THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

Cite This: J. Phys. Chem. B 2018, 122, 3635−3646

pubs.acs.org/JPCB

Protein Hydration Thermodynamics: The Influence of Flexibility and Salt on Hydrophobin II Hydration

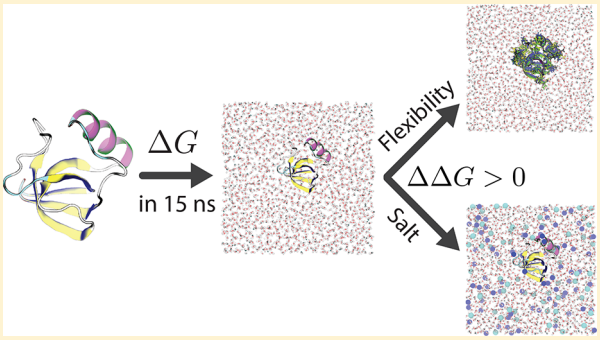
Richard C. Remsing,[†,‡] Erte Xi,[†] and Amish J. Patel*[,†]

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

[‡]Institute for Computational Molecular Science, Temple University, Philadelphia, Pennsylvania 19122, United States

Supporting Information

ABSTRACT: The solubility of proteins and other macromolecular solutes plays an important role in numerous biological, chemical, and medicinal processes. An important determinant of protein solubility is the solvation free energy of the protein, which quantifies the overall strength of the interactions between the protein and the aqueous solution that surrounds it. Here we present an all-atom explicit-solvent computational framework for the rapid estimation of protein solvation free energies. Using this framework, we estimate the hydration free energy of hydrophobin II, an amphiphilic fungal protein, in a computationally efficient manner. We further explore how the protein hydration free energy is influenced by enhancing flexibility and by the addition of sodium chloride, and find that it increases in both cases, making protein hydration less favorable.



INTRODUCTION

Due to the importance of protein solubility in biotechnology, structural biology, and numerous human diseases, characterizing protein solubility and developing strategies to enhance it are active areas of research.[1,2] For example, immuno-therapy often requires solubilizing proteins in high concentrations to maximize therapeutic efficiency while minimizing detrimental effects like high solution viscosity.[3−5] Additionally, low solubility is thought to play an important role in diseases caused by aggregation-prone proteins. Understanding and controlling these phenomena require a fundamental understanding of protein solubility, and its dependence on the properties of both the protein and its surrounding environment; the protein size, shape, polarity, and conformational flexibility, as well as the presence of cosolvents and other additives can all have a significant influence on protein solubility.[2,6,7]

Although protein solubility is influenced by the complex interplay between protein−water and protein−protein interactions (and the possibility of protein unfolding), it has been suggested that the solubility of a protein varies monotonically with its hydration free energy, ΔG.[8] Thus, being able to efficiently estimate ΔG is an important first step toward better understanding protein solubility, and modulating ΔG offers a means to influence protein solubility. The hydration free energy of a protein also sheds light more directly on protein dehydration and rehydration processes, which are important for the storage of biologics as freeze-dried powders and for their subsequent reconstitution into therapeutics.[9−11]

However, computing the hydration free energies of proteins is very expensive due to their large size and complex surfaces, which display nanoscale chemical and topographical heterogeneity. In this work, we introduce a framework for estimating the solvation free energies of large, complex solutes, such as proteins, in a highly efficient manner, from all-atom molecular dynamics (MD) simulations that explicitly account for the solvent. By incorporating the solvent explicitly, we are able to capture the context-dependent hydration behavior that arises at the chemically and topographically heterogeneous protein surfaces,[12−19] and thereby, to describe protein solvation thermodynamics more accurately than implicit solvent approaches.[20] Moreover, such a framework is transferable, and allows us to readily incorporate cosolvents, which can not only alter solution properties in a nontrivial manner, but also interact directly with the protein.[21−26]

To efficiently compute protein hydration free energies, while retaining the accuracy and transferability afforded by explicit solvent, we extend classic potential distribution theory (PDT)[27,28] and quasi-chemical theory (QCT) ideas,[29] and combine these concepts with novel simulation techniques[17,30] to rapidly compute hydration free energies following the thermodynamic cycle in Figure 1. The computational framework allows us to obtain initial estimates of protein hydration

Special Issue: Benjamin Widom Festschrift

Received: December 7, 2017
Revised: February 1, 2018
Published: February 2, 2018

DOI: 10.1021/acs.jpcb.7b12060
J. Phys. Chem. B 2018, 122, 3635−3646

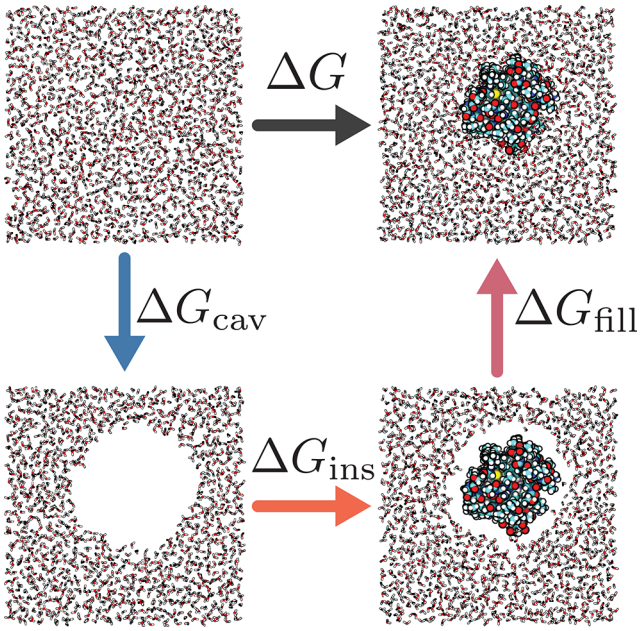


Figure 1. Hydration free energy, ΔG, of the protein, hydrophobin II, is defined as the free energy change upon insertion of the protein (from vacuum) into the solvent. To efficiently compute ΔG, we use the thermodynamic cycle shown here. In the first step, a cavity that is somewhat larger than the protein is created in bulk water, with a free energy change of ΔG_{cav}. In the second step, the solute is inserted into the cavity. The corresponding free energy, ΔG_{ins}, can be computed efficiently when short-ranged solute−solvent interactions are absent. Finally, the cavity is filled by the solvent molecules, with the corresponding free energy, ΔG_{fill}, being dominated by short-ranged solute−solvent interactions, such as hydrogen bonds.

free energies orders of magnitude faster than conventional approaches; the accuracy of those estimates can then be improved, as needed, through further computation. Using this methodology, we study the hydration of the protein hydrophobin II (HFBII), a small fungal protein, which is also utilized in biotechnology and materials science due to its amphiphilicity.[31,32] For simplicity, we first illustrate our approach by solvating a rigid protein molecule in water. We then relax this condition to elucidate the effects of flexibility on protein hydration, and further demonstrate the utility of our approach by studying how the addition of NaCl influences HFBII hydration.

ESTIMATING PROTEIN HYDRATION FREE ENERGIES

Upon hydration, a protein not only displaces water from the region it occupies, it also alters water structure in its vicinity. Consequently, there is no tangible overlap between the ensembles before and after protein hydration, making it unfeasible to compute the protein's hydration free energy, ΔG (or its excess chemical potential, to be precise), by directly inserting the protein into water. To address this challenge, solute−solvent interactions can be turned on in tiny increments, chosen to ensure overlap between adjacent alchemical states. By adequately sampling such a series of alchemical states, the hydration free energies of the solute of interest as well as all the alchemical intermediates can then be accurately estimated.[33−36] However, due to the numerous long simulations required,[37] such approaches can exact a steep computational cost for hydrating even small molecules.[35,36,38−41] Moreover, because the alchemical intermediates are unphysical species, it is challenging to extract physical insight from their hydration free energies, or to ascertain how those hydration free energies might change in response to physical perturbations, such as a change in temperature or the addition of salt to the solution.

Alternatively, approaches based on quasi-chemical theory (QCT) prescribe using the thermodynamic cycle shown in Figure 1 for estimating solvation free energies in three steps.[29,42] In the first step, a cavity somewhat larger than the protein is created, incurring a free energy cost, ΔG_{cav}. Then, the protein is inserted into the cavity with an accompanying free energetic change of ΔG_{ins}. Finally, the cavity is filled by the surrounding water, resulting in a free energy change of ΔG_{fill}. The total protein hydration free energy is then

$$\Delta G = \Delta G_{cav} + \Delta G_{ins} + \Delta G_{fill} \quad (1)$$

Because the protein is inserted in a preformed cavity in step 2 of the QCT approach, insertion does not lead to any short-ranged, rapidly varying interactions, such as hydrogen bonds or repulsions from hard-core overlap; thus, the ensembles before and after protein insertion are not too dissimilar. As we will show below, reasonable estimates of ΔG_{ins} can thus be obtained by sampling the end states alone, with more accurate estimates requiring only a few alchemical steps; in contrast, several hundred to thousand alchemical steps may be required in the absence of the preformed cavity. The majority of the computational burden in the QCT formalism is then transferred to steps 1 and 3, which involve the emptying and filling of v, respectively. Because both steps involve manipulating water density in v, we will leverage an understanding of the free energetics of water density fluctuations in bulk water (step 1)[43−45] and in the protein hydration shell (step 3),[17,46] and utilize computational methods for characterizing such free energetics[17,30,43,47] to efficiently estimate ΔG_{cav} and ΔG_{fill}. Moreover, because steps 1 and 2 of the QCT thermodynamic cycle involve simple physical processes, estimating the effect of perturbations on their free energetics can be relatively straightforward, as we will illustrate.

To compute each of the components of ΔG, we make use of biased simulations which sample the following generalized Hamiltonian,

$$\mathcal{H}_{\lambda,\phi}(\bar{\mathbf{R}}) = \mathcal{H}_0(\bar{\mathbf{R}}) - U(\bar{\mathbf{R}}) + \lambda U(\bar{\mathbf{R}}) + \phi \tilde{N}_v(\bar{\mathbf{R}}) \quad (2)$$

where $\bar{\mathbf{R}}$ represents the positions of all atoms in the system, and $\mathcal{H}_0$ is the Hamiltonian associated with the explicitly solvated protein, which includes intramolecular protein−protein interactions, solvent−solvent interactions, as well as protein−solvent interactions with interaction energy, $U(\bar{\mathbf{R}})$. The parameter, λ, is used to reversibly couple ($\lambda = 1$) or decouple ($\lambda = 0$) the protein and the solvent. $\tilde{N}_v(\bar{\mathbf{R}})$ is the smoothed number of water oxygens in the volume, v, which encompasses the protein and its first hydration shell; the volume, v, is formally defined as the union of spherical subvolumes of radius $r_v = 0.6$ nm centered on every protein heavy atom. Although the choice of r_v is somewhat arbitrary, it exerts a substantial influence on the efficiency of the approach; considerations governing our choice of r_v above are discussed in detail in the Appendix. The parameter, ϕ, is used to modulate the number of waters in v reversibly; as ϕ is increased, waters are systematically displaced from v, resulting in the formation of a cavity as $\phi \to \infty$. To ensure that the forces arising from this biasing potential are continuous,[43] the smoothed number of waters, $\tilde{N}_v(\bar{\mathbf{R}})$, is biased

DOI: 10.1021/acs.jpcb.7b12060
J. Phys. Chem. B 2018, 122, 3635−3646

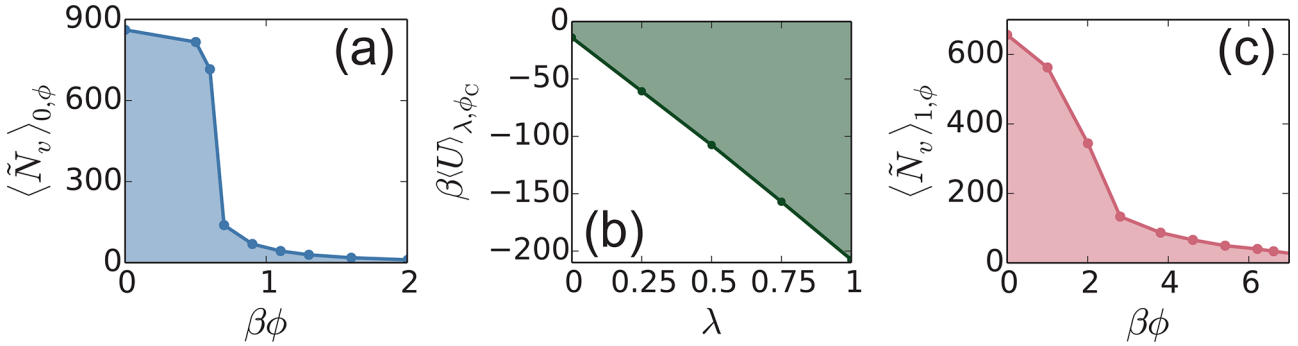


Figure 2. (a) The free energy, ΔG_{cav}, of creating a protein-shaped cavity in water (by choosing r_v = 0.6 nm) is given by the integral of $\langle \tilde{N}_v(\bar{\mathbf{R}}) \rangle_{0,\phi}$ over ϕ, and is indicated by the shaded region (eq 10). (b) In the second step of the thermodynamic cycle shown in Figure 1, the protein is inserted inside the cavity. The insertion free energy, ΔG_{ins}, can be obtained by integrating the dependence of the average energy, $\langle U(\bar{\mathbf{R}}) \rangle_{\lambda,\phi_C}$, over the range of the coupling parameter, $0 \le \lambda \le 1$ (eq 12). Symbols indicate simulation data and the shaded region indicates the integral, ΔG_{ins}. (c) Filling the cavity in the presence of the protein is the reverse of emptying the protein hydration shell. Here we show how the number of waters in the hydration shell, $\langle \tilde{N}_v(\bar{\mathbf{R}}) \rangle_{1,\phi}$, decreases with ϕ, with ΔG_{fill} being the negative of the area that is indicated by the shaded region (eq 15).

following procedures outlined in ref 30. Associated with the above Hamiltonian, $\mathcal{H}_{\lambda,\phi}$, we further define a partition function,

$$\mathcal{Z}_{\lambda,\phi} \equiv \int d\bar{\mathbf{R}} \{ e^{-\beta \mathcal{H}_{\lambda,\phi}(\bar{\mathbf{R}})} \} \tag{3}$$

the corresponding free energy,

$$G_{\lambda,\phi} = -\frac{1}{\beta} \ln \mathcal{Z}_{\lambda,\phi} \tag{4}$$

and the average of an observable $O(\bar{\mathbf{R}})$,

$$\langle O(\bar{\mathbf{R}}) \rangle_{\lambda,\phi} \equiv \frac{1}{\mathcal{Z}_{\lambda,\phi}} \int d\bar{\mathbf{R}} \left\{ O(\bar{\mathbf{R}}) e^{-\beta \mathcal{H}_{\lambda,\phi}(\bar{\mathbf{R}})} \right\} \tag{5}$$

where $\int d\,\bar{\mathbf{R}}$ indicates an integration over configuration space. Then, the cavity creation free energy can be obtained as

$$\Delta G_{cav} = G_{0,\phi_C} - G_{0,0} \tag{6}$$

in the limit of $\phi_C \to \infty$, the protein insertion free energy is given by,

$$\Delta G_{ins} = G_{1,\phi_C} - G_{0,\phi_C} \tag{7}$$

and the filling free energy is

$$\Delta G_{fill} = G_{1,0} - G_{1,\phi_C} \tag{8}$$

The total protein solvation free energy is

$$\Delta G = G_{1,0} - G_{0,0} \tag{9}$$

and can be obtained from the sum of its components (eq 1).

The approach described above is fairly general, and can be used to estimate the solvation free energies of complex solutes in diverse solvation environments, ranging from solvent mixtures to interfaces. To illustrate its utility in computing protein hydration free energies, here we estimate ΔG for the amphiphilic protein hydrophobin II (PDB ID: 2B97).[48] Hydrophobin II is a small protein with large hydrophobic and hydrophilic regions on its surface,[17] and it has served as a model biomolecule for studying protein−water interactions.[14,17,19] For expediency, we will first consider the solvation of a rigid protein molecule, such that the positions of all its heavy atoms are constrained; the particular configuration to which the protein atoms are constrained was drawn randomly from an equilibrium simulation of a solvated (and fully flexible) protein. In the subsequent section, we will then relax this constraint and elucidate how the protein hydration free energy is influenced by its flexibility. We now discuss the estimation of the components of ΔG for the rigid hydrophobin II protein. The simulation methods and the force fields that we use in our calculations are described in the SI.

Cavity Creation Free Energy. To create a cavity that is large enough to accommodate a protein, we make use of the potential $\phi \tilde{N}_v$, which linearly couples to the smoothed number of water oxygen atoms, $\tilde{N}_v$, in the volume, v.[17] The corresponding free energy can then be efficiently obtained through thermodynamic integration

$$\Delta G_{cav} = \int_0^{\phi_C} d\phi \langle \tilde{N}_v(\bar{\mathbf{R}}) \rangle_{0,\phi} \tag{10}$$

Evaluation of eq 10 and sampling using $\tilde{N}_v$-dependent potentials in molecular dynamics (MD) simulations is performed following previous work.[17,30] The integrand in eq 10 involves averages, which can be estimated very efficiently using short simulations, if the underlying distributions are unimodal. Additionally, because $\langle \tilde{N}_v(\bar{\mathbf{R}}) \rangle_{0,\phi}$ is a monotonically decaying function of ϕ,[17] only a few data points are needed to accurately evaluate the integral in eq 10. For hydrophobin II, $\langle \tilde{N}_v(\bar{\mathbf{R}}) \rangle_{0,\phi}$ is shown in Figure 2a as a function of ϕ; the shaded region in the figure corresponds to the integral in eq 10, and yields $\beta \Delta G_{cav}$ = 613 ± 1, where the error bars indicate one standard deviation as determined from block averaging. We note that this error estimate incorporates errors from the finite simulation length used to estimate $\langle \tilde{N}_v(\bar{\mathbf{R}}) \rangle_{0,\phi}$, but not the integration errors arising from the knowledge of $\langle \tilde{N}_v(\bar{\mathbf{R}}) \rangle_{0,\phi}$ at a finite number of ϕ-values.

Note of Caution. Large volumes ($v \gtrsim 1$ nm^3) in bulk water tend to exhibit collective dewetting, which results in $\langle \tilde{N}_v(\bar{\mathbf{R}}) \rangle_{0,\phi}$ decreasing sharply over a small range of ϕ-values (Figure 2a).[43,46] Such collective dewetting can, for certain values of ϕ, lead to bimodal distributions, $P_v^{\phi}(\tilde{N}) \equiv \langle \delta(\tilde{N}_v(\bar{\mathbf{R}}) - \tilde{N}) \rangle_{0,\phi}$, of water numbers in v. The associated hysteresis can lead to inaccuracies in estimating $\langle \tilde{N}_v(\bar{\mathbf{R}}) \rangle_{0,\phi}$, and systematic errors in ΔG_{cav}.[47] To quantify such errors, we suggest performing two sets of simulations to obtain the dependence of $\langle \tilde{N}_v(\bar{\mathbf{R}}) \rangle_{0,\phi}$ on ϕ: one initialized with a dry v devoid of any waters, and another initialized with a wet v that is filled with waters. The true value of ΔG_{cav} is then bounded by the estimates obtained from the two sets of simulations. If more refined estimates are desired,

DOI: 10.1021/acs.jpcb.7b12060
J. Phys. Chem. B 2018, 122, 3635−3646

sufficiently strong parabolic potentials, which result in unimodal biased water number distributions, must be used. We have done so here, using umbrella sampling with biasing potentials that are either linear or harmonic in $\tilde{N}_v$ to sample this order parameter over its entire range. From these calculations, we have obtained a highly accurate estimate of $\beta\Delta G_{cav} = 600 \pm 16$, albeit at roughly 400 times the computational cost of the estimate obtained using eq 10 (2200 ns vs 5 ns), as detailed in the SI. Thus, the thermodynamic integration estimate of ΔG_{cav} obtained from eq 10 gives rise to a relatively small error of 2.2%, which can be improved systematically by performing simulations at additional ϕ-values.

Conversely, a reasonable first estimate for ΔG_{cav} can be obtained even without performing simulations. From a physical standpoint, the first step of the QCT thermodynamic cycle corresponds to the creation of vapor−liquid interfacial area, so that,

$$\Delta G_{cav} \approx \gamma_{SPC/E} A > 0 \tag{11}$$

where $\gamma_{SPC/E}$ is the liquid−vapor surface tension of our water model, and A is the surface area of v. In eq 11, we have ignored the contribution to ΔG_{cav} from $(P - P_{sat})\, v$, which tends to be negligible for the formation of nanoscale cavities in water at ambient conditions,[49] but can become important at high pressures, or for solvation in low-surface tension liquids; here, P_{sat} is the saturation pressure. Although quantitatively accurate determination of ΔG_{cav} is complicated by the presence of nanoscale bumps and crevices on the protein (and thereby on v), and by the nontrivial dependence of the surface tension on such nanoscale curvature, eq 11 can nevertheless be used to obtain a rough estimate of ΔG_{cav}. In particular, we use the zero curvature value[50] of $\gamma_{SPC/E} \approx 60.5$ mJ/m^2, and the solvent accessible surface, $A \approx 41$ nm^2, determined by rolling a water-sized probe over the surface of the protein (see SI for details) using the g_sas tool, which is part of the GROMACS MD simulation package.[51] With these estimates, we obtain $\beta\Delta G_{cav} \approx 599$, in excellent agreement with both the thermodynamic integration ($\beta\Delta G_{cav} = 613$) and umbrella sampling ($\Delta G_{cav} = 600$) estimates. More importantly, eq 11 provides a straightforward approach for anticipating how ΔG_{cav} ought to change in response to physical perturbations, such as changes in temperature, pressure, or the addition of salt to the solution.

Protein Insertion Free Energy. To compute the free energy of inserting the protein into a preformed cavity, ΔG_{ins}, we couple the solute−solvent interaction energy, $U(\bar{\mathbf{R}})$, linearly to a parameter, λ, which varies from zero (noninteracting solute) to one (fully interacting solute). Analogous to the estimation of ΔG_{cav}, the insertion free energy can also be obtained by thermodynamic integration,

$$\Delta G_{ins} = \int_0^1 d\lambda \langle U(\bar{\mathbf{R}}) \rangle_{\lambda,\phi_C} \tag{12}$$

here the subscripts λ and ϕ_C indicate that the ensemble average is obtained by sampling in the presence of the scaled solute−solvent interactions $\lambda U(\bar{\mathbf{R}})$, and a large enough ϕ-value (denoted by ϕ_C) to maintain a cavity devoid of any water molecules. For hydrophobin II, $\langle U(\bar{\mathbf{R}}) \rangle_{\lambda,\phi_C}$ decreases monotonically as λ is increased from 0 to 1, as shown in Figure 2b; the shaded region corresponds to the integral in eq 12 and results in $\beta\Delta G_{ins} = -111.0 \pm 0.3$. As shown in the SI, five λ-states spanning from 0 to 1 in increments of $\Delta\lambda = 0.25$ are sufficient to achieve phase space overlap between neighboring λ-ensembles. Thus, umbrella sampling can be used to obtain highly accurate estimates of ΔG_{ins},[52,53] and it yields $\beta\Delta G_{ins} = -110.8 \pm 0.3$ in excellent agreement with the result obtained from eq 12.

Figure 2b also highlights the roughly linear dependence of $\langle U(\bar{\mathbf{R}}) \rangle_{\lambda,\phi_C}$ on λ; such linear response suggests that a reasonably accurate estimate of ΔG_{ins} could be obtained by sampling the end states alone, using[29,34,37]

$$\Delta G_{ins} \approx \frac{1}{2}[\langle U(\bar{\mathbf{R}}) \rangle_{0,\phi_C} + \langle U(\bar{\mathbf{R}}) \rangle_{1,\phi_C}] \tag{13}$$

The linear response approximation of eq 13 results in $\beta\Delta G_{ins} \approx -112.7 \pm 0.5$, which is in good agreement with the results obtained from thermodynamic integration (eq 12) and umbrella sampling, both of which yield $\beta\Delta G_{ins} \approx -111$.

To rationalize the linear relationship between $\langle U(\bar{\mathbf{R}}) \rangle_{\lambda,\phi_C}$ and λ, we first recognize that linear response is underpinned by the corresponding distributions of the solute−solvent interaction energy, $P_\lambda(U) \equiv \langle \delta(U(\bar{\mathbf{R}}) - U) \rangle_{\lambda,\phi_C}$, following Gaussian statistics. As shown in the Appendix, $P_\lambda(U)$ is indeed Gaussian to a very good approximation when hydrophobin II is inserted into a cavity that is sufficiently large (for $r_v \geq 0.6$ nm). For $r_v \geq 0.6$ nm, cavity creation eliminates the entire first hydration shell of the protein, so that strong, short-ranged protein−water interactions, like hydrogen bonding, do not contribute to ΔG_{ins}; instead, $U(\bar{\mathbf{R}})$ is dominated by long-ranged interactions that vary slowly over molecular length scales, e.g., polarization of the liquid−vapor interface by the partial charges on the protein. Such slowly varying interactions do not alter the liquid structure in a drastic manner, and therefore, they follow Gaussian statistics and display linear response.[29,34,45,54−60]

Due to the attractive nature of the dispersion interactions, and the long-ranged electrostatics that give rise to solvent polarization, we expect that $\Delta G_{ins} < 0$. Moreover, for large r_v, electrostatic interactions are expected to dominate ΔG_{ins}. In this case, a simple and physically motivated estimate of ΔG_{ins} can be obtained even without performing simulations. For the charge-neutral hydrophobin II, we approximate the leading order dipolar contribution to ΔG_{ins} by using Kirkwood's expression for the free energy of inserting a dipole into a spherical cavity that is in a uniform dielectric solvent,[61]

$$\Delta G_{ins} \approx -\frac{\mu^2}{R^3}\frac{(\varepsilon - 1)}{(2\varepsilon + 1)} \tag{14}$$

here μ is the dipole moment of the solute, R is the radius of the spherical cavity, and ε is the dielectric constant of the solvent. The force field used to describe the protein yields $\mu \approx 73$ D, in reasonable agreement with the estimate of 107 D obtained from the PDB structure using the protein dipole moments server.[62] Note that more hydrophilic proteins have significantly larger dipole moments; for example, $\mu \approx 340$ D for cytochrome C.[63] To estimate an effective size for the protein, we use the solvent accessible surface area $A \approx 41$ nm^2, and assume $R \approx \sqrt{A/4\pi} = 1.8$ nm. Using these values of μ and R in eq 14, along with the dielectric constant of SPC/E water ($\varepsilon = 71$), yields $\beta\Delta G_{ins} \approx -107$, in reasonable agreement with our most accurate estimate of $\beta\Delta G_{ins} = -111$. At first glance, such agreement may seem spectacular given the simplicity of the assumptions we make along the way; however, dielectric continuum theories, including eq 14, are expected to provide a reasonable description of solvation processes when long-ranged

DOI: 10.1021/acs.jpcb.7b12060
J. Phys. Chem. B 2018, 122, 3635−3646

electrostatics dominate,[59,60] and changes in local, short-ranged interactions are negligible. Analogous theoretical descriptions for dispersion interactions can also be obtained when the harshly repulsive excluded volume cores of the solute are not involved in the solvation process.[45] Moreover, the success of eq 14 in estimating ΔG_{ins} also provides a means for understanding how certain physical perturbations are likely to influence ΔG_{ins}, for example, how ΔG_{ins} responds to a change in temperature, T, can be readily estimated using a knowledge of $\varepsilon(T)$.

Filling Free Energy. In the final step of the thermodynamic cycle, the cavity is filled by the solvent, as shown in Figure 1. Because filling the cavity is the reverse of creating the cavity around the protein, the filling free energy, ΔG_{fill}, is the negative of the corresponding cavity creation free energy. Therefore, we employ the same formalism that was used in step 1, but with protein−water interactions turned on ($\lambda = 1$),

$$\Delta G_{fill} = -\int_0^\infty d\phi \langle \tilde{N}_v(\bar{\mathbf{R}}) \rangle_{1,\phi} \tag{15}$$

Due to the presence of strong, short-ranged electrostatic interactions between certain regions of the protein and water, removing the last few waters from the protein hydration shell is challenging. To address this challenge and ensure reversible dewetting of the protein hydration shell, we employ a slightly different definition of $\tilde{N}_v(\bar{\mathbf{R}})$, as described in detail in the SI. Briefly, our previous definition of $\tilde{N}_v(\bar{\mathbf{R}})$, which counts the number of waters in the protein hydration shell, is now augmented to include an additional potential on waters that are closest to the protein (within 0.3 nm), and participate in strong, short-ranged interactions with the protein. The response, $\langle \tilde{N}_v(\bar{\mathbf{R}}) \rangle_{1,\phi}$, is shown in Figure 2c, and upon integration, according to eq 15, yields a filling free energy of $\beta\Delta G_{fill} = -1541 \pm 14$. We note that the decrease of $\langle \tilde{N}_v(\bar{\mathbf{R}}) \rangle_{1,\phi}$ with ϕ (Figure 2c) is much more gradual than that of $\langle \tilde{N}_v(\bar{\mathbf{R}}) \rangle_{0,\phi}$ (Figure 2a). The relatively gradual dewetting of protein hydration shells arises from the fact that typical protein hydration shells display the entire spectrum of chemistries from hydrophobic to hydrophilic, as discussed in detail in ref 47. As a result, ΔG_{fill} can be efficiently obtained using simulations at only a few ϕ-values. Moreover, as the fraction of exposed protein residues that are hydrophilic increases, the variation of $\langle \tilde{N}_v(\bar{\mathbf{R}}) \rangle_{1,\phi}$ with ϕ is expected to become increasingly linear,[17,46] so that for a given computational investment, smaller errors in ΔG_{fill} can be expected for more hydrophilic proteins.

From a physical standpoint, the filling free energy corresponds to the work required to form strong, short-ranged, and generally favorable protein−water interactions. As such, we expect $\Delta G_{fill} < 0$. However, because the free energetics of such local interactions depend sensitively on the chemical and topographical patterning of the protein surface,[12−20] we expect that in contrast with ΔG_{cav} and ΔG_{ins}, it will be challenging to develop an approximate analytical expression for ΔG_{fill}. Moreover, short-ranged solute−water interactions also give rise to a nonlinear response and asymmetries in the solvation of charged and polar solutes,[60,64−69] which are not captured by simple solvation models, such as Born theory[70] and Kirkwood's eq 14.[61] Consequently, the dependence of ΔG_{fill} on physical perturbations, such as the introduction of mutations or the inclusion of salt, is expected to be nontrivial; nevertheless, using the methods presented here, it is possible to characterize ΔG_{fill} under conditions that may be of interest.

Total Protein Hydration Free Energy. Following eq 1, the total hydration free energy of the protein is obtained from the sum of the three contributions depicted in Figure 1, and for hydrophobin II, it is $\beta\Delta G = -1039 \pm 14$, which is in agreement with the significantly more expensive umbrella sampling-based estimate of $\beta\Delta G = -1057 \pm 24$. While this indicates that protein−water interactions are favorable for HFBII, the magnitude of $\beta\Delta G$ for HFBII is nearly half of that reported for the similarly sized protein, cytochrome C, which is highly charged.[42] The smaller magnitude of ΔG for HFBII can be attributed to the large number of exposed hydrophobic groups on the surface of the protein, which are not present in cytochrome C, and other proteins which internalize most of their hydrophobic groups in the solvent inaccessible protein core. The surface hydrophobicity of HFBII reduces the magnitude of the favorable term, ΔG_{fill}, and, to a lesser extent, that of ΔG_{ins}, when compared to more hydrophilic proteins.

Finally, we highlight the computational efficiency of our approach for estimating ΔG, and contrast it with the cost associated with typical alchemical transformation-based approaches. In particular, the ΔG estimate for HFBII was obtained from approximately 15 ns of simulation data in total, distributed across approximately 25 simulations. Moreover, if the error associated with such an estimate (roughly 2% here) is unsatisfactory, the ΔG estimate can be progressively refined by adding more simulations in an adaptive manner. For example, the number of simulations used to determine ΔG_{cav}, ΔG_{ins}, or ΔG_{fill} (Figure 2) can be increased, thereby decreasing the numerical integration error, at only a modest increase in the computational expense; we implement such a strategy to study the effects of salt concentration on protein hydration below. In contrast, although an alchemical approach is expected to provide a highly accurate estimate for ΔG, such an estimate can only be obtained upon expending a substantial computational cost. In other words, if only 15 ns of simulation in total were available, it would be challenging to obtain a ΔG-estimate for hydrophobin II, however approximate, using the alchemical route. Thus, with a well chosen r_v, initial estimates of solvation free energies of large macromolecules, such as proteins, can be obtained with a computational effort that is similar to (or even less than) that required for small molecules;[33,71] such estimates can then be systematically refined depending on the available computational budget.

THE ROLE OF PROTEIN FLEXIBILITY

In the previous section, we estimated the solvation free energy of a rigid protein. In this section, we incorporate the effects of protein flexibility on its hydration free energy. We modulate protein flexibility by applying harmonic restraints to all the heavy atoms of the protein; in particular, we apply a potential

$$U_r = \frac{\kappa}{2} \sum_{i=1}^{N_{prot}} |\mathbf{r}_i - \mathbf{r}_{i,0}|^2 \tag{16}$$

where κ is a force constant, N_{prot} is the number of protein heavy atoms, and $\mathbf{r}_{i,0}$ is the restraint position of atom i. For $\kappa = 0$, the protein is fully flexible, whereas in the limit of large κ, the protein becomes rigid; κ thus serves to modulate protein flexibility. To estimate the solvation free energy of a partially flexible protein, we first recognize that whether a protein is solvated or in vacuum, there is a free energetic cost associated with restraining the protein. Moreover, the relative cost of

DOI: 10.1021/acs.jpcb.7b12060
J. Phys. Chem. B 2018, 122, 3635−3646

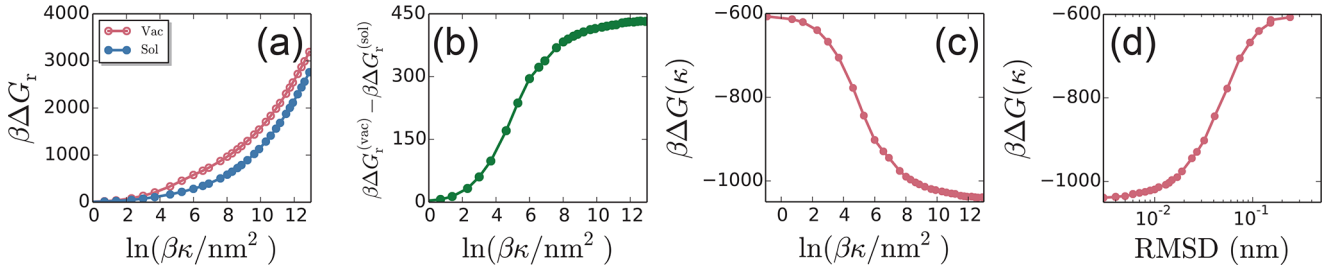


Figure 3. (a) As the force constant, κ, is increased, there is an increase in the free energetic cost of turning on the restraint potential for both the vacuum and the solvated states. (b) However, the difference between these free energetic costs, which determines how the protein hydration free energy varies with κ, plateaus at high κ-values when the protein conformation is essentially rigid. (c) The protein hydration free energy decreases as the restraints on the protein heavy atoms are increased, suggesting that the protein is solvated more favorably as it becomes more rigid. (d) This is confirmed by examining the solvation free energy of a partially flexible protein as a function of its RMSD in the solvated state.

restraining the protein in water and in vacuum ought to determine the dependence of ΔG on κ.

To formalize this notion, we define the free energetic cost of turning on the restraint potential in vacuum (vac) to be $\Delta G_r^{(vac)}(\kappa)$, and that in the solvated (sol) state to be $\Delta G_r^{(sol)}(\kappa)$. We then augment the thermodynamic cycle of Figure 1 with two additional steps. Before inserting the partially flexible protein, we first increase the restraints on the protein in vacuum by increasing the force constant from κ to a large value (say, $\kappa_C \to \infty$) until the protein conformation becomes rigid. The free energy of this step is $\Delta G_r^{(vac)}(\kappa_C) - \Delta G_r^{(vac)}(\kappa)$. We then carry out the insertion and filling steps of the thermodynamic cycle in Figure 1 using the rigid protein, as before. After the rigid protein has been solvated, we can relax the restraints in the solvated state from κ_C to the finite κ of interest. The free energy of this process is $\Delta G_r^{(sol)}(\kappa) - \Delta G_r^{(sol)}(\kappa_C)$. The protein solvation free energy as a function of the restraint strength, $\Delta G(\kappa)$, is then the sum of the solvation free energy of the rigid protein, $\Delta G_{rigid} \equiv \Delta G(\kappa_C)$, evaluated in the previous section, and the free energies of the two additional steps described above.

$$\Delta G(\kappa) = \Delta G_{rigid} + [\Delta G_r^{(vac)}(\kappa_C) - \Delta G_r^{(sol)}(\kappa_C)] - [\Delta G_r^{(vac)}(\kappa) - \Delta G_r^{(sol)}(\kappa)] \quad (17)$$

Moreover, the hydration free energy of the fully flexible protein, $\Delta G_{flex} \equiv \Delta G(\kappa \to 0)$, is then given by

$$\Delta G_{flex} = \Delta G_{rigid} + [\Delta G_r^{(vac)}(\kappa_C) - \Delta G_r^{(sol)}(\kappa_C)] \quad (18)$$

From eqs 17 and 18,

$$\Delta G(\kappa) = \Delta G_{flex} - [\Delta G_r^{(vac)}(\kappa) - \Delta G_r^{(sol)}(\kappa)] \quad (19)$$

From eqs 18 and 19, it is clear that by estimating $\Delta G_r^{(vac)}(\kappa)$ and $\Delta G_r^{(sol)}(\kappa)$ over the entire range of κ-values, the corresponding estimates of $\Delta G(\kappa)$ can be readily obtained. Solvation free energies,[72] as well as free energy differences between conformational states,[73,74] have previously been estimated using similar procedures.

The free energies associated with turning on the restraint potential in the solvated and vacuum states of the protein are shown in Figure 3a. As κ is increased, both $\Delta G_r^{(vac)}$ and $\Delta G_r^{(sol)}$ increase, as expected. However, only the difference between these free energies, shown in Figure 3b, contributes to $\Delta G(\kappa)$ (eq 19). For the protein, hydrophobin II, $\Delta G_r^{(vac)}(\kappa) - \Delta G_r^{(sol)}(\kappa)$ is positive for all κ-values, and plateaus at approximately 432 k_BT for large κ. Then, according to eq 18, the solvation free energy of the fully flexible protein is $\Delta G_{flex} = -607\ k_BT$. In comparison, the solvation free energy of the rigid protein, $\Delta G_{rigid} = -1039\ k_BT$, suggests a substantial influence of flexibility on the protein solvation free energy.

The protein solvation free energy as a function of the restraint strength, $\Delta G(\kappa)$, can also be readily obtained using eq 19, and is shown in Figure 3c. $\Delta G(\kappa)$ decreases monotonically (favorable) with increasing κ, indicating that protein hydration becomes more favorable as the protein is made increasingly rigid. This decrease in the hydration free energy of a protein stems directly from the greater expense of restraining a protein in vacuum than in water (eq 19). In other words, a solvated protein is better able to respond to the inclusion of restraints than a protein in vacuum, likely because when a solvated protein is restrained, it can relax not only through changes to its own structure, but also through reorientation of its hydration waters. To better illustrate the relationship between the hydration free energy of a protein and its flexibility, in Figure 3d we additionally plot ΔG as a function of the root mean squared displacement (RMSD) of the protein computed in the partially flexible hydrated state, and reference it to the corresponding rigid protein. As the protein becomes more flexible, its RMSD increases, and is accompanied by a concomitant increase in ΔG. Our finding that a rigid protein is hydrated more favorably is also consistent with the work of Debenedetti and co-workers, who demonstrated that the thermodynamic drive for water sorption by protein crystals increases when the rigidity of the proteins in the crystal is enhanced, either artificially or physically via disulfide bonds.[10,11]

THE ROLE OF SALT CONTENT

Most proteins can be rendered insoluble by the addition of certain salts; such a "salting out" of proteins is a commonly employed protein purification technique.[2,6] However, the phenomenon of salting-out depends on the nature of both the protein and the salt, and a molecular level understanding of specific ion effects in protein solubility, the so-called Hofmeister effects[7,75] remain elusive.[76−78] To shed light on how salt influences the strength of protein−water interactions for the HFBII protein, we estimate ΔG as well as its components, ΔG_{cav}, ΔG_{ins}, and ΔG_{fill}, as a function of sodium chloride concentration, c. Because the protein hydration shell now contains salt ions in addition to water, we augment the definition of $\tilde{N}_\nu(\bar{\mathbf{R}})$ to additionally include the salt ions in ν. The hydration free energy of the rigid protein as a function of salt concentration is shown in Figure 4a, and consistent with a "salting out" of solution, ΔG indeed increases (unfavorable) upon addition of NaCl. To uncover the molecular origins of

DOI: 10.1021/acs.jpcb.7b12060
J. Phys. Chem. B 2018, 122, 3635−3646

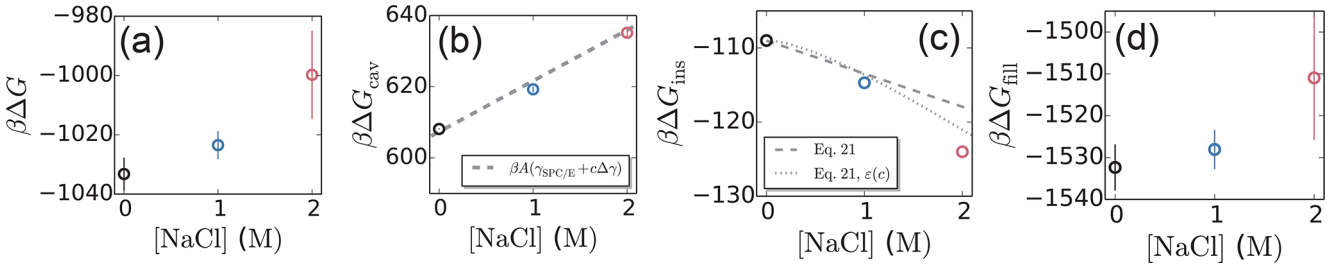


Figure 4. (a) The hydration free energy of hydrophobin II, obtained from the simulations (symbols), increases with increasing NaCl concentration, consistent with the mild salting out effect of this salt. Error bars are determined by block averaging and represent one standard deviation. (b) The cavity creation component of the hydration free energy, ΔG_{cav}, increases approximately linearly with NaCl concentration, consistent with the change in surface tension upon addition of salt. The dashed line is a fit to eq 20, as described in the text. (c) The insertion free energy, ΔG_{ins}, decreases roughly linearly with increasing salt concentration. The dashed line corresponds to $\Delta G_{ins}(c)$ predicted by eq 21 with $\Delta G_{ins}(0)$ determined from simulation and a fixed dielectric constant, whereas the dotted line incorporates a c-dependent dielectric constant, $\varepsilon(c)$, via eq 22. (d) The filling free energy, ΔG_{fill}, increases as NaCl is added to the solution.

this phenomenon, here we further examine the effects of salt concentration on the components of ΔG.

The vapor−liquid surface tension of water increases upon addition of NaCl as ions are excluded from the interface, and correspondingly, we expect ΔG_{cav} to increase with c.[79] As shown in Figure 4b, the dependence of ΔG_{cav} on c can indeed be described by

$$\Delta G_{cav}(c) \approx A(\gamma_{SPC/E} + c\Delta\gamma) \tag{20}$$

where $\Delta\gamma$ is the linear slope of the change in surface tension with NaCl concentration. Moreover, an unconstrained linear fit to the simulated ΔG_{cav}-values yields a reasonable estimate for $A \approx 41.6$ nm^2 (assuming $\gamma_{SPC/E} = 60.6$ mJ/nm^2), and $\Delta\gamma = 1.5$ mN m^{-1}/M, which is in good agreement with the value of roughly 2 mN m^{-1}/M obtained experimentally.[80]

Assuming that the insertion of the protein into the preformed cavity is dominated by long-ranged electrostatic interactions, we can once again use classic continuum treatments to understand how ΔG_{ins} depends on salt concentration. In particular, the theory developed by Kirkwood[61,81] predicts that

$$\Delta G_{ins}(c) \approx \Delta G_{ins}(0) - \frac{3\pi}{R}\left(\frac{\beta\mu}{\varepsilon}\right)^2 c \tag{21}$$

where an estimate for $\Delta G_{ins}(0)$ can be provided by eq 14. Because our focus is on the c-dependence of ΔG_{ins}, here we use the simulated value of $\Delta G_{ins}(0)$ instead. Plugging in previously discussed estimates of the cavity radius, R, the protein dipole moment, μ, and the dielectric constant of SPC/E water, ε, into eq 21, we obtain a functional form for $\Delta G_{ins}(c)$, which is depicted by the dashed line in Figure 4c, and is in qualitative agreement with the simulated data. The agreement is noteworthy given that it was obtained without the use of any fit parameters.

Even more refined estimates of $\Delta G_{ins}(c)$ can be obtained by accounting for the dependence of the dielectric constant on salt concentration, which is provided by the theory of Gavish and Promislow,[82]

$$\varepsilon(c) = \varepsilon(0) - \Delta\varepsilon L(3\alpha c/\Delta\varepsilon) \tag{22}$$

where $\Delta\varepsilon = \varepsilon(0) - \varepsilon_{ms}$, ε_{ms} is the molten salt limit of the dielectric constant, α is the excess polarization of the ions, and $L(x) = \coth(x) - 1/x$ is the Langevin function. Following Gavish and Promislow,[82] we use $\varepsilon_{ms} = 51.42$ and $\alpha = 11.5$ for NaCl solutions. Including this prediction for $\varepsilon(c)$ in eq 21 yields the dotted line in Figure 4c, which is in even better agreement with $\Delta G_{ins}(c)$, particularly at higher salt concentration, suggesting that combining eqs 21 and 22 yields excellent estimates of $\Delta G_{ins}(c)$.

We note that although eq 21 was originally intended to describe the electrostatic portion of the total solvation free energy, ΔG, the linear response assumptions inherent in the theory are expected to be reasonably accurate only for the long-ranged electrostatic interactions captured by ΔG_{ins}, and not for the corresponding short-ranged interactions that contribute to ΔG_{fill}.[59,60] Thus, the success of eqs 14, 21, and 22 in predicting ΔG_{ins}, not just qualitatively, but also quantitatively, is underpinned by the QCT partitioning of long-ranged and short-ranged electrostatic interactions into ΔG_{ins} and ΔG_{fill}, respectively, and by a suitable choice of r_v (i.e., $r_v \geq 0.6$ nm), which enforces that partitioning in practice.

The filling of the cavity in the final step of the QCT thermodynamic cycle is dominated by the formation of strong short-ranged interactions, such as hydrogen-bonding and local charge−charge "ion-pairing" interactions. Figure 4d highlights that the free energetics of this step, ΔG_{fill}, have a nonlinear dependence on salt concentration, with ΔG_{fill} increasing monotonically (less favorable) with increasing c. Thus, local protein−water interactions are more favorable in the absence of salt, and are weakened by the addition of salt. In contrast, ΔG_{ins}, which represents nonlocal long-ranged interactions, becomes more favorable (decreases) with increasing salt content.

The increase in the total solvation free energy of the protein with increasing salt content (Figure 4a), i.e., "salting out" of HFBII upon addition of NaCl, arises from ΔG_{cav} and ΔG_{fill}, both of which become increasingly unfavorable as salt is added and are opposed by ΔG_{ins}, which becomes favorable with the addition of salt. Thus, the increase in water−vapor surface tension upon addition of salt and the weakening of local protein−water interactions by NaCl are responsible for the "salting out" behavior. In contrast, long-ranged protein−solvent interactions, which can be accurately described by continuum theories, tend to instead favor "salting in" of the protein.

CONCLUSIONS

In this work, we introduce an efficient computational framework for estimating protein hydration free energies, ΔG, using all-atom MD simulations with an explicit representation of the solvent. We hydrate the protein in three steps (Figure 1); a cavity slightly larger than the protein is first created in bulk

DOI: 10.1021/acs.jpcb.7b12060
J. Phys. Chem. B 2018, 122, 3635−3646

water; the protein is then introduced into that cavity to establish long-ranged interactions between the protein and water; and finally, the short-ranged protein−water interactions are formed by filling the cavity with water. We show that to a good approximation, the free energetics of cavity formation (ΔG_{cav}) can be estimated by considering the creation of liquid−vapor interfacial area. Similarly, protein insertion free energetics (ΔG_{ins}) can be estimated by considering the polarization of the liquid−vapor interface when a dipole (protein) is introduced into the preformed cavity. In contrast, the free energetics of the short-ranged protein−water interactions (ΔG_{fill}) depend non-trivially on the chemical and topographical context presented by the protein surface.[12−20]

By leveraging an understanding of the free energetics of water density fluctuations, both in bulk water and in the protein hydration shell,[43,46,83] and by employing efficient simulation methods for characterizing such free energetics,[17,30] we are able to mitigate the computational expense of estimating ΔG_{cav} and ΔG_{fill}, thereby allowing us to estimate ΔG for the protein hydrophobin II using only 15 ns of simulation data. A particular strength of the approach is that approximate estimates of ΔG (with errors of less than 5%) can be obtained with minimal computational effort, and if required, the accuracy can be progressively enhanced by spending more computational effort. Due to its efficiency, the approach presented here can serve as a basis for exploring the molecular underpinnings of the diverse determinants of biomolecular solubility, ranging from protein mutations and solution additives to environmental conditions, such as temperature or pH. Here, we explore how the hydration free energy of a protein is influenced by its flexibility and by the inclusion of salt in the solution.

We find that a protein can be hydrated more favorably by enhancing its rigidity; for example, by engineering it to have a higher content of rigidifying residues, like proline,[84−87] a high concentration of strong interactions, like salt bridges and disulfide bonds,[11,85,87−89] or through chemical cross-linking.[90−92] Conversely, enhancing protein flexibility would serve to lower its propensity to remain hydrated, and in principle, it could not only lower protein solubility, but also be used to trigger protein assemblies. Intraprotein interactions, such as disulfide bonds or salt bridges, are also used to stabilize the folded state of a protein.[85,87,88] In addition to stabilizing the folded state directly, our findings suggest that such favorable interactions ought to also favor the folded state indirectly through their role in enhancing protein rigidity. In particular, consider the solvent-induced folding free energy, ΔG_{fold}, that is, the difference between the free energies of folding the protein in water and in vacuum, which is equal to the difference in the solvation free energies of the folded and unfolded states, $\Delta G_{fold} = \Delta G_F - \Delta G_U$. Because disulfide bonds and salt bridges are typically broken when the protein is unfolded, they ought to have little effect on ΔG_U. In contrast, such interactions are present in the folded state, and our results suggest that by increasing protein rigidity, they make ΔG_F, and thereby ΔG_{fold}, more favorable.

We also find that the hydration free energy, ΔG, of the protein hydrophobin II increases with increasing NaCl concentration, c, which is consistent with the known decrease in the solubility of proteins upon addition of NaCl. The increase in ΔG with increasing c can be traced to the corresponding increase in both $\Delta G_{cav}(c)$ and $\Delta G_{fill}(c)$. In contrast, $\Delta G_{ins}(c)$ decreases (favoring hydration) as c is increased. Because only long-ranged interactions contribute to ΔG_{ins}, we do not expect its decrease upon addition of salt to depend on the specific nature of the salt ions. In contrast, we expect both $\Delta G_{cav}(c)$ and $\Delta G_{fill}(c)$ to depend on the particular salt ions being added to the solution. We show that the c-dependence of ΔG_{cav} derives primarily from that of the liquid−vapor surface tension $\gamma(c)$, which usually increases, but can also decrease with increasing c, depending upon which ions are added to the solution.[80,93−95] The c-dependence of ΔG_{fill} is also expected to be ion specific, and $\Delta G_{fill}(c)$ may either increase or decrease with increasing c, depending not only on how protein−water interactions are modulated by the presence of the ions, but also on the direct "ion-pairing" interactions between charged protein residues and salt ions.

APPENDIX: CHOOSING AN OPTIMAL r_v

The physical significance of the free energy components of ΔG (as defined by the QCT thermodynamic cycle of Fig. 1), as well

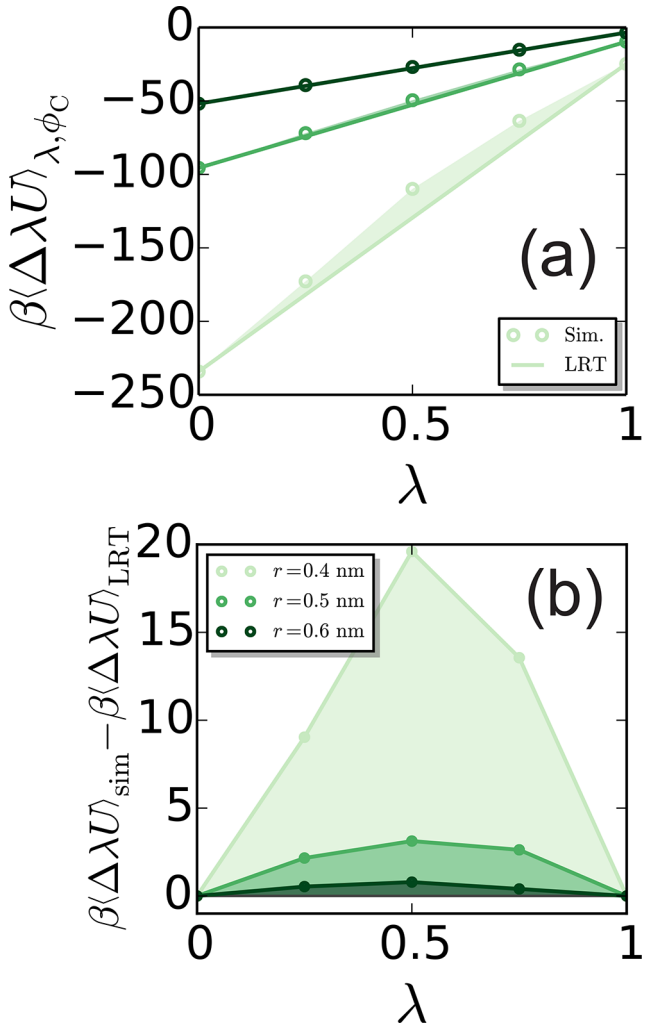


Figure 5. (a) The average insertion energy determined from simulation (Sim, points) and linear response theory (LRT, lines) agrees for large cavity sizes, while LRT fails as the cavity size is decreased. (b) The difference between the actual (sim) and LRT estimated average energies indicates that LRT becomes more accurate as the size of the cavity, r_v, is increased. The shaded areas indicate the error in the insertion free energy by using LRT, which becomes smaller as r_v is increased. Note that the areas of the shaded regions are the same in (a) and (b). Color scheme is the same in both panels.

DOI: 10.1021/acs.jpcb.7b12060
J. Phys. Chem. B 2018, 122, 3635−3646

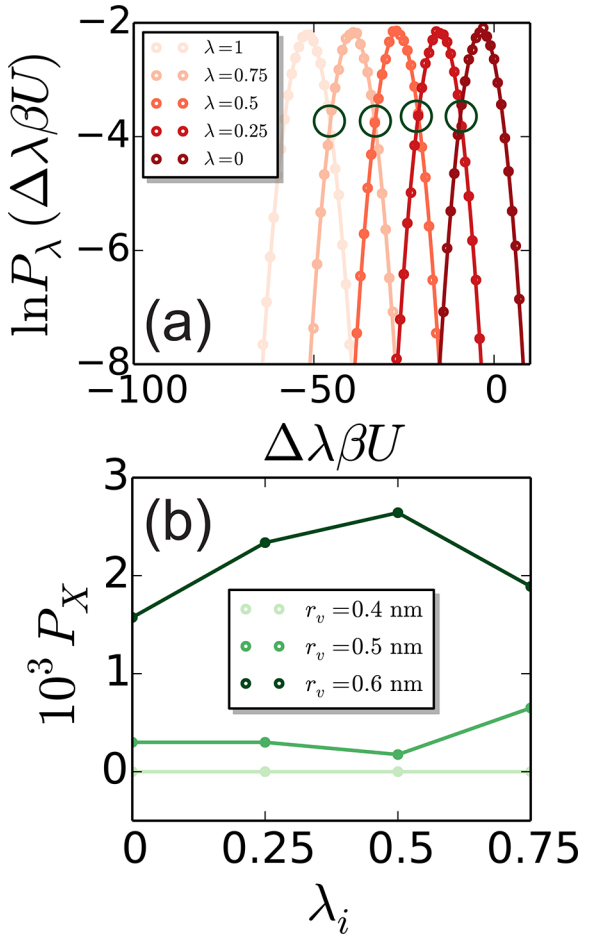


Figure 6. (a) The probability distributions of the protein−water interaction energy, $P_\lambda(\Delta\lambda\beta U)$, follow Gaussian statistics, for a well-chosen cavity, such as those shown here for the r_v = 0.6 nm volume. Symbols indicate simulation data and lines are Gaussian distributions with the same mean and variance as the simulation data. (b) The probability, P_X, where neighboring distributions are in λ-space cross, indicates that phase space overlap also increases as the cavity volume is increased. P_X is schematically illustrated by the circles in (a).

as the overall computational efficiency of the approach are strongly influenced by the size (and shape) of the cavity volume, v, that is employed. We define v as the union of spherical sub-volumes of radius r_v centered on each of the protein heavy atoms, so that it complements the size (and shape) of the protein. For simplicity, here we use the same r_v for all protein heavy atoms, but in principle a different r_v can be chosen for each heavy atom, if needed. When an additional inner volume is employed to reversibly add/remove strongly bound water molecules in the filling step (as detailed in the SI), the r_v discussed here refers to the outer volume; the boundary of the inner volume is not varied.

The use of a cavity in the thermodynamic cycle of Figure 1 serves to partition protein−solvent interactions into short- and long-ranged components,[29,42] so that upon insertion of the protein into the pre-formed cavity, it does not experience short-ranged protein−water interactions, such as repulsions or hydrogen bonds. The radius, r_v, then represents the length-scale over which these interactions are separated, with long-ranged protein−water interactions contributing to ΔG_{ins}, and the corresponding short-ranged interactions being captured by the sum of ΔG_{cav} and ΔG_{fill}.

If r_v is chosen to be too small (e.g., the size of protein heavy atoms or roughly 0.3 nm), the protein experiences short-ranged interactions upon insertion, with the alchemical approach being recovered in the limit of vanishing r_v. In this limit, ΔG_{cav} and ΔG_{fill} vanish, and ΔG is dominated by the protein insertion free energy, ΔG_{ins}. However, to estimate ΔG_{ins}, protein−water interactions must be turned on incrementally, necessitating the simulation of a large number of alchemical states, whose solvation free energies are not particularly meaningful from a physical standpoint.

In contrast, for a sufficiently large value of r_v, because the slowly varying and long-ranged interactions that contribute to ΔG_{ins} perturb the solvent structure in gradual manner, we expect $\langle U(\bar{\mathbf{R}})\rangle_{\lambda,\phi_C}$ to respond linearly to λ, that is

$$\langle U(\bar{\mathbf{R}})\rangle^{LRT}_{\lambda,\phi_C} = \langle U(\bar{\mathbf{R}})\rangle_{0,\phi_C} + \lambda[\langle U(\bar{\mathbf{R}})\rangle_{1,\phi_C} - \langle U(\bar{\mathbf{R}})\rangle_{0,\phi_C}] \quad (23)$$

Correspondingly, we also expect $P_\lambda(\Delta\lambda U)$ to display Gaussian statistics, with increased overlap between adjacent λ-ensembles as r_v is increased, enabling estimation of ΔG_{ins} with only a few (or no) intermediate alchemical states. However, the cancellation between the free energetics of cavity creation ($\Delta G_{cav} \geq 0$) and cavity filling ($\Delta G_{fill} \leq 0$) increases as r_v is increased, resulting in substantial numerical uncertainties in the free energetics of the short-ranged solute−solvent interactions.

We thus recommend that r_v (and correspondingly, the cavity size) be chosen to be large enough to yield distributions of $P_\lambda(\Delta\lambda U)$ that are Gaussian, but no longer to reduce cancellation between ΔG_{cav} and ΔG_{fill}, and the associated uncertainty in $\Delta G_{cav} + \Delta G_{fill}$. The results in Figures 5 and 6 suggest that the appropriate balance is provided by r_v = 0.6 nm, which is roughly the size of a protein heavy atom plus that of a water molecule. Thus, the optimal choice of v corresponds to it encompassing roughly the first hydration shell of the protein.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.7b12060.

Simulation details, additional data, and discussion of the determination of the free energies using umbrella sampling, and details pertaining to the reversible sampling of protein hydration and dehydration (PDF)

■ AUTHOR INFORMATION

Corresponding Author

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

ORCID

Amish J. Patel: 0000-0001-6482-543X

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

A.J.P. gratefully acknowledges financial support from the National Science Foundation (UPENN MRSEC DMR 11-20901, CBET 1652646, and CHE 1665339) and a fellowship from the Alfred P. Sloan Research Foundation. The authors would also like to thank Pablo Debenedetti and Sang Kim for helpful discussions.

DOI: 10.1021/acs.jpcb.7b12060
J. Phys. Chem. B 2018, 122, 3635−3646

REFERENCES

(1) Shen, V. K.; Cheung, J. K.; Errington, J. R.; Truskett, T. M. Coarse-Grained Strategy for Modeling Protein Stability in Concentrated Solutions. II: Phase Behavior. *Biophys. J.* **2006**, *90*, 1949−1960.

(2) Kramer, R. M.; Shende, V. R.; Motl, N.; Pace, C. N.; Scholtz, J. M. Toward a Molecular Understanding of Protein Solubility: Increased Negative Surface Charge Correlates with Increased Solubility. *Biophys. J.* **2012**, *102*, 1907−1915.

(3) Hung, J. J.; Borwankar, A. U.; Dear, B. J.; Truskett, T. M.; Johnston, K. P. High Concentration Tangential Flow Ultrafiltration of Stable Monoclonal Antibody Solutions with Low Viscosities. *J. Membr. Sci.* **2016**, *508*, 113−126.

(4) Johnston, K. P.; Maynard, J. A.; Truskett, T. M.; Borwankar, A. U.; Miller, M. A.; Wilson, B. K.; Dinin, A. K.; Khan, T. A.; Kaczorowski, K. J. Concentrated Dispersions of Equilibrium Protein Nanosclusters That Reversibly Dissociate into Active Monomers. *ACS Nano* **2012**, *6*, 1357−1369.

(5) Shire, S. J.; Shahrokh, Z.; Liu, J. Challenges in the Development of High Protein Concentration Formulations. *J. Pharm. Sci.* **2004**, *93*, 1390−1402.

(6) Trevino, S. R.; Scholtz, J. M.; Pace, C. N. Measuring and Increasing Protein Solubility. *J. Pharm. Sci.* **2008**, *97*, 4155−4166.

(7) Zhang, Y.; Cremer, P. S. The Inverse and Direct Hofmeister Series for Lysozyme. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 15249−15253.

(8) Chong, S.-H.; Ham, S. Interaction with the Surrounding Water Plays a Key Role in Determining the Aggregation Propensity of Proteins. *Angew. Chem., Int. Ed.* **2014**, *53*, 3751−3751.

(9) Kim, S. B.; Palmer, J. C.; Debenedetti, P. G. A Computational Study of the Effect of Matrix Structural Order on Water Sorption by Trp-Cage Miniproteins. *J. Phys. Chem. B* **2015**, *119*, 1847−1856.

(10) Palmer, J. C.; Debenedetti, P. Computer Simulation of Water Sorption on Flexible Protein Crystals. *J. Phys. Chem. Lett.* **2012**, *3*, 2713−2718.

(11) Kim, S. B.; Singh, R. S.; Paul, P. K. C.; Debenedetti, P. G. Effects of Disulfide Bridges and Backbone Connectivity on Water Sorption by Protein Matrices. *Sci. Rep.* **2017**, *7*, 7957.

(12) Giovambattista, N.; Debenedetti, P. G.; Rossky, P. J. Hydration Behavior Under Confinement by Nanoscale Surfaces with Patterned Hydrophobicity and Hydrophilicity. *J. Phys. Chem. C* **2007**, *111*, 1323−1332.

(13) Giovambattista, N.; Lopez, C. F.; Rossky, P. J.; Debenedetti, P. G. Hydrophobicity of Protein Surfaces: Separating Geometry from Chemistry. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 2274−2279.

(14) Jamadagni, S. N.; Godawat, R.; Garde, S. Hydrophobicity of Proteins and Interfaces: Insights from Density Fluctuations. *Annu. Rev. Chem. Biomol. Eng.* **2011**, *2*, 147−171.

(15) Wang, J.; Bratko, D.; Luzar, A. Probing Surface Tension Additivity on Chemically Heterogeneous Surfaces by a Molecular Approach. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 6374−6379.

(16) Ritchie, J. A.; Yazdi, J. S.; Bratko, D.; Luzar, A. Metastable Sessile Nanodroplets on Nanopatterned Surfaces. *J. Phys. Chem. C* **2012**, *116*, 8634−8641.

(17) Patel, A. J.; Garde, S. Efficient Method to Characterize the Context-Dependent Hydrophobicity of Proteins. *J. Phys. Chem. B* **2014**, *118*, 1564−1573.

(18) Factorovich, M. H.; Molinero, V.; Scherlis, D. A. Hydrogen-Bond Heterogeneity Boosts Hydrophobicity of Solid Interfaces. *J. Am. Chem. Soc.* **2015**, *137*, 10618−10623.

(19) Remsing, R. C.; Weeks, J. D. Hydrophobicity Scaling of Aqueous Interfaces by an Electrostatic Mapping. *J. Phys. Chem. B* **2015**, *119*, 9268−9277.

(20) Xi, E.; Venkateshwaran, V.; Li, L.; Rego, N.; Patel, A. J.; Garde, S. Hydrophobicity of Proteins and Nanostructured Solutes Is Governed by Topographical and Chemical Context. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 13345−13350.

(21) O'Brien, E. P.; Dima, R. I.; Brooks, B.; Thirumalai, D. Interactions Between Hydrophobic and Ionic Solutes in Aqueous Guanidinium Chloride and Urea Solutions: Lessons for Protein Denaturation Mechanism. *J. Am. Chem. Soc.* **2007**, *129*, 7346−7353.

(22) O'Brien, E. P.; Ziv, G.; Haran, G.; Brooks, B. R.; Thirumalai, D. Effects of Denaturants and Osmolytes on Proteins Are Accurately Predicted by the Molecular Transfer Model. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 13403−13408.

(23) Athawale, M. V.; Sarupria, S.; Garde, S. Enthalpy-Entropy Contributions to Salt and Osmolyte Effects on Molecular-Scale Hydrophobic Hydration and Interactions. *J. Phys. Chem. B* **2008**, *112*, 5661−5670.

(24) Berne, B. J.; Weeks, J. D.; Zhou, R. Dewetting and Hydrophobic Interaction in Physical and Biological Systems. *Annu. Rev. Phys. Chem.* **2009**, *60*, 85−103.

(25) England, J. L.; Haran, G. Role of Solvation Effects in Protein Denaturation: From Thermodynamics to Single Molecules and Back. *Annu. Rev. Phys. Chem.* **2011**, *62*, 257−277.

(26) Canchi, D. R.; García, A. E. Cosolvent Effects on Protein Stability. *Annu. Rev. Phys. Chem.* **2013**, *64*, 273−293.

(27) Widom, B. Some Topics in the Theory of Fluids. *J. Chem. Phys.* **1963**, *39*, 2808−2812.

(28) Widom, B. Potential-Distribution Theory and the Statistical Mechanics of Fluids. *J. Phys. Chem.* **1982**, *86*, 869−872.

(29) Beck, T. L.; Paulaitis, M. E.; Pratt, L. R. *The Potential Distirbution Theorem and Models of Molecular Solutions*; Cambridge University Press, 2006.

(30) Patel, A. J.; Varilly, P.; Chandler, D.; Garde, S. Quantifying Density Fluctuations in Volumes of All Shapes and Sizes Using Indirect Umbrella Sampling. *J. Stat. Phys.* **2011**, *145*, 265−275.

(31) Kallio, J. M.; Rouvinen, J. Amphiphilic Nanotubes in the Crystal Structure of a Biosurfactant Protein Hydrophobin HFBII. *Chem. Commun.* **2011**, *47*, 9843−9845.

(32) Khalesi, M.; Gebruers, K.; Derdelinckx, G. Recent Advances in Fungal Hydrophobin Towards Using in Industry. *Protein J.* **2015**, *34*, 243−255.

(33) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. Extremely Precise Free Energy Calculations of Amino Acid Side Chain Analogs: Comparison of Common Molecular Mechanics Force Fields for Proteins. *J. Chem. Phys.* **2003**, *119*, 5740−5761.

(34) *Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Chipot, C., Pohorille, A., Eds.; Springer, 2007.

(35) Shirts, M. R.; Mobley, D. L.; Chodera, J. D. *Annu. Rep. Comput. Chem.*; Elsevier Ltd., 2007; Vol. 3; Chapter 4, pp 41−59.

(36) Mobley, D. L.; Klimovich, P. V. Perspective: Alchemical Free Energy Calculations for Drug Discovery. *J. Chem. Phys.* **2012**, *137*, 230901.

(37) Pohorille, A.; Jarzynski, C.; Chipot, C. Good Practices in Free-Energy Calculations. *J. Phys. Chem. B* **2010**, *114*, 10235−10253.

(38) Wang, J.; Deng, Y.; Roux, B. Absolute Binding Free Energy Calculations Using Molecular Dynamics Simulations with Restraining Potentials. *Biophys. J.* **2006**, *91*, 2798−2814.

(39) Deng, Y.; Roux, B. Calculation of Standard Binding Free Energies: Aromatic Molecules in the T4 Lysozyme L99a Mutant. *J. Chem. Theory Comput.* **2006**, *2*, 1255−1273.

(40) Shivakumar, D.; Deng, Y.; Roux, B. Computations of Absolute Solvation Free Energies of Small Molecules Using Explicit and Implicit Solvent Model. *J. Chem. Theory Comput.* **2009**, *5*, 919−930.

(41) Pham, T. T.; Shirts, M. R. Optimal Pairwise and Non-Pairwise Alchemical Pathways for Free Energy Calculations of Molecular Transformation in Solution Phase. *J. Chem. Phys.* **2012**, *136*, 124120.

(42) Weber, V.; Asthagiri, D. Regularizing Binding Energy Distributions and the Hydration Free Energy of Protein Cytochrome C from All-Atom Simulations. *J. Chem. Theory Comput.* **2012**, *8*, 3409−3415.

(43) Patel, A. J.; Varilly, P.; Chandler, D. Fluctuations of Water Near Extended Hydrophobic and Hydrophilic Surfaces. *J. Phys. Chem. B* **2010**, *114*, 1632−1637.

(44) Patel, A. J.; Varilly, P.; Jamadagni, S. N.; Acharya, H.; Garde, S.; Chandler, D. Extended Surfaces Modulate Hydrophobic Interactions

DOI: 10.1021/acs.jpcb.7b12060
J. Phys. Chem. B 2018, 122, 3635−3646

of Neighboring Solutes. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 17678−17683.
(45) Remsing, R. C.; Patel, A. J. Water Density Fluctuations Relevant to Hydrophobic Hydration Are Unaltered by Attractions. *J. Chem. Phys.* **2015**, *142*, 024502.
(46) Patel, A. J.; Varilly, P.; Jamadagni, S. N.; Hagan, M. F.; 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.
(47) Xi, E.; Remsing, R. C.; Patel, A. J. Sparse Sampling of Water Density Fluctuations in Interfacial Environments. *J. Chem. Theory Comput.* **2016**, *12*, 706−713.
(48) Hakanpaa, J.; Linder, M.; Popov, A.; Schmidt, A.; Rouvinen, J. Hydrophobin HFBII in Detail: Ultrahigh-Resolution Structure at 0.75 A. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2006**, *62*, 356−367.
(49) Ashbaugh, H. S.; Truskett, T. M. Putting the Squeeze on Cavities in Liquids: Quantifying Pressure Effects on Solvation Using Simulations and Scaled-Particle Theory. *J. Chem. Phys.* **2011**, *134*, 014507.
(50) Vega, C.; de Miguel, E. Surface Tension of the Most Popular Models of Water by Using the Test-Area Simulation Method. *J. Chem. Phys.* **2007**, *126*, 154707.
(51) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(52) Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* **2008**, *129*, 124105.
(53) Tan, Z.; Gallicchio, E.; Lapelosa, M.; Levy, R. M. Theory of Binless Multi-State Free Energy Estimation with Applications to Protein-Ligand Binding. *J. Chem. Phys.* **2012**, *136*, 144102.
(54) Widom, B. Intermolecular Forces and the Nature of the Liquid State. *Science* **1967**, *157*, 375−382.
(55) Weeks, J. D.; Chandler, D.; Andersen, H. C. Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids. *J. Chem. Phys.* **1971**, *54*, 5237−5247.
(56) Chandler, D.; Weeks, J. D.; Andersen, H. C. Van Der Waals Picture of Liquids, Solids, and Phase Transformations. *Science* **1983**, *220*, 787−794.
(57) Raineri, F. O.; Stell, G.; Ben-Amotz, D. New Mean-Energy Formulae for Free Energy Differences. *Mol. Phys.* **2005**, *103*, 3209−3221.
(58) Chandler, D. Interfaces and the Driving Force of Hydrophobic Assembly. *Nature* **2005**, *437*, 640−647.
(59) Remsing, R. C.; Liu, S.; Weeks, J. D. Long-Ranged Contributions to Solvation Free Energies from Theory and Short-Ranged Models. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 2819−2826.
(60) Remsing, R. C.; Weeks, J. D. Role of Local Response in Ion Solvation: Born Theory and Beyond. *J. Phys. Chem. B* **2016**, *120*, 6238−6249.
(61) Kirkwood, J. G. Theory of Solutions of Molecules Containing Widely Separated Charges with Special Application to Zwitterions. *J. Chem. Phys.* **1934**, *2*, 351−361.
(62) Felder, C. E.; Prilusky, J.; Silman, I.; Sussman, J. L. A Server and Database for Dipole Moments of Proteins. *Nucleic Acids Res.* **2007**, *35*, W512−W521.
(63) Koppenol, W. H.; Rush, J. D.; Mills, J. D.; Margoliash, E. The Dipole Moment of Cytochrome C. *Mol. Biol. Evol.* **1991**, *8*, 545−558.
(64) Ashbaugh, H. S. Convergence of Molecular and Macroscopic Continuum Descriptions of Ion Hydration. *J. Phys. Chem. B* **2000**, *104*, 7235−7238.
(65) Rajamani, S.; Ghosh, T.; Garde, S. Size Dependent Ion Hydration, Its Asymmetry, and Convergence to Macroscopic Behavior. *J. Chem. Phys.* **2004**, *120*, 4457.
(66) Mobley, D. L., II; Barber, A. E.; Fennell, C. J.; Dill, K. A. Charge Asymmetries in Hydration of Polar Solutes. *J. Phys. Chem. B* **2008**, *112*, 2405−2414.
(67) Bardhan, J. P.; Jungwirth, P.; Makowski, L. Affine-Response Model of Molecular Solvation of Ions: Accurate Predictions of Asymmetric Charging Free Energies. *J. Chem. Phys.* **2012**, *137*, 124101.
(68) Remsing, R. C.; Baer, M. D.; Schenter, G. K.; Mundy, C. J.; Weeks, J. D. The Role of Broken Symmetry in Solvation of a Spherical Cavity in Classical and Quantum Water Models. *J. Phys. Chem. Lett.* **2014**, *5*, 2767−2774.
(69) Duignan, T. T.; Baer, M. D.; Schenter, G. K.; Mundy, C. J. Electrostatic Solvation Free Energies of Charged Hard Spheres Using Molecular Dynamics with Density Functional Theory Interactions. *J. Chem. Phys.* **2017**, *147*, 161716.
(70) Born, M. Volumes and Hydration Warmth of Ions. *Eur. Phys. J. A* **1920**, *1*, 45−48.
(71) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations. *J. Chem. Theory Comput.* **2009**, *5*, 350−358.
(72) Esque, J.; Cecchini, M. Accurate Calculation of Conformational Free Energy Differences in Explicit Water: The Confinement-Solvation Free Energy Approach. *J. Phys. Chem. B* **2015**, *119*, 5194−5207.
(73) Cecchini, M.; Krivov, S. V.; Spichty, M.; Karplus, M. Calculation of Free-Energy Differences by Confinement Simulations. Application to Peptide Conformers. *J. Phys. Chem. B* **2009**, *113*, 9728−9740.
(74) Tyka, M. D.; Clarke, A. R.; Sessions, R. B. An Efficient, Path-Independent Method for Free-Energy Calculations. *J. Phys. Chem. B* **2006**, *110*, 17212−17220.
(75) Collins, K. D. Ions from the Hofmeister Series and Osmolytes: Effects on Proteins in Solution and in the Crystallization Process. *Methods* **2004**, *34*, 300−311.
(76) Collins, K. D. Ion Hydration: Implications for Cellular Function, Polyelectrolytes, and Protein Crystallization. *Biophys. Chem.* **2006**, *119*, 271−281.
(77) Collins, K. D.; Neilson, G. W.; Enderby, J. E. Ions in Water: Characterizing the Forces That Control Chemical Processes and Biological Structure. *Biophys. Chem.* **2007**, *128*, 95−104.
(78) Collins, K. D. Why Continuum Electrostatics Theories Cannot Explain Biological Structure, Polyelectrolytes or Ionic Strength Effects in Ion-Protein Interactions. *Biophys. Chem.* **2012**, *167*, 43−49.
(79) Melander, W.; Horváth, C. Salt Effects on Hydrophobic Interactions in Precipitation and Chromatography of Proteins: An Interpretation of the Lyotropic Series. *Arch. Biochem. Biophys.* **1977**, *183*, 200−215.
(80) Weissenborn, P. K.; Pugh, R. J. Surface Tension of Aqueous Solutions of Electrolytes: Relationship with Ion Hydration, Oxygen Solubility, and Bubble Coalescence. *J. Colloid Interface Sci.* **1996**, *184*, 550−563.
(81) Kirkwood, J. G. In *Proteins, Amino Acids, and Peptides As Ions and Dipolar Ions*; American Chemical Society Monograph Series 90;Cohn, E. J., Edsall, J. T., Eds.; Reinhold Publishing Corporation: New York, NY, 1943; Chapter 12, pp 276−303.
(82) Gavish, N.; Promislow, K. Dependence of the Dielectric Constant of Electrolyte Solutions on Ionic Concentration: A Microfield Approach. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **2016**, *94*, 012611.
(83) Lum, K.; Chandler, D.; Weeks, J. D. Hydrophobicity at Small and Large Length Scales. *J. Phys. Chem. B* **1999**, *103*, 4570−4577.
(84) Jaenicke, R. Stability and Stabilization of Globular Proteins in Solution. *J. Biotechnol.* **2000**, *79*, 193−203.
(85) Yu, H.; Huang, H. Engineering Proteins for Thermostability Through Rigidifying Flexible Sites. *Biotechnol. Adv.* **2014**, *32*, 308−315.
(86) Yu, H.; Zhao, Y.; Guo, C.; Gan, Y.; Huang, H. The Role of Proline Substitutions Within Flexible Regions on Thermostability of Luciferase. *Biochim. Biophys. Acta, Proteins Proteomics* **2015**, *1854*, 65−72.
(87) Karshikoff, A.; Nilsson, L.; Ladenstein, R. Rigidity Versus Flexibility: The Dilemma of Understanding Protein Thermal Stability. *FEBS J.* **2015**, *282*, 3899−3917.

DOI: 10.1021/acs.jpcb.7b12060
J. Phys. Chem. B 2018, 122, 3635−3646

(88) Fass, D. Disulfide Bonding in Protein Biophysics. *Annu. Rev. Biophys.* **2012**, *41*, 63−79.
(89) Liu, T.; Wang, Y.; Luo, X.; Li, J.; Reed, S. A.; Xiao, H.; Young, T. S.; Schultz, P. G. Enhancing Protein Stability with Extended Disulfide Bonds. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 5910−5915.
(90) Sinz, A. Chemical Cross-Linking and Mass Spectrometry to Map Three-Dimensional Protein Structures and Protein-Protein Interactions. *Mass Spectrom. Rev.* **2006**, *25*, 663−682.
(91) Leitner, A.; Walzthoeni, T.; Kahraman, A.; Herzog, F.; Rinner, O.; Beck, M.; Aebersold, R. Probing Native Protein Structures by Chemical Cross-Linking, Mass Spectrometry, and Bioinformatics. *Mol. Cell. Proteomics* **2010**, *9*, 1634−1649.
(92) Leitner, A.; Joachimiak, L. A.; Unverdorben, P.; Walzthoeni, T.; Frydman, J.; Förster, F.; Aebersold, R. Chemical Cross-Linking/Mass Spectrometry Targeting Acidic Residues in Proteins and Protein Complexes. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 9455−9460.
(93) Netz, R. R.; Horinek, D. Progress in Modeling of Ion Effects at the Vapor/Water Interface. *Annu. Rev. Phys. Chem.* **2012**, *63*, 401−418.
(94) Geissler, P. L. Water Interfaces, Solvation, and Spectroscopy. *Annu. Rev. Phys. Chem.* **2013**, *64*, 317−337.
(95) Tobias, D. J.; Stern, A. C.; Baer, M. D.; Levin, Y.; Mundy, C. J. Simulation and Theory of Ions at Atmospherically Relevant Aqueous Liquid-Air Interfaces. *Annu. Rev. Phys. Chem.* **2013**, *64*, 339−359.

DOI: 10.1021/acs.jpcb.7b12060
J. Phys. Chem. B 2018, 122, 3635−3646