and the Cartesian coordinate system (red line).

EXPERIMENTAL DATA

The standard energies, entropies, and free energies of solvation of solvation for benzene, isobutane, methane, and methanol are taken from the same source.[35] The standard free energies of solvation for acetamide and *N*-methylacetamide are derived from Wolfenden[36] and taken from the same source.[37] The enthalpies of solvation for acetamide and *N*-methylacetamide are taken from the same source[38] and then modified to yield the standard enthalpies of solvation.[39] The entropies of solvation were determined from the standard energies and free energies.

RESULTS AND DISCUSSION

The simulation of bulk water is a key part of this work, as ΔE_{IFST} depends critically upon the calculated value of E_{bulk}. This is because the volume of solvent we are considering (voxels within 12.0 Å of the origin) contains approximately 240 water molecules. Thus, a difference of 0.01 kcal/mol in E_{bulk} leads to a difference of 2.4 kcal/mol in ΔE_{IFST} and ΔE_{FEP}. This is significant in the context of predicting solvation free energies. A simple calculation suggests that E_{bulk} must be correctly converged to within 0.0002 kcal/mol for ΔE_{IFST} and ΔE_{FEP} to be correctly calculated to within 0.1 kcal/mol. This extreme dependence means that it is thus vital that the calculation of E_{bulk} is converged. Figure 1 shows the convergence of E_{bulk} for the 100.0 ns NPT simulation. After approximately 75 ns, the running average remains within 0.0002 kcal/mol of the final calculated value, yielding a final E_{bulk} of −11.5813 kcal/mol for the TIP4P-2005 water model.

The simulation of bulk water also allows calculation of the excess entropy (S_{bulk}). Numerous works have reported the excess translational entropy of bulk water based on IFST

dx.doi.org/10.1021/jp4042233 | J. Phys. Chem. B 2013, 117, 8232−8244

Table 2. Results of IFST Calculations on Bulk Water and the Six Solutes[a]

thermodynamic quantity (kcal/mol)	bulk water	acetamide	benzene	isobutane	methane	methanol	*N*-methylacetamide
$-TS_{sw}$ trans	0.1	3.7	4.7	4.6	2.3	3.1	4.5
$-TS_{sw}$ orient	0.8	2.9	2.7	2.4	1.6	2.5	3.2
$-T\Delta S_{ww}$ trans	0.1	−0.4	−0.5	−0.4	−0.4	−0.4	−0.6
$-T\Delta S_{ww}$ orient	0.1	1.2	1.9	1.8	0.9	1.3	1.6
$-T\Delta S$ trans	0.1	3.3	4.1	4.1	2.0	2.7	3.9
$-T\Delta S$ orient	0.8	4.1	4.6	4.2	2.6	3.8	4.8
$\boldsymbol{-T\Delta S_{IFST}}$	**1.0**	**7.4**	**8.8**	**8.3**	**4.5**	**6.5**	**8.6**
E_{sw}	0.0	−29.4	−15.8	−9.8	−3.6	−19.7	−29.1
ΔE_{ww}	0.2	14.8	8.3	4.8	1.2	9.7	14.2
$\boldsymbol{\Delta E_{IFST}}$	**0.2**	**−14.7**	**−7.4**	**−5.1**	**−2.5**	**−10.0**	**−14.9**
$\boldsymbol{\Delta G_{IFST}}$	**1.2**	**−7.2**	**1.3**	**3.2**	**2.1**	**−3.4**	**−6.2**

[a]The calculated IFST quantities for bulk water and the six solutes from the 100 ns simulations. The total translational and orientational contributions are the sums of the solute−water and water−water terms.

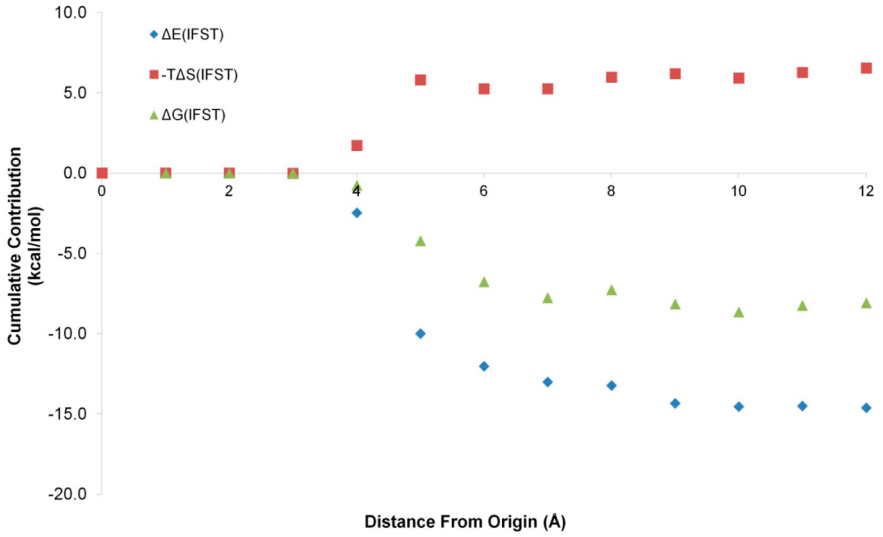


Figure 4. Cumulative contributions to ΔE_{IFST}, ΔS_{IFST}, and ΔG_{IFST} for acetamide at increasing distances from the origin: the cumulative contributions to ΔE_{IFST} (blue diamonds), $T\Delta S_{IFST}$ (red squares), and ΔG_{IFST} (green triangles) for acetamide between 0.0 and 12.0 Å from the origin.

calculations using spherical polar coordinates.[25,26,40−43] Due to bulk water being isotropic, the excess translational entropy can be calculated from the radial distribution function (RDF). Using the RDF with a radial bin size of 0.1 Å, the excess translational entropy in the first solvation shell (between 0.0 and 3.6 Å) is calculated to be −3.29 cal/mol/K for the TIP4P-2005 water model. This makes a contribution of +0.98 kcal/mol to the excess free energy. It is interesting to compare this with the translational entropy based on IFST calculations using a Cartesian coordinate system. Using the default grid resolution of 0.5 Å, the excess translational entropy for voxels within the first solvation shell (between 0.0 and 3.6 Å) is calculated to be −2.10 cal/mol/K. This makes a contribution of +0.63 kcal/mol to the excess free energy. The true RDF and an "effective RDF" using the Cartesian coordinate system can be seen in Figure 2.

The curve is much flatter for the Cartesian coordinate system, with a significantly reduced contribution from the excluded volume. This suggests that a Cartesian grid of 0.5 Å may be too coarse to fully capture the translation probability density function. However, similar results are observed for the solute water RDFs (data not shown) and the errors in the translational entropy for the bulk system and the solute−water systems largely cancel. A similar approach can also be taken for the orientational entropy. Using spherical polar coordinates with a radial bin size of 0.1 Å and an angular bin size of 10°, the excess orientational entropy in the first solvation shell (between 0.0 and 3.6 Å) is calculated to be −7.28 cal/mol/K for the TIP4P-2005 water model. This makes a contribution of +2.17 kcal/mol to the excess free energy. For the Cartesian coordinate system using a grid resolution of 0.5 Å and an angular bin size of 10°, the excess orientational entropy in the first solvation shell (between 0.0 and 3.6 Å) is calculated to be −4.70 cal/mol/K. This makes a contribution of +1.40 kcal/mol to the excess free energy. Figure 3 shows the cumulative orientational contributions to S_{bulk} between 0.0 and 3.6 Å for the radial and Cartesian coordinate calculations. These findings show that S_{bulk} is underestimated using a grid resolution of 0.5 and this suggests that S_{ww} is likely to be misestimated.

After calculating the necessary values for the bulk water simulation, we tested the method by using the bulk simulation

dx.doi.org/10.1021/jp4042233 | J. Phys. Chem. B 2013, 117, 8232−8244

as a basis for IFST calculations. The IFST quantities were calculated for all voxels within 12.0 Å of the origin. In all the following work, we report the entropic contributions to the free energy ($-T\Delta S_{IFST}$) in kcal/mol rather than the entropies (ΔS_{IFST}) in cal/mol/K, as this allows a direct comparison with the enthalpic contributions to the free energy. The second column of Table 2 shows the contributions from the different IFST quantities for the bulk water simulation, which should all be zero when there is no perturbation from a solute. It is clear that the orientational contributions to $-TS_{sw}$ are overestimated by this method and this is due to the inherent bias in the histogram method. The overestimate is very small for each voxel (approximately 0.000 015 kcal/mol) but there are approximately 60 000 voxels and this leads to a significant overestimate in total. For this reason, we expect that the orientational contributions to S_{sw} will be overestimated for the solute systems. The other contributions to the free energy are close enough to zero to be acceptable.

We then moved on to consider the solute−water systems. A key question that we have previously attempted to address is the size of the solvation shell around a solute that is affected by its presence.[13,23] However, previous studies were incomplete because they considered the per voxel values of ΔE_{IFST}, ΔS_{IFST}, and ΔG_{IFST} with increasing distance from the origin rather than the total ΔE_{IFST}, ΔS_{IFST}, and ΔG_{IFST}. These total values are affected by the volume increasing with the cube of the distance from the origin and are more revealing. We have calculated these values for the six solutes to more thoroughly address this issue. The results can be seen in Figure 4 for acetamide.

The total ΔS_{IFST} approaches its final value at around 5.0 Å from the origin whereas the total ΔE_{IFST} and thus the total ΔG_{IFST} approach their final value at around 9.0 Å from the origin. The slight peaks in ΔS_{IFST} at 5.0 and 9.0 Å are due to the first and second solvation shells, as the plot is of the distance from the origin and acetamide has an approximate "radius" of 3.0 Å. When the size of the solutes is considered, the results indicate that the solvent is perturbed to a distance of approximately 7.0 Å from the surface of a solute. This corresponds to the first two solvation shells, with the first solvation shell contributing around 80% to ΔG_{IFST}. As discussed previously, these distances may be much larger for charged systems and they may also be different for larger solutes. Indeed, studies on a 16-residue peptide suggest that significant perturbations may extend to the third solvation shell at around 10 Å from the surface.[44]

We also wished to study the convergence of the IFST quantities. As previously, we distinguish between the convergence with increased simulation time for a given sampling frequency and the convergence with increased sampling frequency for a given simulation time.[13,23] The convergence with increased simulation time can be seen for acetamide in Table 3 for a sampling frequency of 10 fs. The result for ΔE_{IFST} after 20.0 ns differs from the result after 100 ns by only 0.1 kcal/mol but the result for $T\Delta S_{IFST}$ differs by more than 0.1 kcal/mol, even after 50.0 ns. It is the difference in the orientational component of S_{sw} that is notably different as the simulation time is decreased. This is not unexpected, as there are 128 times as many histogram bins for the orientational component of S_{sw} as for the translational component. These calculations suggest that converged predictions for IFST require at least 100 ns of simulation with a grid resolution of 0.5 Å and an angular bin size of 45° but that simulations of 20 ns may be sufficient to predict

Table 3. Convergence of IFST Predictions with Increasing Simulation Time[a]

time (ns)	10	20	50	100
$-TS_{sw}$ trans (kcal/mol)	4.0	3.9	3.7	3.7
$-TS_{sw}$ orient (kcal/mol)	9.2	5.7	3.6	2.9
$-T\Delta S_{ww}$ trans (kcal/mol)	−0.3	−0.3	−0.4	−0.4
$-T\Delta S_{ww}$ orient (kcal/mol)	1.7	1.5	1.3	1.2
$-T\Delta S_{IFST}$ (kcal/mol)	14.5	10.8	8.2	7.4
E_{sw} (kcal/mol)	−29.4	−29.4	−29.4	−29.4
ΔE_{ww} (kcal/mol)	15.4	14.9	14.9	14.8
ΔE_{IFST} (kcal/mol)	−14.0	−14.5	−14.5	−14.6
ΔG_{IFST} (kcal/mol)	0.6	−3.8	−6.3	−7.2

[a]The calculated enthalpic and entropic contributions to the free energy of acetamide calculated using all snapshots within 10, 20, 50, and 100 ns. Snapshots are taken every 10 fs.

ΔE_{IFST}. The convergence with increased sampling frequency for acetamide can be seen in Table 4 for a simulation time of 100

Table 4. Convergence of IFST Predictions with Increased Sampling[a]

sampling freq (fs)	1000	500	200	100	50	20	10
$-TS_{sw}$ trans (kcal/mol)	3.8	3.8	3.7	3.7	3.7	3.7	3.7
$-TS_{sw}$ orient (kcal/mol)	13.8	7.8	4.6	3.6	3.1	2.9	2.9
$-T\Delta S_{ww}$ trans (kcal/mol)	−0.4	−0.4	−0.4	−0.4	−0.4	−0.4	−0.4
$-T\Delta S_{ww}$ orient (kcal/mol)	1.2	1.2	1.2	1.2	1.2	1.2	1.2
$-T\Delta S_{IFST}$ (kcal/mol)	18.5	12.4	9.1	8.1	7.7	7.5	7.4
E_{sw} (kcal/mol)	−29.4	−29.4	−29.4	−29.4	−29.4	−29.4	−29.4
ΔE_{ww} (kcal/mol)	14.9	14.9	14.8	14.8	14.8	14.8	14.8
ΔE_{IFST} (kcal/mol)	−14.6	−14.6	−14.6	−14.6	−14.6	−14.6	−14.6
ΔG_{IFST} (kcal/mol)	3.9	−2.2	−5.5	−6.5	−6.9	−7.1	−7.2

[a]The calculated enthalpic and entropic contributions to the free energy of acetamide calculated using 100 ns of data with sampling frequencies of 1000, 500, 200, 100, 50, 20, and 10 fs.

ns. The result for ΔE_{IFST} using a sampling frequency of 1000 fs is the same as the result using a sampling frequency of 10 fs to one decimal place, and the result for $T\Delta S_{IFST}$ using a sampling frequency of 20 fs is the same as the result using a sampling frequency of 10 fs to one decimal place. Again, it is the orientational component of TS_{sw} that is notably different as the sampling frequency is decreased. These calculations suggest that converged predictions for IFST using a histogram method require at least 5 000 000 samples from a 100 ns simulation with a grid resolution of 0.5 Å and an angular bin size of 45°. Other methods, such as the *k*-nearest neighbors (k-NN) algorithm,[45] may require shorter simulations and less sampling.[9,42]

We also wished to consider the effect of the grid resolution and angular bin size on the orientational and translational components of S_{sw}. The results in Tables 3 and 4 indicate that the translational component of S_{sw} is well converged for a grid resolution of 0.5 Å. We also studied the results for grid resolutions of 0.25 and 0.75 Å to assess whether the translational component of S_{sw} is fully quantified with the default grid resolution of 0.5 Å. The results for acetamide are

dx.doi.org/10.1021/jp4042233 | J. Phys. Chem. B 2013, 117, 8232−8244

reported in Table 5 and indicate that the answer is no. A grid resolution of 0.25 Å yields a converged value of 4.3 kcal/mol

Table 5. Effect of the Grid Resolution on the Translational Component of $-TS_{sw}$[a]

thermodynamic quantity (kcal/mol)	grid resolution (Å)	samples			
		1000000	2000000	5000000	10000000
$-TS_{sw}$ trans	0.25	4.4	4.3	4.3	4.3
$-TS_{sw}$ trans	0.5	3.7	3.7	3.7	3.7
$-TS_{sw}$ trans	0.75	3.1	3.1	3.1	3.1

[a]The translational component of $-TS_{sw}$ for acetamide, calculated from 1 000 000, 2 000 000, 5 000 000, and 10 000 000 samples with grid resolutions of 0.25, 0.5, and 0.75 Å..

for the translational component of $-TS_{sw}$ in comparison with a value of 3.7 kcal/mol for a grid resolution of 0.5 Å. Smaller grid resolutions, which may yield a higher prediction, were not tested. Testing the orientational component of $-TS_{sw}$ in this fashion is more complicated because it requires a grid resolution in addition to an angular bin size. In this work, we have only explored the variation in entropy with respect to the angular bin size at the default grid resolution of 0.5 Å. We studied the results for angular bin sizes of 36° and 60° to assess whether the translational component of $-TS_{sw}$ is fully quantified with the default angular bin size of 45°. The results for acetamide are reported in Table 6. An angular bin size of 36° yields a value of 3.6 kcal/mol for the orientational component of $-TS_{sw}$ in comparison with a value of 2.9 kcal/mol for an angular bin size of 45°. However, it is not clear that the results for an angular bin size of 36° are fully converged from 10 000 000 samples over 100 ns and thus the question cannot be answered. Indeed, the results from the simulation of bulk water presented in the first column of Table 2 suggest that the orientational component of $-TS_{sw}$ is not fully converged from 10 000 000 samples over 100 ns for an angular bin size of 45°. Due to computational restrictions, we have not tested effect of the grid resolution and angular bin size on the orientational and translational components of ΔS_{ww}. However, the results from the simulation of bulk water suggest that the effects on S_{sw} and ΔS_{ww} may counterbalance one another.

The IFST predictions can be seen in Table 2. Calculations suggest that the translational and orientational contributions to the solvation entropy are of a similar magnitude in all cases, though the orientational contributions are always slightly larger. In addition, the solute−water terms are predicted to be 5−8-fold larger than the water−water terms. This is partly because the contribution of the translational water−water term is negative due to the solute excluded volume, but the orientational water−water term is also small. However, there are two issues to be considered. The first is that the water−water terms depend on the validity of the KSAs, which have not been thoroughly explored. The second is that our calculations suggest that the water−water terms are misestimated using a Cartesian grid with a resolution of 0.5 Å. It is interesting to note that the solute−water interactions (E_{sw}) are highly favorable in all cases, at the expense of an unfavorable change in water−water interactions (ΔE_{ww}). This leads to a favorable change in overall energy (ΔE_{IFST}) that is approximately half the magnitude of E_{sw} for the majority of solutes.

Assessing the data generated using IFST, it is clear that there are a number of issues with the method. The selection of grid resolution and angular bin size affects all the calculated entropies and in addition the use of the KSAs has not been comprehensively validated. However, these issues affect all solutes and thus we might expect a systematic misestimation of the entropies in comparison with FEP. We proceed, with this in mind. The results from experiment along with the predictions from FEP and IFST can be seen in Table 7. The correlation plots between experiment, IFST, and FEP can be seen in Figure 5.

Of primary interest is the agreement between FEP and IFST. The mean unsigned difference in the solvation free energies is 0.7 kcal/mol, with the IFST predictions tending to be less favorable than the FEP predictions. FEP and IFST both yield reasonable predictions of the experimental quantities with mean unsigned errors for the solvation free energies of 1.3 and 1.8 kcal/mol, respectively. *N*-Methylacetamide is particularly poorly predicted by both methods. Agreement of IFST with the experimental enthalpic and entropic contributions is reasonable, with mean unsigned errors of 0.9 and 1.2 kcal/mol, respectively. However, the results suggest that the entropic component $-T\Delta S$ tends to be overestimated by IFST, with a mean signed error of +1.2 kcal/mol. These results may explain why the IFST predictions of the hydration free energies tend to be less favorable than the FEP predictions. They are also commensurate with the overestimation of $T\Delta S$ by 1.0 kcal/mol for the simulation of bulk water.

One of the main advantages of IFST calculations is the ability to break the solvation free energies into contributions from different regions and visualize these contributions. This promotes understanding of the system being studied and an insight into the thermodynamics of solvation. Figure 6 shows the contributions from different regions to the hydration free energy of acetamide at different contour levels, visualized using VMD.[46]

The region with the highest relative free energy density is a ring surrounding the carbonyl oxygen (Figure 6A). The carbonyl group is actually predicted to contribute approximately −6.0 kcal/mol to the hydration free energy of acetamide (Figure 6B). At lower contour levels, the contributions from the amide hydrogens are visible, with each

Table 6. Effect of the Angular Bin Size on the Orientational Component of $-TS_{sw}$[a]

thermodynamic quantity (kcal/mol)	bins	angular bin size (deg)	samples			
			1000000	2000000	5000000	10000000
$-TS_{sw}$ orient	6	60	2.3	2.1	2.0	2.0
$-TS_{sw}$ orient	8	45	3.6	3.1	2.9	2.9
$-TS_{sw}$ orient	10	36	5.1	4.2	3.7	3.6

[a]The orientational component of $-TS_{sw}$ for acetamide, calculated from 1 000 000, 2 000 000, 5 000 000, and 10 000 000 samples with angular bin sizes of 30°, 45°, and 60°.

dx.doi.org/10.1021/jp4042233 | J. Phys. Chem. B 2013, 117, 8232−8244

Table 7. Results of the FEP and IFST Calculations Compared to Experiment[a]

solute	thermodynamic quantity	expt (kcal/mol)	FEP prediction (kcal/mol)	IFST prediction (kcal/mol)
acetamide	ΔG	−9.7	−8.3 ± 0.0	−7.2
	ΔH	−16.3		−14.7
	$-T\Delta S$	6.6		7.4
benzene	ΔG	−0.8	0.3 ± 0.1	1.3
	ΔH	−7.1		−7.4
	$-T\Delta S$	6.3		8.8
isobutane	ΔG	2.3	3.1 ± 0.1	3.2
	ΔH	−4.8		−5.1
	$-T\Delta S$	7.1		8.3
methane	ΔG	2.0	2.5 ± 0.0	2.1
	ΔH	−2.7		−2.5
	$-T\Delta S$	4.8		4.5
methanol	ΔG	−5.1	−4.6 ± 0.0	−3.5
	ΔH	−10.3		−10.0
	$-T\Delta S$	5.2		6.5
N-methylacetamide	ΔG	−10.1	−6.8 ± 0.0	−6.3
	ΔH	−17.1		−14.9
	$-T\Delta S$	7.0		8.6

[a]The experimental hydration enthalpy, entropy, and free energy for the six solutes, along with the predictions of FEP and IFST. The statistical errors are reported for the FEP calculations.

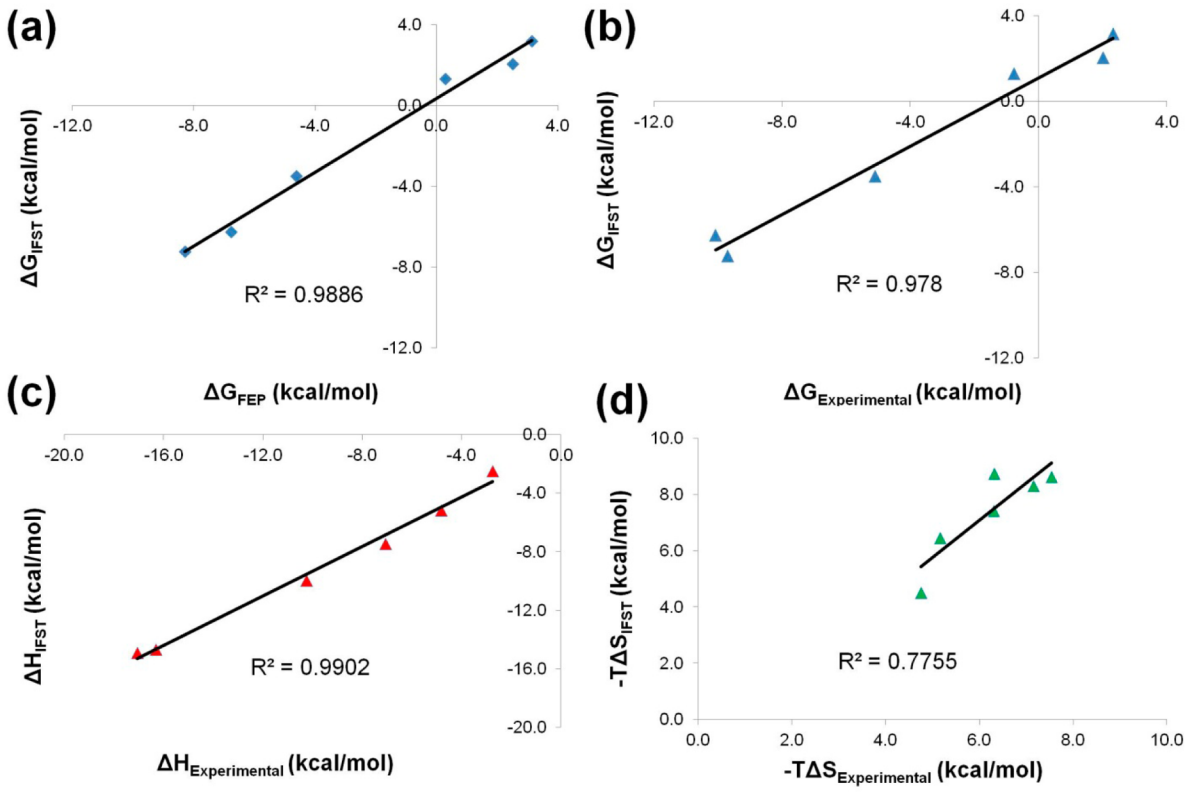


Figure 5. Correlation plots between ΔG, ΔH, and $T\Delta S$ for experiment and the predictions of FEP and IFST. Plots of the correlation between (a) ΔG_{IFST} and ΔG_{FEP} with blue diamonds, (b) ΔG_{IFST} and $\Delta G_{Experimental}$ with blue triangles, (c) ΔH_{IFST} and $\Delta H_{Experimental}$ with red triangles, and (d) $T\Delta S_{IFST}$ and $T\Delta S_{Experimental}$ with green triangles.

contributing approximately −1.5 kcal/mol to the hydration free energy (Figure 6C). It is worth noting that the favorable regions surrounding the amide hydrogens would be well described by spherical hydration sites but the favorable regions surrounding the carbonyl oxygen would not. Contouring from positive values, the regions contributing most unfavorably to the hydration free energy are above and below the plain of the amide bond (Figure 6D). This effect has been noted in previous studies using WaterMap[47] and agrees with observations from the Cambridge Crystallographic Data Centre on the hydrophobic character above and below the plane of amide bonds (personal communication with Oliver Korb). It may also explain why defectively packed backbone hydrogen bonds, which have been termed dehydrons, lead to hotspots at protein surfaces.[48] At lower contour levels, the contributions from the methyl group are visible, again contributing unfavorably to the

dx.doi.org/10.1021/jp4042233 | J. Phys. Chem. B 2013, 117, 8232−8244

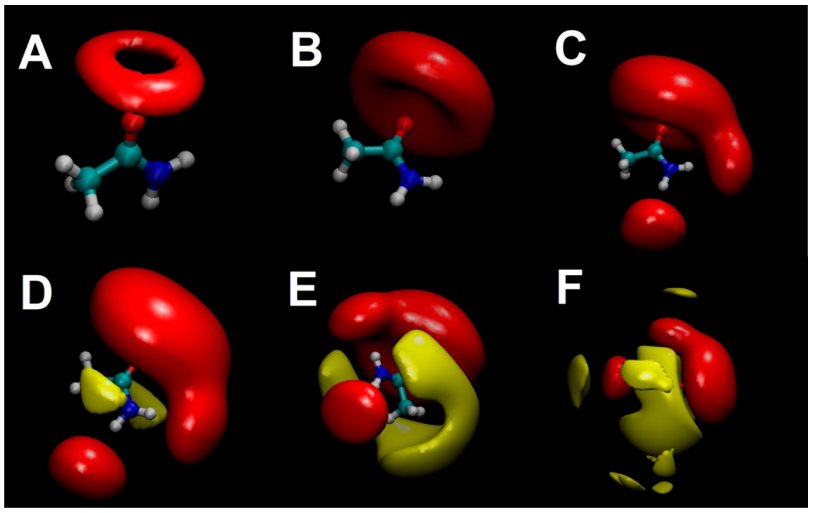


Figure 6. Visualizations of the relative free energy density relative to bulk water for regions surrounding acetamide. Six views of acetamide, with the relative free energy density contoured at different levels. Favorable relative free energy density is contoured in red and unfavorable relative free energy density is contoured in yellow. The contouring for the favorable relative free energy increases from (A) to (B) to (C). The contouring for the unfavorable relative free energy decreases from (D) to (E) to (F).

solvation free energy (Figure 6E). At very low contour levels, one can observe the small and unfavorable contributions from water molecules in the second solvation shell (Figure 6F).

The presence of regions with high number density that contribute unfavorably to a favorable hydration free energy may seem counterintuitive. The reason why a vacuum is not present within such regions can be understood using a simple example. Consider water molecules in a voxel with weak solute−water interactions (E_{sw} = −1.0 kcal/mol) and water−water interactions that are more favorable than in bulk (ΔE_{ww} = −2.0 kcal/mol) that is strongly ordered ($-TS_{sw}$ = +3.0 kcal/mol and $-T\Delta S_{ww}$ = +0.5 kcal/mol). For this voxel, ΔG_{IFST} = +0.5 kcal/mol and the region contributes unfavorably to the solvation free energy. One could view these water molecules as "icelike" given their unfavorable entropy and favorable enthalpy compared with bulk water. However, if such a region is vacated (without allowing solvent rearrangement) then all surrounding voxels also lose their pair interactions and correlations with water molecules inside the voxel. The total change in free energy (going to this unphysical state) can thus be unfavorable. If subsequent solvent rearrangement around the vacuum region cannot overcome this loss, then the voxel will be occupied despite contributing unfavorably to the solvation free energy. Clearly, there are cases where water can rearrange around a vacuum, but water molecules at the vacuum interface are likely to show reduced water−water interactions. It is worth noting that the unfavorable contributions are smaller in magnitude than the favorable contributions. Considering all voxels in the case of acetamide, the most favorable relative free energy density is −0.28 kcal/mol/Å^3 whereas the least favorable relative free energy density is +0.04 kcal/mol/Å^3. Because the exact free energy densities calculated will depend on the grid resolution and angular bin size used, it will be important to test this principle in future work. If these results accurately reflect the thermodynamics, an interesting consequence is the potential for hydration sites in protein binding sites that contribute unfavorably to the hydration free energy. Such sites would be major hotspots of binding.

There is one additional issue which must be mentioned. ΔE_{IFST}, $-T\Delta S_{IFST}$, and ΔG_{IFST} are calculated as the sum of contributions from each voxel inside a given volume V. The contributions from each voxel include pairwise contributions due to correlations with every other voxel. In the IFST framework, the pairwise contributions are split equally between the two voxels of the pair. Importantly, there are pairs where one of the voxels is not inside the volume V and thus there are regions outside the volume V which contribute to ΔE_{IFST}, $-T\Delta S_{IFST}$, and ΔG_{IFST}. This must be considered when implementing IFST by counting pair contributions consistently. However, the contribution of voxels outside the volume V is expected to be very small if the volume V is significantly larger than the perturbing solute. In this case, a voxel outside V will only have significant correlations with voxels inside V that are close to the periodic boundary. These will have bulklike properties, leading to $E_{sw} \approx 0$, $\Delta E_{ww} \approx 0$, $S_{sw} \approx 0$, $\Delta S_{ww} \approx 0$, and $\Delta G_{IFST} \approx 0$.

CONCLUSIONS

IFST is a useful method to model solvation and has found particular application in understanding hydration phenomena in biological systems.[3,7,10,12] However, there has been little analysis to assess the quantitative predictions of IFST.[3] This study addresses this issue by comparing hydration free energies from IFST with hydration free energies from FEP. The findings of this study are that IFST calculations show a good agreement with FEP calculations, with a mean unsigned difference of 0.7 kcal/mol in the predicted hydration free energies. Comparison of IFST predictions with the experimental hydration enthalpies and entropies suggests that there is a systematic overestimation of the entropies, which leads to a systematic underestimation of the free energies. For the CHARMM36 force field and the TIP4P-2005 water model, the agreement with the experimental hydration free energies is reasonable for FEP and IFST, with mean unsigned errors of 1.3 and 1.8 kcal/mol, respectively. Both IFST and FEP rely on the force field and are thus subject to the same problems in this respect. For IFST, the fine dependence of ΔE_{IFST} on E_{bulk} means that it is thus vital that the calculation of E_{bulk} is converged correctly. E_{bulk} will depend on the force field, water model, simulation package, and simulation parameters employed. Due to cancellation, this fine

dx.doi.org/10.1021/jp4042233 | J. Phys. Chem. B 2013, 117, 8232−8244

dependence will be of less importance in calculations of hydration free energy differences or binding energies, but only if the number of water molecules considered is similar. In addition, the systematic overestimations of $-T\Delta S_{IFST}$ are expected to be similar if the grid resolution and bin sizes are the same. Thus, predictions for different molecules and different regions are directly comparable for a given IFST implementation, but not necessarily between different IFST implementations.

In addition to these results, there are a number of important points raised by the study. IFST has a number of limitations that should be addressed before widespread quantitative application. Current implementations are only rigorously valid for a fixed solute and this limits the utility. Flexible solutes can be considered in the IFST framework by performing an analysis on a number of solute conformations,[9] but this will require yet more data. One of the major approximations in IFST is the use of the KSAs for the water−water entropy calculations. Previous work on the solvation of methane suggests that these KSAs lead to reasonable results due to the cancellation of errors between more and less structured regions. However, this may not be the case for polar solutes and further testing of the KSAs should be a focus of future work. Sampling is also a problem for IFST. Previous work has explored the data requirements for convergence of the solvation energy and entropy.[13,23] However, the comparison with FEP in this work allows a more thorough analysis. The results suggest that converged energies can be effectively sampled from 10 000 snapshots over a 20 ns simulation whereas the orientational component of the solute−water entropy term requires at least 5 000 000 snapshots over a 100 ns simulation to reach a converged prediction even with a modest angular bin size of 45°. However, the histogram method introduces a systematic overestimating bias to the entropies and this will scale with the volume of the system. Other work has shown that it is also difficult to estimate entropies using perturbation methods such as FEP and thermodynamic integration (TI).[49] In the context of IFST, assessment of the entropies may be more feasible using the k-NN algorithm,[9,42] particularly for the increased sampling requirements of considering a highly flexible solute. It is difficult to fairly assess the computation times required, because the FEP calculations are integrated into NAMD whereas the IFST calculations are not. This requires postprocessing of large-trajectory files at considerable computational cost. Comparing just the simulation times, each solute was simulated for a total of 336 ns for FEP and 100 ns for IFST. While FEP is inherently more parallelizable than IFST in its current format, it should be possible to apply IFST to multiple independent MD simulations, which would parallelize it effectively.

While suffering a number of drawbacks noted above, IFST does have a number of advantages over perturbation methods such as FEP and TI. Because IFST calculations are performed on a single simulation, there is no need for a lambda window schedule and no need to monitor overlap between lambda windows. In addition, there are no end points and thus there are no end-point catastrophes and no need for soft-core potentials. Importantly, there is no pathway and thus small and large perturbations can be considered with the same degree of sampling. However, perhaps the most useful feature of IFST is the spatial decomposition of the solvation free energies. This yields results that are more insightful than the total solvation free energies. Such visualization has also proved useful in 3D-RISM.[50] Indeed, it would be interesting to compare IFST results with the results of 3D-RISM[51] and cell theory,[52,53] which can provide spatially decomposed predictions of solvation free energies using different methods.

This work represents a preliminary study on the quantitative application of IFST using a very small set of solutes. Future work should explore a number of avenues. A wide range of solutes would provide a sterner test for the methods. The inclusion of charged solutes may necessitate the use of a polarizable force field. Prior to this, it would be very useful to test the validity of the KSAs for a number of solutes, using the true inhomogeneous water−water correlation functions or via comparison with lower order entropy approximations. This should be done in concert with a thorough determination of the time scales and sampling requirements that are necessary for effective convergence of the thermodynamic predictions at the desired level of accuracy. In addition, care must be taken to employ methods that effectively capture the underlying probability distributions. The combination of ΔG_{IFST} with an assessment of protein−ligand interactions will facilitate the calculation of binding free energies. This will augment binding affinity predictions based on a single unbound state simulation,[54] albeit at increased computational cost. To achieve this, it will be necessary to incorporate solute flexibility into IFST.[9]

In conclusion, the agreement between hydration free energy predictions for IFST and FEP is very encouraging. It is clear that the entropic contributions are overestimated by IFST and further work is needed. However, the insight gained by visualizing the relative free energy densities provides a major advantage of IFST that make this a worthwhile objective.

AUTHOR INFORMATION

Corresponding Author

*E-mail: djh210@cam.ac.uk. Tel.: +44(0)1223763367.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Thanks for support go to Professor Ashok Venkitaraman, Professor Chris Abell, and Pembroke College Cambridge. Acknowledgments go to Stuart Rankin for technical help and the NVIDIA CUDA Centre of Excellence at the Cambridge HPCS for use of the CUDA-accelerated GPUs. All calculations were performed using the Darwin Supercomputer of the University of Cambridge High Performance Computing Service (http://www.hpc.cam.ac.uk/) provided by Dell Inc. using Strategic Research Infrastructure Funding from the Higher Education Funding Council for England and were funded by the EPSRC under grants EP/F032773/1 and EP/J017639/1. Thanks for financial support go to the MRC and Wellcome Trust.

REFERENCES

(1) Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory. *J. Phys. Chem. B* **1998**, *102* (18), 3531−3541.

(2) Lazaridis, T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 2. Applications to simple fluids. *J. Phys. Chem. B* **1998**, *102* (18), 3542−3550.

(3) Li, Z.; Lazaridis, T. Thermodynamics of Buried Water Clusters at a Protein-Ligand Binding Interface. *J. Phys. Chem. B* **2006**, *110* (3), 1464−1475.
(4) Czapiewski, D.; Zielkiewicz, J. Structural Properties of Hydration Shell Around Various Conformations of Simple Polypeptides. *J. Phys. Chem. B* **2010**, *114* (13), 4536−4550.
(5) Lazaridis, T. Solvent Reorganization Energy and Entropy in Hydrophobic Hydration. *J. Phys. Chem. B* **2000**, *104* (20), 4964−4979.
(6) Esposito, R.; Saija, F.; Saitta, A. M.; Giaquinta, P. V. Entropy-Based Measure of Structural Order in Water. *Phys. Rev. E* **2006**, *73* (4), 040502.
(7) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. Motifs for Molecular Recognition Exploiting Hydrophobic Enclosure in Protein-Ligand Binding. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (3), 808−813.
(8) Li, Z.; Lazaridis, T. Computing the Thermodynamic Contributions of Interfacial Water. *Methods Mol. Biol.* **2012**, *819*, 393−404.
(9) Nguyen, C. N.; Young, T. K.; Gilson, M. K. Grid Inhomogeneous Solvation Theory: Hydration Structure and Thermodynamics of the Miniature Receptor Cucurbit[7]uril. *J. Chem. Phys.* **2012**, *137* (4), 044101.
(10) Beuming, T.; Farid, R.; Sherman, W. High-Energy Water Sites Determine Peptide Binding Affinity and Specificity of PDZ Domains. *Protein Sci.* **2009**, *18* (8), 1609−1619.
(11) Abel, R.; Salam, N. K.; Shelley, J.; Farid, R.; Friesner, R. A.; Sherman, W. Contribution of Explicit Solvent Effects to the Binding Affinity of Small-Molecule Inhibitors in Blood Coagulation Factor Serine Proteases. *Chemmedchem* **2011**, *6* (6), 1049−1066.
(12) Huggins, D. J.; Marsh, M.; Payne, M. C. Thermodynamic Properties of Water Molecules at a Protein Protein Interaction Surface. *J. Chem. Theory Comput.* **2011**, *7* (11), 3514−3522.
(13) Huggins, D. J. Benchmarking the Thermodynamic Analysis of Water Molecules Around a Model Beta Sheet. *J. Comput. Chem.* **2012**, *33* (15), 1383−1392.
(14) Hess, B.; van der Vegt, N. F. Hydration Thermodynamic Properties of Amino Acid Analogues: A Systematic Comparison of Biomolecular Force Fields and Water Models. *J. Phys. Chem. B* **2006**, *110* (35), 17616−17626.
(15) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22*, 1420.
(16) Abascal, J. L. F.; Vega, C. A General Purpose Model for the Condensed Phases of Water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123* (23), 234505.
(17) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D., Jr.; Pastor, R. W. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B* **2010**, *114* (23), 7830−7843.
(18) Grubmüller, H. *Solvate: A Program to Create Atomic Solvent Models*, 1996.
(19) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant-Pressure Molecular-Dynamics Algorithms. *J. Chem. Phys.* **1994**, *101* (5), 4177−4189.
(20) Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R. Constant-Pressure Molecular-Dynamics Simulation - the Langevin Piston Method. *J. Chem. Phys.* **1995**, *103* (11), 4613−4621.
(21) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G.; Smooth, A. Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577−8593.
(22) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781−1802.
(23) Huggins, D. J. Application of Inhomogeneous Fluid Solvation Theory to Model the Distribution and Thermodynamics of Water Molecules around Biomolecules. *Phys. Chem. Chem. Phys.* **2012**, *14* (43), 15106−15117.
(24) Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3* (5), 300−313.
(25) Zielkiewicz, J. Structural Properties of Water: Comparison of the SPC, SPCE, TIP4P, and TIP5P Models of Water. *J. Chem. Phys.* **2005**, *123* (10), 104501.
(26) Huggins, D. J. Correlations in Liquid Water for the TIP3P-Ewald, TIP4P-2005, TIP5P-Ewald, and SWM4-NDP Models. *J. Chem. Phys.* **2012**, *136* (6), 064518.
(27) Beutler, T. C.; Mark, A. E.; Vanschaik, R. C.; Gerber, P. R.; Vangunsteren, W. F. Avoiding Singularities and Numerical Instabilities in Free-Energy Calculations Based on Molecular Simulations. *Chem. Phys. Lett.* **1994**, *222* (6), 529−539.
(28) Zacharias, M.; Straatsma, T. P.; Mccammon, J. A. Separation-Shifted Scaling, a New Scaling Method for Lennard-Jones Interactions in Thermodynamic Integration. *J. Chem. Phys.* **1994**, *100* (12), 9025−9031.
(29) Pohorille, A.; Jarzynski, C.; Chipot, C. Good Practices in Free-Energy Calculations. *J. Phys. Chem. B* **2010**, *114* (32), 10235−10253.
(30) Bennett, C. H. Efficient Estimation of Free-Energy Differences from Monte-Carlo Data. *J. Comput. Phys.* **1976**, *22* (2), 245−268.
(31) Liu, P.; Dehez, F.; Cai, W. S.; Chipot, C. A Toolkit for the Analysis of Free-Energy Perturbation Calculations. *J. Chem. Theory Comput.* **2012**, *8* (8), 2606−2616.
(32) Gallicchio, E.; Kubo, M. M.; Levy, R. M. Enthalpy-Entropy and Cavity Decomposition of Alkane Hydration Free Energies: Numerical Results and Implications for Theories of Hydrophobic Solvation. *J. Phys. Chem. B* **2000**, *104* (26), 6271−6285.
(33) Sharp, K. A.; Nicholls, A.; Friedman, R.; Honig, B. Extracting Hydrophobic Free-Energies from Experimental-Data - Relationship to Protein Folding and Theoretical-Models. *Biochemistry* **1991**, *30* (40), 9686−9697.
(34) Ben-Naim, A. Standard Thermodynamics of Transfer. Uses and Misuses. *J. Chem. Phys.* **1978**, *82* (7), 792−803.
(35) Ben-Naim, A.; Marcus, Y. Solvation Thermodynamics of Nonionic Solutes. *J. Chem. Phys.* **1984**, *81* (4), 2016−2027.
(36) Wolfenden, R. Interaction of the Peptide Bond with Solvent Water: A Vapor Phase Analysis. *Biochemistry* **1978**, *17* (1), 201−204.
(37) Ding, Y. B.; Bernardo, D. N.; Kroghjespersen, K.; Levy, R. M. Solvation Free-Energies of Small Amides and Amines from Molecular-Dynamics Free-Energy Perturbation Simulations Using Pairwise Additive and Many-Body Polarizable Potentials. *J. Phys. Chem.* **1995**, *99* (29), 11575−11583.
(38) Dellagatta, G.; Barone, G.; Elia, V. Enthalpies of Solvation for N-Alkylamides in Water and in Carbon-Tetrachloride at 25-Degrees-C. *J. Solution Chem.* **1986**, *15* (2), 157−167.
(39) Kubo, M. M.; Gallicchio, E.; Levy, R. M. Thermodynamic Decomposition of Hydration Free Energies by Computer Simulation: Application to Amines, Oxides, and Sulfides. *J. Phys. Chem. B* **1997**, *101* (49), 10527−10534.
(40) Lazaridis, T.; Karplus, M. Orientational Correlations and Entropy in Liquid Water. *J. Chem. Phys.* **1996**, *105* (10), 4294−4316.
(41) Giuffre, E.; Prestipino, S.; Saija, F.; Saitta, A. M.; Giaquinta, P. V. Entropy from Correlations in TIP4P Water. *J. Chem. Theory Comput.* **2010**, *6* (3), 625−636.
(42) Wang, L.; Abel, R.; Friesner, R. A.; Berne, B. J. Thermodynamic Properties of Liquid Aater: An Application of a Nonparametric Approach to Computing the Entropy of a Neat Fluid. *J. Chem. Theory Comput.* **2009**, *5* (6), 1462−1473.
(43) Sharma, R.; Chakraborty, S. N.; Chakravarty, C. Entropy, Diffusivity, and Structural Order in Liquids with Waterlike Anomalies. *J. Chem. Phys.* **2006**, *125* (20), 204501.
(44) Agarwal, M.; Kushwaha, H. R.; Chakravarty, C. Local Order, Energy, and Mobility of Water Molecules in the Hydration Shell of Small Peptides. *J. Phys. Chem. B* **2010**, *114* (1), 651−659.
(45) Singh, H.; Misra, N.; Hnizdo, V.; Fedorowicz, A.; Demchuk, E. Nearest Neighbor Estimates of Entropy. *Am. J. Math. Manage. Sci.* **2003**, *23* (3−4), 301−322.
(46) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14* (1), 33−38.

(47) Beuming, T.; Che, Y.; Abel, R.; Kim, B.; Shanmugasundaram, V.; Sherman, W. Thermodynamic Analysis of Water Molecules at the Surface of Proteins and Applications to Binding Site Prediction and Characterization. *Proteins* **2012**, *80* (3), 871−83.
(48) Fernández, A.; Scott, R. Dehydron: A Structurally Encoded Signal for Protein Interaction. *Biophys. J.* **2003**, *85* (3), 1914−1928.
(49) Peter, C.; Oostenbrink, C.; van Dorp, A.; van Gunsteren, W. F. Estimating Entropies from Molecular Dynamics Simulations. *J. Chem. Phys.* **2004**, *120* (6), 2652−2661.
(50) Yamazaki, T.; Kovalenko, A. Spatial Decomposition of Solvation Free Energy Based on the 3D Integral Equation Theory of Molecular Liquid: Application to Miniproteins. *J. Phys. Chem. B* **2011**, *115* (2), 310−318.
(51) Kovalenko, A.; Hirata, F. Three-Dimensional Density Profiles of Water in Contact with a Solute of Arbitrary Shape: A RISM Approach. *Chem. Phys. Lett.* **1998**, *290* (1−3), 237−244.
(52) Henchman, R. H. Free Energy of Liquid Water from a Computer Simulation via Cell Theory. *J. Chem. Phys.* **2007**, *126* (6), 064504.
(53) Irudayam, S. J.; Henchman, R. H. Solvation Theory to Provide a Molecular Interpretation of the Hydrophobic Entropy Loss of Noble-Gas Hydration. *J. Phys.: Condens. Matter* **2010**, *22* (28), 284108.
(54) Abel, R.; Wang, L.; Friesner, R. A.; Berne, B. J.; Displaced-Solvent, A. Functional Analysis of Model Hydrophobic Enclosures. *J. Chem. Theory Comput.* **2010**, *6* (9), 2924−2934.