JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Regularizing Binding Energy Distributions and the Hydration Free Energy of Protein Cytochrome C from All-Atom Simulations

Valéry Weber[†] and D. Asthagiri*[,‡]

[†]IBM Research, Zurich, Switzerland

[‡]Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland 21218, United States

ABSTRACT: By introducing an external field to temper short-range protein water interactions, we regularize the statistical problem of calculating the hydration free energy, μ^{ex}, of the protein cytochrome C using the potential distribution theorem. Using this approach, we calculate the nonelectrostatic (dispersion) and electrostatic contributions to μ^{ex}. The nonelectrostatic contribution interpreted within an accessible surface area approach leads to a surface energy parameter that is about twice the value based on the hydration of small alkanes: at the size scale of the protein, hydrophobic hydration is more stronger relative to small alkanes. The electrostatic contribution does not obey linear response behavior. Further, depending on the choice of the protein dielectric constant, continuum dielectric calculations of the electrostatic contribution differ from the all-atom result by between 6%−12% (in a net value of about −2000 kcal/mol). We conclude by indicating potential applications of the present physically transparent approach toward illuminating the role of water, ions, and osmolytes in protein solution thermodynamics, including in protein folding and aggregation.

1. INTRODUCTION

Protein structure, function, and organization is determined by the solvent water and the dissolved ions and osmolytes that together constitute the solvent medium. The excess free energy of the protein due to its interaction with this solvent medium, the hydration free energy of the protein when the solvent is pure water, is a fundamental quantity characterizing the solution thermodynamics of the protein. Being able to characterize the excess free energy of the protein, for example, can illuminate the role of water and the solvent additive(s) in determining the protein conformation or state of aggregation, aspects that are of fundamental scientific[1−7] and technological[8,9] importance.

Computer simulations of a protein in a bath of water molecules can provide molecular level insight into the structural and dynamical aspects of hydration, and they can, in principle, be used to calculate the protein's hydration free energy. However, while all-atom calculations of hydration free energies are routine for small molecular solutes, except for small peptides containing on the order of ten amino acid residues,[10,11] such calculations are daunting on the scale of a folded protein. Perhaps this is one reason why implicit solvent models continue to be widely used for modeling proteins in solution,[12] although the consequences of an implicit description of hydration are not always apparent. Moreover, such implicit models are inadequate in exploring ion and osmolyte effects in protein solution thermodynamics.

Formally, the excess free energy of a solute in solution, μ^{ex}, is given by the potential distribution theorem[13,14]

$$\mu^{ex} = k_B T \ln\langle e^{\Delta U/k_B T}\rangle \quad (1)$$

where ΔU is the interaction energy of the solute with the solution, T is the temperature, and $\langle ...\rangle$ indicates averaging over all solute−solvent configurations. The challenge in calculating μ^{ex} stems from the wide range of energies with which the solute interacts with the solvent: relative to thermal energy, the short-range (first-hydration shell) solute−solvent interactions are typically strong, even rivaling covalent bonds in the case of ionic groups. Farther away from the solute−solvent interface, the interaction energy is weaker. Using eq 1 requires being able to sample the binding energy distribution and, in particular, the high ΔU values effectively, a difficult challenge in the face of the disparate energies of interaction. (To better appreciate this difficulty, it helps to consider the conjugate of eq 1 that requires sampling low-ΔU states by inserting a solute in pure solvent configurations, an approach that usually fails for molecular solutes.) The traditional approach thus relies on alchemically transforming the solute from a noninteracting point to the physical solute,[15] a challenging proposition on the scale of a globular protein. Here, we present an approach that obviates the need for alchemically transforming the solute. Our approach is founded on the quasichemical organization of the potential distribution theorem[14,16−18] and has the added virtue of providing information on the hydrophobic and hydrophilic contributions to hydration.

The quasichemical (QC) organization of the potential distribution theorem (PDT) is based on appreciating the existence of the disparate energy scales noted above. In QC/PDT, one defines a coordination domain around the solute, the spatial demarcation serving to separate the strong, short-range solute−solvent interactions from the weaker, longer-range interactions, and the PDT is recast to reveal the hydrophobic, local hydrophilic (chemical) and long-range contributions to the hydration free energy. In essence, the spatial exclusion regularizes the statistical problem of calculating the hydration free energy using the PDT.[19,20]

Received: June 18, 2012
Published: August 21, 2012

   dx.doi.org/10.1021/ct300505b | J. Chem. Theory Comput. 2012, 8, 3409−3415

The QC/PDT approach has illuminated the physics of hydration of water,[21,22] aqueous ions,[23−26] and hydrophobic solutes[27−29] and helped elucidate results from *ab initio* molecular dynamics simulations.[30−36] A virtue of the theory is that it transparently reveals the underlying physical contributions to hydration, making it particularly apposite for our goal of understanding ion and osmolyte effects in protein solution thermodynamics. In pursuit of this goal, here we use the QC/PDT-inspired regularization approach[19,20] to calculate the hydration free energy, μ^{ex}, of cytochrome C with an explicit account of water molecules. These studies also help evaluate surface-area and continuum dielectric models of hydration for this protein.

Below, we sketch the theory (section 2) and then describe the simulation methods (section 3). We next present results on the hydration of cytochrome C (section 4). [In Appendix A, we present calculations on $Ca^{2+}(aq)$ to highlight the importance of finite size thermodynamic corrections, an aspect that is necessarily treated approximately for the protein.] We conclude the paper by discussing the potential applications of the present approach in exploring fundamental questions in protein solution thermodynamics.

2. THEORY

Imagine applying an external field, $\phi(r;\lambda)$, centered on the hydrated solute of interest. Here, r is the distance of the water molecule from the center of the field, and λ is the parameter characterizing the range of the field, in particular $\phi(r;\lambda) = 0$ for $r \geq \lambda$. Then, μ^{ex} is given by[20]

$$\begin{aligned}\beta\mu^{ex} &= \ln\langle e^{-\beta\phi}\rangle - \ln\langle e^{-\beta\phi}\rangle_0 - \ln\langle e^{-\beta\Delta U}\rangle_\phi \\ &= \ln x(\lambda) - \ln p(\lambda) + \beta\mu_{lr}^{ex}(\lambda)\end{aligned} \tag{2}$$

where each member on the first line defines the corresponding term in the second. As usual, $\beta = 1/k_BT$, $\langle ...\rangle$ denotes ensemble averaging in the coupled solute−solvent system, $\langle ...\rangle_0$ denotes averaging in the neat solvent system, and $\langle ...\rangle_\phi$ denotes averaging with the solute uncoupled from the solvent-field system. Note that μ^{ex} is independent of λ, whereas the individual contributions on the right-hand side of eq 2 are not.

The field ϕ serves to exclude solvent contact up to a range λ away from the solute, thereby regularizing the solute−solvent binding energy distribution.[19,20,29] In the general case, ϕ need not impose a hard exclusion; we need only have a soft boundary at λ. With this understanding, note that λ defines a local coordination domain around the solute. The free energy gained when solute−solvent contact is restored is $\ln x(\lambda)$, and in the context of the quasichemical organization of the potential distribution theorem,[14,17,18] this quantity is related to the free energy of forming solute−solvent clusters within the specified coordination domain. We will follow earlier terminology[14,17,18] and call this quantity the chemical contribution to hydration. The quantity $\ln p(\lambda)$ is independent of the solute and measures the free energy to create an empty space the size of the coordination domain. We will call this quantity the packing contribution. The general soft boundary can be corrected to give the free energy for imposing a hard exclusion;[20] in this, case $-\ln p$ is precisely the hydrophobic contribution to hydration. $\ln x/p$ thus defines the local or inner-shell contribution to hydration.

$\beta\mu_{lr}^{ex}(\lambda)$ is the contribution due to solute−solvent interactions (in the presence of ϕ). Since ϕ excludes close solute−solvent contact, $\beta\mu_{lr}^{ex}(\lambda)$ necessarily accounts for the effects of long-range (lr) interactions on μ^{ex}. $\beta\mu_{lr}^{ex}(\lambda)$ can be further decomposed into nonelectrostatic (dispersion or van der Waals) and electrostatic contributions:

$$\mu_{lr}^{ex}(\lambda) = \mu_{lr,elec}^{ex}(\lambda) + \mu_{lr,vdW}^{ex}(\lambda) \tag{3}$$

$\ln x_s(\lambda)$ and $-\ln p_s(\lambda)$ are obtained by integrating the average force $\langle\partial\phi/\partial\lambda\rangle$ with respect to λ.[20] ϕ can always be chosen such that solute−solvent binding energies are Gaussian distributed, in which case, $\beta\mu_{lr}^{ex}(\lambda)$ is determined by the mean and variance of the binding energy distribution.[20] Here, we calculate $\mu_{vdW}^{ex}(\lambda)$ using the Gaussian model. Although the Gaussian model can also be used for electrostatics, we use Gauss−Legendre quadratures for maintaining consistency with electrostatic calculations when $\lambda = 0$ (section 3.1).

Following earlier work,[20] we use

$$\phi_{ramp}(r;\lambda) = a\left(\sqrt{(r-\lambda)^2 + b^2} - b\right) \tag{4a}$$

$$\phi_{lj}(r;\lambda) = 4a\left[\left(\frac{b}{r-\lambda+\sqrt[6]{2}\,b}\right)^{12} - \left(\frac{b}{r-\lambda+\sqrt[6]{2}\,b}\right)^{6}\right] + a \tag{4b}$$

where a and b are positive constants and $(r < \lambda)$ and $\phi(r;\lambda) = 0$ for $r \geq \lambda$. For $\phi_{lj}(r;\lambda)$, $a = 0.155$ kcal/mol and $b = 3.1655$ Å, the LJ energy and size parameters for SPC/E water. For $\phi_{ramp}(r;\lambda)$, $b = 0$ and $a = 5.0$ kcal/mol. For the protein, both the fields yield similar final results. Hence, only results with ϕ_{lj} are reported. [In Appendix B, we provide examples of the field for the specific case of $\lambda = 5.0$ Å.]

3. METHODS

The simulations were performed using NAMD[37] with external forces applied using the Tcl interface. The solvent system contains 16 384 SPC/E water[38] molecules. For the protein (PDB ID: 1hrc[39]), we added hydrogens and set charges (at pH = 7.0) for ionizable residues consistent with the standard ionization values. The net charge of the protein was $+7e$ (the protein without the HEME was $+9e$). CHARMM[40] force-field parameters were used for the protein. The heavy atoms of the protein were held fixed throughout the simulation. (Thus mixing CHARMM for the protein with SPC/E for water is not an issue here, but see also ref 41.) The average of the minimum and maximum values of the coordinates of all the atoms was set as the center of the protein about which we apply the field ϕ.

The simulations were performed at a temperature of 298.15 K and a pressure of 1 bar as described earlier.[20] The equations of motion were propagated using the Verlet algorithm with a time step of 2.0 fs. SHAKE[42] was used to constrain the geometry of water molecules. The Lennard-Jones interactions were terminated at 10.43 Å by smoothly switching to zero starting at 9.43 Å. Electrostatic interactions were treated with the particle mesh Ewald method with a grid spacing of 0.5 Å.

We first equilibrated states with $\lambda_m = (2.5 + 5.0{\cdot}m)$ Å, with $m = 0,...,6$, each for 2.0 ns. From each λ_m, we next generated states spanning $\lambda_m \pm 2.5$ Å in increments of 0.1 Å. At each λ point, we performed 600 ps of simulations and calculated the average force $\langle\partial\phi/\partial\lambda\rangle$ from the last 500 ps.

For calculating $\mu_{lr}^{ex}(\lambda)$, trajectories of the neat solvent with $\lambda = 25.0$, 27.5, 30.0, 32.5, and 35.0 Å were extended by an additional 1 ns, saving configurations every 1 ps. For each of the 1000 frames, we performed test particle insertions of the

protein using four randomly chosen orientations per frame. $\mu^{ex}_{lr,vdW}$ was then calculated using a Gaussian model[22,28,29] for the binding energy distribution.

We calculate $\mu^{ex}_{lr,elec}$ using a three-point Gauss−Legendre quadrature,[43] with protein charges scaled by $\eta = 0.5, 0.5 \pm (3/20)^{1/2}$. For $\lambda = 0$, we also used $\eta = 0,1$. Electrostatic self-interaction corrections, μ^{ex}_{self}, were applied following Hummer et al.[44]

In a simulation of a charged particle with periodic boundary conditions, one must also consider finite-size (fs) thermodynamic corrections. Conceptually, μ^{ex}_{fs} is given by the difference in the hydration free energy of the solute in an infinite dielectric medium and the value obtained with a dielectric in a periodic simulation. For a spherical solute with a point charge, to leading order[44]

$$\mu^{ex}_{fs} = \frac{1}{2}q^2\left[\frac{-\xi}{\varepsilon} + \frac{4\pi(\varepsilon - 1)R_B^2}{3\varepsilon L^3}\right] \tag{5}$$

where $\xi = -2.837297/L$ is the Wigner potential, L is the length of the (cubic) simulation cell, ε is the dielectric constant of the solvent (70 for SPC/E water[45]), and R_B is the Born radius. For convenience in presentation, we set

$$\mu^{ex}_{sim}(\lambda) = \ln\frac{x(\lambda)}{p(\lambda)} + \mu^{ex}_{lr,elec}(\lambda) + \mu^{ex}_{lr,vdW}(\lambda) + \mu^{ex}_{self}(\lambda) \tag{6}$$

and thus (eq 2)

$$\mu^{ex} = \mu^{ex}_{sim}(\lambda) + \mu^{ex}_{fs}(\lambda) \tag{7}$$

Since R_B is not known *a priori*, we calculate R_B from the fluctuation in electrostatic potential experienced by the solute. (Within the Born model, the free energy depends solely on the fluctuations in the electrostatic potential experienced by the solute.) To simplify the discussion, we assume a solute with a point charge; extension to a solute with a distributed charge follows the same logic. The electrostatic contribution to the hydration free energy after self-interaction corrections are applied is $\mu^{ex}_{lr,elec} + 1/2\xi q^2$, where q is the charge on the solute. We equate this quantity to a second order cumulant expansion about the neutral charge state and obtain

$$-\frac{\beta q^2}{2}\langle\delta\psi^2\rangle_0 = \mu^{ex}_{lr,elec} + \frac{1}{2}\xi q^2 - q\langle\psi\rangle_0 \tag{8}$$

where $\langle\psi\rangle_0$ is the potential at the center of the uncharged solute and $\langle\delta\psi^2\rangle_0$ is the fluctuation in the electrostatic potential. From the Born model of hydration,

$$\mu^{ex}_{Born} = -\frac{1}{2}\frac{q^2}{R_B}\left(1 - \frac{1}{\varepsilon}\right) = -\frac{\beta q^2}{2}\langle\delta\psi^2\rangle_0 \tag{9}$$

we can thus relate R_B to $\langle\delta\psi^2\rangle_0$.

For the protein, calculating μ^{ex}_{fs} is further complicated by the presence of a charge distribution. (In Appendix A, we present the case of Ca^{2+} to highlight the importance of μ^{ex}_{fs} in a case where μ^{ex}_{fs} can be calculated without the complication of a distributed charge.) One could estimate μ^{ex}_{fs} using a numerical approach, for example, by adapting a boundary element approach[46] to periodic electrostatics. Tentative attempts in this direction emphasized the necessity of implementing Ewald summations within the boundary element code. Hence for simplicity, we use eq 6 for the protein as well, but we approximate the influence of the charge distribution in the calculation of the Born radius. Within the dielectric model, we calculate the hydration free energy of the physical protein charge distribution in spheres of varying radii.[47] Then by interpolation we find the radius that matches the calculated fluctuation contribution (eq 9). This gives $R_B(\{q\})$, the Born radius based on the distributed charge model. Alternatively, we can assume that the entire protein charge is concentrated at the origin, giving the point charge approximation $R_B(q)$. Note that for a fixed radius, the hydration free energy of the distributed charge model will always be lower, hence $R_B(\{q\}) > R_B(q)$.

3.1. μ^{ex} with $\lambda = 0$. To assess limitations in calculating μ^{ex}_{fs}, we calculate $\mu^{ex}_{lr,elec}$ with $\lambda = 0$, i.e. without any regularization. We use the three-point Gauss Legendre quadrature as well as lower-order quadratures to assess the variation with order. Since the electrostatic contribution without regularization is large, R_B is very small (about 4 Å, for a protein net charge of $+7e$), and we can safely ignore μ^{ex}_{fs}. We will denote by $\mu^{ex}(\lambda = 0)$ the estimate of μ^{ex} based on the sum of $\mu^{ex}_{lr,elec}(\lambda = 0)$ and the dispersion contribution (obtained using regularization).

3.2. Statistical Uncertainities. We use the Friedberg−Cameron algorithm[48,49] to estimate the standard error of the mean. In integrating the force $\langle\partial\phi/\partial\lambda\rangle$, or in adding multiple quantities, the errors were propagated using standard rules. Throughout, statistical uncertainties are quoted at 2σ.

4. RESULTS

Nonpolar Contribution. From Figure 1, for the uncharged protein, we find that $k_BT \ln x_s(\lambda)/P_s(\lambda) = 93 \pm 3$ kcal/mol for λ

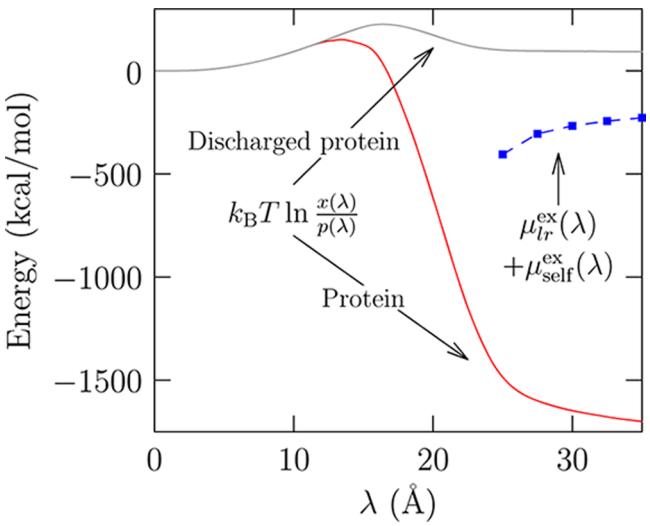

Figure 1. Components of eq 6 for the protein and its analogue with no partial charges. μ^{ex} of the latter system gives the dispersion (nonelectrostatic) contribution to hydration.

$\geq$ 30 Å. Clearly, then $\mu^{ex}_{lr}(\lambda \geq 30\ \text{Å}) = 0$, and thus the nonelectrostatic (dispersion) contribution to the hydration free energy is 93 ± 3 kcal/mol.

Often, the approximation $\gamma_s \cdot A_s + c$, where γ_s is a surface-energy parameter, A_s is the solvent-accessible surface area, and c is some constant, is used to estimate the dispersion contribution to hydration.[50−52] Using standard values of atomic radii, we calculated[53] $A_s = 6715$ Å^2. (For purposes of calculating A_s, the HEME group, which is mostly buried in the protein, was treated as a collection of carbon atoms.) Ignoring c (which is typically small), we thus find $\gamma_s \approx 14$ cal/mol·Å^2, which can be

dx.doi.org/10.1021/ct300505b | J. Chem. Theory Comput. 2012, 8, 3409−3415

contrasted with the 5−7 cal/mol·Å^2 typically used in surface-area models.[54,55] The smaller value of γ_s is based on the transfer free energy of small alkanes from vapor to water.

The present comparison thus suggests that γ_s = 5 cal/mol·Å^2 will substantially underestimate the dispersion (or non-electrostatic) contribution to hydration. A recent calculation on hydrocarbon clusters also suggests similar large deviations.[56]

Electrostatic Contribution. We obtained $\mu^{ex}_{lr,elec}(\lambda = 0)$ using quadratures of different orders. The linear response (η = 0.5) result is −1570 kcal/mol. Using η = 0, 0.5, 1 the result is −1655 kcal/mol, correct to fourth order in perturbation theory.[57] The three-point Gauss-Legendre rule gives −1665 kcal/mol, correct to sixth order in perturbation theory. The fourth-order result differs by less than 1% from the sixth-order estimate, which is our reference value for $\mu^{ex}_{lr,elec}(\lambda = 0)$. Together with $\mu^{ex}_{self}(\lambda = 0)$ = −271 kcal/mol, we find that $\mu^{ex}_{fs}(\lambda = 0)$ is negligible. Combining these results with the dispersion contribution obtained above, we find that the net hydration free energy of the protein is −1843 ± 4 kcal/mol. This value, denoted by $\mu^{ex}(\lambda = 0)$, is indicated by the horizontal line in Figure 2.

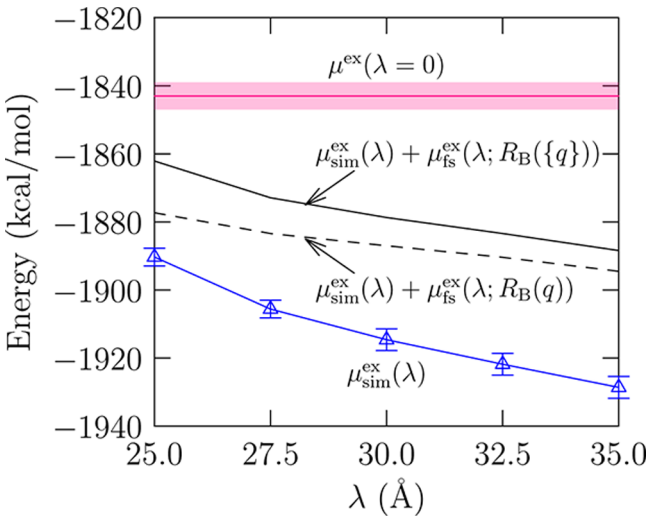

Figure 2. $\mu^{ex}(\lambda = 0)$ (horizontal line; see text) and estimated μ^{ex} for $\lambda \neq 0$. Two estimates of the net hydration free energy based on the two definitions of the Born radius (section 3) are shown. For the distributed charge $R_B(\{q\})$, $4\pi R_B^3/3L^3$ various between 9% (λ = 25 Å) and 16% (λ = 35 Å). Statistical uncertainties at the 2σ level are shown.

We next consider the results for $\lambda \neq 0$. From Figure 1, between λ = 25 Å and 35 Å, the decrease of about 215 kcal/mol in $k_BT \ln x_s/p_s$ is compensated by an increase of about 177 kcal/mol in $\mu^{ex}_{lr}(\lambda) + \mu^{ex}_{self}(\lambda)$. Thus, $\mu^{ex}_{sim}(\lambda)$ (eq 6) changes by about 38 kcal/mol in a total magnitude of about 1900 kcal/mol. Including $\mu^{ex}_{fs}(\lambda)$ dampens the variation in μ^{ex} and moves the net free energy closer to the λ = 0 estimate noted above (Figure 2). An encouraging aspect in Figure 2 is that μ^{ex}_{sim} is itself well estimated, and as Figure 1 shows, when $\lambda \geq 25$ Å, the chemical and packing contributions make the larger contribution to this quantity.

Neither approximation of R_B is entirely satisfactory in the context of the protein, although incorporating the influence of the distributed charge is somewhat better. Further, with increasing λ the Born cavity occupies a larger fraction of the system volume, and the leading order correction (eq 5) can be expected to fail, as is indeed seen in the downward drift of $\mu^{ex}_{sim} + \mu^{ex}_{fs}$. Despite these limitations, it is encouraging that for λ = 25 Å, the $\lambda \neq 0$ result differs from the λ = 0 value by less than 20 kcal/mol. (Appendix A shows the contrasting case of Ca^{2+} where the Born cavity occupies a smaller volume fraction and R_B is also well-defined.)

5. DISCUSSIONS AND CONCLUSIONS

Comparison with Continuum Solvent and Integral Equation Methods. The self-interaction corrected linear response estimate −1841 kcal/mol differs from the sixth-order result (−1936 kcal/mol) by about 5%. We should therefore expect a continuum solvent approximation to do no better than this. Removing the potential at zero charge, we get −1996 kcal/mol as the value to compare with a continuum dielectric model. (Note that continuum electrostatic models do not capture fundamental physical effects such as a nonzero potential at the center of an uncharged solute; the consequences of such effects are typically incorporated into the radius assigned to the atom.) For a continuum model,[58−60] we find the electrostatic contribution to be ∼−2250 kcal/mol. (To reflect the simulation study, the protein dielectric was set at 1.0, and care was taken to use the same charge distribution as the one used in the atomistic simulation; using a dielectric of 2.0 leads to −1860 kcal/mol.) Either estimate differs by between 6% and 12% from the molecularly detailed calculation.

Thus, both the surface area model for dispersion contributions and continuum dielectric estimate of electrostatic contributions model differ from the molecular simulations. For the dispersion (nonpolar/hydrophobic) contribution, similar to a recent study,[56] we find that a molecularly detailed description of the solvent predicts a more positive value of the free energy than that obtained using a surface area model with a surface energy parameter around 5 cal/mol·Å^2. On the basis of this and the earlier study,[56] we can thus expect simulations based on GBSA[55] to underestimate the consequences of hydrophobic effects.

However, in contrast to the case for dispersion contribution, for the electrostatic contribution, we are unable to infer a general trend in the deviation between the all-atom and continuum dielectric results, since the direction of deviation appears to depend on the magnitude of the protein dielectric constant. For a dielectric of 1, we expect a continuum model to overestimate the hydration free energy of an individual protein, a feature that can be expected to cause the strength of protein−protein electrostatic interactions to be underestimated.

For cytochrome C, an integral equation approach predicts $\mu^{ex} \approx$ −1500 kcal/mol.[61] The difference in the solvent model (TIP3 vs SPCE used here) could account for some, but definitely not all, of the difference with the present result. But we note that even for solute water, the integral equation approach predicts free energies that differ from the well-established value by more than 1 kcal/mol,[62] in contrast to results obtained by the regularization approach.[20] So there are likely inherent limitations in the integral equation method that could explain the difference with the present results.

Summarizing the above discussion, we emphasize that the comparison is solely between the all-atom approach and alternative approaches that describe the solvent at varying levels of detail. Since μ^{ex} of cytochrome C has not been experimentally established, no claims can be made as to which of these methods is closer to the experimental truth. However, we hope the effort presented here can spur experimental efforts to obtain either μ^{ex} of the protein or

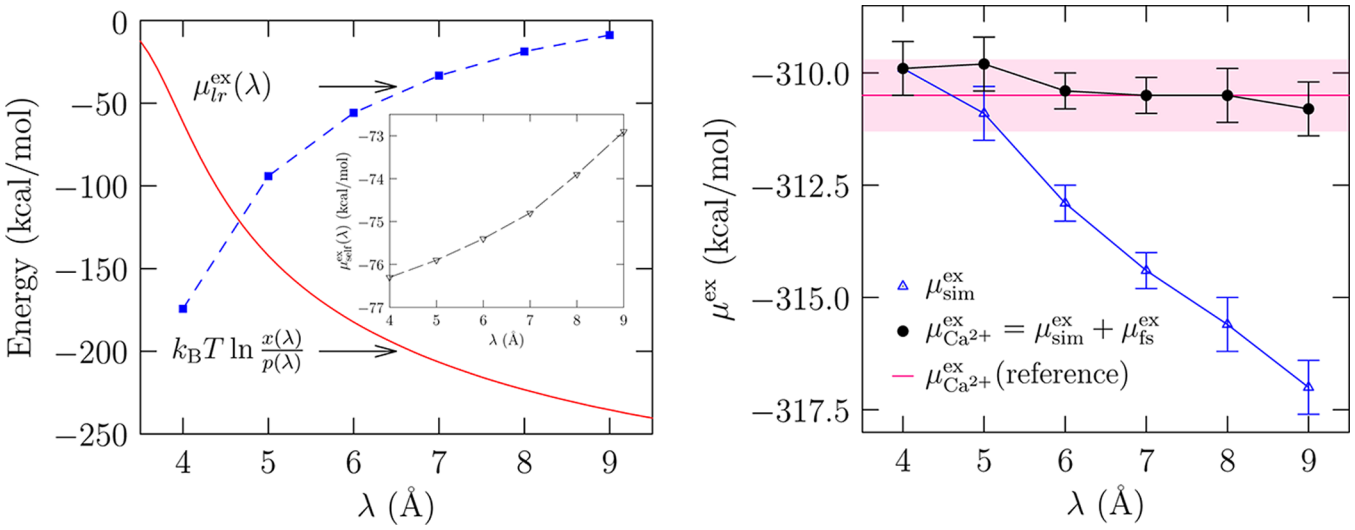

Figure 3. Left panel: The local and long-range contributions to μ^{ex}_{sim} obtained using ϕ_{lj}. The inset gives the average electrostatic self-interaction correction for various λ. Right panel: μ^{ex}_{sim} obtained from the data in the left panel, the net hydration free energy of Ca^{2+} (•), and the reference estimate obtained without regularization. The reference value, obtained using a three-point Gauss−Legendre quadrature for electrostatics and particle insertions for the nonpolar contribution, is in excellent agreement with the value reported by Hummer et al.[44] $4\pi R_B^3/3L^3$ varies between 0.3% (λ = 4 Å) and 7.5% (λ = 9 Å). Statistical uncertainties at the 2σ level are indicated by the error bars and by the shading.

perhaps the free energy of diluting a concentrated protein solution. The latter may be obtainable from osmotic pressure measurements (V. A. Parsegian, personal communication). While calculations to model such processes are still daunting, these could be cases where a realistic assessment of the different methods could potentially be made.

Relevance in Understanding Hydration Effects in Protein Solution Thermodynamics. In the present study, for simplicity we chose a spherical regularization domain. The downside of this choice is the loss of fidelity in assessing the effects of local solute−solvent interactions: a spherical exclusion domain for a nonspherical object will necessarily overestimate the magnitude of the chemical and packing contributions. An exclusion that follows the topography of the molecule, for example, one limited to the first hydration shell of the protein, is expected to be a better representation of the local protein−solvent interactions. (See, for example, ref 28.) In this instance, information about local chemical, packing, and long-range contributions to hydration can prove useful in understanding how changes in hydration influence folding or the association of proteins.

The local packing contribution obtained in the presence of the field can be readily corrected to obtain the hydrophobic contribution to hydration.[20] The importance of such a development in the context of protein biophysics needs no emphasis. (For example, see ref 63 for one example of a still unresolved, foundational question in hydrophobic hydration and protein solution thermodynamics.) The local chemical contribution can be parsed into solvent (including possible ions and osmolytes) specific effects on defined regions or chemical groups of the protein. Such studies can help illuminate the role of hydration of the peptide bond versus the hydration of residues in protein stability. In effect, with a topographically detailed regularization, the entire quasichemical machinery that has so far illuminated the solution thermodynamics of small molecular solutes, for example,[14,18,23,27−29] now can be brought to bear in studying the hydration of a protein.

The present framework has the virtue of cleanly separating the role of intrinsic solvent properties (the $-\ln p_s$ term) from effects due to solute−solvent interactions. The solute−solvent interactions are themselves separated into specific local and nonspecific long-range contributions. We expect only the local contributions to be sensitive to the type of the ion[23] or osmolyte in the medium. Thus the present approach offers a clear way to explore ion- and osmolyte-specific effects in protein hydration and folding. Results of such studies in the folding of a deca-alanine peptide will be reported separately.

In summary, we have shown that a physically transparent framework to interrogate the hydration thermodynamics of a protein using atomically detailed simulations is now available. The usefulness of the approach to understand various theoretical and practical questions in protein solution thermodynamics remains to be explored.

APPENDIX

A. Hydration of Ca^{2+}

The regularization approach has been used to study the hydration of water, using both classical and *ab initio* simulations,[20] simple ions such as Na^+ and K^+, and molecular ions such as imidazolium[57] (unpublished data). The method has also been successfully tested for calculating the hydration free energy of a deca-alanine peptide[11] in various conformations and in the presence of solvent additives (manuscript in preparation). Here we present calculations on Ca^{2+} to highlight the consistency of the framework and the importance of thermodynamic finite size correction (eq 5). In contrast to the calculations on the protein, we do not incur additional approximations in using eq 5, beyond the obvious limitation that the correction is only asymptotically correct. For Ca^{2+} the simulation procedure was exactly as it was for the protein, except that the simulation contained 512 water molecules. Ca^{2+} parameters were from ref 44. In Figure 3 we show the results for the hydration of Ca^{2+} using ϕ_{lj}. From Figure 3 (right panel), it is clear that after including the finite size thermodynamic corrections, the estimated chemical potential of Ca^{2+} is in excellent agreement with the reference result and is also largely independent of λ, as it should be. For all the λ values

considered, the ratio of the volume of the Born cavity to the system volume varies between 0.3% and 7.5%, and thus the correction (eq 5) is expected to be satisfactory, as indeed the results indicate.

B. Structure of the Fields

Figure 4 below shows the fields modeled by eq 4 for the specific case of λ = 5.0 Å. μ^{ex} (eq 7) is relatively indifferent to the choice

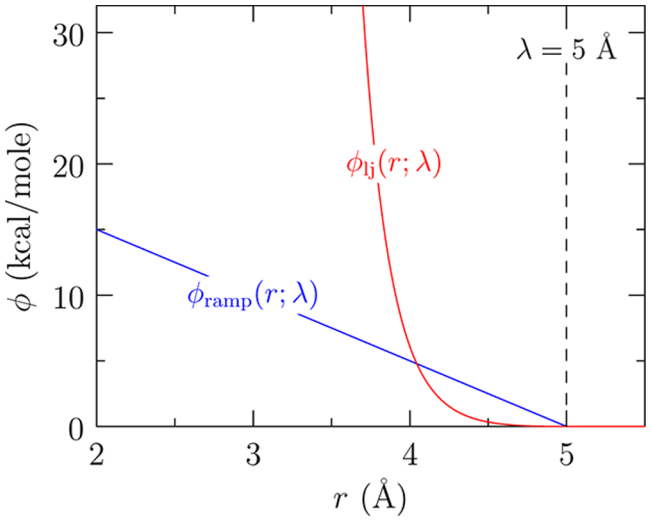

Figure 4. $\phi_{ramp}(r;\lambda)$ and $\phi_{lj}(r;\lambda)$ for the specific case of λ = 5.0 Å.

of the field, as it must be. (See also the example of water in ref 20.)

AUTHOR INFORMATION

Corresponding Author

*E-mail: dilipa@jhu.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Safir Merchant and Purushottam Dixit for helpful comments and a critical reading of the manuscript. We acknowledge helpful discussions with S. Tonddast-Navaei, A. Kravats, B. Baird, T. L. Beck, G. Stan, and V. A. Parsegian. D.A. thanks the donors of the American Chemical Society Petroleum Research Fund. This research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

REFERENCES

(1) Yancey, P. H.; Clark, M. E.; Hand, S. C.; Bowlus, R. D.; Somero, G. N. *Science* **1982**, *217*, 1214−1222.
(2) Dill, K. A. *Biochemistry* **1990**, *29*, 7133−7155.
(3) Hochachka, P. W.; Somero, G. N. Chapter 6. Water-solute adaptations: The evolution and regulation of the internal milieu. *Biochemical Adaptation. Mechanism and Process in Physiological Evolution*; Oxford University Press: New York, 2002; pp 217−289.
(4) Yancey, P. *J. Exp. Biol.* **2005**, *208*, 2819−2830.
(5) Baldwin, R. L. *J. Mol. Biol.* **2007**, *371*, 283−301.
(6) Bolen, D. W.; Rose, G. D. *Annu. Rev. Biochem.* **2008**, *77*, 339−362.
(7) Bandyopadhyay, A.; Saxena, K.; Kasturia, N.; Dalal, V.; Bhatt, N.; Rajkumar, A.; Maity, S.; Sengupta, S.; Chakraborty, K. *Nat. Chem. Biol.* **2012**, *8*, 238−245.
(8) Frokjaer, S.; Otzen, D. E. *Nat. Rev. Drug Discovery* **2005**, *4*, 298−306.
(9) Arakawa, T.; Tsumoto, K.; Kita, Y.; Chang, B.; Ejima, D. *Amino Acids* **2007**, *33*, 587−605.
(10) Staritzbichler, R.; Gu, W.; Helms, V. *J. Phys. Chem. B* **2005**, *109*, 19000−19007.
(11) Kokubo, H.; Hu, C. Y.; Pettitt, B. M. *J. Am. Chem. Soc.* **2011**, *133*, 1849−1858.
(12) Chen, J.; Brooks, C. L., III. *Phys. Chem. Chem. Phys.* **2008**, *10*, 471−481.
(13) Widom, B. *J. Phys. Chem.* **1982**, *86*, 869−872.
(14) Beck, T. L.; Paulaitis, M. E.; Pratt, L. R. *The Potential Distribution Theorem and Models of Molecular Solutions*; Cambridge University Press: Cambridge, U.K., 2006.
(15) Chipot, C.; Pohorille, A. Chapter 2. Calculating free energy differences using perturbation theory. In *Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Chipot, C., Pohorille, A., Eds.; Springer: Berlin, 2007; Springer Series in Chemical Physics, Vol. *86*, pp 33−75.
(16) Pratt, L. R.; Rempe, S. B. Chapter 9. Quasi-Chemical Theory and Implicit Solvent Models for Simulations. In *Simulation and Theory of Electrostatic Interactions in Solution. Computational Chemistry, Biophysics, and Aqueous Solutions*; Pratt, L. R., Hummer, G., Eds.; American Institute of Physics: Melville, NY, 1999; AIP Conference Proceedings, Vol. *492*, pp 172−201.
(17) Paulaitis, M. E.; Pratt, L. R. *Adv. Protein Chem.* **2002**, *62*, 283−310.
(18) Pratt, L. R.; Asthagiri, D. Chapter 9. Potential distribution methods and free energy models of molecular solutions. In *Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Chipot, C., Pohorille, A., Eds.; Springer: Berlin, 2007; Springer Series in Chemical Physics, Vol. *86*, pp 323−351.
(19) Merchant, S. Regularizing free energy calculations to study ion specific effects in biology. Ph.D. thesis, Johns Hopkins University, Baltimore, MD, 2011.
(20) Weber, V.; Merchant, S.; Asthagiri, D. *J. Chem. Phys.* **2011**, *135*, 181101.
(21) Paliwal, A.; Asthagiri, D.; Pratt, L. R.; Ashbaugh, H. S.; Paulaitis, M. E. *J. Chem. Phys.* **2006**, *124*, 224502.
(22) Shah, J. K.; Asthagiri, D.; Pratt, L. R.; Paulaitis, M. E. *J. Chem. Phys.* **2007**, *127*, 144508.
(23) Merchant, S.; Asthagiri, D. *J. Chem. Phys.* **2009**, *130*, 195102.
(24) Merchant, S.; Dixit, P. D.; Dean, K. R.; Asthagiri, D. *J. Chem. Phys.* **2011**, *135*, 054505.
(25) Rogers, D. M.; Beck, T. L. *J. Chem. Phys.* **2010**, *132*, 014505.
(26) Beck, T. L. *J. Phys. Chem. B* **2011**, *115*, 9776−9781.
(27) Ashbaugh, H. S.; Asthagiri, D.; Pratt, L. R.; Rempe, S. B. *Biophys. Chem.* **2003**, *105*, 321−336.
(28) Asthagiri, D.; Ashbaugh, H. S.; Piryatinski, A.; Paulaitis, M. E.; Pratt, L. R. *J. Am. Chem. Soc.* **2007**, *129*, 10133−10140.
(29) Asthagiri, D.; Merchant, S.; Pratt, L. R. *J. Chem. Phys.* **2008**, *128*, 244512.
(30) Rempe, S. B.; Pratt, L. R.; Hummer, G.; Kress, J. D.; Martin, R. L.; Redondo, A. *J. Am. Chem. Soc.* **2000**, *122*, 966−967.
(31) Rempe, S. B.; Pratt, L. R. *Fluid Phase Equilib.* **2001**, *183*−*184*, 121−132.
(32) Rempe, S. B.; Asthagiri, D.; Pratt, L. R. *Phys. Chem. Chem. Phys.* **2004**, *6*, 1966−1969.
(33) Asthagiri, D.; Pratt, L. R.; Kress, J. D.; Gomez, M. A. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 7229−7233.
(34) Asthagiri, D.; Pratt, L. R.; Kress, J. D. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6704−6708.
(35) Weber, V.; Merchant, S.; Dixit, P. D.; Asthagiri, D. *J. Chem. Phys.* **2010**, *132*, 204509.
(36) Weber, V.; Asthagiri, D. *J. Chem. Phys.* **2010**, *133*, 141101.
(37) Kale, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. *J. Comput. Phys.* **1999**, *151*, 283.

(38) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269−6271.
(39) Bushnell, G. W.; Louie, G. V.; Brayer, G. D. *J. Mol. Biol.* **1990**, *214*, 585−595.
(40) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(41) Mackerell, A. D., Jr. *J. Comput. Chem.* **2004**, *25*, 1584−1604.
(42) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(43) Hummer, G.; Szabo, A. *J. Chem. Phys.* **1996**, *105*, 2004−2010.
(44) Hummer, G.; Pratt, L. R.; Garcia, A. E. *J. Phys. Chem.* **1996**, 1206−1215.
(45) Aragones, J. L.; MacDowell, L. G.; Vega, C. *J. Phys. Chem. A* **2011**, *115*, 5745−5758.
(46) Yoon, B. J.; Lenhoff, A. M. *J. Comput. Phys.* **1990**, *11*, 1080−1086.
(47) Neal, B. L.; Asthagiri, D.; Lenhoff, A. M. *Biophys. J.* **1998**, *75*, 2469−2477.
(48) Friedberg, R.; Cameron, J. E. *J. Chem. Phys.* **1970**, *52*, 6049−6058.
(49) Allen, M. P.; Tildesley, D. J. Chapter 6. How to analyze the results. *Computer Simulation of Liquids*; Oxford University Press: New York, 1987; pp 192−195.
(50) Hermann, R. B. *J. Phys. Chem.* **1972**, *76*, 2754−2759.
(51) Eisenberg, D.; McLachlan, A. D. *Nature* **1986**, *319*, 199−203.
(52) Ooi, T.; Oobatake, M.; Némethy, G.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 3086−3090.
(53) Sanner, M.; Olson, A. J.; Spehner, J. C. *Biopolymers* **1996**, *38*, 305−320.
(54) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978−1988.
(55) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(56) Varilly, P.; Patel, A. J.; Chandler, D. *J. Chem. Phys.* **2011**, *134*, 074109.
(57) Hummer, G.; Pratt, L. R.; Garcia, A. E. *J. Phys. Chem. A* **1998**, *102*, 7885−7895.
(58) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. *Nucleic Acids Res.* **2004**, *32*, W665−W667.
(59) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. *Nucleic Acids Res.* **2007**, *35*, W522−W525.
(60) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10037−10041.
(61) Karino, Y.; Matubayasi, N. *J. Chem. Phys.* **2011**, *134*, 041105.
(62) Matubayasi, N.; Nakahara, M. *J. Chem. Phys.* **2002**, *117*, 3605−3616.
(63) Baldwin, R. L. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 7310−7313.