Journal of Theoretical and Computational Chemistry
Vol. 17, No. 3 (2018) 1840003 (22 pages)

DOI: 10.1142/S0219633618400035

World Scientific
www.worldscientific.com

Alchemical free energy calculations and umbrella sampling with local molecular field theory

Richard C. Remsing*
Institute for Computational Molecular Science
Department of Chemistry
Temple University, Philadelphia, PA 19122, USA
rremsing@temple.edu

John D. Weeks*
Institute for Physical Science and Technology
Department of Chemistry and Biochemistry
University of Maryland, College Park, MD 20742, USA
jdw@umd.edu

Received 21 September 2017
Accepted 14 March 2018
Published 18 April 2018

Understanding the thermodynamic driving forces underlying any chemical process requires a description of the underlying free energy surface. However, computation of free energies is difficult, often requiring advanced sampling techniques. Moreover, these computations can be further complicated by the evaluation of any long-ranged interactions in the system of interest, such as Coulomb interactions in charged and polar media. Local molecular field theory is a promising approach to avoid many of the conceptual and computational difficulties associated with long-ranged interactions. We present frameworks for performing alchemical free energy calculations and non-Boltzmann sampling with local molecular field theory. We demonstrate that local molecular field theory can be used to perform these free energy calculations with accuracy comparable to traditional methodologies while eliminating the need for explicit treatment of long-ranged interactions in simulations.

Keywords: Mean field theory; solvation; density fluctuations; hydrophobic effects.

1. Introduction

A thorough understanding of any equilibrium chemical process requires quantification of the thermodynamic driving forces that drive it. The free energy is therefore of central importance, and remaining thermodynamic quantities can typically be computed by differentiating the free energy. However, estimating free energies involves computing ratios of partition functions, which is a highly nontrivial and difficult task.[1] Recent theoretical and computational developments have made free

*Corresponding authors.

energy calculations more routine, significantly advancing our understanding of molecular, biochemical, and materials systems. Despite these advances, the computation of thermodynamic properties is still hampered by many issues, including problems posed by the evaluation of long-ranged interactions.

The simulation of nearly all molecular systems requires the evaluation of long-ranged interactions,[2] arising from van der Waals or Coulomb (charge–charge) interactions, for example. The accurate evaluation of these interactions, especially when using periodic boundary conditions in systems with Coulombic potentials, usually involves complex lattice summations that lead to poor scaling in large-scale systems.[2,3] Moreover, typical lattice summation techniques, such as Ewald summation, can sometimes generate significant artifacts that arise from the imposed periodic nature of the cell images on which the infinite sums are evaluated and the finite nature of simulation cells.[4,5] A promising alternative for describing long-ranged interactions is local molecular field (LMF) theory,[6–8] which essentially relates properties of a full system with long-ranged interactions to those of a simpler "mimic system" with renormalized short-ranged interactions chosen to capture the major effects of the long-ranged forces. LMF theory accurately describes the structure of a wide variety of nonuniform systems with long-ranged interactions, and in recent work, we showed how LMF theory can be used to give a very accurate description of solvation-free energies as well.[8]

In this work, we extend LMF theory to two other major classes of free energy calculations[1]: alchemical transformations and potentials of mean force. We first review LMF theory and its application to solvation free energies in the next section. The extension of LMF theory to alchemical transformations is performed in Sec. 4. We then further extend the LMF formalism to calculate free energies as a function of an order parameter describing a process of interest. The calculation of such potentials of mean force involves combining LMF theory with umbrella sampling,[1,9] and this is accomplished in Sec. 5 before concluding in Sec. 6.

2. LMF Theory

Consider a system of solvent molecules that interact through the pair potential $u(r)$, in the presence of the one-body field $\phi(\mathbf{r})$. The field $\phi(\mathbf{r})$ can arise from solute–solvent interactions, as is the focus here, but is in general not limited to these situations and could represent interactions from confining walls or electric fields, for example. In general, $u(r)$ and $\phi(\mathbf{r})$ can involve long-ranged interactions, which present difficulties for both theory and simulation, as discussed above.

LMF theory seeks to map this long-ranged system onto a simpler *mimic* system, whereby the solvent molecules interact with the purely short-ranged potential $u_0(r)$, but the structure of the full system is reproduced by introducing an effective or *renormalized* one-body potential $\phi_R(\mathbf{r})$. The LMF equation, obtained from an approximate integration of the exact Yvon–Born–Green hierarchy relating intermolecular forces to induced structure,[6–8,10] enables the determination of the

renormalized potential,

$$\phi_R(\mathbf{r}) = \phi(\mathbf{r}) + \int d\mathbf{r}'[\rho_R(\mathbf{r}') - \rho_B]u_1(|\mathbf{r} - \mathbf{r}'|), \tag{1}$$

where ρ_B is the bulk density of the solvent, $u_1(r) = u(r) - u_0(r)$ is the long-ranged, slowly-varying portion of the solvent–solvent pair potential, and the subscript R indicates a quantity evaluated in the renormalized (mimic) system. Here, slowly-varying refers to the fact that spatial variations in $u_1(r)$ are small over typical correlation lengths in the solvent. For Lennard–Jones interaction potentials, $u_0(r)$ and $u_1(r)$ correspond to the parts of $u(r)$ that yield repulsive and attractive *forces*, as prescribed by WCA theory.[11]

If the interparticle interactions are electrostatic in nature, we must split the Coulomb potential into a long-ranged, uniformly slowly-varying part and the remaining short-ranged, rapidly-varying part. This is accomplished by introducing a length-scale σ over which the $1/r$ dependence of the Coulomb potential can be separated,

$$v(r) = \frac{1}{r} = \frac{\mathrm{erfc}(r/\sigma)}{r} + \frac{\mathrm{erf}(r/\sigma)}{r} \equiv v_0(r) + v_1(r), \tag{2}$$

where we have denoted $v_0(r)$ and $v_1(r)$ as the short-ranged, rapidly-varying and long-ranged, slowly-varying portions of the potential. The smoothing length σ is chosen on the scale of typical nearest-neighbor distances between charges to ensure that $v_1(r)$ indeed varies slowly over molecular scales. Moreover, the same value of σ should be used for all charges. This enables us to derive an electrostatic LMF equation, where the number density in Eq. (1) is simply replaced by the charge density $\rho_R^q(\mathbf{r})$ and $u_1(r)$ by $v_1(r)$,

$$\mathcal{V}_R(\mathbf{r}) = \mathcal{V}(\mathbf{r}) + \int d\mathbf{r}'\rho_R^q(\mathbf{r}')v_1(|\mathbf{r} - \mathbf{r}'|), \tag{3}$$

where $\mathcal{V}_R(\mathbf{r})$ is the renormalized electrostatic LMF field and $\mathcal{V}(\mathbf{r})$ is the original bare electrostatic field of the full system.[7,8]

Previous work has shown that LMF theory can reproduce the structure of many systems with long-ranged Coulomb interactions,[6–8,12–14] and in recent work, we have shown how one can integrate over this structure to obtain accurate free energies.[8] In particular, the difference in free energy of the full system, $\Omega[\phi]$, and that of the mimic system, $\Omega[\phi_R]$, for the systems considered in Eq. (1) can be expressed analytically as

$$\Delta\Omega_{\mathrm{LMF}}[\phi_R] \equiv \Omega[\phi] - \Omega[\phi_R] = -\frac{1}{2}\int d\mathbf{r}[\rho_R(\mathbf{r}) + \rho_B][\phi_R(\mathbf{r}) - \phi(\mathbf{r})], \tag{4}$$

where we have noted the functional dependence of the free energies on the external field. For electrostatic interactions, the equivalent expression can again be readily

obtained by substituting charge density for number densities and replacing the relevant fields by their electrostatic counterparts,

$$\Delta\Omega_{\mathrm{LMF}}[\mathcal{V}_R] = -\frac{1}{2}\int d\mathbf{r}\rho_R^q(\mathbf{r})[\mathcal{V}_R(\mathbf{r}) - \mathcal{V}(\mathbf{r})]. \tag{5}$$

In addition to the LMF mimic system, we will also consider the *strong coupling approximation* (SCA), which assumes that the long-ranged interactions play a negligible role. The interactions in this system consist of short-ranged solvent–solvent interactions, $u_0(r)$ and/or $v_0(r)$, in addition to short-ranged solute–solvent interactions, $\phi_0(\mathbf{r})$ and/or $\mathcal{V}_0(\mathbf{r})$, which result from an analogous separation of the one-body external fields into short- and long-ranged components. The SCA system serves to illustrate the role of long-ranged interactions in determining the structure and thermodynamics of molecular systems, when compared to the full system that includes both short- and long-ranged interactions.

3. Short-Ranged Water Models

We now discuss the models we use for the solvent. While these concepts can be extended to long ranged interactions in more general models of the solvent, here we focus on the widely used extended simple point charge (SPC/E) model of water,[15] which gives a good qualitative description of the thermodynamic properties of water that are needed to describe the hydrophobic solvation processes at the focus of this work. The SPC/E model is a rigid model consisting of an LJ core centered at the oxygen atom and partial charges on the oxygen and hydrogen atoms. Therefore, the solvent–solvent interactions in the full system consist only of LJ interactions, $u(r)$, and Coulomb interactions, $q_iq_jv(r)$, between atomic sites i and j.

Because the full model contains both LJ and Coulombic interactions, we can define two classes of truncated models.[13,16,17] The first class of short-ranged water model is the *Gaussian-truncated* (GT) model. In the GT model, water molecules interact via the full LJ potential $u(r)$, but $v(r)$ is replaced by the short-ranged $v_0(r)$, such that the charges interact through $q_iq_jv_0(r)$. With this replacement, we have truncated the point charges by surrounding them with neutralizing Gaussian distributions of width σ, which removes the long-ranged portion of the Coulomb potential. We will use the GT model to study the effects of long-ranged Coulomb interactions in the next section.

In addition to the GT model, we can further truncate the solvent–solvent LJ interactions. This yields the *GT repulsive core* (GTRC) variant of the water model, whereby the charge–charge interactions are given by $q_iq_jv_0(r)$ between sites i and j, and the LJ interactions are replaced by the short-ranged WCA core $u_0(r)$. This model gives an accurate minimal representation of the hydrogen bond network of liquid water,[16] but neglects the LJ attractions necessary to reproduce drying associated with hydrophobic effects, for example Ref. 17. We will use this model to

illustrate the utility of LMF theory in the context of hydrophobic solvation (and drying) in Sec. 5.

4. Alchemical Free Energy Calculations

In "computational alchemy," we are interested in the free energy required to transform one molecule into another,[1,18] e.g. a hydrophobic solute into a hydrophilic one. This class of free energy calculations is often used in protein–ligand binding,[18–22] for example, where one is often interested in the differences in binding free energies among many molecules that may bind to a protein, such that one can rank candidate molecules in order of binding strength. Computing such differences can be accomplished by calculating the free energy change upon transforming one bound molecule into another, while it is still bound to the protein. Alchemical transformations are not limited to protein–ligand binding, and have been widely used to examine a range of chemical processes, including solvent effects on crystal nucleation from solution,[23] the stability of crystal polymorphs,[24] molecular isomerization in bulk and confined systems,[25] partition coefficients,[26] conformational changes,[27] and the impact of protonation state on ion selectivity in transmembrane systems.[28]

We now describe how to compute alchemical free energy changes using LMF theory as follows, which simplifies these computations by eliminating the need for the explicit evaluation of long-ranged interactions.

4.1. *Thermodynamic cycle for computational alchemy*

We now consider an arbitrary fixed solute in solution, interacting with a solvent. This solvent could be water, a mixture of a number of solvents, a protein, etc. Here, solvent simply refers to the environment surrounding the solute of interest. The solvent–solvent interactions can be both electrostatic and nonelectrostatic in nature, and the solvent molecules also interact with the solute via the potential $\psi(\mathbf{r};\xi)$. This solute–solvent potential, fixed at the origin, is coupled to a parameter ξ which controls the alchemical transformation. Note that this can be readily generalized to the case of a set of potentials $\{\psi_i(\mathbf{r};\{\xi_j\})\}$ coupled to numerous parameters $\{\xi_j\}$ that control various aspects of the potential. This is what would be needed when fixing a molecule, for example, and transforming this molecule into another.

When the coupling parameter ξ is zero, the solute–solvent interaction is that of the initial solute of interest, schematically illustrated in the lower left panel of Fig. 1. We are interested in the free energy change $\Omega^{(0\to1)}$ upon transforming this solute ($\xi = 0$) to another ($\xi = 1$), as indicated by the lower path in Fig.1. In general, this path may require many nonphysical intermediate states, each requiring the accurate evaluation of long-ranged interactions at every state, and these difficulties are schematically indicated by the red arrow connecting the lower panels in Fig. 1. However, we wish to calculate this free energy difference using short-ranged systems

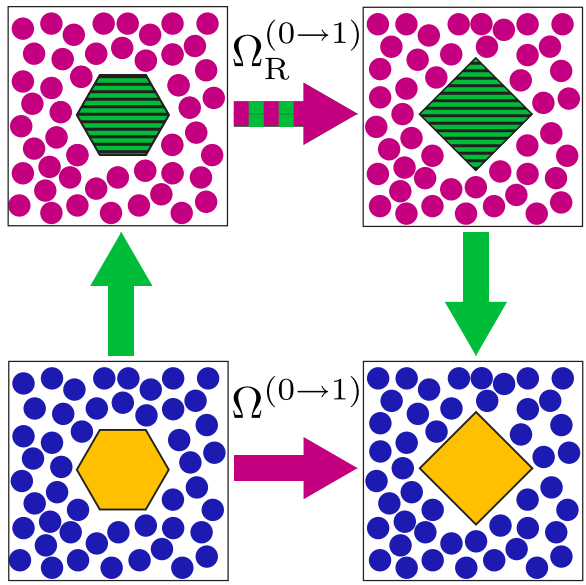


Fig. 1. (Color online) Thermodynamic cycle for calculating alchemical free energy changes via LMF theory. The bottom leg of the cycle corresponds to changing the solute from state $\xi = 0$ to state $\xi = 1$, and in general can involve a change in the shape of the solute and the nature of the solute–solvent interactions. This transformation occurs in the full system, where both the solute–solvent and solvent–solvent interactions can be long ranged, and is therefore described by the free energy $\Omega^{(0\to1)}$. The top leg of the cycle depicts the analogous transformation being performed in the mimic system, with a short-ranged reference solvent and a renormalized solute at both state $\xi = 1$ and state $\xi = 0$. Performing the alchemical transformation in the mimic system is characterized by the free energy $\Omega_R^{(0\to1)}$. The sum of the free energies of the paths depicted by green arrows is a difference of analytic LMF corrections derived for solvation free energies, as described in the text.

only, and to do so, we will use LMF theory by defining an alternative path to obtain $\Omega^{(0\to1)}$.

The alternative path to compute $\Omega^{(0\to1)}$ corresponds to the upper panel in Fig. 1, in conjunction with the green arrows connecting the upper and lower transformations. In the LMF alchemical transformation illustrated in the upper panel, we begin with a mimic system that has the same structure as that of the fully interacting initial state. This LMF system is described by short-ranged solvent–solvent interactions $v_0(r)$ and renormalized solute–solvent interactions $\psi_R(\mathbf{r}; \xi = 0)$, such that the renormalization induces the same solute–solvent structure as in the full system. We can then consider performing the alchemical transformation in the mimic system, where the renormalized solute potential $\psi_R(\mathbf{r}; \xi = 0)$ is slowly transformed into a different renormalized solute with potential $\psi_R(\mathbf{r}; \xi = 1)$. This process is illustrated by the upper leg of the thermodynamic cycle in Fig. 1 and is associated with the free energy $\Omega_R^{(0\to1)}$. This process will in general require as many steps as performing the transformation in the full system. However, this process is more efficient, since we have eliminated much of the evaluation of long-ranged interactions at each step of

the transformation process. Hence, we have connected the upper panels with a red and green striped arrow.

The free energy associated with the green vertical arrows in Fig. 1 can be evaluated analytically using Eq. (4).[8] Therefore, we find that the free energy of the alchemical transformation is given by

$$\begin{aligned}\beta\Omega^{(0\to1)} &= \beta\Omega_R^{(0\to1)} + \beta\Omega_{\mathrm{LMF}}^{(\xi=1)} - \beta\Omega_{\mathrm{LMF}}^{(\xi=0)} \\ &= \beta\Omega_R^{(0\to1)} - \frac{\beta}{2}\int d\mathbf{r}[\psi_R(\mathbf{r};\xi=1) - \psi(\mathbf{r};\xi=1)][\rho_R(\mathbf{r};\xi=1) + \rho_B] \\ &\quad + \frac{\beta}{2}\int d\mathbf{r}[\psi_R(\mathbf{r};\xi=0) - \psi(\mathbf{r};\xi=0)][\rho_R(\mathbf{r};\xi=0) + \rho_B]. \qquad (6)\end{aligned}$$

The free energy of the transformation in the mimic system $\Omega_R^{(0\to1)}$ can be obtained by using standard techniques.[1] We can therefore calculate general free energy changes in a fully interacting, long-ranged system from simulations involving mimic systems with only short-ranged interactions.

4.2. *Alchemical addition of solute–solvent attractions*

We now demonstrate the LMF formalism for alchemical transformations involving the hydration of uncharged spherical solutes of various sizes with attractive solute–solvent van der Waals interactions, generalizing previous work on purely repulsive hydrophobic solutes.[17,29] In this section, we truncate solvent–solvent electrostatic interactions only, so in the mimic system, the full SPC/E water model is replaced by its GT variant; other examples using the GTRC model will be discussed later. LMF theory will be used to account for the effects of the long-ranged electrostatic interactions neglected in the GT model (with $\sigma = 4.5$ Å) given by the green arrows in Fig. 1.

For small purely hydrophobic solutes, the hydrogen bond network of water remains intact, and the solvation structure and thermodynamics are determined by short-ranged interactions alone.[29–32] In this regime, the effects of the electrostatic LMF corrections are negligible, and the SCA system very accurately reproduces the solvation properties of the full system.[8,16,17] When the solute size is increased beyond a crossover radius, for solute volumes of roughly 1 nm^3, the hydrogen bond network of water can no longer be maintained around the solute.[17,29–32] Water instead nucleates a soft, liquid–vapor-like interface around the hydrophobic solute, such that the solvation-free energy scales as γA, where γ is the surface tension of water and A is the solute surface area. In previous work,[8,16,17] we showed that GT water qualitatively reproduces the crossover in the scaling behavior of the solvation-free energy, but slightly underestimates the free energies in the large solute regime because it has a lower surface tension than the full water model. We additionally showed that the small disparities between the full and GT solvation-free energies for large solutes could be corrected using Eq. (4), leading to quantitatively accurate predictions.[8]

4.2.1. *Solute models and definition of the alchemical transformation*

We now consider adding solute–solvent van der Waals attractions to the purely repulsive spherical solutes studied previously,[17,29] a simple but general demonstration of LMF theory for alchemical transformations. We write the solute–water interaction potential with the attractive interactions coupled linearly by the parameter ξ,

$$U_{\mathrm{sw}}^{(\xi)} = U_{0,sw} + \xi U_{1,sw}, \tag{7}$$

where $U_{0,\mathrm{sw}}$ and $U_{1,\mathrm{sw}}$ are the repulsive and attractive portions of the solute–water interaction energy, respectively, such that $U_{\mathrm{sw}}^{(\xi)} = \sum_{i=1}^{N} u_{\mathrm{sw}}(r_i;\xi)$ and N is the number of water molecules in the system. Here, $u_{\mathrm{sw}}(r)$ is an integrated 9-3 LJ potential,

$$\begin{aligned} u_{\mathrm{sw}}(r) = \pi\varepsilon_{sw}\rho\sigma_{\mathrm{sw}}^3 \Bigg[\frac{4}{5}\sigma_{\mathrm{sw}}^9 \left(\frac{1}{8rr_+^8} - \frac{1}{9r_+^9} - \frac{1}{8rr_-^8} + \frac{1}{9r_-^9} \right) \\ - 2\sigma_{\mathrm{sw}}^3 \left(\frac{1}{2rr_+^2} - \frac{1}{3r_+^3} - \frac{1}{2rr_-^2} + \frac{1}{3r_-^3} \right) \Bigg], \end{aligned} \tag{8}$$

where $r_\pm = r \pm R_S$, R_S tunes the size of the particle, $\rho = 0.0240$ Å^{-3}, $\varepsilon_{sw} = 0.882$ kJ/mol, and $\sigma_{\mathrm{sw}} = 3.468$ Å, respectively.[17,29] The effective hard-sphere radius of the particle, R_{HS}, can be estimated as[33]

$$R_{\mathrm{HS}} \approx \int_0^\infty dr \{1 - \exp[-\beta u_{0,sw}(r)]\}, \tag{9}$$

where $u_{0,sw}(r)$ is the purely repulsive part of the potential.

This type of solute–solvent potential is associated with the nonelectrostatic portion of the external field, $\phi(\mathbf{r};\xi) = u_{\mathrm{sw}}(\mathbf{r};\xi)$, because no solute–solvent electrostatic interactions are present. The renormalized potential acting on solvent site α can then be written as

$$\psi_{R,\alpha}(\mathbf{r};\xi) = \phi_\alpha(\mathbf{r};\xi) + q_\alpha \mathcal{V}_R(\mathbf{r};\xi), \tag{10}$$

where the LMF treatment has only been applied to electrostatic interactions and ϕ_α is the nonelectrostatic solute–solvent potential acting on solvent site α. If we separate only the electrostatic portion of the renormalized potential into short- and long-ranged components for use with LMF theory, this field can be written as $\psi_{R,\alpha}(\mathbf{r};\xi) = \psi_{0,\alpha}(\mathbf{r};\xi) + \psi_{R1,\alpha}(\mathbf{r};\xi)$, where the short-ranged portion is

$$\psi_{0,\alpha}(\mathbf{r};\xi) = \phi_\alpha(\mathbf{r};\xi) + q_\alpha \mathcal{V}_0(\mathbf{r};\xi) \tag{11}$$

and the long-ranged component is simply given by

$$\psi_{R1,\alpha}(\mathbf{r};\xi) = q_\alpha \mathcal{V}_{R1}(\mathbf{r};\xi). \tag{12}$$

For the solute–solvent potentials considered here, given by Eq. (7), the local, short-ranged electrostatic component of the renormalized solute field is zero, $\mathcal{V}_0(\mathbf{r}) = 0$,

but $\mathcal{V}_{R1}(\mathbf{r}) \neq 0$, because this contains the averaged effects of the long-ranged solvent–solvent electrostatic interactions.

4.2.2. *Performing the solute transformations*

In order to compute the free energy of the alchemical transformation in the full and mimic systems, we first need to compute the free energy of turning on the solute–solvent attractive potential. For the full system, this corresponds to the lower leg of the cycle in Fig. 1. The free energy change along the transformation $\xi = 0 \rightarrow \xi = 1$ can be obtained by thermodynamic integration,[1,34,35]

$$\Delta G^{(0\rightarrow 1)} = \int_0^1 d\xi \left\langle \frac{\partial \mathcal{H}^{(\xi)}}{\partial \xi} \right\rangle_\xi = \int_0^1 d\xi \langle U_{1,sw} \rangle_\xi, \tag{13}$$

where $\mathcal{H}^{(\xi)}$ is the ξ-dependent Hamiltonian of the system and $\langle \cdots \rangle_\xi$ indicates an ensemble average over configurations of the system interacting with coupling parameter ξ. Note that the remainder of the paper will use ΔG to indicate free energy changes because the simulations are performed in the isothermal–isobaric ensemble. The upper leg of the thermodynamic cycle in Fig. 1 corresponds to performing the alchemical transformation in the mimic system with the same solute–solvent structure. The free energy of this process can be decomposed into two parts:

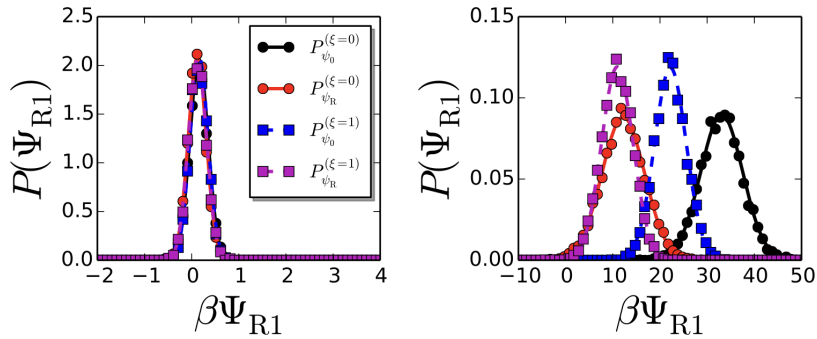
$$\Delta G_R^{(0\rightarrow 1)} = \Delta G_{\text{SCA}}^{(0\rightarrow 1)} + (\Delta G_{R1}^{(1)} - \Delta G_{R1}^{(0)}), \tag{14}$$

where $\Delta G_{\text{SCA}}^{(0\rightarrow 1)}$ is the free energy of turning on the solute–solvent attractions in the SCA system, where all structural corrections from the truncated Coulomb interactions in GT water are ignored, and $\Delta G_{R1}^{(\xi)}$ is the free energy of turning on the slowly-varying portion of the LMF corrected solute–solvent attractive potential at a specific value of the solute–solvent coupling parameter ξ. The free energy $\Delta G_{\text{SCA}}^{(0\rightarrow 1)}$ is determined by using Eq. (13).

We determined the free energy of turning on the slowly-varying portion of the LMF potential, $\Delta G_{R1}^{(\xi)}$, using the Gaussian approximation[1,34]

$$\beta \Delta G_{R1}^{(\xi)} = \frac{1}{2} [\langle \Psi_{R1}(\bar{\mathbf{R}}; \xi) \rangle_{\text{SCA}} + \langle \Psi_{R1}(\bar{\mathbf{R}}; \xi) \rangle_R], \tag{15}$$

where $\langle \cdots \rangle_{\text{SCA}}$ and $\langle \cdots \rangle_R$ indicate ensemble averages in the SCA and mimic systems, respectively, and $\bar{\mathbf{R}}$ indicates a point in the configuration space of the system. To verify the accuracy of this approximation, we calculated the probability distributions $P_0^{(\xi)}(\Psi_{R1})$ and $P_R^{(\xi)}(\Psi_{R1})$, where $\Psi_{R1}(\bar{\mathbf{R}}) = \sum_{i=1}^N \psi_{R1}(\mathbf{r}_i; \bar{\mathbf{R}})$. These distributions are shown in Fig. 2 for representative solute sizes and for both values of ξ in the corresponding SCA and mimic systems. The symbols in Fig. 2 are data calculated from simulation, while solid lines are Gaussian distributions with the same mean and variance as the calculated distribution. All distributions studied are found to follow Gaussian statistics, verifying the use of Eq. (15) to estimate the free energy $\Delta G_{R1}^{(\xi)}$.

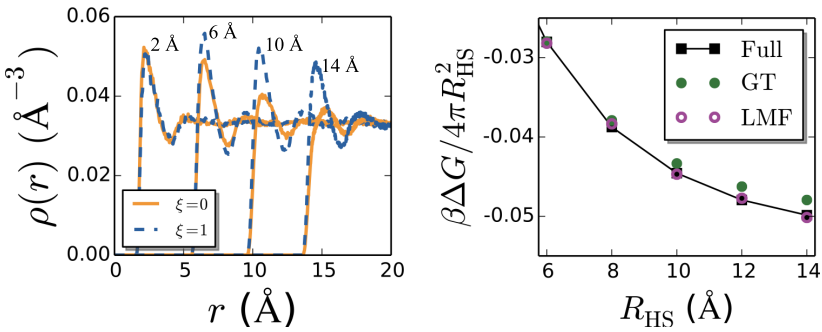


(a) (b)

Fig. 2. Probability distributions of the potential energy due to the slowly-varying portion of the electrostatic LMF, $\mathcal{V}_{R1}$, for solutes with effective hard-sphere radii of (a) $R_{\mathrm{HS}} = 4$ Å and (b) $R_{\mathrm{HS}} = 14$ Å. Symbols are data obtained from simulation and lines are Gaussian distributions with the same mean and variance of the corresponding computed distributions.

4.2.3. *Results*

We performed alchemical transformations for spherical solutes of varying size. Before discussing the free energetics of the transformation, we discuss the solvation structure around the solutes, and how this structure changes upon the addition of solute–solvent attractions. Density distributions of water, $\rho(r)$, around representative solutes spanning the small to large solute size regimes are shown in Fig. 3(a), in the presence ($\xi = 1$) and absence ($\xi = 0$) of solute–solvent attractive interactions in the full system; results for the SCA and mimic systems are similar to those shown here. For small solutes, the solvation structure is determined entirely by the need for water to maintain its hydrogen bond network around the solute.[17,29–32] Therefore,



(a) (b)

Fig. 3. (a) Nonuniform density of water around solutes with HS radii $R_{\mathrm{HS}} = 2, 6, 10$, and 14 Å (indicated in the figure) with ($\xi = 1$) and without ($\xi = 0$) solute–solvent LJ attractions. (b) Free energy change per unit solute surface area upon turning on the attractive portion of the solute–water potential $U_{1,sw}$ calculated in the full, SPC/E water system, the short-ranged GT water system, and from LMF theory.

for R_{HS} less than roughly 5 Å, attractive interactions have a negligible effect on the solvation structure.

The hydrogen bond network of water cannot be maintained around large solutes, and a soft, liquid–vapor-like interface is nucleated in the vicinity of the solute surface. This results in the phenomena of drying in the interfacial region, which becomes more pronounced as the solute size is increased,[17,29–32] and is particularly notable around repulsive solutes ($\xi = 0$). This decreases the height of the first peak of $\rho(r)$ as R_{HS} increases (Fig. 3(a)). This soft interface is highly responsive to perturbations.[36–41] The attractive solute–solvent interactions with $\xi = 1$ significantly alter the solvation structure, effectively "pulling" water molecules back toward the solute surface and increasing the magnitude of the first peak in $\rho(r)$ with respect to that observed in the absence of solute–solvent attractions. Note, however, that, for large solutes, the height of the first peak in $\rho(r)$ decreases with increasing solute size for both $\xi = 0$ and $\xi = 1$.

The free energy differences calculated using the full SPC/E model are compared with those calculated using the short-ranged GT water model and with those computed from LMF theory in Fig. 3(b) for a range of solute radii. For small solutes, GT water reproduces the alchemical free energies of the full system and the LMF corrections are negligible. In contrast, the magnitude of the free energy change for large solutes is underestimated in the SCA system. Physically, the change in the free energy for large solutes corresponds to a change in the solute–water interfacial tension, and GT underestimates the magnitude of this change. Using the LMF formalism for alchemical transformation, Eq. (6) corrects this solute–water interfacial tension and quantitatively reproduces the free energies of the full system for all solute sizes. These results illustrate that the LMF theory framework can be applied to alchemical transformations with quantitative accuracy.

5. Umbrella Sampling: Density Fluctuations and Solvation

The preceding section describes LMF-based methods for computing free energy changes of a solute in solution as it is transformed from one state to another. In general, one would like to compute a free energy along some relevant order parameter of interest. However, to compute such free energy profiles, one often needs to employ non-Boltzmann sampling techniques.[2]

One type of non-Boltzmann sampling often employed is umbrella sampling,[1,9] in which many simulations of a system are performed, each with a unique "umbrella" potential that biases the system toward a particular value of the order parameter of interest. The results of this set of simulations can then be stitched together through a number of means, such as the multi-state Bennett acceptance ratio[42] (MBAR) and the weighted histogram analysis method (WHAM).[43,44] Therefore, in order to use the LMF framework to calculate general free energy landscapes as a function of an order parameter of interest, we can combine LMF theory with non-Boltzmann sampling

techniques and the rigorous methods for obtaining free energies from sets of biased simulations.

We have developed such a formalism, and the technical details are described in the appendix. We combine LMF theory, umbrella sampling, and MBAR to calculate ensemble averages in a full system from sampling performed in a mimic or SCA system. Umbrella sampling is performed with an appropriate biasing potential for each window *and* an accompanying window-specific LMF potential. This ensures an accurate representation of the full system structure in *each window*. Free energy differences between these windows, and the associated free energy landscape, can then be determined for the mimic system using MBAR. Finally, the analytic LMF contribution to the free energy given by Eq. (4) can be used to *reweight* mimic system ensemble averages in order to obtain the corresponding averages in the full system. An analogous procedure can be performed for the SCA system when linear response theory (LRT) is accurate,[14] however, the reweighting must also account for the free energy differences between the SCA and mimic systems, as discussed in Appendix B.

As a demonstration of these theoretical developments, we study density fluctuations in bulk water. Specifically, we seek to calculate the probability of observing N water molecules within a spherical volume v. This probability $P_v(N)$ is intimately related to the free energy of solvation ΔG_{HS} of a hard sphere of the same volume through the potential distribution theorem,[45,46]

$$e^{-\beta \Delta G_{\text{HS}}} = \left\langle e^{-\beta U_{\text{HS}}} \right\rangle = \left\langle \delta_{N_v,0} \right\rangle = P_v(0), \tag{16}$$

where the ensemble averages are performed over the bulk solvent ensemble in the absence of a hard sphere, $U_{\text{HS}} = \sum_{i=1}^{N} u_{\text{HS}}(r_i)$ is the total solute–solvent interaction energy, $u_{\text{HS}}(r)$ is the hard-sphere potential, and $\delta_{N,N'}$ is the Kronecker delta function. Therefore, $\beta \Delta G_{\text{HS}} = -\ln P_v(0)$, where $P_v(0)$ is the probability of observing a cavity the same size and shape of the solute volume v. It was shown earlier that one can use LMF theory to determine ΔG_{HS},[8] and therefore $-\ln P_v(0)$, from a purely short-ranged system with quantitative accuracy for both small and large solute volumes. In this section, we demonstrate that the full distribution $P_v(N)$, and therefore the free energy $-\ln P_v(N)$, can also be obtained from short-ranged systems with high accuracy by combining LMF theory and non-Boltzmann sampling techniques.

For relatively small probe volumes, the distributions $P_v(N)$ can be calculated using the Widom particle insertion method.[2,45–47] However, large volume density fluctuations are rare events not accurately sampled by conventional simulations, and some type of biased, non-Boltzmann sampling technique is needed to calculate $P_v(N)$ accurately. In this case, we use umbrella sampling. At first glance, one would like a biasing potential in terms of the variable of interest, N. However, N is a discrete variable, leading to problems when biasing within MD simulations. Therefore, we use the indirect umbrella sampling (INDUS) method, which dictates that N be coarse-grained by smoothing it with a truncated and shifted Gaussian distribution, resulting in the continuous coarse-grained variable $\tilde{N}$.[48] We can then bias the simulation

toward desired values of $\tilde{N}$, which, with correctly tuned parameters, will closely follow N. After performing a set of n_w simulations that adequately sample N and $\tilde{N}$, with sufficient overlap of the $P_v(N, \tilde{N})$ distributions between neighboring windows, we can reconstruct the desired $P_v(N)$ using MBAR.[42] In all cases, a simple harmonic potential was used to bias the system toward the desired value of $\tilde{N}$,

$$V_k(\tilde{N}) = \frac{\kappa}{2}\left(\tilde{N} - \eta_k\right)^2, \tag{17}$$

where η_k is the desired value of $\tilde{N}$ in window k, and the force constant κ was chosen independently for each window to yield sufficient overlap between neighboring windows. In order to maintain a constant bulk density far from the probe volume across all biased ensembles, we perform simulations in the isothermal–isobaric (NPT) ensemble. For the GTRC model, the van der Waals theory-corrected pressure was used to maintain the correct bulk density, as described previously.[16]

For small volumes, long-ranged interactions have a negligible influence on the solvation behavior. Therefore, density fluctuations in the bulk fluid on small length scales are determined by local, hydrogen bonding interactions, and these fluctuations will not be altered upon the inclusion of an LMF in the short-ranged reference system. In contrast, for large volumes, $P_v(N)$ displays non-Gaussian tails at low N when long-ranged LJ attractions are present in the system. This is illustrated by the free energy $-\ln P_v(N)$ shown in Fig. 4 for GT water, which will we use as the "full" system here. The appearance of fat tails in $P_v(N)$ is a manifestation of the nucleation of a nanoscale liquid–vapor interface at the surface of a large cavity, and is consistent with the appearance of interfacial drying due to unbalanced attractive LJ forces.[6,31,32,38,49,50] Indeed, removal of the LJ attractions in GTRC water eliminates drying and this model does not display the crossover in scaling of free energies with solute size at the same density as GT water.[17] Because this crossover is absent in GTRC water, its $P_v(N)$ does not contain fat tails at low N, as shown in Fig. 4. In addition, the GTRC $P_v(N)$ is nearly Gaussian for all N, similar to what is observed in a hard-sphere fluid.[51]

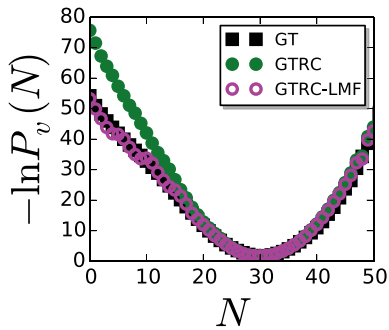


Fig. 4. Free energies, $-\ln P_v(N)$, for finding N water molecules in a spherical observation volume v with a radius of 6 Å determined in GT and GTRC water. Also shown are the results obtained from a combination of LRT and LMF theory (GTRC-LMF), as described in Appendix B.

Using LRT,[14] we determine the densities and renormalized fields in each biased ensemble necessary for performing the reweighting described in Appendix B without additional mimic system simulations. Upon doing so, averages obtained from configurations obtained in the GTRC system are appropriately reweighted, yielding a free energy $-\ln P_v(N)$ consistent with that of GT water, the desired result, as shown in Fig. 4. By including the averaged effects of the unbalanced LJ forces *in each biased ensemble*, LMF theory can capture the nontrivial fat tails observed at low N with high accuracy.

6. Conclusions

We have extended LMF theory to two major classes of free energy calculations, alchemical transformations and potentials of mean force.[1] The accuracy of this theoretical framework was demonstrated by computing free energies for model systems and processes related to the hydration of nonpolar solutes. These are challenging test cases for LMF theory, because unbalanced interfacial forces arising from long-ranged interactions are important for their structure and thermodynamics. Despite this challenge, LMF theory accurately predicts the free energetics of these systems using only a short-ranged mimic system, removing the need to include explicit treatments of long-ranged interactions. In addition to circumventing the explicit treatment of long-ranged interactions in simulations, the energetic partitioning inherent within LMF theory enables physical decompositions to interpret thermodynamic driving forces.[52–55] The frameworks presented here will extend the utility of these decompositions to general processes of interest.

LMF theory is not limited to the relatively simple model systems studied here. In fact, the LMF framework for free energy calculations should be of significant importance to large-scale molecular systems, like those typical of biomolecular and materials research. Simulations involving extremely large numbers of particles and long timescales are limited by the scaling of the algorithms involved in the computation. While the best algorithms for handling long-ranged electrostatic interactions scale as $N\log(N)$,[3] often with a large prefactor, LMF-based approaches are short-ranged and scale linearly with N. The ability of LMF theory to avoid computationally expensive and theoretically complex long-ranged interactions should therefore enable thermodynamic calculations in massively large-scale systems.[56–60] Moreover, the LMF framework also alleviates conceptual burdens associated with artifacts induced by periodic boundaries and lattice summation techniques,[4,5] enabling reductions in systems size necessary to achieve convergence in many calculations.

LMF theory is not limited to the order parameters used here and can in general be extended to perform non-Boltzmann sampling with general order parameters of interest.[61] Similarly, the LMF framework should also be readily extendable to other types of enhanced sampling methods beyond umbrella sampling,[1,34,62] because LMF theory was not employed to sample configuration space. In fact, our results in Sec. 5

show that, even in a situation where the SCA and mimic systems have significant structural differences, LMF theory can be used to simply reweight the results obtained from SCA configurations and explicit sampling of the mimic system is not needed. Such large structural differences between the SCA and mimic systems are typically not observed in many applications, however, especially when only truncating Coulomb interactions. Therefore, we expect that in most situations of interest, enhanced sampling methods will only be needed to sample the phase-space of the SCA system, which can be connected to the mimic and full systems using a combination of linear response and LMF theories. This will enable the use of LMF theory largely as a post-processing tool, avoiding the need to significantly modify existing molecular simulation software.

We conclude with a discussion of the limitations of LMF theory in the form used here. As discussed previously,[7] the physically-justified approximations permitting a molecular scale choice for the smoothing length σ in the derivation of the LMF equation are expected to fail in systems with significant intrinsic long-ranged pair correlations, like those found along a liquid–vapor interface with capillary waves, and near a critical point due to a diverging correlation length. In addition, the current LMF theory framework may not be appropriate for computing thermodynamic properties of systems driven significantly far from equilibrium, because it is inherently an equilibrium theory. Nonequilbrium extensions are needed in these contexts, although simple approximations can be made to enable the use of the equilibrium LMF theory when there exists a separation of timescales and the solvent is in local equilibrium, as discussed previously.[63] This does not prohibit the use of nonequilbrium work relations[34,64,65] to determine free energy differences between equilibrium states involving short-ranged mimic systems; only the LMF-based long-ranged correction to the free energy, $\Delta\Omega_{\mathrm{LMF}}$, requires equilibrium quantities as input.

Acknowledgments

This work was supported by National Science Foundation Grants CHE0848574 and CHE1300993.

Appendix A. Ensemble Averages in the Full System from Biased Sampling Performed in the Mimic System

We consider performing umbrella sampling in the mimic system. Each of the K biased ensembles will have an associated LMF. Therefore, we will have a set of renormalized fields $\{\phi_{\mathrm{R},k}(\mathbf{r})\}$ associated with each of the K windows. Omitting (trivial) momentum space terms and focusing on configuration space, the partition function in the unbiased mimic ensemble is given by

$$Z_R = \int d\,\bar{\mathbf{R}} e^{-\beta[U_0(\bar{\mathbf{R}})+\Phi_{R1}(\bar{\mathbf{R}})]} \tag{A.1}$$

and the partition function of the kth biased window is

$$Z_{R,k} = \int d\bar{\mathbf{R}} e^{-\beta[U_0(\bar{\mathbf{R}})+\Phi_{R1,k}(\bar{\mathbf{R}})+V_k(\bar{\mathbf{R}})]}, \tag{A.2}$$

where $\Phi_{R1,k} = \sum_{i=1}^{N} \phi_{R1,k}(\mathbf{r}_i)$. The free energy difference between the kth window and the unbiased mimic ensemble is given by

$$\Delta F_{R,k} = F_{R,k} - F_R = \beta^{-1} \ln\left(\frac{Z_R}{Z_{R,k}}\right). \tag{A.3}$$

Estimates for the normalization constants Z_R and $Z_{R,k}$ are given, respectively, by[42,66]

$$\mathcal{Z}_R \approx \sum_{j=1}^{K} \sum_{n=1}^{N_j} p_{j,n}^{(R)} \tag{A.4}$$

and

$$\mathcal{Z}_{R,k} \approx \sum_{j=1}^{K} \sum_{n=1}^{N_j} p_{j,n}^{(R)} e^{-\beta V_k(\bar{\mathbf{R}}_{j,n})} e^{-\beta \Delta\Phi_{R1,k}(\bar{\mathbf{R}}_{j,n})}, \tag{A.5}$$

where $\Delta\Phi_{R1,k} \equiv \Phi_{R1,k} - \Phi_{R1}$.

The set of free energy differences $\{\Delta F_{R,k}\}$ can then be obtained self-consistently, consistent with typical MBAR, by replacing $V_k(\bar{\mathbf{R}})$ with $\tilde{V}_k(\bar{\mathbf{R}}) \equiv V_k(\bar{\mathbf{R}}) + \Delta\Phi_{R1,k}(\bar{\mathbf{R}})$ and $p_{j,n}$ with $p_{j,n}^{(R)}$:

$$e^{-\beta \Delta F_{R,i}} = \sum_{j=1}^{K} \sum_{n=1}^{N_j} \frac{e^{-\beta \tilde{V}_i(\bar{\mathbf{R}}_{j,n})}}{\sum_{k=1}^{K} N_k e^{\beta \Delta F_{R,k} - \beta \tilde{V}_k(\bar{\mathbf{R}}_{j,n})}}. \tag{A.6}$$

The $\Delta\Phi_{R1,k}$ term presents a problem, since we may not know $\phi_{R1}(\mathbf{r})$ of the unbiased ensemble in advance (if we could obtain this easily, we would not need to overcome sampling issues in the first place!). However, the free energy differences obtained from the MBAR calculation are known up to an undetermined constant. Therefore, we can choose this constant such that it exactly cancels the Φ_{R1} term in each window, and we can write

$$e^{-\beta \Delta F_{R,i}} = \sum_{j=1}^{K} \sum_{n=1}^{N_j} \frac{e^{-\beta[V_i(\bar{\mathbf{R}}_{j,n})+\Phi_{R1,i}(\bar{\mathbf{R}}_{j,n})]}}{\sum_{k=1}^{K} N_k e^{\beta[\Delta F_{R,k} - V_k(\bar{\mathbf{R}}_{j,n}) - \Phi_{R1,k}(\bar{\mathbf{R}}_{j,n})]}}, \tag{A.7}$$

without any loss of generality. Ensemble averages in the mimic system can then be estimated following

$$\langle A \rangle_R \approx \sum_{j=1}^{K} \sum_{n=1}^{N_j} \frac{A_R(\bar{\mathbf{R}}_{j,n})}{\sum_{k=1}^{K} N_k e^{\beta[\Delta F_{R,k} - V_k(\bar{\mathbf{R}}_{j,n}) - \Phi_{R1,k}(\bar{\mathbf{R}}_{j,n})]}}, \tag{A.8}$$

where the notation A_R emphasizes that the observable A is evaluated over configurations in the mimic system.

However, averages in the mimic system are not of interest, and we want to calculate averages in the *full* system. The free energy differences between the unbiased full ensemble and the K biased full ensembles is enough to estimate the probability of *every* sample in the unbiased ensemble of the *full* system. From the LMF theory-based framework for free energy calculations, we know that the free energy difference between the full and mimic systems is given by[8]

$$F - F_R = F_{\mathrm{LMF}} = -\frac{1}{2}\int d\mathbf{r}[\rho_R(\mathbf{r}) + \rho_B][\phi_R(\mathbf{r}) - \phi(\mathbf{r})]. \tag{A.9}$$

Therefore, the free energy difference between the kth biased mimic ensemble and the kth biased full ensemble is given by an analogous expression

$$F_k - F_{R,k} = F_{\mathrm{LMF},k} = -\frac{1}{2}\int d\mathbf{r}[\rho_{R,k}(\mathbf{r}) + \rho_B][\phi_{R,k}(\mathbf{r}) - \phi_k(\mathbf{r})]. \tag{A.10}$$

Now, note that $\Delta F_k = F_k - F$ are the desired free energy differences, and we can define $\tilde{F}_k \equiv F_{R,k} + F_{\mathrm{LMF},k}$ to be the LMF theory-based estimate of the free energy in the kth window of the full system. By again taking advantage of the fact that the free energies determined using MBAR have an undetermined constant, averages in the *full*, *unbiased* ensemble can be obtained from

$$\langle A\rangle \approx \sum_{j=1}^{K}\sum_{n=1}^{N_j}\frac{A_R(\bar{\mathbf{R}}_{j,n})}{\sum_{k=1}^{K} N_k e^{\beta[\Delta F_{R,k} + F_{\mathrm{LMF},k} - V_k(\bar{\mathbf{R}}_{j,n})]}}. \tag{A.11}$$

Equation (A.11) is the main result of this subsection, and states that averages in the unbiased, full ensemble can be obtained from calculations in the mimic system by simply *reweighting* the free energy differences between windows by a simple, analytic correction, $F_{\mathrm{LMF},k}$.

Appendix B. Ensemble Averages in the Full System from Biased Sampling Performed in the Strong Coupling System

Within the linear response regime, we can seek to evaluate the full free energy directly from the SCA system. We again consider K biasing potentials and the corresponding biased ensembles in the strong-coupling (SCA) system. The unbiased SCA ensemble has a partition function

$$Z_0 = \int d\,\bar{\mathbf{R}} e^{-\beta U_0(\bar{\mathbf{R}})}, \tag{B.1}$$

while the kth biased SCA ensemble has the partition function

$$Z_{0,k} = \int d\,\bar{\mathbf{R}} e^{-\beta[U_0(\bar{\mathbf{R}}) + V_k(\bar{\mathbf{R}})]}. \tag{B.2}$$

The free energy difference between the kth biased SCA ensemble and the unbiased SCA ensemble is

$$\Delta F_{0,k} = F_{0,k} - F = \beta^{-1} \ln \left(\frac{Z_0}{Z_{0,k}} \right). \tag{B.3}$$

This set of free energy differences can be obtained using MBAR, in analogy with the full system, as

$$e^{-\beta \Delta F_{0,i}} = \sum_{j=1}^{K} \sum_{n=1}^{N_j} \frac{e^{-\beta V_i(\bar{\mathbf{R}}_{j,n})}}{\sum_{k=1}^{K} N_k e^{\beta \Delta F_{0,k} - \beta V_k(\bar{\mathbf{R}}_{j,n})}} \tag{B.4}$$

and averages in the SCA system can be obtained as follows:

$$\langle A \rangle_0 \approx \sum_{j=1}^{K} \sum_{n=1}^{N_j} \frac{A_0(\bar{\mathbf{R}}_{j,n})}{\sum_{k=1}^{K} N_k e^{\beta \Delta F_{0,k} - \beta V_k(\bar{\mathbf{R}}_{j,n})}}, \tag{B.5}$$

where the notation A_0 is used to emphasize that the observable A is evaluated over configurations in the SCA system.

However, as stated above, we are really interested in averages in the full system. These averages can be calculated if we have an estimate for the set of free energy differences $\{\Delta F_k\}$. To obtain this set of free energy differences, we note that

$$\begin{aligned} \Delta F_k &= F_k - F \approx F_{R,k} + F_{\mathrm{LMF},k} - F \\ &= F_{0,k} + F_{R1,k} + F_{\mathrm{LMF},k} - F, \end{aligned} \tag{B.6}$$

where $F_{R1,k}$ is the free energy of turning on the LMF $\phi_{R1,k}(\mathbf{r})$ in the kth biased ensemble. We now need to determine $F_{\mathrm{LMF},k}$ and $F_{R1,k}$.

Using LRT,[14] we can estimate the set of renormalized fields $\{\phi_{R,k}(\mathbf{r})\}$ from the set of mimic system densities, given for each k by

$$\rho_R(\mathbf{r}) \equiv \big\langle \rho(\mathbf{r}; \bar{\mathbf{R}}) \big\rangle_{\phi_R} \simeq \big\langle \rho(\mathbf{r}; \bar{\mathbf{R}}) \big\rangle_{\phi_{R'}} - \beta \big\langle \delta\rho(\mathbf{r}; \bar{\mathbf{R}}) \delta\Phi_{R1}(\bar{\mathbf{R}}) \big\rangle_{\phi_{R'}}, \tag{B.7}$$

where

$$\delta\rho(\mathbf{r}; \bar{\mathbf{R}}) \equiv \rho(\mathbf{r}; \bar{\mathbf{R}}) - \big\langle \rho(\mathbf{r}; \bar{\mathbf{R}}) \big\rangle_{\phi_{R'}} \tag{B.8}$$

and

$$\delta\Phi_{R1}(\bar{\mathbf{R}}) \equiv \Phi_{R1}(\bar{\mathbf{R}}) - \big\langle \Phi_{R1}(\bar{\mathbf{R}}) \big\rangle_{\phi_{R'}}. \tag{B.9}$$

Here, the densities and fields are iterated to self-consistency using a combination of the LMF equation (1) and the above LRT expression for the density. Thus, the subscript $\phi_{R'}$ indicates the renormalized field of the previous iteration, or a guess for the field if it is the first iteration.

The free energy of turning on each field in the corresponding biased ensemble is given by the Gaussian approximation,[1,34]

$$F_{R1,k} \approx \frac{1}{2}\int d\mathbf{r}[\rho_{R,k}(\mathbf{r}) + \rho_{0,k}(\mathbf{r})]\phi_{R1,k}(\mathbf{r}), \tag{B.10}$$

which is accurate because the renormalized part of the LMF potential varies slowly over molecular length scales. With the renormalized potentials determined through LRT, we can also calculate the LMF free energy term,[8]

$$F_{\mathrm{LMF},k} = -\frac{1}{2}\int d\mathbf{r}[\rho_{R,k}(\mathbf{r}) + \rho_B][\phi_{R,k}(\mathbf{r}) - \phi_k(\mathbf{r})]. \tag{B.11}$$

By again taking advantage of the fact that the free energy differences $\{\Delta F_{0,k}\}$ are determined up to some unknown constant, we can estimate averages in the full system from umbrella simulations of the SCA system through reweighting,

$$\langle A\rangle \approx \sum_{j=1}^{K}\sum_{n=1}^{N_j} \frac{A_0(\bar{\mathbf{R}}_{j,n})}{\sum_{k=1}^{K} N_k e^{\beta[\Delta F_{0,k}+F_{R1,k}+F_{\mathrm{LMF},k}-V_k(\bar{\mathbf{R}}_{j,n})]}}. \tag{B.12}$$

If $F_{R1,k}$ can be obtained accurately with the Gaussian approximation for all k, this is the most efficient way to estimate $\langle A\rangle$. We have found this to typically be true, due to the long-ranged slowly-varying nature of the perturbation. If this is not true, then $\{\Delta F_{R,k}\}$ needs to be determined from biased simulations in the mimic system. We finally note that the above formalism can be trivially extended to electrostatic interactions by replacing number densities with charge densities, $\rho_R^q(\mathbf{r})$, which is zero in the bulk, and by replacing the nonelectrostatic fields $\phi_R(\mathbf{r})$ and $\phi(\mathbf{r})$ by the electrostatic fields $\mathcal{V}_R(\mathbf{r})$ and $\mathcal{V}(\mathbf{r})$.

References

1. Chipot C, Pohorille A (eds.), *Free Energy Calculations: Theory and Applications in Chemistry and Biology*, Springer, Berlin, 2007.
2. Allen MP, Tildesley DJ, *Computer Simulation of Liquids*, Oxford, New York, 1987.
3. Schulz R, Lindner B, Petridis L, Smith JC, Scaling of multimillion-atom biological molecular dynamics simulation on a petascale supercomputer, *J Chem Theory Comput* **5**:2798–2808, 2009.
4. Figueirido F, Del Buono GS, Levy RM, On finite-size effects in computer simulations using the ewald potential, *J Chem Phys* **103**:6133–6142, 1995.
5. Hünenberger PH, McCammon JA, Effect of artificial periodicity in simulations of biomolecular under ewald boundary conditions: A continuum electrostatics study, *Biophys Chem* **78**:69–88, 1999.
6. Weeks JD, Connecting local structure to interface formation: A molecular scale van der waals theory of nonuniform liquids, *Annu Rev Phys Chem* **53**:533–562, 2002.
7. Rodgers JM, Weeks JD, Local molecular field theory for the treatment of electrostatics, *J Phys Condens Matter* **20**:494206, 2008.
8. Remsing RC, Liu S, Weeks JD, Long-ranged contributions to solvation free energies from theory and short-ranged models, *Proc Natl Acad Sci* **113**(11):2819–2826, 2016.

9. Torrie GM, Valleau JP, Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling, *J Comput Phys* **23**:187–199, 1977.
10. Hansen JP, McDonald IR, *Theory of Simple Liquids*, Elsevier Ltd., Amsterdam, 2006.
11. Weeks JD, Chandler D, Andersen HC, Role of repulsive forces in determining the equilibrium structure of simple liquids, *J Chem Phys* **54**:5237–5247, 1971.
12. Chen YG, Kaur C, Weeks JD, Connecting systems with short and long ranged interactions: Local molecular field theory for ionic fluids, *J Phys Chem B* **108**:19874, 2004.
13. Rodgers JM, Weeks JD, Interplay of local hydrogen-bonding and long-ranged dipolar forces in simulations of confined water, *Proc Natl Acad Sci USA* **105**:19136, 2008.
14. Hu Z, Weeks JD, Efficient solutions of self-consistent mean field equations for dewetting and electrostatics in nonuniform liquids, *Phys Rev Lett* **105**:140602, 2010.
15. Berendsen HJC, Grigera JR, Straatsma TP, The missing term in effective pair potentials, *J Phys Chem* **91**:6269–6271, 1987.
16. Remsing RC, Rodgers JM, Weeks JD, Deconstructing classical water models at interfaces and in bulk, *J Stat Phys* **145**:313–334, 2011.
17. Remsing RC, Weeks JD, Dissecting hydrophobic hydration and association, *J Phys Chem B* **117**(49):15479–15491, 2013.
18. Mobley DL, Klimovich PV, Perspective: Alchemical free energy calculations for drug discovery, *J Chem Phys* **137**(23):230901, 2012.
19. Deng Y, Roux B, Calculation of standard binding free energies: Aromatic molecules in the t4 lysozyme l99a mutant, *J Chem Theory Comput* **2**:1255–1273, 2006.
20. Wang L, Berne BJ, Friesner RA, On achieving high accuracy and reliability in the calculation of relative protein-ligand binding affinities, *Proc Natl Acad Sci* **109**(6):1937–1942, 2012.
21. Baron R, McCammon JA, Molecular recognition and ligand association, *Annu Rev Phys Chem* **64**(1):151–175, 2013.
22. Gallichio E, Lapelosa M, Levy RM, Binding energy distribution analysis method (bedam) for estimation of protein-ligand binding affinities, *J Chem Theory Comput* **6**:2961–2977, 2010.
23. Duff N, Dahal YR, Schmit JD, Peters B, Salting out the polar polymorph: Analysis by alchemical solvent transformation, *J Chem Phys* **140**(1):014501, 2014.
24. Dybeck EC, Abraham NS, Schieber NP, Shirts MR, Capturing entropic contributions to temperature-mediated polymorphic transformations through molecular modeling, *Cryst Growth Des* **17**(4):1775–1787, 2017.
25. Mullen RG, Maginn EJ, Reaction ensemble monte carlo simulation of xylene isomerization in bulk phases and under confinement, *J Chem Theory Comput* **13**(9):4054–4062, 2017.
26. Procacci P, Cardelli C, Fast switching alchemical transformations in molecular dynamics simulations, *J Chem Theory Comput* **10**:2813–2823, 2014.
27. Fuller JC, Jackson RM, Shirts MR, Configurational preferences of arylamide αhelix mimetics via alchemical free energy calculations of relative binding affinities, *J Phys Chem B* **116**:10856–10869, 2012.
28. Razavi AM, Delemotte L, Berlin JR, Carnevale V, Voelz VA, Molecular simulations and free-energy calculations suggest conformation-dependent anion binding to a cytoplasmic site as a mechanism for na^+/k^+-atpase ion selectivity, *J Biol Chem* **292**(30):12412–12423, 2017.
29. Huang DM, Chandler D, The hydrophobic effect and the influence of solute-solvent attractions, *J Phys Chem B* **106**:2047–2053, 2002.
30. Stillinger FH, Structure in aqueous solutions of nonpolar solutes from the standpoint of scaled-particle theory, *J Solution Chem* **2**:141–158, 1973.

31. Lum K, Chandler D, Weeks JD, Hydrophobicity at small and large length scales, *J Phys Chem B* **103**:4570–4577, 1999.
32. Chandler D, Interfaces and the driving force of hydrophobic assembly, *Nature* **437**:640–647, 2005.
33. Andersen HC, Weeks JD, Chandler D, Relationship between the hard-sphere fluid and fluids with realistic repulsive forces, *Phys Rev A* **4**:1597–1607 (1971).
34. Pohorille A, Jarzynski C, Chipot C, Good practices in free-energy calculations, *J Phys Chem B* **114**:10235–10253, 2010.
35. Kirkwood JG, Statistical mechanics of fluid mixtures, *J Chem Phys* **3**:300–313, 1935.
36. Godawat R, Jamadagni SN, Garde S, Characterizing hydrophobicity of interfaces by using cavity formation, solute binding, and water correlations, *Proc Natl Acad Sci USA* **106**:15119–15124, 2009.
37. Sarupria S, Garde S, Quantifying water density fluctuations and compressibility of hydration shells of hydrophobic solutes and proteins, *Phys Rev Lett* **103**:037803, 2009.
38. Patel AJ, Varilly P, Jamadagni SN, Hagan MF, Chandler D, Garde S, Sitting at the edge: How biomolecules use hydrophobicity to tune their interactions and function, *J Phys Chem B* **116**:2498–2503, 2012.
39. Willard AP, Chandler D, The molecular structure of the interface between water and a hydrophobic substrate is liquid-vapor like, *J Chem Phys* **141**(18):18C519, 2014.
40. Evans R, Wilding NB, Quantifying density fluctuations in water at a hydrophobic surface: Evidence for critical drying, *Phys Rev Lett* **115**(1):016103, 2015.
41. Evans R, Stewart MC, The local compressibility of liquids near nonadsorbing substrates: A useful measure of solvophobicity and hydrophobicity? *J Phys Condens Matter* **27**:194111, 2015.
42. Shirts MR, Chodera JD, Statistically optimal analysis of samples from multiple equilibrium states, *J Chem Phys* **129**:124105, 2008.
43. Souaille M, Roux B, Extension to the weighted histogram analysis method: Combining umbrella sampling with free energy calculations, *Comput Phys Commun* **135**:40–57, 2001.
44. Tan Z, Gallichio E, Lapelosa M, Levy RM, Theory of binless multi-state free energy estimation with applications to protein-ligand binding, *J Chem Phys* **136**:144102, 2012.
45. Hummer G, Garde S, García AE, Pohorille A, Pratt LR, An information theory model of hydrophobic interactions, *Proc Natl Acad Sci USA* **93**:8951–8955, 1996.
46. Beck TL, Paulaitis ME, Pratt LR, *The Potential Distirbution Theorem and Models of Molecular Solutions*, Cambridge University Press, Cambridge, 2006.
47. Widom B, Some topics in the theory of fluids, *J Chem Phys* **39**:2808–2812, 1963.
48. Patel AJ, Varilly P, Chandler D, Garde S, Quantifying density fluctuations in volumes of all shapes and sizes using indirect umbrella sampling, *J Stat Phys* **145**:265–275, 2011.
49. Patel AJ, Varilly P, Chandler D, Fluctuations of water near extended hydrophobic and hydrophilic surfaces, *J Phys Chem B* **114**:1632–1637, 2010.
50. Remsing RC, Xi E, Vembanur S, Sharma S, Debenedetti PG, Garde S, Patel AJ, Pathways to dewetting in hydrophobic confinement, *Proc Natl Acad Sci USA* **112** (27):8181–8186, 2015.
51. Crooks GE, Chandler D, Gaussian statistics of the hard-sphere fluid, *Phys Rev E* **56**:4217–4221, 1997.
52. Beck TL, A local entropic signature of specific ion hydration, *J Phys Chem B* **115**:9776–9781, 2011.

53. Arslanargin TL, Beck TL, Free energy partitioning analysis of the driving forces that determine ion density profiles near the water liquid–vapor interface, *J Chem Phys* **136**(10):104503, 2012.
54. Beck TL, Hydration free energies by energetic partitioning of the potential distribution theorem, *J Stat Phys* **145**(2):335–354, 2011.
55. Beck TL, The influence of water interfacial potentials on ion hydration in bulk water and near interfaces, *Chem Phys Lett* **561–562**:1–13, 2013.
56. Hyeon C, Thirumalai D, Capturing the essence of folding and functions of biomolecules using coarse-grained models, *Nat Commun* **2**:487, 2011.
57. Chung HS, Piana-Agostinetti S, Shaw DE, Eaton WA, Structural origin of slow diffusion in protein folding, *Science* **349**:1504–1510, 2015.
58. Kawamoto S, Klein ML, Shinoda W, Coarse-grained molecular dynamics study of membrane fusion: Curvature effects on free energy barriers along the stalk mechanism, *J Chem Phys* **143**(24):243112, 2015.
59. MacDermaid CM, DeVane CM, Klein ML, Fiorin G, Dehydration of multilamellar fatty acid membranes: Towards a computational model of the stratum corneum, *J Chem Phys* **141**(22):22D526, 2014.
60. MacDermaid CM, Kashyap HK, DeVane RH, Shinoda W, Klauda JB, Klein ML, Fiorin G, Molecular dynamics simulations of cholesterol-rich membranes using a coarse-grained force field for cyclic alkanes, *J Chem Phys* **143**(24):243144, 2015.
61. Fiorin G, Klein ML, Hénin J, Using collective variables to drive molecular dynamics simulations, *Mol Phys* **111**(22–23):3345–3362, 2013.
62. Valsson O, Tiwary P, Parrinello M, Enhancing important fluctuations: Rare events and metadynamics from a conceptual viewpoint, *Annu Rev Phys Chem* **67**(1):159–184, 2016.
63. Denesyuk NA, Weeks JD, A new approach for efficient simulation of coulomb interactions in ionic fluids, *J Chem Phys* **128**:124109, 2008.
64. Jarzynski C, Nonequilibrium equality for free energy differences, *Phys Rev Lett* **78**:2690–2693, 1997.
65. Crooks GE, Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences, *Phys Rev E* **60**:2721–2726, 1999.
66. Varilly PS, *Fluctuations in Water and their Relation to the Hydrophobic Effect*, PhD Thesis, University of California, Berkeley, 2011.