

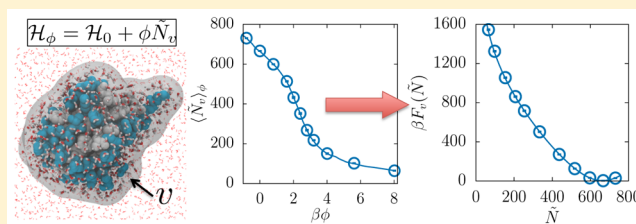
# Sparse Sampling of Water Density Fluctuations in Interfacial Environments

Erte Xi, Richard C. Remsing, and Amish J. Patel\*

Department of Chemical and Biomolecular Engineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104, United States

**S** Supporting Information

**ABSTRACT:** The free energetics of water density fluctuations near a surface, and the rare low-density fluctuations in particular, serve as reliable indicators of surface hydrophobicity; the easier it is to displace the interfacial waters, the more hydrophobic the underlying surface is. However, characterizing the free energetics of such rare fluctuations requires computationally expensive, non-Boltzmann sampling methods like umbrella sampling. This inherent computational expense associated with umbrella sampling makes it challenging to investigate the role of polarizability or electronic structure effects in influencing interfacial fluctuations. Importantly, it also limits the size of the volume, which can be used to probe interfacial fluctuations. The latter can be particularly important in characterizing the hydrophobicity of large surfaces with molecular-level heterogeneities, such as those presented by proteins. To overcome these challenges, here we present a method for the sparse sampling of water density fluctuations, which is roughly 2 orders of magnitude more efficient than umbrella sampling. We employ thermodynamic integration to estimate the free energy differences between biased ensembles, thereby circumventing the umbrella sampling requirement of overlap between adjacent biased distributions. Further, a judicious choice of the biasing potential allows such free energy differences to be estimated using short simulations, so that the free energetics of water density fluctuations are obtained using only a few, short simulations. Leveraging the efficiency of the method, we characterize water density fluctuations in the entire hydration shell of the protein, ubiquitin, a large volume containing an average of more than 600 waters.



## 1. INTRODUCTION

An understanding of density fluctuations in bulk water and at interfaces has played a central role in the description of hydrophobic effects,<sup>1–21</sup> which drive biomolecular<sup>22–27</sup> and other aqueous assemblies.<sup>28–32</sup> For example, the fact that  $P_\nu(N)$ , the probability of observing  $N$  water molecules in a small observation volume ( $\nu \lesssim 1 \text{ nm}^3$ ), obeys Gaussian statistics<sup>3,33</sup> has provided molecular-level insights into the pressure-induced denaturation of proteins,<sup>34–36</sup> as well as the convergence of protein unfolding entropies at a particular temperature.<sup>20,33</sup> Similarly, the Lum–Chandler–Weeks (LCW) theory<sup>5</sup> prediction that  $P_\nu(N)$  should develop fat low- $N$  tails for large volumes ( $\nu \gtrsim 1 \text{ nm}^3$ ) in bulk water and near hydrophobic surfaces<sup>5,15,37</sup> and its verification by simulations<sup>13,14,19,38,39</sup> has clarified that water near hydrophobic surfaces is sensitive to perturbations<sup>19,40–45</sup> and led to the insight that extended hydrophobic surfaces could generically serve as catalysts for the assembly and disassembly of small hydrophobic solutes.<sup>14,46</sup>

Importantly, the fact that low- $N$  fluctuations are enhanced near hydrophobic surfaces also makes them a suitable metric for quantifying the hydrophobicity of a surface or the strength of its interactions with water.<sup>13,16,19,21</sup> Fluctuations are particularly useful in characterizing the hydrophobicity of complex surfaces with molecular-level heterogeneities, wherein conventional macroscopic measures, such as the water droplet contact angle, break down.<sup>47</sup> Indeed, for the rugged, heterogeneous

surfaces of proteins, hydrophobicity depends not only on the chemistry of the underlying residues,<sup>12,19,21,48</sup> but also on the particular topography<sup>49–51</sup> and chemical pattern<sup>12,52–55</sup> presented by the protein and can depend nontrivially on the specific combination of the two.<sup>41,42,56–58</sup> Fluctuations have previously been used to characterize the hydrophobicity of such complex surfaces using small (e.g., methane-sized)  $\nu$ ,<sup>12,48</sup> wherein the fluctuations of interest can be estimated using equilibrium molecular simulations.<sup>3,59</sup> However, the likelihood of low- $N$  fluctuations decreases roughly exponentially with the size of  $\nu$ , and estimating  $P_\nu(N)$  for larger  $\nu$  requires computationally expensive biased sampling techniques, such as umbrella sampling.<sup>13,60,61</sup> In particular, the larger the volume of interest is, the larger the number of biased simulations required. As a result, using large volumes, such as the hydration shells of entire proteins, to characterize surface hydrophobicity would be very expensive. Further, using umbrella sampling in conjunction with more computationally expensive treatments, which incorporate the polarizability of interfacial waters or electronic structure effects,<sup>62,63</sup> to characterize density fluctuations in large volumes would also be prohibitively expensive.

Here, we present a method that enables estimation of  $P_\nu(N)$  at a number of sparsely separated  $N$ -values, albeit with a

Received: November 2, 2015

Published: January 8, 2016

roughly 2 orders of magnitude increase in computational efficiency, as compared with conventional techniques, such as umbrella sampling or free energy perturbation. To facilitate sparse sampling, we employ thermodynamic integration to estimate free energy differences between biased ensembles, thereby circumventing the umbrella sampling requirement that adjacent biased distributions overlap. Building upon recent work by Patel and Garde,<sup>21</sup> we employ a linear biasing potential, which enables efficient estimation of free energy differences between biased ensembles using averages that converge rapidly and require only short simulations. In the following section, we first derive the central equations underlying our method, followed by details pertaining to the systems studied and the simulations employed. In the subsequent section, we demonstrate how the method works using a small volume in bulk water. We then apply the method to characterize the hydrophobicity of interfaces: first, uniform self-assembled monolayer surfaces, and then the entire hydration shell of the protein, ubiquitin. We then discuss the underpinnings of the method's efficiency and its limitations and conclude with a discussion of scenarios where the method may find broader applicability.

## 2. THEORY AND METHODS

**2.1. Derivation of Central Equations.** Consider an observation volume,  $v$ , described by its size, shape, and location. We are interested in characterizing the statistics of water density fluctuations in  $v$ , as quantified by the probability,  $P_v(N) \equiv \langle \delta(N_v - N) \rangle_0$ , of observing  $N$  water molecules in  $v$ . Here,  $\langle \dots \rangle_0$  corresponds to an average in the presence of a generalized Hamiltonian,  $\mathcal{H}_0$ , and the operator,  $N_v(\{\mathbf{r}_i\})$ , depends on the positions,  $\{\mathbf{r}_i\}$ , of all the water oxygens. To circumvent issues related to the sampling of the discrete operator,  $N_v(\{\mathbf{r}_i\})$ , using molecular dynamics (MD) simulations, here we will focus on quantifying the closely related probability,  $P_v(\tilde{N}) \equiv \langle \delta(\tilde{N}_v - \tilde{N}) \rangle_0$ , of observing  $\tilde{N}$  coarse-grained water molecules in  $v$ . The operator,  $\tilde{N}_v(\{\mathbf{r}_i\})$ , is chosen to be strongly correlated with the (discrete) operator,  $N_v(\{\mathbf{r}_i\})$ , but is a continuous function of  $\{\mathbf{r}_i\}$ , so that  $\tilde{N}_v$ -dependent biasing potentials do not give rise to impulsive forces in MD simulations. The functional forms that we employ for  $\tilde{N}_v(\{\mathbf{r}_i\})$ , can be found in ref 61. Importantly, as shown in the [Supporting Information](#), in addition to enabling sparse sampling of  $P_v(\tilde{N})$ , the method we present below can also be used to obtain  $P_v(N)$ .

To estimate  $P_v(\tilde{N})$  over the entire range of  $\tilde{N}$ -values, including those that are highly improbable, standard approaches such as umbrella sampling prescribe the application of a biasing potential that is a function of  $\tilde{N}_v$ .<sup>60,64</sup> Consider the biasing potential,  $U_\phi(\tilde{N}_v)$ , parametrized by  $\phi$ , so that the Hamiltonian of the biased system becomes  $\mathcal{H}_\phi = \mathcal{H}_0 + U_\phi(\tilde{N}_v)$ ; let  $Q_\phi$  be the partition function associated with  $\mathcal{H}_\phi$ . The unbiased probability,  $P_v(\tilde{N})$ , of observing  $\tilde{N}$  coarse-grained waters in  $v$  can be readily related to the corresponding probability in the biased ensemble,  $P_v^\phi(\tilde{N})$ , as

$$P_v(\tilde{N}) = P_v^\phi(\tilde{N}) e^{\beta U_\phi(\tilde{N})} \left( \frac{Q_\phi}{Q_0} \right) \quad (1)$$

where  $\beta \equiv 1/k_B T$ , with  $k_B$  being the Boltzmann constant and  $T$  being the temperature. A derivation for eq 1 is included in the [Supporting Information](#). Taking the logarithm of both sides and defining the free energy,  $F_v(\tilde{N})$ , of observing  $\tilde{N}$  coarse-

grained waters in  $v$  through  $\beta F_v(\tilde{N}) \equiv -\ln P_v(\tilde{N})$ , we get the standard result that serves as a starting point for analyzing umbrella sampling results,<sup>60,64</sup>

$$F_v(\tilde{N}) = F_v^\phi(\tilde{N}) - U_\phi(\tilde{N}) + F_\phi \quad (2)$$

Here,  $\beta F_v^\phi(\tilde{N})$  is similarly defined as  $-\ln P_v^\phi(\tilde{N})$ , and  $\beta F_\phi \equiv -\ln \left( \frac{Q_\phi}{Q_0} \right)$  is the free energy difference between the biased and unbiased ensembles. The first term in eq 2 can be readily estimated from biased simulations, and the second term is trivial to evaluate, because the functional form of the biasing potential is known. In contrast, the third term,  $F_\phi$ , is not known a priori; umbrella sampling relies on overlap in adjacent biased distributions to estimate  $F_\phi$ , usually with the help of algorithms such as WHAM or MBAR.<sup>65–67</sup> Such a prescription provides accurate estimates of  $F_v(\tilde{N})$  over the entire range of sampled  $\tilde{N}_v$ -values.

Nevertheless, obtaining a continuous  $F_v(\tilde{N})$  distribution is not always necessary; instead, a knowledge of its functional form at  $\tilde{N}$ -values that are well separated can often be sufficient. However, the umbrella sampling requirement of overlap between adjacent biased distributions is incommensurate with the spirit of such sparse sampling. Importantly, the overlap requirement also necessitates a large number of biased simulations and can be computationally expensive, in particular for a large  $v$ . The proposed method aims to circumvent this overlap requirement by estimating  $F_\phi$  using thermodynamic integration. We will demonstrate that such an approach can be both accurate and highly efficient for a particular choice of the biasing potential,  $U_\phi(\tilde{N}_v)$ . In particular, we build upon recent work by Patel and Garde,<sup>21</sup> which employed a linear biasing potential to estimate the free energy of emptying  $v$ , as quantified by  $F_v(\tilde{N} = 0)$ , with a roughly 2 orders of magnitude decrease in computational effort. Here, we show how the entire  $P_v(\tilde{N})$  distribution can be sparsely sampled with a similar increase in computational efficiency through the use of such a linear potential,  $U_\phi(\tilde{N}_v) = \phi \tilde{N}_v$ , whose strength is determined by the parameter,  $\phi$ . In the presence of the linear potential,  $U_\phi$ , eq 2 becomes

$$F_v(\tilde{N}) = F_v^\phi(\tilde{N}) - \phi \tilde{N} + F_\phi \quad (3)$$

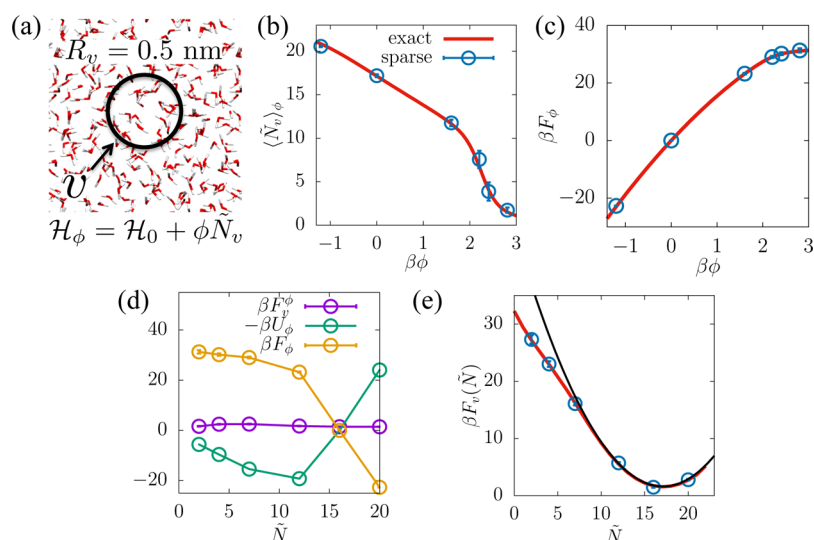
As discussed above, the first two terms of the above equation are readily determined. Here, we determine  $F_\phi$  using thermodynamic integration, that is, by integrating  $\partial F_\phi / \partial \phi = \langle \tilde{N}_v \rangle_\phi$  as

$$F_\phi = \int_0^\phi \langle \tilde{N}_v \rangle_{\phi'} d\phi' \quad (4)$$

The integration is performed over a range of  $\phi$ -values from zero (i.e., the unbiased system) to the particular  $\phi$ -ensemble of interest. Increasing  $\phi$  penalizes the presence of water molecules in  $v$ , indeed, because  $\partial^2 F_\phi / \partial \phi^2 = \partial \langle \tilde{N}_v \rangle_\phi / \partial \phi = -\beta \langle \delta \tilde{N}_v^2 \rangle_\phi < 0$ ,  $\langle \tilde{N}_v \rangle_\phi$  decreases monotonically as  $\phi$  is increased. To illustrate how the method works, consider employing eq 3 for  $\tilde{N} = \langle \tilde{N}_v \rangle_\phi$ ,

$$F_v(\langle \tilde{N}_v \rangle_\phi) = F_v^\phi(\langle \tilde{N}_v \rangle_\phi) - \phi \langle \tilde{N}_v \rangle_\phi + \int_0^\phi \langle \tilde{N}_v \rangle_{\phi'} d\phi' \quad (5)$$

Thus, from estimates of the average values,  $\langle \tilde{N}_v \rangle_{\phi_k}$ , obtained from a series of biased simulations with potentials of strength,  $\{\phi_k\}$ , the corresponding estimates of  $F_v(\tilde{N} = \langle \tilde{N}_v \rangle_{\phi_k})$  can be readily obtained at  $k$  distinct  $\tilde{N}$ -values. In addition to  $F_v(\tilde{N} = \langle \tilde{N}_v \rangle_{\phi_k})$ , this method can also be used to obtain



**Figure 1.** Use of the sparse sampling method (eq 5) to estimate the free energetics of water density fluctuations,  $\beta F_v(\tilde{N}) \equiv -\ln P_v(\tilde{N})$ , in a sphere of radius,  $R_v = 5 \text{ \AA}$ , in bulk water. (a) Simulation snapshot highlighting the observation volume,  $v$ . (b) Response of the average number of waters,  $\langle \tilde{N}_v \rangle_\phi$  in  $v$ , to the strength,  $\phi$ , of a linear biasing potential,  $U_\phi = \phi \tilde{N}_v$ , is estimated using biased simulations. (c) Free energy difference between the biased and unbiased ensembles,  $F_\phi$ , is estimated by integrating the response of  $\langle \tilde{N}_v \rangle_\phi$  to  $\phi$  (eq 4). (d) Three components of  $\beta F_v(\tilde{N})$  according to eq 5. (e) Sparsely sampled water density fluctuations obtained using the  $\langle \tilde{N}_v \rangle_\phi$ -values in panel b agree well with the exact fluctuation spectrum obtained using umbrella sampling. Comparison with the corresponding Gaussian distribution (black line) highlights the presence of a low- $\tilde{N}$  fat tail in  $P_v(\tilde{N})$ .

estimates of  $F_v(\tilde{N})$  in the vicinity of  $\tilde{N} = \langle \tilde{N}_v \rangle_{\phi k}$ ; in particular, using eqs 3 and 4,  $F_v(\tilde{N})$  can be estimated for each  $\tilde{N}$ -value sampled by any of the biased simulations.

We note that while we have chosen to focus on the fluctuations in  $\tilde{N}_v$ , the method presented here can readily be generalized to characterize fluctuations in any order parameter. Further, as mentioned above and shown in the Supporting Information, this approach can also be used to indirectly sample discrete order parameters; in particular,  $P_v(N)$ , the probability of observing  $N$  water molecules in  $v$ , can be readily estimated. Finally, we note that while there are certain similarities between our method and the Umbrella Integration (UI) method, proposed by Kästner and Thiel,<sup>68</sup> there are important fundamental and practical differences between the two methods; a detailed discussion of these issues is included in the Supporting Information.

**2.2. Simulation Details.** We illustrate the application of eq 5 to compute  $F_v(\tilde{N})$  for three different classes of systems: (1) a small spherical volume of radius  $R_v = 0.5 \text{ nm}$  in bulk water, (2) a cylindrical disk of width,  $w = 0.3 \text{ nm}$  and radius,  $R_v = 2 \text{ nm}$ , in the hydration shell of self-assembled monolayers (SAMs) of alkyl chains with two different head groups,  $-\text{OH}$  and  $-\text{CH}_3$ , which correspond to hydrophilic and hydrophobic SAM–water interfaces, respectively, and (3) the entire hydration shell of the protein, ubiquitin (PDB: 1UBQ),<sup>69</sup> as defined by the union of spherical subvolumes of radii,  $r_v = 0.6 \text{ nm}$ , centered on each of the protein heavy atoms.

To study each of these systems, we use all-atom MD simulations using the GROMACS package<sup>70</sup> suitably modified to incorporate the biasing potentials of interest. We perform biased simulations at different  $\phi$ -values and use our central eq 5 to estimate  $F_v(\tilde{N})$  at sparsely separated  $\tilde{N}$ -values. In all cases, a simulation box with periodic boundary conditions was employed, and the leapfrog algorithm<sup>64</sup> with a 2 fs time-step was used to integrate the equations of motion. For the bulk water and SAM systems, the SPC/E model of water was used,<sup>71</sup> whereas the protein was hydrated using TIP4P water.<sup>72</sup> The

short-range Lennard–Jones and electrostatic interactions were truncated at 1 nm, whereas the long-range electrostatic interactions were treated using the Particle Mesh Ewald algorithm.<sup>73</sup> Bonds involving hydrogen atoms were constrained. The SHAKE algorithm<sup>74</sup> was employed for the bulk water system, and the LINCS algorithm<sup>75</sup> was used for the SAM and protein systems. In all cases, a constant temperature of 300 K was maintained by using the canonical velocity-rescaling thermostat.<sup>76</sup> For the bulk water system, the pressure was additionally maintained at 1 bar using the Parrinello–Rahman barostat.<sup>77</sup> While the SAM and protein systems are simulated in the canonical ensemble, a buffering water–vapor interface was employed to ensure that the system remains at coexistence pressure.<sup>13,61,78</sup> Additional system specific details including those pertaining to the placement of observation volumes are given below.

**Bulk Water.** The spherical observation volume is situated at the center of a cubic simulation box with a side of length 6 nm.

**SAM Surfaces.** The simulation setup is similar to that used in ref 14. The disk-shaped observation volumes are placed adjacent to the SAM surfaces, with their axes perpendicular of the surface. For the  $\text{CH}_3$ -terminated SAM surface, the observation volume is centered at the first peak of the water density distribution in the direction perpendicular of the SAM surface. For the  $\text{OH}$ -terminated SAM surface, the proximal edge of the observation volume is placed at a location where the averaged density of the SAM heavy atoms drops below 2% of the bulk water density.

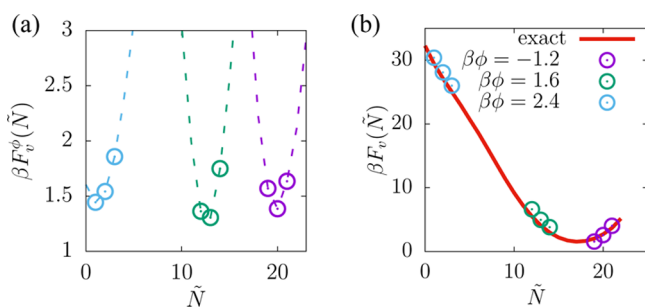
**Protein Hydration Shell.** We used the CHARMM27 force field<sup>79</sup> to simulate ubiquitin, which was hydrated by roughly 13,000 TIP4P water molecules.<sup>72</sup> A 3 ns unbiased simulation was first run to equilibrate the hydrated protein. To ensure that protein atoms remain in the observation volume, all protein heavy atoms are position restrained harmonically with a relatively soft spring constant of 1000  $\text{kJ}/\text{nm}^2$  in each dimension. Each biased simulation was then run for 3 ns. The first 100 ps is discarded as equilibration in response to the

biasing potential, and the subsequent 100 ps is used to estimate  $\langle \tilde{N}_v \rangle_\phi$ , whereas the subsequent 2.9 ns were used to estimate  $\langle \delta \tilde{N}_v^2 \rangle_\phi$  and to obtain accurate estimates of  $P_v(\tilde{N})$  using umbrella sampling in conjunction with WHAM. To obtain  $P_v(\tilde{N})$  using umbrella sampling, several additional biased simulations are also needed to ensure overlap between adjacent biased distributions.

### 3. RESULTS AND DISCUSSION

**3.1. Bulk Water.** Figure 1 illustrates the sparse sampling method for estimating  $F_v(\tilde{N})$  using a small spherical volume,  $v$ , of radius,  $R_v = 0.5$  nm, in bulk water (Figure 1a).  $v$  contains roughly  $\langle \tilde{N}_v \rangle_0 \approx 16$  water molecules on average. Figure 1b shows how the average number of waters in  $v$ ,  $\langle \tilde{N}_v \rangle_\phi$ , responds to the linear biasing potential,  $U_\phi = \phi \tilde{N}_v$ .  $\langle \tilde{N}_v \rangle_\phi$  decreases monotonically as the strength of the unfavorable potential,  $\phi$ , is increased; the decrease is linear in the vicinity of  $\phi = 0$  but becomes more pronounced at larger  $\phi$ -values. Integrating this response according to eq 4 enables us to estimate the free energies,  $F_\phi$ , of the biased ensembles relative to that of the unbiased ensemble (Figure 1c). Each of the three terms that go into the estimation of  $F_v(\tilde{N})$  using eq 2 can then be readily obtained at  $\tilde{N} = \langle \tilde{N}_v \rangle_\phi$ , and are shown in Figure 1d. The biased distribution free energies,  $F_v^\phi(\tilde{N} = \langle \tilde{N}_v \rangle_\phi)$ , are small because typical water number distributions are peaked at their means; consequently,  $P_v^\phi(\tilde{N})$  is highest for  $\tilde{N} = \langle \tilde{N}_v \rangle_\phi$ , and  $\beta F_v^\phi(\tilde{N}) \equiv -\ln P_v^\phi(\tilde{N})$  is correspondingly small at  $\tilde{N} = \langle \tilde{N}_v \rangle_\phi$ . The sparse sampled  $F_v(\tilde{N})$  resulting from the addition of the three terms in Figure 1d displays a fat low- $\tilde{N}$  tail as shown in Figure 1e and is in excellent agreement with the exact  $F_v(\tilde{N})$  obtained by umbrella sampling.

**Obtaining  $F_v(\tilde{N})$  in the Vicinity of  $\tilde{N} = \langle \tilde{N}_v \rangle$ .** In addition to estimating  $F_v(\tilde{N})$  at  $\tilde{N} = \langle \tilde{N}_v \rangle_\phi$  using eq 5,  $F_v(\tilde{N})$  can be accurately estimated at other well-sampled  $\tilde{N}$ -values using eqs 3 and 4. In Figure 2, we demonstrate this by calculating  $F_v(\tilde{N})$  at

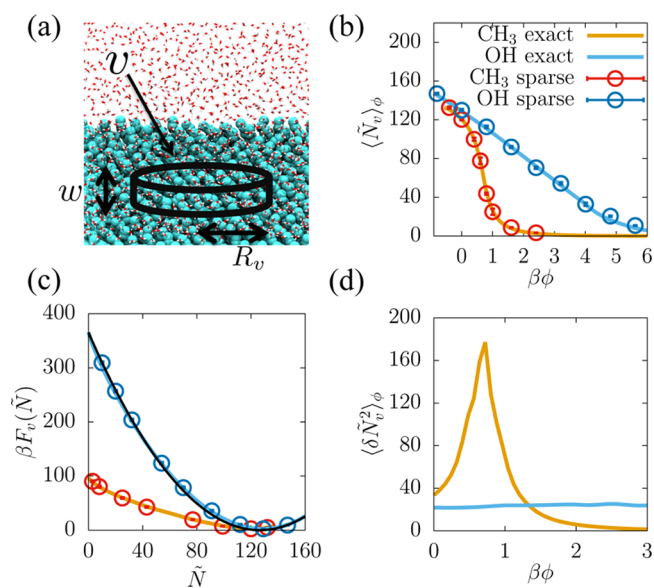


**Figure 2.** Computing  $F_v(\tilde{N})$  for all sampled  $\tilde{N}$ -values (in addition to  $\tilde{N} = \langle \tilde{N}_v \rangle_\phi$ ). (a) Free energetics of  $\tilde{N}$ -fluctuations obtained from three biased simulations ( $\beta\phi = -1.2, 1.6, 2.4$ ) are shown (dashed lines). Three representative  $\tilde{N}$ -values in the vicinity of  $\tilde{N} = \langle \tilde{N}_v \rangle_\phi$  are highlighted for each biased distribution (symbols). (b)  $F_v(\tilde{N})$  is estimated at each of these  $\tilde{N}$ -values using eqs 3 and 4 and agrees well with the exact results obtained by umbrella sampling.

three different  $\tilde{N}$ -values each for three of the biased simulations shown in Figure 1. Figure 2a shows the biased distributions,  $F_v^\phi(\tilde{N})$ , obtained at all the  $\tilde{N}$ -values sampled from the biased simulations (dashed lines). Also highlighted with symbols, are the  $F_v^\phi(\tilde{N})$ -values at three well-sampled  $\tilde{N}$ -values chosen in the vicinity of  $\tilde{N} = \langle \tilde{N}_v \rangle_\phi$ . Using these  $F_v^\phi(\tilde{N})$ -values and the  $F_\phi$ -values shown in Figure 1c, the sparse sampled  $F_v(\tilde{N})$ -values obtained from the three biased simulations are shown in

Figure 2b and agree well with the umbrella sampling results. In principle,  $F_v(\tilde{N})$  can thus be estimated at every  $\tilde{N}$ -value sampled in a biased simulation. However,  $\tilde{N}$ -values far from the mean,  $\langle \tilde{N}_v \rangle_\phi$ , may be insufficiently sampled in practice, leading to significant uncertainties in estimates of  $F_v^\phi(\tilde{N})$  and correspondingly of  $F_v(\tilde{N})$ .

**3.2. SAM–Water Interfaces.** The low- $\tilde{N}$  tail of water density fluctuations, estimated in the vicinity of a surface, is expected to depend strongly on surface hydrophobicity or the strength of surface water interactions.<sup>5,9–16,19,21,38,58</sup> To illustrate that the sparse sampling method can capture such marked differences in the respective  $F_v(\tilde{N})$  profiles, here we use it to estimate  $F_v(\tilde{N})$  in a cylindrical disk-shaped volume,  $v$ , adjacent to a hydrophobic  $\text{CH}_3$ -terminated SAM surface (Figure 3a) and a hydrophilic OH-terminated SAM surface; in both cases,  $v$  contains roughly  $\langle \tilde{N}_v \rangle_0 \approx 120$  waters on average. As shown in Figure 3b, the simulated  $\langle \tilde{N}_v \rangle_\phi$ -values decrease



**Figure 3.** Estimating interfacial water density fluctuations using sparse sampling. (a) Representative simulation snapshot of the interface between a  $\text{CH}_3$ -terminated self-assembled monolayer (SAM) surface and water. The alkyl chains (cyan) are shown in space-fill representation, whereas the water molecules (red/white) are shown as sticks. The observation volume,  $v$  (black lines), is a cylindrical disk of radius  $R_v = 2$  nm and thickness  $w = 0.3$  nm that is situated at the SAM–water interface. (b) Response of the average number of interfacial waters,  $\langle \tilde{N}_v \rangle_\phi$ , to the strength of the linear biasing potential,  $\phi$ , depends on the hydrophobicity of the SAM–water interface. While the response is roughly linear for the hydrophilic OH-terminated SAM surface, it is sigmoidal for the hydrophobic  $\text{CH}_3$ -terminated SAM surface.<sup>19</sup> (c) Disparate responses arise from a fundamental difference in the statistics of the underlying interfacial density fluctuations; while fluctuations near hydrophilic surfaces are roughly Gaussian (black line), those near hydrophobic surfaces display fat low- $\tilde{N}$  tails. The disparate free energetics of density fluctuations adjacent to the OH and  $\text{CH}_3$  SAM surfaces are readily captured by the sparse sampling method and agree well with those obtained from umbrella sampling; however, the sparse results were obtained at a tiny fraction of the computational cost of the umbrella sampling results. (d) The susceptibility,  $-\partial \langle \tilde{N}_v \rangle_\phi / \partial (\beta\phi) = \langle \delta \tilde{N}_v^2 \rangle_\phi$ , also shows distinct features for the two SAM surfaces. While the susceptibility of the OH-terminated SAM surface is roughly constant, a sharp peak is observed in the susceptibility of the  $\text{CH}_3$ -terminated SAM surface due to the corresponding sigmoidal response of  $\langle \tilde{N}_v \rangle_\phi$  to  $\phi$ .

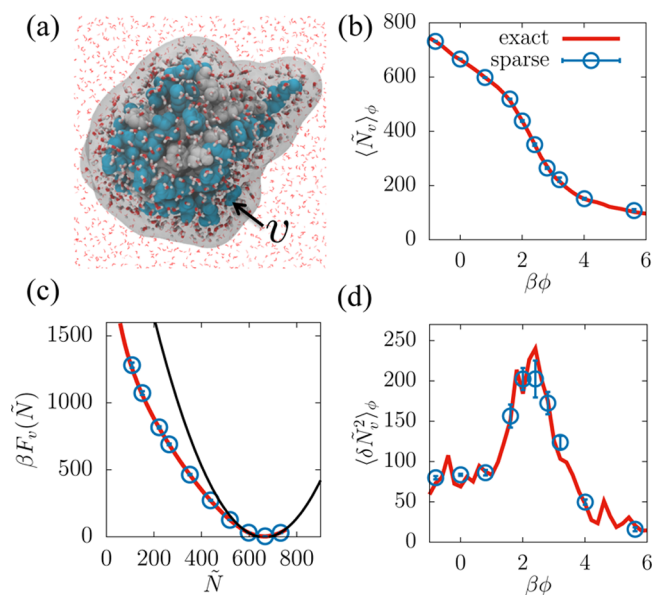
with increasing  $\phi$  for both of the SAM surfaces; however, the decrease in  $\langle \tilde{N}_v \rangle_\phi$  is more rapid for the hydrophobic SAM surface. Using these  $\langle \tilde{N}_v \rangle_\phi$ -values in conjunction with eq 5 allows us to estimate  $F_v(\tilde{N})$  adjacent to the two SAM surfaces, as shown in Figure 3c. While the sparse sampled estimates of  $F_v(\tilde{N})$  for the two SAM surfaces are very different from one another, they are nevertheless in excellent agreement with the exact results obtained from umbrella sampling.

$F_v(\tilde{N})$  adjacent to the hydrophilic SAM surface is parabolic (black line) to a good approximation, consistent with the underlying density fluctuations being Gaussian. In contrast, a marked fat tail in water density fluctuations is observed adjacent to the hydrophobic SAM surface in agreement with previous findings.<sup>13,15,19</sup> Such a difference has previously been demonstrated to arise from the fact that water near hydrophobic surfaces is situated at the edge of a dewetting transition.<sup>19,80</sup> This difference is also reflected in the sensitive response of interfacial water to perturbations, that is, in the sigmoidal decrease in  $\langle \tilde{N}_v \rangle_\phi$  with increasing  $\phi$  (Figure 3b) and in a peak in the corresponding susceptibility,  $-\partial \langle \tilde{N}_v \rangle_\phi / \partial(\beta\phi) = \langle \delta \tilde{N}_v^2 \rangle_\phi$  (Figure 3d). In contrast, the Gaussian fluctuations adjacent to the hydrophilic surface are associated with a linear decrease in  $\langle \tilde{N}_v \rangle_\phi$  (Figure 3b) and a constant susceptibility (Figure 3d).

It is clear that there is a correspondence between the functional forms of the free energetics of water number fluctuations,  $F_v(\tilde{N})$ , and the corresponding response of the average water number,  $\langle \tilde{N}_v \rangle_\phi$ , to the strength of the potential,  $\phi$ . In particular, if  $F_v(\tilde{N})$  is known over the entire range of  $\tilde{N}$ -values of interest,  $\langle \tilde{N}_v \rangle_\phi$  can be readily obtained at all values of  $\phi$  through reweighting.<sup>21</sup> Indeed, the curves shown in Figure 3b and d were obtained in that manner. In contrast, the sparse sampling method introduced above not only allows us to perform the inverse operation, but it does so with only a select few values of  $\langle \tilde{N}_v \rangle_\phi$ , that is, given well-separated  $\langle \tilde{N}_v \rangle_\phi$  estimates at a select few  $\phi$ -values, the sparse sampling method allows us to estimate  $F_v(\tilde{N})$  at those  $\tilde{N} = \langle \tilde{N}_v \rangle_\phi$ -values.

**3.3. Entire Protein Hydration Shell.** Water density fluctuations have previously been used to characterize the hydrophobicity of regions within protein hydration shells in volumes containing a few to roughly 10 waters on average. Here, we leverage the efficiency of the sparse sampling method to estimate water density fluctuations in the entire hydration shell of the protein, ubiquitin (Figure 4a), which is significantly larger and contains roughly 660 waters on average. In contrast with the uniform SAM surfaces studied in the previous section, the protein–water interface is chemically and topographically heterogeneous; such complexity influences the corresponding water density fluctuations in a nontrivial manner.<sup>12,16</sup> Using our central eq 5, the response,  $\langle \tilde{N}_v \rangle_\phi$ , to the biasing potential,  $\phi$ , shown in Figure 4b, can readily be transformed into the free energy,  $F_v(\tilde{N})$ . As shown in Figure 4c,  $F_v(\tilde{N})$  thus obtained is once again in excellent agreement with the umbrella sampling results, albeit at a fraction of the computational cost.

Interestingly, the fluctuations display a marked low- $\tilde{N}$  fat tail relative to Gaussian statistics (black line), suggesting the presence of extended hydrophobic regions in the hydration shell of ubiquitin. Indeed, ubiquitin is known to have a hydrophobic patch, which facilitates its targeting of proteins for proteasome degradation.<sup>81</sup> The ubiquitin hydration shell also displays a peak in the susceptibility, as shown in Figure 4d, suggesting a collective dewetting of water molecules from the protein hydration shell. We note certain commonalities



**Figure 4.** Estimating the free energetics of water density fluctuations in the entire hydration shell of a protein. (a) Simulation snapshot illustrating the protein ubiquitin in a space-fill representation, with the nonpolar residues colored white, and the polar and charged residues colored in blue. Water molecules in the hydration shell (gray) are shown as sticks (red/white), while those outside the hydration shell are shown as lines. (b) Response of the average number of coarse-grained waters,  $\langle \tilde{N}_v \rangle_\phi$ , in the protein hydration shell,  $v$ , to the strength of the biasing potential,  $\phi$ . While the response is linear near  $\phi = 0$ , it becomes sigmoidal at higher  $\phi$ -values, suggesting the presence of a hydrophobic patch on the ubiquitin surface. (c) Free energetics of water density fluctuations,  $F_v(\tilde{N})$ , are Gaussian (black line) near the mean but display a marked fat tail at lower  $\tilde{N}$ -values. The sparsely sampled  $F_v(\tilde{N})$ -values (symbols) are not only in excellent agreement with umbrella sampling (red line), but capture the functional form of  $F_v(\tilde{N})$  at a 42-fold reduction in the computational cost. (d) The susceptibility,  $-\partial \langle \tilde{N}_v \rangle_\phi / \partial(\beta\phi) = \langle \delta \tilde{N}_v^2 \rangle_\phi$ , displays a peak near  $\beta\phi = 2$ , suggesting a collective dewetting of water molecules from the ubiquitin hydration shell.

between our results and reports of percolation transitions on partially hydrated protein surfaces. It has been shown that when proteins are hydrated with insufficient waters to form a complete protein hydration shell an interconnected network of waters appears abruptly over a narrow range of hydration levels.<sup>82–84</sup> When a protein is partially hydrated, its hydration waters are essentially in vacuum; in contrast, the hydration shell waters in our simulations are surrounded by and interact with other waters. However, in both cases, a collective (percolation or dewetting) transition is facilitated by a competition between protein–water and water–water interactions, with the protein providing a heterogeneous surface that displays a wide range of surface chemistries and thereby protein–water interaction strengths.

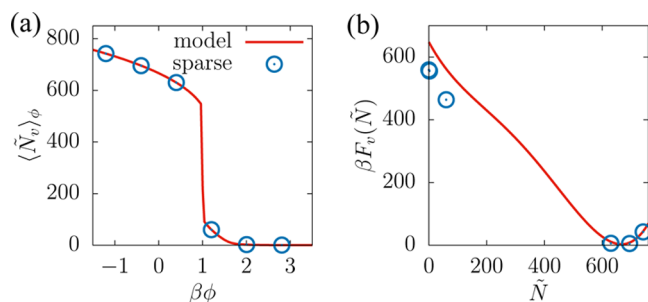
**3.4. Efficiency and Limitations of the Method.** We note that  $F_v(\tilde{N})$ , shown in Figure 4c, was obtained using 10 simulations run for 0.2 ns per simulation (including equilibration in response to the biasing potential) for a total simulation time of 2 ns. In contrast, the exact umbrella sampling results employed 28 windows run for 3 ns per window for a total simulation time of 84 ns. Thus, a dramatic speed-up in computational efficiency can be achieved if only sparse estimates of  $F_v(\tilde{N})$  are desired. At the heart of this remarkable efficiency of the method is the fact that not just

fewer but shorter simulations are needed. Because  $\langle \tilde{N}_v \rangle_\phi$  decreases monotonically with  $\phi$ , and can even be linear in  $\phi$ ,  $F_\phi$  can be estimated accurately with estimates of  $\langle \tilde{N}_v \rangle_\phi$  at only a few  $\phi$ -values. Additionally, because the averages,  $\langle \tilde{N}_v \rangle_\phi$ , are typically dominated by the most probable regions of the underlying unimodal biased distributions (Figure 2a), they converge rapidly, and short simulations are sufficient to accurately estimate them. To further understand the source of the method's efficiency and for details on how to implement it optimally (for example, how to adaptively pick the set of  $\phi$ -values for running biased simulations), the reader is referred to ref 21.

The efficient estimation of  $\langle \tilde{N}_v \rangle_\phi$  relies on the most probable region(s) of the corresponding biased  $\tilde{N}_v$ -distribution being well-sampled; this is readily achieved when the distribution is unimodal. Because accurate estimates of  $\langle \tilde{N}_v \rangle_\phi$  are required at a number of  $\phi$ -values in order to obtain  $F_v(\tilde{N})$  accurately, it is important that every biased  $\tilde{N}_v$ -distribution be adequately sampled. While this is true for the results shown in Figures 1–4, it may not always be the case. To illustrate what happens when the biased  $\tilde{N}_v$ -distributions are not unimodal, and highlight an important limitation of the sparse sampling method, we consider an analytical model for  $F_v(\tilde{N})$  designed to yield a bimodal biased distribution for a certain biasing strength  $\phi_0$ ,

$$F_v(\tilde{N}) = \kappa_0(\tilde{N} - 100)^2(\tilde{N} - 540)^2 - \phi_0\tilde{N} \quad (6)$$

We choose  $\beta\phi_0 = 1$ , and to facilitate comparison with the ubiquitin  $F_v(\tilde{N})$  (Figure 4), we choose  $\beta\kappa_0 = 10^{-8}$ , so that the two distributions have roughly the same mean and variance. As shown in Figure 5a,  $\langle \tilde{N}_v \rangle_\phi$  for the analytical model features a



**Figure 5.** (a) For the analytical model discussed in the text,  $\langle \tilde{N}_v \rangle_\phi$  decreases sharply over a narrow range of  $\phi$ -values, making it challenging to sample intermediate  $\tilde{N}_v$ -values. (b) The sparse sampling method displays substantive errors in  $F_v(\tilde{N})$  for the lower  $\tilde{N}$ -values and is unable to estimate  $F_v(\tilde{N})$  over a wide range of intermediate  $\tilde{N}_v$ -values.

sharp cliff, wherein the value of  $\langle \tilde{N}_v \rangle_\phi$  decreases dramatically over a narrow range of  $\phi$ -values. The corresponding  $F_v(\tilde{N})$  obtained using the sparse sampling method is shown in Figure 5b and highlights the limitations of the method. Due to the sharp cliff in  $\langle \tilde{N}_v \rangle_\phi$  and the associated inability to sample intermediate  $\tilde{N}$ -values, estimates of  $F_v(\tilde{N})$  cannot be obtained for a large range of  $\tilde{N}$ -values. In addition, an inability to capture the precise location of the cliff leads to substantive errors in  $F_\phi$  for higher  $\phi$ -values and thereby in  $F_v(\tilde{N})$  for the lower  $\tilde{N}$ -values. A sharply decreasing response function akin to that for the analytical model is thus a strong indicator of bistability in the underlying free energetics; in such cases, the sparse sampling method should not be used in conjunction with linear

potentials as proposed here. Strategies for overcoming such limitations, which are likely to be encountered when water is at or close to coexistence with its vapor in the volume of interest (e.g., in hydrophobic confinement), will be the subject of a future publication.

We additionally note that the computational cost for quantifying  $F_v(\tilde{N})$  is nearly independent of the size of the observation volume,  $v$ , subject to the caveat that the underlying biased distributions are unimodal. Because overlap between adjacent biased distributions is not required, the spacing between biased simulation windows can be scaled with the size of the probe volume, resulting in the number of requisite simulations being the same for observation volumes of different sizes. In particular, consider the fact that roughly the same number of simulation windows are used to obtain  $F_v(\tilde{N})$  in the protein hydration shell and at the SAM surfaces, even though  $v$  for the protein hydration shell is roughly five times larger than that for the SAM surfaces.

#### 4. OUTLOOK

By circumventing the umbrella sampling requirement of overlap between adjacent biased distributions, and instead using thermodynamic integration to estimate free energy differences between the biased and unbiased ensembles, the method presented here enables sparse sampling of the free energetics of water density fluctuations,  $F_v(\tilde{N})$ . Furthermore, a judicious choice of the functional form of the biasing potential, that is, one which is linear in the order parameter of interest,  $\tilde{N}_v$ , enables estimation of  $F_v(\tilde{N})$  in a computationally efficient manner. The low- $\tilde{N}$  behavior of  $F_v(\tilde{N})$  serves to characterize the hydrophobicity of complex surfaces with nanoscale heterogeneities, such as those presented by proteins; here, we use the method to characterize  $F_v(\tilde{N})$  in a large volume, constituting the entire hydration shell of the protein, ubiquitin. Such a characterization of protein hydration shells not only provides an overall measure of protein hydrophobicity, but also quantifies the free energy required to displace water molecules from the protein hydration shell, which could inform its propensity to bind hydrophobic ligands.<sup>85</sup> In addition to facilitating characterization of  $F_v(\tilde{N})$  in large volumes, the efficiency of our method may also enable estimation of  $F_v(\tilde{N})$ , using more detailed treatments of water in the bulk or at interfaces, which are inherently expensive from a computational standpoint. Such treatments include force fields that explicitly account for molecular polarizability,<sup>62,86</sup> or ab initio molecular dynamics simulations, which incorporate electronic structure effects.<sup>63,87</sup>

The method presented here is fairly general and can be readily generalized to other order parameters besides  $\tilde{N}_v$ , as well as to higher dimensions. A straightforward generalization of the method could facilitate characterization of the free energetics of concentrations (as opposed to number density) fluctuations in multicomponent aqueous solutions and mixtures.<sup>88</sup> Biasing potentials that couple linearly to an order parameter of interest are also employed in a variety of other contexts; examples include constant electrostatic potential simulations used to study charge fluctuations in capacitors,<sup>89</sup> simulations that bias trajectory space using a dynamical order parameters such as activity,<sup>90</sup> simulations of semiflexible chains under external tension,<sup>91</sup> and alchemical methods for estimating binding free energies.<sup>92,93</sup> While free energy perturbation is typically employed in alchemical calculations to estimate the free energy differences between the ensembles of interest, the method

introduced here could additionally provide the statistics of the energy differences between those ensembles; an understanding of such statistics may further inform optimal strategies for accurately and efficiently estimating the corresponding free energies. Such an understanding may also lead to the development of analytical expressions for the estimation of free energies. Due to the dramatic increase in computational efficiency that the method provides, we believe that it will also be well suited for the characterization of free energies in multiple dimensions, wherein ensuring overlap in all order parameters becomes very expensive and increases exponentially with the number of dimensions. In particular, the method may find use in the characterization of two-dimensional free energetic landscapes that serve as a starting point for recent spatial coarse-graining schemes.<sup>94</sup> The method could also be used to characterize spatial correlations between density fluctuations in distinct observation volumes.<sup>17,38</sup>

## ■ ASSOCIATED CONTENT

### ■ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b01037.

Detailed derivation of eq 1; demonstration on how the sparse sampling method can be used obtain the distribution of a discrete order parameter; discussion of the salient differences between the sparse sampling method and umbrella integration method; employment of the sparse sampling method to characterize water density fluctuations in the hydration shell of the protein, hydrophobin II. (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: amish.patel@seas.upenn.edu.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

A.J.P. and R.C.R. were supported in part by a Seed Grant from the University of Pennsylvania Materials Research Science and Engineering Center (NSF UPENN MRSEC DMR 11-20901). A.J.P. also acknowledges financial support from the National Science Foundation (CBET 1511437) and thanks Shekhar Garde and Talid Sinno for numerous helpful discussions.

## ■ REFERENCES

- (1) Stillinger, F. H. *J. Solution Chem.* **1973**, *2*, 141–158.
- (2) Pratt, L. R.; Chandler, D. *J. Chem. Phys.* **1977**, *67*, 3683–3704.
- (3) Hummer, G.; Garde, S.; Garcia, A. E.; Pohorille, A.; Pratt, L. R. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 8951–8955.
- (4) Ashbaugh, H. S.; Pratt, L. R. *Rev. Mod. Phys.* **2006**, *78*, 159.
- (5) Lum, K.; Chandler, D.; Weeks, J. D. *J. Phys. Chem. B* **1999**, *103*, 4570–4577.
- (6) Huang, D. M.; Geissler, P. L.; Chandler, D. *J. Phys. Chem. B* **2001**, *105*, 6704–6709.
- (7) Huang, D. M.; Chandler, D. *J. Phys. Chem. B* **2002**, *106*, 2047–2053.
- (8) Chandler, D. *Nature* **2005**, *437*, 640–647.
- (9) Mittal, J.; Hummer, G. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 20130–20135.
- (10) Sarupria, S.; Garde, S. *Phys. Rev. Lett.* **2009**, *103*, 037803.
- (11) Godawat, R.; Jamadagni, S. N.; Garde, S. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 15119–15124.
- (12) Acharya, H.; Vembanur, S.; Jamadagni, S. N.; Garde, S. *Faraday Discuss.* **2010**, *146*, 353–365.
- (13) Patel, A. J.; Varilly, P.; Chandler, D. *J. Phys. Chem. B* **2010**, *114*, 1632–1637.
- (14) Patel, A. J.; Varilly, P.; Jamadagni, S. N.; Acharya, H.; Garde, S.; Chandler, D. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 17678–17683.
- (15) Varilly, P.; Patel, A. J.; Chandler, D. *J. Chem. Phys.* **2011**, *134*, 074109.
- (16) Jamadagni, S. N.; Godawat, R.; Garde, S. *Annu. Rev. Chem. Biomol. Eng.* **2011**, *2*, 147–171.
- (17) English, N. J.; Tse, J. S. *Phys. Rev. Lett.* **2011**, *106*, 037801.
- (18) Chandler, D.; Varilly, P. Lectures on Molecular- and Nano-Scale Fluctuations in Water. *Proceedings of the International School of Physics “Enrico Fermi”*, 2012; pp 75–111.
- (19) Patel, A. J.; Varilly, P.; Jamadagni, S. N.; Hagan, M. F.; Chandler, D.; Garde, S. *J. Phys. Chem. B* **2012**, *116*, 2498–2503.
- (20) Remsing, R. C.; Weeks, J. D. *J. Phys. Chem. B* **2013**, *117*, 15479–15491.
- (21) Patel, A. J.; Garde, S. *J. Phys. Chem. B* **2014**, *118*, 1564–1573.
- (22) Southall, N. T.; Dill, K. A.; Haymet, A. D. J. *J. Phys. Chem. B* **2002**, *106*, 521–533.
- (23) Dobson, C. M. *Nature* **2003**, *426*, 884–890.
- (24) Levy, Y.; Onuchic, J. N. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 389–415.
- (25) Krone, M. G.; Hua, L.; Soto, P.; Zhou, R.; Berne, B. J.; Shea, J.-E. *J. Am. Chem. Soc.* **2008**, *130*, 11066–11072.
- (26) Ball, P. *Chem. Rev.* **2008**, *108*, 74–108.
- (27) Thirumalai, D.; Reddy, G.; Straub, J. E. *Acc. Chem. Res.* **2012**, *45*, 83–92.
- (28) Tanford, C. *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*; John Wiley & Sons, New York, 1973.
- (29) Whitesides, G. M.; Grzybowski, B. *Science* **2002**, *295*, 2418–21.
- (30) Rabani, E.; Reichman, D. R.; Geissler, P. L.; Brus, L. E. *Nature* **2003**, *426*, 271–274.
- (31) Maibaum, L.; Dinner, A. R.; Chandler, D. *J. Phys. Chem. B* **2004**, *108*, 6778–6781.
- (32) Morrone, J. A.; Li, J.; Berne, B. J. *J. Phys. Chem. B* **2012**, *116*, 378–389.
- (33) Garde, S.; Hummer, G.; Garcia, A. E.; Paulaitis, M. E.; Pratt, L. R. *Phys. Rev. Lett.* **1996**, *77*, 4966–4968.
- (34) ten Wolde, P. R. *J. Phys.: Condens. Matter* **2002**, *14*, 9445–9460.
- (35) Pratt, L. R. *Annu. Rev. Phys. Chem.* **2002**, *53*, 409–436.
- (36) Hummer, G.; Garde, S.; Garcia, A.; Paulaitis, M.; Pratt, L. R. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 1552–1555.
- (37) Huang, D. M.; Chandler, D. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **2000**, *61*, 1501–1506.
- (38) Remsing, R. C.; Patel, A. J. *J. Chem. Phys.* **2015**, *142*, 024502.
- (39) Evans, R.; Wilding, N. B. *Phys. Rev. Lett.* **2015**, *115*, 016103.
- (40) Kitchen, D. B.; Reed, L. H.; Levy, R. M. *Biochemistry* **1992**, *31*, 10083–10093.
- (41) Zhou, R.; Huang, X.; Margulis, C. J.; Berne, B. J. *Science* **2004**, *305*, 1605–1609.
- (42) Liu, P.; Huang, X.; Zhou, R.; Berne, B. J. *Nature* **2005**, *437*, 159–162.
- (43) Choudhury, N.; Pettitt, B. M. *J. Am. Chem. Soc.* **2007**, *129*, 4847–4852.
- (44) Choudhury, N.; Pettitt, B. M. *Mol. Simul.* **2005**, *31*, 457–463.
- (45) Rasaiah, J. C.; Garde, S.; Hummer, G. *Annu. Rev. Phys. Chem.* **2008**, *59*, 713–740.
- (46) Vembanur, S.; Patel, A. J.; Sarupria, S.; Garde, S. *J. Phys. Chem. B* **2013**, *117*, 10261–10270.
- (47) Granick, S.; Bae, S. C. *Science* **2008**, *322*, 1477–1478.
- (48) Siebert, X.; Hummer, G. *Biochemistry* **2002**, *41*, 2956–2961.
- (49) Giovambattista, N.; Lopez, C. F.; Rosky, P. J.; Debenedetti, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 2274–2279.
- (50) Mittal, J.; Hummer, G. *Faraday Discuss.* **2010**, *146*, 341–352.
- (51) Daub, C. D.; Wang, J.; Kudesia, S.; Bratko, D.; Luzar, A. *Faraday Discuss.* **2010**, *146*, 67–77.

- (52) Giovambattista, N.; Debenedetti, P. G.; Rosicky, P. J. *J. Phys. Chem. C* **2007**, *111*, 1323–1332.
- (53) Hua, L.; Zangi, R.; Berne, B. J. *J. Phys. Chem. C* **2009**, *113*, 5244–5253.
- (54) Argyris, D.; Cole, D. R.; Striolo, A. *Langmuir* **2009**, *25*, 8025–8035.
- (55) Wang, J.; Bratko, D.; Luzar, A. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 6374–6379.
- (56) Giovambattista, N.; Debenedetti, P. G.; Rosicky, P. J. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 15181–15185.
- (57) Surblyis, D.; Yamaguchi, Y.; Kuroda, K.; Nakajima, T.; Fujimura, H. *J. Chem. Phys.* **2011**, *135*, 014703.
- (58) Rotenberg, B.; Patel, A. J.; Chandler, D. *J. Am. Chem. Soc.* **2011**, *133*, 20521–20527.
- (59) Widom, B. *J. Chem. Phys.* **1963**, *39*, 2808–2812.
- (60) Chandler, D. *Introduction to Modern Statistical Mechanics*; Oxford University Press, New York, 1987.
- (61) Patel, A. J.; Varilly, P.; Chandler, D.; Garde, S. *J. Stat. Phys.* **2011**, *145*, 265–275.
- (62) Warren, G. L.; Patel, S. J. *J. Phys. Chem. B* **2008**, *112*, 11679–11693.
- (63) Remsing, R. C.; Baer, M. D.; Schenter, G. K.; Mundy, C. J.; Weeks, J. D. *J. Phys. Chem. Lett.* **2014**, *5*, 2767–2774.
- (64) Frenkel, D.; Smit, B. *Understanding Molecular Simulations: From Algorithms to Applications*, 2nd ed.; Academic Press, New York, 2002.
- (65) Souaille, M.; Roux, B. *Comput. Phys. Commun.* **2001**, *135*, 40–57.
- (66) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
- (67) Shirts, M. R.; Chodera, J. D. *J. Chem. Phys.* **2008**, *129*, 124105.
- (68) Kästner, J.; Thiel, W. *J. Chem. Phys.* **2005**, *123*, 144104.
- (69) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531–544.
- (70) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (71) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
- (72) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (73) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (74) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (75) Hess, B. *J. Chem. Theory Comput.* **2008**, *4*, 116–122.
- (76) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.
- (77) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (78) Miller, T.; Vanden-Eijnden, E.; Chandler, D. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 14559–14564.
- (79) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (80) Remsing, R. C.; Xi, E.; Vembanur, S.; Sharma, S.; Debenedetti, P. G.; Garde, S.; Patel, A. J. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 8181–8186.
- (81) Sloper-Mould, K. E.; Jemc, J. C.; Pickart, C. M.; Hicke, L. *J. Biol. Chem.* **2001**, *276*, 30483–30489.
- (82) Smolin, N.; Oleinikova, A.; Brovchenko, I.; Geiger, A.; Winter, R. *J. Phys. Chem. B* **2005**, *109*, 10995–11005.
- (83) Oleinikova, A.; Smolin, N.; Brovchenko, I.; Geiger, A.; Winter, R. *J. Phys. Chem. B* **2005**, *109*, 1988–1998.
- (84) Cui, D.; Ou, S.; Patel, S. *Proteins: Struct., Funct., Genet.* **2014**, *82*, 3312–3326.
- (85) Freed, A. S.; Garde, S.; Cramer, S. M. *J. Phys. Chem. B* **2011**, *115*, 13320–13327.
- (86) Soniat, M.; Rick, S. W. *J. Chem. Phys.* **2012**, *137*, 044511.
- (87) Hassanali, A.; Giberti, F.; Cuny, J.; Kühne, T. D.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 13723–13728.
- (88) Gupta, R.; Patey, G. *J. Chem. Phys.* **2014**, *141*, 064502.
- (89) Limmer, D. T.; Merlet, C.; Salanne, M.; Chandler, D.; Madden, P. A.; Van Roij, R.; Rotenberg, B. *Phys. Rev. Lett.* **2013**, *111*, 106102.
- (90) Elmatad, Y. S.; Jack, R. L.; Chandler, D.; Garrahan, J. P. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 12793–12798.
- (91) Zhang, W.; Gomez, E. D.; Milner, S. T. *Macromolecules* **2015**, *48*, 1454–1462.
- (92) Mobley, D. L.; Chodera, J. D.; Dill, K. A. *J. Chem. Phys.* **2006**, *125*, 084902.
- (93) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150–160.
- (94) Liu, X.; Crocker, J. C.; Sinno, T. *J. Chem. Phys.* **2013**, *138*, 244111.