



Published in final edited form as:

*J Phys Chem B*. 2018 May 03; 122(17): 4700–4707. doi:10.1021/acs.jpcc.8b02666.

## The Excess Chemical Potential of Water at the Interface with a Protein from Endpoint Simulations

Bin W. Zhang<sup>†</sup>, Di Cui<sup>†</sup>, Nobuyuki Matubayasi<sup>‡</sup>, and Ronald M. Levy<sup>\*,†</sup>

<sup>†</sup>Center for Biophysics and Computational Biology, Department of Chemistry, and Institute for Computational Molecular Science, Temple University, Philadelphia, Pennsylvania 19122, United States

<sup>‡</sup>Division of Chemical Engineering, Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka 560-8531, Japan and Elements Strategy Initiative for Catalysts and Batteries, Kyoto University, Katsura, Kyoto 615-8520, Japan

### Abstract

We use endpoint simulations to estimate the excess chemical potential of water in the homogeneous liquid and at the interface with a protein in solution. When the pure liquid is taken as the reference, the excess chemical potential of interfacial water is the difference between the solvation free energy of a water molecule at the interface and in the bulk. Using the homogeneous liquid as an example, we show that the solvation free energy for growing a water molecule can be estimated by applying UWHAM to the simulation data generated from the initial and final states (i.e., “the endpoints”) instead of multi-state free energy perturbation simulations because of the possible overlaps of the configurations sampled at the endpoints. Then endpoint simulations are used to estimate the solvation free energy of water at the interface with a protein in solution. The estimate of the solvation free energy at the interface from two simulations at the endpoints agrees with the benchmark using 32 states within a 95% confidence interval for most interfacial locations. The ability to accurately estimate the excess chemical potential of water from end point simulations facilitates the statistical thermodynamic analysis of diverse interfacial phenomena. Our focus is on analyzing the excess chemical potential of water at protein receptor binding sites with the goal of using this information to assist in the design of tight binding ligands.

### Introduction

The excess chemical potential of a solvent molecule at the interface with a solute is the difference between the free energy of insertion of the solvent molecule at the interface and insertion in the bulk far from the solute. It is equivalent to the potential of mean force (pmf) to move a solvent molecule from the bulk to the interface. The excess chemical potential has

\*To whom correspondence should be addressed: ronlevy@temple.edu.

#### Supporting Information Available

Estimates of the solvation free energy for growing a water molecule in a solution containing one FXa molecule by using endpoint simulations with different amounts of data points. The Probability density of  $U_{v\bar{v}}$  for the inhomogeneous liquid. Estimates of the solvation free energy for growing a water molecule in a solution containing one FXa molecule by using three  $\gamma$ -states. Uncertainty evaluation. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

a direct part which corresponds to the potential energy of interaction of the solvent molecule with the solute, and an indirect part, a free energy, corresponding to the difference between the pmf and the direct part. Analysis of the excess chemical potential is key to understanding interfacial phenomena and to the statistical thermodynamics of solutions.<sup>1-27</sup> Recently we have shown how knowledge of the excess chemical potential of hydrating waters at the interface of protein-ligand binding sites can be used to inform the design of tighter binding ligands.<sup>28</sup> In principle the evaluation of the excess chemical potential of interfacial water molecules requires sampling over intermediate states as a tagged water molecule with fixed position and orientation is coupled into the solution. In this article we show that the excess chemical potential of water can often be estimated accurately using data from just the two endpoints of the coupling process (the pure liquid and the solution) obviating the need for simulating the intermediate states. This observation will facilitate the use of the methods described in reference [28] to analyze solvent effects on protein-ligand binding, as well as the further development of end point methods based on density functional theory<sup>7,27,29</sup> for estimating the excess chemical potential of solutes in solution. Although this paper focuses on analyzing the excess chemical potential of water at the protein-water interface, the endpoint simulations can be applied to facilitate the statistical thermodynamic analysis of diverse interfacial phenomena, which is essential to understand many chemical and biophysical phenomena such as ion channel gating, protein folding and self-assembly of membrane proteins.<sup>14,30,31</sup>; and is also relevant to applications in the energy industry including such phenomena as the transport of electrolytes through pores.<sup>32</sup>

## Methodology and Simulations

### Methodology

The potential of mean force to move a tagged water molecule from the bulk to the position  $\mathbf{x}$  in solution  $WT(\mathbf{x})$ , namely, the excess chemical potential of a water molecule at  $\mathbf{x}$ , can be estimated by the difference between the solvation free energy for growing a water molecule at  $\mathbf{x}$ ,  $F(\mathbf{x})$  and the solvation free energy for growing a water molecule in the bulk  $F(\infty)$  (or the pure liquid  $F^{(0)}$ )

$$WT(\mathbf{x}) = \Delta F(\mathbf{x}) - \Delta F(\infty) \approx \Delta F(\mathbf{x}) - \Delta F^{(0)} \quad (1)$$

Here  $\mathbf{x}$  includes the coordinates of the oxygen atom of the tagged water molecule ( $x, y, z$ ) and the orientation of the tagged water molecule ( $\alpha, \beta, \gamma$ ). The solvation free energy for growing a water molecule in solution (or the bulk) can be estimated by free energy perturbation (FEP) simulations.<sup>33-40</sup> As shown in Fig. 1,  $M$  independent parallel simulations are run for the solution with a tagged water molecule fixed at  $\mathbf{x}$ . Each simulation follows the Hamiltonian (potential) function

$$H_i(\mathbf{x}, \{\vec{x}\}) = U_{uv}(\{\vec{x}\}) + E_{u\vec{v}}(\mathbf{x}, \gamma_i) + U_{v\vec{v}}(\{\vec{x}\}) + E_{v\vec{v}}(\mathbf{x}, \{\vec{x}\}, \gamma_i), \quad (2)$$

where  $\{\vec{x}\}$  are the coordinates of the other water molecules;  $E_{uv}(\mathbf{x}, \gamma_i)$  is the soft-core interaction energy between the tagged water molecule fixed at  $\mathbf{x}$  and the solute at the  $i$ th  $\gamma$ -state; and  $E_{v\bar{v}}(\mathbf{x}, \{\vec{x}\}, \gamma_i)$  is the soft-core total interaction energy between the tagged water molecule and the other water molecules at the  $i$ th  $\gamma$ -state.  $E_{uv}(\mathbf{x}, \gamma_i)$  and  $E_{v\bar{v}}(\mathbf{x}, \{\vec{x}\}, \gamma_i)$  change from zero to  $U_{uv}(\mathbf{x})$  and  $U_{v\bar{v}}(\mathbf{x}, \{\vec{x}\})$  respectively when  $\gamma_i$  changes from zero to one, where  $U_{uv}(\mathbf{x}) \equiv E_{uv}(\mathbf{x}, 1)$  and  $U_{v\bar{v}}(\mathbf{x}, \{\vec{x}\}) \equiv E_{v\bar{v}}(\mathbf{x}, \{\vec{x}\}, 1)$  are the full interaction energies.  $U_{uv}(\{\vec{x}\})$  is the interaction energy between the solute and the other (untagged) water molecules; and  $U_{v\bar{v}}(\{\vec{x}\})$  is the total interaction energy of all the other (untagged) water molecules with each other. In this study the solute is rigid and always fixed in the solution, therefore the full interaction energy between the solute and the fixed tagged water molecule  $U_{uv}(\mathbf{x})$  is a constant.

The unbinned weighted histogram analysis method (UWHAM) is an algorithm to estimate the free energy differences and density of states from the data generated by multi-state simulations.<sup>41-44</sup> Suppose  $M$  parallel simulations in the canonical ensemble are run at  $M$  states and  $X_j$  is the  $j$ th observation. The probability of observing  $X_j$  at the  $\alpha$ th state is

$$P_\alpha(X_j) \sim \frac{q_\alpha(\{\vec{x}\}_j)}{Z_\alpha} = \frac{\exp\{-\beta_\alpha E_\alpha(\{\vec{x}\}_j)\}}{Z_\alpha}, \quad (3)$$

where  $q_\alpha(\{\vec{x}\}_j) = \exp\{-\beta_\alpha E_\alpha(\{\vec{x}\}_j)\}$  is the unnormalized probability;  $\{\vec{x}\}_j$  are the coordinates of the microstate  $X_j$ ;  $\beta_\alpha = 1/(k_B T_\alpha)$  is the inverse temperature of the  $\alpha$ th state;  $E_\alpha(\{\vec{x}\}_j)$  is the potential energy of the microstate  $X_j$  at the  $\alpha$ th state; and  $Z_\alpha$  is the partition function of the  $\alpha$ th state. The UWHAM estimate of the density of states  $\Omega(u_j)$  and the partition function  $Z_\alpha$  are obtained by solving the coupled equations:

$$\hat{Z}_\alpha = \sum_{i=1}^N q_\alpha(u_i) \hat{\Omega}(u_i) \quad (4)$$

$$\hat{\Omega}(u_i) = \frac{1}{\sum_{\kappa=1}^M N_\kappa \hat{Z}_\kappa^{-1} q_\kappa(u_i)},$$

where  $N_\kappa$  is the number of observations observed at the  $\kappa$  state;  $N = \sum_{\kappa=1}^M N_\kappa$  is the total number of observations;  $u_j$  is the reduced (energy) coordinate of the microstate  $X_j$ ; and the hat on  $\hat{Z}_\alpha$  and  $\hat{\Omega}$  denotes the most likely estimate of the true value given the discrete data set sampled from the distributions at each of the  $\gamma$ -states.

To study the statistical thermodynamics of growing a tagged water molecule in solution, we define the *effective* density of states  $\Omega(\mathbf{x}; U^{tot})$  as the total probability of observing the microstates that satisfy:

- the sum of interaction energies  $U_{u\bar{v}} + U_{v\bar{v}}$  equals  $U^{tot}$ .

at the reference state ( $\gamma = 0$  state) when the tagged water molecule is fixed at  $\mathbf{x}$ . Namely,

$$\Omega(\mathbf{x}, U^{tot}) = \int d\{\vec{x}\} \delta(U^{tot} - [U_{u\bar{v}}(\mathbf{x}) + U_{v\bar{v}}(\mathbf{x}, \{\vec{x}\})]) \exp\{-\beta H_0(\{\vec{x}\})\}, \quad (5)$$

where  $\delta(U^{tot} - [U_{u\bar{v}}(\mathbf{x}) + U_{v\bar{v}}(\mathbf{x}, \{\vec{x}\})])$  is the delta function which satisfies the identity

$$\int_{-\infty}^{\infty} dU^{tot} \delta(U^{tot} - [U_{u\bar{v}}(\mathbf{x}) + U_{v\bar{v}}(\mathbf{x}, \{\vec{x}\})]) = 1, \quad (6)$$

and  $H_0$  is the Hamiltonian function of the reference state ( $\gamma = 0$  state)

$$H_0(\{\vec{x}\}) = U_{u\bar{v}}(\{\vec{x}\}) + U_{v\bar{v}}(\{\vec{x}\}). \quad (7)$$

Here  $\gamma$  is the coupling parameter for the intermolecular interaction of the tagged water molecule with the others. The biasing potential for the  $i$ th  $\gamma$ -state is

$E_{u\bar{v}}(\mathbf{x}, \gamma_i) + E_{v\bar{v}}(\mathbf{x}, \{\vec{x}\}, \gamma_i)$ . The free energy difference between the  $\gamma = 0$  and  $\gamma = 1$  states is

$$\begin{aligned} \Delta F(\mathbf{x}) &= -k_B T \ln \left( \frac{Z_1}{Z_0} \right) \quad (8) \\ &= -k_B T \ln \frac{\int d\{\vec{x}\} \exp\{-\beta[H_1(\mathbf{x}, \{\vec{x}\}) - H_0(\{\vec{x}\})]\} \exp\{-\beta H_0(\{\vec{x}\})\}}{\int d\{\vec{x}\} \exp\{-\beta H_0(\{\vec{x}\})\}} \\ &= -k_B T \ln \frac{\int d\{\vec{x}\} \int_{-\infty}^{\infty} dU^{tot} e^{-\beta[H_1 - H_0]} \delta(U^{tot} - [U_{u\bar{v}}(\mathbf{x}) + U_{v\bar{v}}(\mathbf{x}, \{\vec{x}\})]) e^{-\beta H_0}}{\int d\{\vec{x}\} \int_{-\infty}^{\infty} dU^{tot} \delta(U^{tot} - [U_{u\bar{v}}(\mathbf{x}) + U_{v\bar{v}}(\mathbf{x}, \{\vec{x}\})]) e^{-\beta H_0}} \\ &= -k_B T \ln \frac{\int \Omega(\mathbf{x}, U^{tot}) \exp(-\beta U^{tot}) dU^{tot}}{\int \Omega(\mathbf{x}, U^{tot}) dU^{tot}}, \end{aligned}$$

where  $Z_0$  and  $Z_1$  are the partition functions of the  $\gamma = 0$  and  $\gamma = 1$  states respectively. Because both the solute and the tagged water molecule are fixed, the interaction energy between the tagged water molecule and the solute at the  $\gamma = 1$  state  $U_{u\bar{v}} \equiv \varepsilon$  is a constant. Eq. (8) can be simplified as

$$\Delta F(\mathbf{x}) = \varepsilon - k_B T \ln \frac{\int \Omega(\mathbf{x}, U_{v\bar{v}}) \exp(-\beta U_{v\bar{v}}) dU_{v\bar{v}}}{\int \Omega(\mathbf{x}, U_{v\bar{v}}) dU_{v\bar{v}}} \quad (9)$$

The UWHAM estimates of the probability of observing the reduced coordinate  $U_{v\bar{v}}$  at the  $\gamma = 0$  and  $\gamma = 1$  states are

$$\hat{P}_0(U_{v\bar{v}}) = \frac{\hat{\Omega}(U_{v\bar{v}})}{\hat{Z}_0} \quad (10)$$

and

$$\hat{P}_1(U_{v\bar{v}}) = \frac{\hat{\Omega}(U_{v\bar{v}}) \exp\{-\beta(U_{v\bar{v}} + \varepsilon)\}}{\hat{Z}_1} \quad (11)$$

respectively.

## Simulations

The simulations in this study were performed using the GROMACS 5.1.2 simulation package with the Amber99SB force field. The soft-core interactions implemented in GROMACS (see Chapter 4.5.1 of the Reference Manual Version 5.1.2 of GROMACS) were used for the free energy perturbation simulations. All the simulations were run at 300 K with constant volume and used the leap-frog (SD) integrator as the thermostat. The step size is 1 fs although the SHAKE constraint algorithm was applied. One data point was recorded every 0.1 ps. When estimating the solvation free energy for growing a water molecule in pure solvent, the Coulombic interaction and van der Waals interaction between the tagged water molecule and the other water molecules were gradually turned off together using 24  $\gamma$ -states in the FEP simulations. The chosen  $\gamma$  values are (0.0, 0.005, 0.02, 0.05, 0.075, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1.0). For the pure solvent with TIP3P water model, the system contains 1,410 water molecules in a cubic box of side 3.484 nm. Each independent simulation lasted 20 ns. For the pure solvent with the SPC/E water model, the system contains 1,000 water molecules in a cubic box of side 3.106 nm. Each independent simulation lasted 20 ns. Before estimating the solvation free energy for growing a water molecule in a solution containing one Factor Xa (FXa) molecule,<sup>51</sup> the chain A of FXa molecule and the coordinated calcium ion (PDB: 1MQ5) were solvated in a cubic box of side 8.029 nm with 15,951 water molecules. To neutralize the system, three chloride ions have been added to the system. In the FEP simulations, the Coulombic interactions between the tagged water molecule and the other molecules were turned off first with 12  $\gamma_e$  values: (0.0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0). Then the van der Waals interactions were turned off with 21  $\gamma_v$  values: (0.0, 0.05, 0.1, 0.15,

0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1.0). Each independent simulation lasted 10 ns except the one at the  $\gamma = 0$  state, which has been kept running for 100 ns so that the simulation data can also be used for endpoint calculations.

## Results and Discussion

By setting  $U_{uv}(\mathbf{x})$  and  $U_{uv}(\{\vec{x}\})$  to zero, Eq.(2)–(9) can also be applied to the homogeneous solution to estimate the solvation free energy for growing a solvent molecule in pure solvent  $F^{(0)}$ , which is required for estimating the excess chemical potential of water in inhomogeneous solutions (see Eq.(1)). To estimate the solvation free energy of inserting (growing) a water molecule in a box of pure solvent  $F^{(0)}$ , we ran simulations of pure water with a tagged water molecule at a fixed location at each of the 24  $\gamma$ -states, and each simulation lasted 20 ns. (Note that fixing the tagged water molecule is not necessary in this calculation because the tagged water molecule is in a homogeneous environment). Then the simulation data were UWHAMed to obtain the free energy difference between the  $\gamma = 0$  and  $\gamma = 1$  states. Since a water molecule is relatively small compared with the simulation box, water-size cavities are observed around position  $\mathbf{x}$  during the independent simulation at the  $\gamma = 0$  state because of the fluctuations of water density.<sup>23,45–48</sup> In those configurations, the virtual water molecule (the tagged water molecule at the  $\gamma = 0$  state) does not overlap with any other water molecules. Therefore, the probabilities of observing those configurations at the  $\gamma = 1$  state are non-trivial. In other words, the configurations generated at the  $\gamma = 0$  and  $\gamma = 1$  states (endpoints) have some (small) overlap without introducing any intermediate  $\gamma$ -states. Because of this, the UWHAM estimate of the solvation free energy calculated from two independent simulations at the endpoints is a close approximation compared with the UWHAM estimate calculated from the 24 independent simulations when the tagged water molecule is progressively inserted into the homogeneous solution. Note that UWHAM (or MBAR)<sup>49</sup> is equivalent to the Bennett acceptance ratio method (BAR) when there are only two  $\gamma$ -states in the system. And the UWHAM equations are the asymptotically unbiased estimators of the equilibrium distributions and free energy differences.<sup>41</sup> As shown in Table. 1, for the TIP3P water model, the solvation free energy for growing a tagged water molecule in pure solvent estimated from FEP simulations using 24  $\gamma$ -states is  $-6.18 \pm 0.02$  kcal/mol. The estimate is  $-6.03 \pm 0.19$  kcal/mol when only the data from the  $\gamma = 0$  and  $\gamma = 1$  states are UWHAMed. When using the SPC/E water model, the difference between these two estimates is also as small as 0.16 kcal/mol.

In Fig. 2, we plot the probability densities of the interaction energy between the tagged water molecule and the other water molecules  $U_{v\bar{v}}$  in pure solvent at the  $\gamma = 0$  and  $\gamma = 1$  states when the TIP3P water model is used. The histograms in Fig. 2 are the UWHAM estimates based on the 24 independent simulations according to Eq.(10) and (11). As shown in these pictures,  $U_{v\bar{v}}$  at the  $\gamma = 1$  state has visible probability density in the region from  $-35$  kcal/mol to  $-5$  kcal/mol, and the probability density peaks at a value  $\sim -20$  kcal/mol. In contrast, the probability density of  $U_{v\bar{v}}$  at the  $\gamma = 0$  state spreads over a much boarder region with a peak at a value  $\sim 10$  kcal/mol and a long decreasing tail on the positive side. The insert in Fig. 2b shows the overlap of the probability density of  $U_{v\bar{v}}$  at the  $\gamma = 0$  and the  $\gamma =$

1 states. The overlap is critical for UWHAM to converge and obtain the free energy difference between these two endpoints.

Fig. 3a shows the estimates of the density of states  $\Omega^{(0)}(u_{v\bar{v}})$  of the TIP3P water model obtained by the UWHAM procedure using the data from 24  $\gamma$ -states compared with the corresponding result obtained from the two endpoint simulations. The two estimates are in very good agreement. Notice that the right side of the density of states histogram beyond  $-5$  *kcal/mol*. Compared with Fig. 2a, the microstates in these bins have very small probabilities to be observed at the  $\gamma = 1$  state, but they constitute the large majority of states in the histogram of the density of states. The last data point on the right side of the density of states shown in Fig. 3a, which includes all the microstates for which the interaction energy between the tagged water molecule and the other water molecules  $u_{v\bar{v}}$  is larger than 50 *kcal/mol*, has total probability as large as 93.8%.

As discussed above, when we decrease the number of  $\gamma$ -states from 24 to 2, the estimate of the solvation free energy for growing a tagged water molecule in pure water  $F^{(0)}$  only changes from  $-6.18$  *kcal/mol* to  $-6.03$  *kcal/mol* for the TIP3P water model. However, the estimate of  $F^{(0)}$  becomes significantly worse when using only the data from one  $\gamma$ -state and Zwanzig's free energy perturbation equation<sup>33</sup>

$$\Delta F(A \rightarrow B) = F_B - F_A = -k_B T \ln \left\langle \exp \left( -\frac{E_B - E_A}{k_B T} \right) \right\rangle_A \quad (12)$$

In Eq.(12),  $E_A$  and  $E_B$  denote the potential energies of a configuration evaluated by the Hamiltonian function of state A and state B respectively. The angle brackets denote an ensemble average. As shown in Table. 1, we found  $F^{(0)} = -3.3$  *kcal/mol* by only using the simulation data obtained at the  $\gamma = 0$  state, and  $F^{(0)} = -9.8$  *kcal/mol* by only using the data obtained at the  $\gamma = 1$  state. This phenomenon is well known.<sup>50</sup>

It is easy to understand this phenomenon by rewriting Eq.(9) in its discrete form

$$\Delta F^{(0)} = -k_B T \ln \frac{\sum_{i=1}^{N_0+N_1} \Omega^{(0)}(U_{v\bar{v}}[i]) \exp(-\beta U_{v\bar{v}}[i])}{\sum_{i=1}^{N_0+N_1} \Omega^{(0)}(U_{v\bar{v}}[i])}, \quad (13)$$

and supposing there are  $N_0$  and  $N_1$  observations observed at the  $\gamma = 0$  and  $\gamma = 1$  states respectively. Note that the probability density of the numerator is plotted in Fig. 2a. Therefore, the data points obtained from the simulation at the  $\gamma = 1$  state are the major contribution of the numerator in Eq. (13). The data in the first and second group together determine the normalization, namely, the factor  $\Omega^{(0)}(U_{v\bar{v}}[i]) / \sum_{i=1}^{N_0+N_1} \Omega^{(0)}(U_{v\bar{v}}[i])$ . According to Eq.(10), we used the probability density of  $U_{v\bar{v}}$  obtained from the independent

simulation at the  $\gamma = 0$  state  $P_0(U_{v\bar{v}})$  to estimate the density of state  $\Omega^{(0)}(u_{v\bar{v}})$ , and the results are plotted in Fig. 3b. As can be seen, the density of states estimated from the  $\gamma = 0$  state agrees with the density of states estimated from 24 independent simulations, but no data were sampled where the interaction energy  $U_{v\bar{v}}$  is smaller than  $-10 \text{ kcal/mol}$  so the density of states is estimated to be zero in this region. According to Eq.(11), we reweighted the probability density of  $U_{v\bar{v}}$  obtained from the independent simulation at the  $\gamma = 1$  state  $P_1(U_{v\bar{v}})$  to estimate the density of states  $\Omega^{(0)}(u_{v\bar{v}})$ , and the results are plotted in Fig. 3b. As can be seen, the density of states estimated from the  $\gamma = 1$  states has data points where the interaction energy  $U_{v\bar{v}}$  is smaller than  $-10 \text{ kcal/mol}$ , but it does not agree with the density of states estimated from 24  $\gamma$ -states because of the incorrect normalization. The shaded area in Fig. 3 shows the overlap of the density of states estimated from the  $\gamma = 0$  and the density of states estimated from the  $\gamma = 1$  state.

Next we have applied endpoint simulations to estimate the solvation free energy for growing a water molecule in a solution contains one protein molecule Factor Xa (FXa)<sup>51</sup> and 15, 951 TIP3P water molecules. Unlike in pure solvent, the water density in inhomogeneous liquids depends on the position and orientation of the water molecule  $\mathbf{x}$ :

$$\rho(\mathbf{x}) = \rho(\infty)\exp\left\{-\frac{WT(\mathbf{x})}{k_B T}\right\} = \rho(\infty)\exp\left\{-\frac{\Delta F(\mathbf{x}) - \Delta F(\infty)}{k_B T}\right\}, \quad (14)$$

where  $\rho(\infty)$  is the water density in the bulk. As explained previously, water-size cavities generated by water density fluctuations around  $\mathbf{x}$  at the  $\gamma = 0$  state are essential for the overlap of the configurations observed at the  $\gamma = 0$  and  $\gamma = 1$  states. For the favorable positions ( $F(\mathbf{x}) < F(\infty)$ ) in solution, it usually requires longer simulations at the  $\gamma = 0$  state than for the unfavorable positions ( $F(\mathbf{x}) > F(\infty)$ ) to obtain converged estimates of  $F(\mathbf{x})$  by endpoint simulations.

Fig. 4 shows the dependence of the average number of water molecules on the direct interaction between the water molecule and FXa. The histogram is constructed based on the 500 snapshots of a 5 ns MD simulation. As can be seen, the direct interaction  $\epsilon$  ranges from  $\sim -25 \text{ kcal/mol}$  to  $\sim +10 \text{ kcal/mol}$ . We chose 21 positions at the interface from the MD trajectory to test endpoint calculations of the solvation free energy of water molecules at those locations. The direct interactions between the tagged water molecules and the solute at those positions are approximately evenly distributed between the minimum and the maximum values observed. To obtain benchmarks with better precision, we ran FEP simulations using 32  $\gamma$ -states for each position. Each independent simulation lasted 10 ns except the one at the  $\gamma = 0$  state which lasted 100 ns. For each chosen position, a total number of 4, 100, 000 data points were UWHAMed to obtain the free energy difference between the  $\gamma = 0$  and  $\gamma = 1$  states as the benchmark using all 32  $\gamma$ -states. The data generated at the  $\gamma = 0$  and  $\gamma = 1$  states were used for the endpoint calculations. We also measured the overlap between the  $\gamma = 0$  and  $\gamma = 1$  states based on the endpoint calculations. The overlap is defined as



$$S = \sum_{i=1}^N \min \{w_0(u_i), w_1(u_i)\}, \quad (15)$$

where  $N = 1, 100, 000$  is the total number of data points observed at the  $\gamma = 0$  and  $\gamma = 1$  states;  $u_i$  is the  $i$ th data point;  $w_0(u_i)$  and  $w_1(u_i)$  are the normalized UWHAM weight of the observation  $u_i$  at the  $\gamma = 0$  and  $\gamma = 1$  state respectively. The summation in Eq.(15) is a quantification of the overlap between the probability density of observing each observation at the  $\gamma = 0$  and the  $\gamma = 1$  states (see the insert of Fig. 2b).<sup>52,53</sup>

The 32-states benchmark, the overlap of the data ensembles at the endpoints, and the results of endpoint calculations are shown in Table. 2. The positions shown with white background in Table. 2 have relatively large overlap (about two orders larger than  $1.0e - 6$ ) between the data ensembles at the  $\gamma = 0$  and  $\gamma = 1$  states. For these positions, the solvation free energy of the tagged water molecule can be estimated by only using the first 10 ns data points at the  $\gamma = 0$  and  $\gamma = 1$  states (see Supporting Information for the results). The positions shown with gray background in Table. 2 have relatively small overlap (about the same order as  $1.0e - 6$ ). The data set which includes 100 ns of sampling at the  $\gamma = 0$  states and 10 ns of sampling at the  $\gamma = 1$  state are required to estimate the solvation free energy of growing a water molecule at these positions. As can be seen, for positions in both categories, the UWHAM estimates of the solvation free energy of the tagged water molecule agree with the benchmark within 95% confidence interval. However, even UWHAMing the 10 ns data points at the  $\gamma = 1$  state and the 100 ns data points at the  $\gamma = 0$  state is insufficient sampling to estimate the solvation free energy for the two positions (#4 and #20). For these two challenging positions, we found that UWHAM analysis converges after including the data of one intermediate  $\gamma$ -state (see Supporting Information for details). Note that these 21 positions are chosen so that their direct interactions  $\epsilon$  are approximately evenly distributed between the minimum and the maximum values observed. They are not chosen randomly from the frames of the 5 ns long MD trajectory. Table. 3 shows the estimates of the excess chemical potential  $WT$  and the indirect part  $\omega = (WT - \epsilon)$  based on the solvation free energies estimated by end point calculations.

The direct and indirect parts of the excess chemical potential of water carry the thermodynamic signatures characteristic of hydrophobic and hydrophilic hydration; this information can be used in a semi-quantitative way to assist in the process of designing tighter binding ligands to proteins. Two of the interfacial water locations listed in Table. 2 and 3 (waters #11 and #13) are located at the active binding site of Factor Xa. We recently reported the results of an analysis of the thermodynamic signatures of these and other active site waters (see Eq.[A11] in the appendix of reference [28]). While waters #11 and #13 both have quite favorable interaction energies with the protein receptor site which are hydrophilic in nature, for water #13 the indirect solvent-solvent contribution to the PMF almost completely cancels the direct interaction, and therefore the density is close to the bulk at this location. The consequences of this for ligand design are discussed in reference [28]. The remaining waters listed in Table. 2 and 3 are at the interface with Factor Xa, but not at the

receptor binding site. We expect that many features of the thermodynamic signatures of interfacial waters shown in Table. 2 and 3 are not specific to Factor Xa, but are general features which characterize the protein-water interface; including the range of direct interaction energies, both the magnitude and sign, and the extent to which the indirect term, partially or completely cancels the direct contribution (e.g. waters # 13, #14, and #18 have direct and indirect terms which almost completely cancel). It is interesting to note that waters are hydrating charged residues at both the most favorable ( $\epsilon = -22 \text{ kcal/mol}$ ) and least favorable ( $\epsilon = +8.9 \text{ kcal/mol}$ ) ends of the distribution. At the most favorable end of the distribution of direct interaction energies (e.g. water #1), the thermodynamic signature of the water is characteristic of a high density solvent region in proximity to a charged site on a solute. At the most repulsive end of the distribution of direct interaction energies (e.g. waters #18 - #21), these waters are in proximity to a pair of charged residues or a very polar and charged residue. The indirect contribution to the pmf is very favorable, while the direct interaction is dominated by a short range electrostatic repulsion. Water #21 as an example, interacts unfavorably with both the negatively charged Glu 76 and Glu 80. The solvation characteristics of these waters are interesting in that they are not typical of either hydrophilic or hydrophobic hydration; they appear to correspond to locations which bridge the hydration shells of two charged residues or a charged and polar residue. We will provide a more detailed analysis of the thermodynamic signatures of waters at the protein interface in a future communication.

## Conclusion

In summary, this article introduced the use of endpoint simulations together with UWHAM to estimate the excess chemical potential of a tagged water molecule  $WT(\mathbf{x})$ , in the pure liquid and at the protein-water interface. First, we reviewed how UWHAM is used to estimate the solvation free energy of a tagged water molecule when it is coupled into the solution by simulating a series of intermediate states explicitly. Next using the homogeneous solution (pure water) as an example, we showed that endpoint simulations can be used to obtain the solvation free energy of a tagged water molecule without the need to simulate the intermediate states explicitly. This is possible because the relatively small size of a water molecule facilitates overlaps in the phase space between the simulations at the  $\gamma = 0$  and  $\gamma = 1$  states. Then we showed that endpoint simulations can be used to estimate the excess chemical potential of solvating waters at the interface of a protein. The solute we chose as an example, the protein FXa, was one we recently used<sup>28</sup> to illustrate how knowledge of the excess chemical potential of interfacial waters can be used to help design tighter binding ligands. We found that for most of the interfacial water locations, the solvation free energies and excess chemical potentials estimated based on a 10 ns simulation at the  $\gamma = 1$  state and a 100 ns simulation at the  $\gamma = 0$  state agree with the 32-states benchmark within the 95% confidence interval. For two locations, it was challenging to obtain the solvation free energy of a tagged water molecule just from the endpoint simulations alone. For those positions, accurate estimates can be obtained by inserting one intermediate  $\gamma$ -state.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

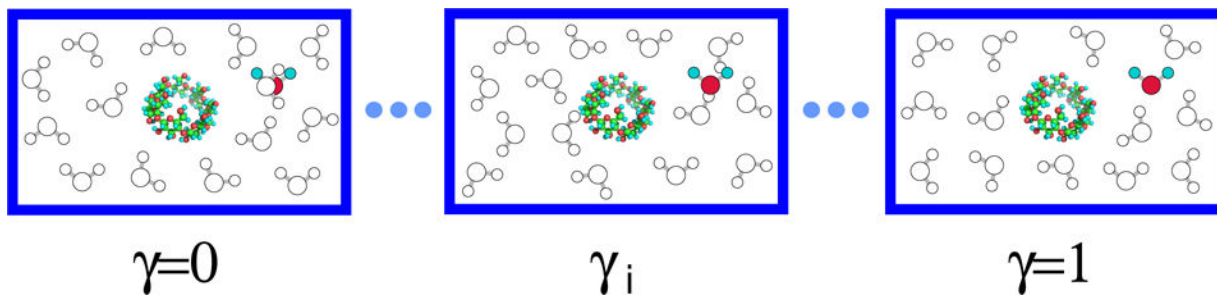
This work was supported by NIH grant (GM30580), NSF grant (1665032), and by an NIH computer equipment grant OD020095. This work also used Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation (ACI-1053575). This work is also supported by the Grants-in-Aid for Scientific Research (Nos. JP15K13550 and JP26240045) from the Japan Society for the Promotion of Science and by the Elements Strategy Initiative for Catalysts and Batteries and the Post-K Supercomputing Project from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

## References

1. Widom B. Some Topics in the Theory of Fluids. *J Chem Phys.* 1963; 39:2808–2812.
2. Widom B. Potential-distribution theory and the statistical mechanics of fluids. *J Phys Chem.* 1982; 86:869–872.
3. Lee SH, Rossky PJ. A comparison of the structure and dynamics of liquid water at hydrophobic and hydrophilic surfaces—a molecular dynamics simulation study. *J Chem Phys.* 1994; 100:3334–3345.
4. Hummer G, Szabo A. Calculation of free-energy differences from computer simulations of initial and final states. *J Chem Phys.* 1996; 105:2004–2010.
5. Cheng Y-K, Rossky PJ. Surface topography dependence of biomolecular hydrophobic hydration. *Nature.* 1998; 392:696–699. [PubMed: 9565030]
6. Hummer G, Garde S, García AE, Pratt LR. New perspectives on hydrophobic effects. *Chem Phys.* 2000; 258:349–370.
7. Matubayasi N, Nakahara M. Theory of solutions in the energetic representation. I. Formulation. *J Chem Phys.* 2000; 113:6070.
8. Gallicchio E, Kubo MM, Levy RM. Enthalpy-Entropy and Cavity Decomposition of Alkane Hydration Free Energies: Numerical Results and Implications for Theories of Hydrophobic Solvation. *J Phys Chem B.* 2000; 104:6271–6285.
9. Wallqvist A, Gallicchio E, Levy RM. A Model for Studying Drying at Hydrophobic Interfaces: Structural and Thermodynamic Properties. *J Phys Chem B.* 2001; 105:6745–6753.
10. Pratt LR. Molecular theory of hydrophobic effects: “She is too mean to have her name repeated”. *Annu Rev Phys Chem.* 2002; 53:409–436. [PubMed: 11972014]
11. Pratt LR, Pohorille A. Hydrophobic effects and modeling of biophysical aqueous solution interfaces. *Chem Rev.* 2002; 102:2671–2692. [PubMed: 12175264]
12. Zhou R, Huang X, Margulis CJ, Berne BJ. Hydrophobic collapse in multidomain protein folding. *Science.* 2004; 305:1605–1609. [PubMed: 15361621]
13. Li Z, Lazaridis T. The Effect of Water Displacement on Binding Thermodynamics: Concanavalin A. *J Phys Chem B.* 2005; 109:662–670. [PubMed: 16851059]
14. Chandler D. Interfaces and the driving force of hydrophobic assembly. *Nature.* 2005; 437:640–647. [PubMed: 16193038]
15. Ben-Amotz D, Raineri FO, Stell G. Solvation Thermodynamics: Theory and Applications. *J Phys Chem B.* 2005; 109:6866–6878. [PubMed: 16851773]
16. Li Z, Lazaridis T. Thermodynamics of buried water clusters at a protein-ligand binding interface. *J Phys Chem B.* 2006; 110:1464–1475. [PubMed: 16471698]
17. Beck, TL., Paulaitis, ME., Pratt, LR. *The Potential Distribution Theorem and Models of Molecular Solutions.* Cambridge University Press; New York: 2006.
18. Li Z, Lazaridis T. Water at biomolecular binding interfaces. *Phys Chem Chem Phys.* 2007; 9:573–581. [PubMed: 17242738]
19. Shah JK, Asthagiri D, Pratt LR, Paulaitis ME. Balancing local order and long-ranged interactions in the molecular theory of liquid water. *J Chem Phys.* 2007; 127:144508. [PubMed: 17935410]
20. Giovambattista N, Lopez CF, Rossky PJ, Debenedetti PG. Hydrophobicity of protein surfaces: Separating geometry from chemistry. *Proc Natl Acad Sci U S A.* 2008; 105:2274–2279. [PubMed: 18268339]
21. Berne BJ, Weeks JD, Zhou R. Dewetting and hydrophobic interaction in physical and biological systems. *Annu Rev Phys Chem.* 2009; 60:85–103. [PubMed: 18928403]

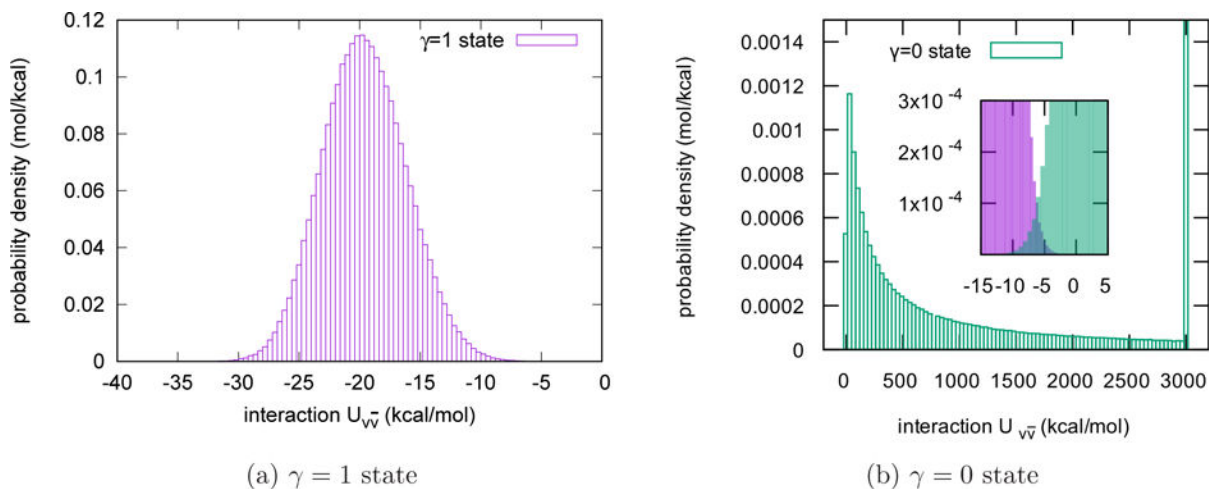
22. Chempath S, Pratt LR. Distribution of binding energies of a water molecule in the water liquid-vapor interface. *J Phys Chem B*. 2009; 113:4147–4151. [PubMed: 19006274]
23. Jamadagni SN, Godawat R, Garde S. Hydrophobicity of proteins and interfaces: insights from density fluctuations. *Annu Rev Chem Biomol Eng*. 2011; 2:147–171. [PubMed: 22432614]
24. Patel AJ, Varilly P, Jamadagni SN, Hagan MF, Chandler D, Garde S. Sitting at the Edge: How Biomolecules use Hydrophobicity to Tune Their Interactions and Function. *J Phys Chem B*. 2012; 116:2498–2503. [PubMed: 22235927]
25. Davis JG, Gierszal KP, Wang P, Ben-Amotz D. Water structural transformation at molecular hydrophobic interfaces. *Nature*. 2012; 491:582–585. [PubMed: 23172216]
26. Remsing RC, Xi E, Vembanur S, Sharma S, Debenedetti PG, Garde S, Patel AJ. Pathways to dewetting in hydrophobic confinement. *Proc Natl Acad Sci U S A*. 2015; 112:8181–8186. [PubMed: 26100866]
27. Levy RM, Cui D, Zhang BW, Matubayasi N. Relationship between Solvation Thermodynamics from IST and DFT Perspectives. *J Phys Chem B*. 2017; 121:3825–3841. [PubMed: 28186751]
28. Cui D, Zhang BW, Matubayasi N, Levy RM. The Role of Interfacial Water in Protein-Ligand Binding: Insights from the Indirect Solvent Mediated PMF. *J Chem Theory Comput*. 2017; 14:512–526.
29. Sakuraba S, Matubayasi N. ERmod: Fast and versatile computation software for solvation free energy with approximate theory of solutions. *J Comput Chem*. 2014; 35:1592–1608. [PubMed: 24923817]
30. Tanford C. How protein chemists learned about the hydrophobic factor. *Protein Sci*. 1997; 6:1358–1366. [PubMed: 9194199]
31. Rasaiah JC, Garde S, Hummer G. Water in Nonpolar Confinement: From Nanotubes to Proteins and Beyond. *Annu Rev Phys Chem*. 2008; 59:713–740. [PubMed: 18092942]
32. Striolo A. Interfacial water studies and their relevance for the energy sector. *Mol Phys*. 2016; 114:2615–2626.
33. Zwanzig RW. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J Chem Phys*. 1954; 22:1420.
34. Bennett CH. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J Comput Phys*. 1976; 22:245–268.
35. Shing K, Gubbins K. The chemical potential in dense fluids and fluid mixtures via computer simulation. *Mol Phys*. 1982; 46:1109–1128.
36. Sakuraba S, Matubayasi N. Distribution-function approach to free energy computation. *J Chem Phys*. 2011; 135:114108. [PubMed: 21950851]
37. Zuckerman DM, Woolf TB. Systematic Finite-Sampling Inaccuracy in Free Energy Differences and Other Nonlinear Quantities. *J Stat Phys*. 2004; 114:1303–1323.
38. Ytreberg FM, Swendsen RH, Zuckerman DM. Comparison of free energy methods for molecular systems. *J Chem Phys*. 2006; 125:184114. [PubMed: 17115745]
39. Chipot, C., Pohorille, A., editors. *Free Energy Calculations: Theory and Applications in Chemistry and Biology*. Springer; Berlin Heidelberg: 2007.
40. Klimovich PV, Shirts MR, Mobley DL. Guidelines for the analysis of free energy calculations. *J Comput-Aided Mol Des*. 2015; 29:397–411. [PubMed: 25808134]
41. Tan Z, Gallicchio E, Lapelosa M, Levy RM. Theory of Binless Multi-State Free Energy Estimation with Applications to Protein-Ligand Binding. *J Chem Phys*. 2012; 136:144102. [PubMed: 22502496]
42. Zhang BW, Xia J, Tan Z, Levy RM. A Stochastic Solution to the Unbinned WHAM Equations. *J Phys Chem Lett*. 2015; 6:3834–3840. [PubMed: 26722879]
43. Tan Z, Xia J, Zhang BW, Levy RM. Locally Weighted Histogram Analysis and Stochastic Solution for Large-Scale Multi-State Free Energy Estimation. *J Chem Phys*. 2016; 144:034107. [PubMed: 26801020]
44. Zhang BW, Deng N, Tan Z, Levy RM. Stratified UWHAM and Its Stochastic Approximation for Multicanonical Simulations Which Are Far from Equilibrium. *J Chem Theory Comput*. 2017; 13:4660–4674. [PubMed: 28902500]

45. Pohorille A, Pratt LR. Cavities in molecular liquids and the theory of hydrophobic solubilities. *J Am Chem Soc.* 1990; 112:5066–5074. [PubMed: 11540917]
46. Pratt LR, Pohorille A. Theory of hydrophobicity: transient cavities in molecular liquids. *Proc Natl Acad Sci U S A.* 1992; 89:2995–2999. [PubMed: 11537863]
47. Patel AJ, Varilly P, Chandler D. Fluctuations of Water near Extended Hydrophobic and Hydrophilic Surfaces. *J Phys Chem B.* 2010; 114:1632–1637. [PubMed: 20058869]
48. Cui D, Ou S, Peters E, Patel S. Ion-Specific Induced Fluctuations and Free Energetics of Aqueous Protein Hydrophobic Interfaces: Toward Connecting to Specific-Ion Behaviors at Aqueous Liquid–Vapor Interfaces. *J Phys Chem B.* 2014; 118:4490–4504. [PubMed: 24701961]
49. Shirts MR, Chodera JD. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J Chem Phys.* 2008; 129:124105. [PubMed: 19045004]
50. Lu N, Singh JK, Kofke DA. Appropriate methods to combine forward and reverse free-energy perturbation averages. *J Chem Phys.* 2003; 118:2977.
51. Turpie AG. Oral, Direct Factor Xa Inhibitors in Development for the Prevention and Treatment of Thromboembolic Diseases. *Arterioscler Thromb Vasc Biol.* 2007; 27:1238–1247. [PubMed: 17379841]
52. Wu D, Kofke DA. Phase-Space Overlap Measures. I. Fail-Safe Bias Detection in Free Energies Calculated by Molecular Simulation. *J Chem Phys.* 2005; 123:054103. [PubMed: 16108627]
53. Wu D, Kofke DA. Phase-Space Overlap Measures. II. Design and Implementation of Staging Methods for Free-Energy Calculations. *J Chem Phys.* 2005; 123:084109. [PubMed: 16164284]



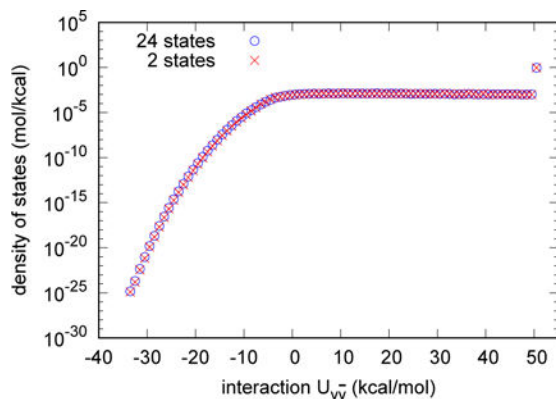
**Figure 1.**

Slow growth of a tagged water molecule in solution. During the simulation at the  $\gamma = 0$  state, the interaction energy between the tagged water molecule and the other water molecules  $E_{v\bar{v}}(\mathbf{x}, \{\vec{x}\}, \gamma_i = 0)$  and the interaction energy between the tagged water molecule and the solute  $E_{u\bar{v}}(\mathbf{x}, \gamma_i = 0)$  are both zero. Therefore, the tagged water molecule can overlap with other water molecules. When the value of  $\gamma$  increases from zero to one,  $E_{v\bar{v}}(\mathbf{x}, \{\vec{x}\}, \gamma_i)$  and  $E_{u\bar{v}}(\mathbf{x}, \gamma_i)$  increase from zero to  $U_{v\bar{v}}(\mathbf{x}, \{\vec{x}\})$  and  $U_{u\bar{v}}(\mathbf{x})$  respectively. In this study, the simulations at different  $\gamma$ -states are independent. Both the solute and the tagged water molecule are fixed in the solution.

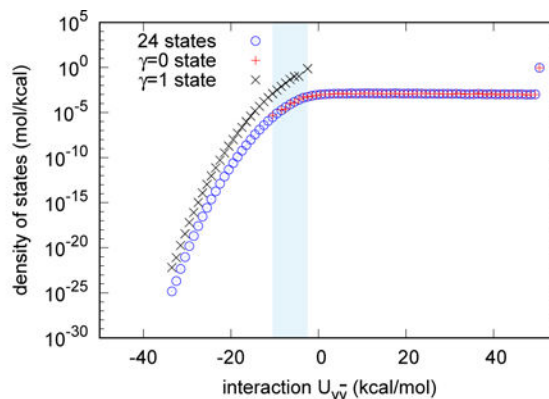
**Figure 2.**

Probability density of  $U_{v\bar{v}}$  at the  $\gamma = 0$  state and the  $\gamma = 1$  state for the homogeneous liquid.

The probability density of  $U_{v\bar{v}}$  at the  $\gamma = 0$  state spreads over a much boarder region than at the  $\gamma = 1$  state. Note that  $U_{v\bar{v}}$  at the  $\gamma = 0$  state has a long decreasing tail. The last bin on the right side of the figure (b) includes all the observations whose  $U_{v\bar{v}}$  is larger than 3000 *kcal/mol*. The total probability in the last bin is 52.0%. The insert shows the overlap of the probability density of  $U_{v\bar{v}}$  at the  $\gamma = 0$  and the  $\gamma = 1$  states.



(a) 24 states vs. 2 states

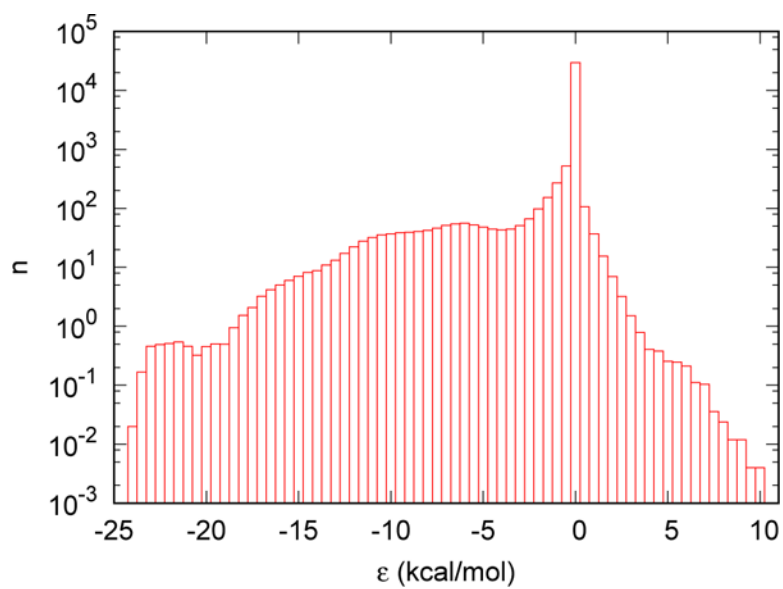


(b) 24 states vs. 1 state

**Figure 3.**

Comparison of the estimates of the density of states for the homogeneous liquid (using the TIP3P water model). In both pictures the density of states estimated from the simulations at 24  $\gamma$ -states is the benchmark. (a) The density of states estimated from the two endpoint simulations agrees with the benchmark very well. (b) The density of states estimated from the simulation at the  $\gamma=0$  state matches the benchmark for more repulsive values of  $U_{vw}$  but does not agree with the benchmark data where  $U_{vw}$  is smaller than  $-10$  kcal/mol. The density of states estimated from the simulation at the  $\gamma=1$  state does not agree with the benchmark. The shady area shows the overlap of density of states estimated from the  $\gamma=0$  and that estimated from  $\gamma=1$  states.





**Figure 4.** The average number of water molecules with direct interaction between the tagged water molecule and the solute in a solution containing 1 protein molecule (Factor Xa) and 15951 TIP3P water molecules.

**Table 1**Solvation free energy (*kcal/mol*) for growing a water molecule in pure solvent

# of states	24	2	$\gamma=0$	$\gamma=1$
$F^0(\text{TIP3P})$	-6.18±0.02	-6.03±0.19	-3.3	-9.8
$F^0(\text{SPC/E})$	-6.81±0.02	-6.97±0.14	-3.7	-11.9

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2**

Solvation free energy (*kcal/mol*) for growing a water molecule in a solution containing one protein molecule (Factor Xa) and 15951 water molecules. The data in the *F*(32 states) column are the benchmarks. The *F*(2 states) column shows the estimates when the 10 *ns* simulation data at the  $\gamma = 1$  state and the 100 *ns* simulation data at the  $\gamma = 0$  state are UWHAMed. The uncertainties marked with star are the lower limits of the uncertainty. The details of uncertainty evaluation are provided in Supporting Information.

#	$\epsilon$	<i>F</i> (32 states)	<i>F</i> (2 states)	Overlap ( $10^{-6}$ )
1	-22.68	-13.84 ± 0.09	-14.0 ± 0.5*	7.5e+00
2	-21.01	-12.7 ± 0.17*	-13.3 ± 0.4*	3.4e+00
3	-19.09	-11.0 ± 0.4*	-11.8 ± 0.4*	2.1e+00
4	-17.47	-12.7 ± 0.13*	-	~ 0
5	-16.16	-11.40 ± 0.07	-11.9 ± 0.4*	2.0e+00
6	-14.44	-10.36 ± 0.03	-10.51 ± 0.11	3.7e+02
7	-12.76	-11.43 ± 0.08	-10.7 ± 0.6*	3.2e+00
8	-11.11	-9.033 ± 0.014	-9.00 ± 0.15	9.1e+01
9	-9.63	-8.782 ± 0.016	-8.87 ± 0.12	1.4e+02
10	-7.83	-9.095 ± 0.013	-9.06 ± 0.07	4.9e+02
11	-6.40	-7.774 ± 0.017	-7.67 ± 0.13	8.1e+01
12	-5.78	-8.895 ± 0.012	-8.92 ± 0.09	1.6e+02
13	-4.74	-5.765 ± 0.011	-5.85 ± 0.08	1.3e+02
14	-4.19	-5.66 ± 0.02	-5.82 ± 0.14	8.2e+01
15	-2.08	-9.33 ± 0.06	-9.26 ± 0.09	1.2e+03
16	-0.015	-6.013 ± 0.010	-6.02 ± 0.08	2.9e+02
17	1.04	-4.731 ± 0.015	-4.61 ± 0.12	9.9e+01
18	3.14	-6.200 ± 0.013	-6.19 ± 0.13	1.6e+02
19	5.29	-4.578 ± 0.014	-4.46 ± 0.08	5.9e+02
20	7.04	-11.0 ± 0.12*	-	~ 0
21	8.94	-8.60 ± 0.14	-7.7 ± 0.6*	3.5e+00

**Table 3**

Excess chemical potential  $WT$  and indirect part  $\omega$ . The excess chemical potentials are estimated based on the end point calculations, where  $WT = F(2 \text{ states}) - (-6.18 \text{ kcal/mol})$ . The indirect part is the difference between the excess chemical potential  $WT$  and the direct part  $\epsilon$ , namely,  $\omega = (WT - \epsilon)$ .

#	$\epsilon$	$F(2 \text{ states})$	$WT$	$\omega$
1	-22.68	$-14.0 \pm 0.5^*$	-7.82	14.86
2	-21.01	$-13.3 \pm 0.4^*$	-7.12	13.89
3	-19.09	$-11.8 \pm 0.4^*$	-5.62	13.47
4	-17.47	-	-	-
5	-16.16	$-11.9 \pm 0.4^*$	-5.72	10.44
6	-14.44	$-10.51 \pm 0.11$	-4.33	10.11
7	-12.76	$-10.7 \pm 0.6^*$	-4.52	8.24
8	-11.11	$-9.00 \pm 0.15$	-2.82	8.29
9	-9.63	$-8.87 \pm 0.12$	-2.69	6.94
10	-7.83	$-9.06 \pm 0.07$	-2.88	4.95
11	-6.40	$-7.67 \pm 0.13$	-1.49	4.91
12	-5.78	$-8.92 \pm 0.09$	-2.74	3.04
13	-4.74	$-5.85 \pm 0.08$	0.33	5.07
14	-4.19	$-5.82 \pm 0.14$	0.36	4.55
15	-2.08	$-9.26 \pm 0.09$	-3.08	-1.00
16	-0.015	$-6.02 \pm 0.08$	0.16	0.18
17	1.04	$-4.61 \pm 0.12$	1.57	0.53
18	3.14	$-6.19 \pm 0.13$	-0.01	-3.15
19	5.29	$-4.46 \pm 0.08$	1.72	-3.57
20	7.04	-	-	-
21	8.94	$-7.7 \pm 0.6^*$	-1.52	-10.46