A smoothly decoupled particle interface: New methods for coupling explicit and implicit solvent

Jason A. Wagoner[1,a)] and Vijay S. Pande[2]
[1]*Department of Chemistry, Stanford University, Stanford, California 94305, USA*
[2]*Departments of Structural Biology and Computer Science, Stanford University, Stanford, California 94305, USA*

(Received 22 February 2011; accepted 8 May 2011; published online 1 June 2011)

A common theme of studies using molecular simulation is a necessary compromise between computational efficiency and resolution of the forcefield that is used. Significant efforts have been directed at combining multiple levels of granularity within a single simulation in order to maintain the efficiency of coarse-grained models, while using finer resolution in regions where such details are expected to play an important role. A specific example of this paradigm is the development of hybrid solvent models, which explicitly sample the solvent degrees of freedom within a specified domain while utilizing a continuum description elsewhere. Unfortunately, these models are complicated by the presence of structural artifacts at or near the explicit/implicit boundary. The presence of these artifacts significantly complicates the use of such models, both undermining the accuracy obtained and necessitating the parameterization of effective potentials to counteract the artificial interactions. In this work, we introduce a novel hybrid solvent model that employs a smoothly decoupled particle interface (SDPI), a switching region that gradually transitions from fully interacting particles to a continuum solvent. The resulting SDPI model allows for the use of an implicit solvent model based on a simple theory that needs to only reproduce the behavior of bulk solvent rather than the more complex features of local interactions. In this study, the SDPI model is tested on spherical hybrid domains using a coarse-grained representation of water that includes only Lennard-Jones interactions. The results demonstrate that this model is capable of reproducing solvent configurations absent of boundary artifacts, as if they were taken from full explicit simulations. © *2011 American Institute of Physics*. [doi:10.1063/1.3595262]

I. INTRODUCTION

The desire to obtain converged sampling within a biomolecular simulation often demands a compromise between computational efficiency and resolution of the applied forcefield. Solutions to quantum mechanics (QM) calculations are only possible for relatively small collections of molecules, and simulations using polarizable forcefields[1–4] can be intractable for most systems of interest. Current technology has allowed the use of all-atom molecular mechanics (MM) forcefields with explicit solvent representations to be used on very large systems and long timescales.[5–7] However, forcefields with an even coarser representation of molecular components have arisen from the desire to model larger systems than those that are manageable with an all-atom representation.[8–23] Similarly, implicit solvent models aim to reduce computational demand by replacing the explicitly sampled solvent degrees of freedom with an approximate, continuum description of the bulk solvent potential of mean force (PMF). Implicit solvent models are commonly used at all levels of molecular granularity, from QM (Refs. 24–27) and polarizable methods[28–30] to MM forcefields.[31–34] Unsurprisingly, lower resolution forcefields and solvent models often incur penalties on accuracy and require the development of model parameters that may not be transferable under different thermodynamic conditions or chemical environments.[35,36]

The development of multiscale methodologies have allowed for a union between the efficiency of coarse-grained models and the accuracy of more detailed techniques within a single simulation. Examples of such methods include QM/MM models[27,37–39] and MM methods that contain varying levels of all-atom and combined-atom forcefields across the simulation domain.[40–43] Similarly, hybrid solvent models have been developed that solvate a region of interest with explicit molecules and use an implicit solvent model elsewhere. Such hybrid models capture the salient features of both techniques: the explicit solvent molecules near a solute interface are capable of sampling relevant local interactions, while the implicit solvent model is used to reproduce the effects of bulk solvent. This hybrid solvation both reduces computational expense and removes the artifacts that may arise from the use of periodic simulations.[44] A number of such hybrid models have been previously developed,[23,24,45–63] many of which take advantage of particularly efficient implicit solvent models for the simple geometries of the defined explicit regions.[45,46,54,55]

Hybrid solvent models are subject to artifacts due to the inability of simple continuum methods to accurately reproduce the bulk solvent PMF near the explicit/implicit boundary. Similar problems exist in other methods that couple varying levels of atomic detail, and the boundaries for any

a)Author to whom correspondence should be addressed. Electronic mail: jwagoner@stanford.edu.

such model must be handled carefully in order to ensure appropriate chemical behavior in transition regions.[38–40,42] The boundary artifacts present in hybrid solvent models require the use of large enough explicit regions so that the thermodynamic properties of the solute remain hopefully unpolluted. Even with large explicit regions, the presence of unrealistic solvent configurations is troubling, especially under consideration of the significance of long-ranged solvent mediation that occurs in confined systems.[64–66]

Significant efforts have been directed at alleviating this problem with effective potentials[45,51,53,57,59,61–63] that reproduce forces from the bulk region as faithfully as possible. While these methods have demonstrated considerable success, the full removal of boundary artifacts is not expected unless the solvent PMF is accurate within a very small tolerance. Such a task has thus far only been accomplished using methods unsuitable for molecular dynamics simulations, such as integral equation or other calculations that generate the solvent response using a self-consistent approach.[67–72] In this work, we introduce a model that attempts to remove artificial boundary interactions using a smoothly decoupled particle interface (SDPI), a switching region over which the explicit particles are slowly annihilated and the implicit representation is "grown" in. This method *slowly* removes the local structure of the solvent. As a result, the requirement of the implicit solvent model is simplified from a many-body function that estimates the more complex features of specific interactions to a single-body term that reproduces bulk behavior according to an elementary theory. We test this method on uncharged systems using the MARTINI coarse-grained forcefield,[22,23] for which water is modeled as a monatomic Lennard-Jones (LJ) fluid, in spherical hybrid models. The result is an explicit region void of boundary artifacts that effectively samples solvent configurations as if they were taken from an infinite bath.

II. TRADITIONAL HYBRID MODEL

We will refer to hybrid models that instantaneously switch from explicit to implicit solvent as "traditional" hybrid models. In Sec. III, we introduce the details of the SDPI formalism.

A. Supersystem partition function

Consider a solute immersed in bulk solvent. The partition function of this system is

$$Z = \frac{1}{\Lambda_U^3 \Lambda^{3N} N!} \int_\Omega d\left(\mathbf{X}_U\right) \int d(\mathbf{x}_1), \ldots, d\left(\mathbf{x}_N\right) \exp\left[-\beta U\right], \tag{1}$$

where U is the total potential energy of the system, β is the inverse thermal energy $1/(k_B T)$, Λ_U and Λ^N are the thermal de Broglie wavelengths of solute and N solvent molecules, and $(\mathbf{X}_U, \mathbf{x}_1, \ldots, \mathbf{x}_N)$ represent the degrees of freedom of the solute and N (identical) solvent molecules.

If we partition our domain Ω into arbitrary interior and exterior portions ($\Omega = \Omega_\mathrm{i} + \Omega_\mathrm{o}$), we can rewrite this integral in terms of the number of molecules confined to the inner portion. Following the notation of previous developments,[49] we define

$$\delta_{nn'(\mathbf{X}_N)} = \begin{cases} 1, & \text{if } n = n'(\mathbf{X}_N) \\ 0, & \text{otherwise} \end{cases}, \tag{2}$$

where $(\mathbf{X}_N) = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$ and

$$n'(\mathbf{X}_N) = \sum_{i=1}^{N} C\left(\mathbf{x}_i\right), \tag{3}$$

$$C\left(\mathbf{x}_i\right) = \begin{cases} 0, & \text{if } \mathbf{x}_i \in \Omega_\mathrm{o} \\ 1, & \text{if } \mathbf{x}_i \in \Omega_\mathrm{i} \end{cases}. \tag{4}$$

The function $n'(\mathbf{X}_N)$ thus returns the number of solvent molecules within the inner domain. We can now rewrite the partition function as

$$Z = \frac{1}{\Lambda_U^3 \Lambda^{3N}} \sum_n \frac{1}{N!} \times \int d\left(\mathbf{X}_U\right) \int d\left(\mathbf{x}_1\right) \cdots d\left(\mathbf{x}_N\right) \delta_{nn'(\mathbf{X}_N)} \exp\left[-\beta U\right]. \tag{5}$$

Noting that the use of Eq. (2) introduces restricted integral domains in Eq. (5),

$$Z = \sum_n \frac{1}{\Lambda_U^3 \Lambda^{3N}} \int d\left(\mathbf{X}_U\right) \int_{\Omega_\mathrm{i}} \frac{1}{n!} d\left(\mathbf{X}_n\right) \times \int_{\Omega_\mathrm{o}} \frac{1}{(N-n)!} d\left(\mathbf{Y}_{N-n}\right) \exp\left[-\beta U\right], \tag{6}$$

where the vectors $\mathbf{X}_n$ and $\mathbf{Y}_{N-n}$ correspond to the configurations of molecules located in the inner (Ω_i) and outer (Ω_o) regions, respectively, and the numbered subscripts indicate the number of solvent molecules restricted to these regions.

B. Subsystem configurational probability

We are in a position to write the joint probability of finding n solvent particles within the inner, or explicit, domain and the configuration of *interior* coordinates $(\mathbf{X}_U, \mathbf{X}_n)$ as

$$P\left(\mathbf{X}_U, \mathbf{X}_n, n\right) = \left[Z \Lambda_U^3 \Lambda^{3N} n! (N-n)!\right]^{-1} \times \int_{\Omega_\mathrm{o}} d\left(\mathbf{Y}_{N-n}\right) \exp[-\beta U\left(\mathbf{X}_U, \mathbf{X}_n, \mathbf{Y}_{N-n}\right)]. \tag{7}$$

It is useful to decompose the total potential into inner-inner ($U_\mathrm{ii}\left(\mathbf{X}_U, \mathbf{X}_n\right)$), inner-outer ($U_\mathrm{io}\left(\mathbf{X}_U, \mathbf{X}_n, \mathbf{Y}_{N-n}\right)$), and outer-outer ($U_\mathrm{oo}\left(\mathbf{Y}_{N-n}\right)$) components and then define the PMF W as[73]

$$\exp\left[-\beta W\left(\mathbf{X}_U, \mathbf{X}_n\right)\right] = \frac{\int_{\Omega_\mathrm{o}} d\left(\mathbf{Y}_{N-n}\right) \exp\left[-\beta U\left(\mathbf{X}_U, \mathbf{X}_n, \mathbf{Y}_{N-n}\right)\right]}{\int_\Omega d\left(\mathbf{Y}_{N-n}\right) \exp\left[-\beta U\left(\mathbf{Y}_{N-n}\right)\right]}$$

$$= \frac{\exp\left[-\beta U_{\mathrm{ii}}\left(\mathbf{X}_U, \mathbf{X}_n\right)\right] \int_{\Omega_{\mathrm{o}}} d\left(\mathbf{Y}_{N-n}\right) \exp\left[-\beta\left(U_{\mathrm{io}}\left(\mathbf{X}_U, \mathbf{X}_n, \mathbf{Y}_{N-n}\right) + U_{\mathrm{oo}}\left(\mathbf{Y}_{N-n}\right)\right)\right]}{\int_{\Omega} d\left(\mathbf{Y}_{N-n}\right) \exp\left[-\beta U\left(\mathbf{Y}_{N-n}\right)\right]}$$

$$= \exp\left[-\beta\left(U_{\mathrm{ii}}\left(\mathbf{X}_U, \mathbf{X}_n\right) + \Delta W\left(\mathbf{X}_U, \mathbf{X}_n\right)\right)\right], \tag{8}$$

where ΔW is the bulk solvent PMF. Inserting this into Eq. (7):

$$P\left(\mathbf{X}_U, \mathbf{X}_n, n\right) = \frac{\exp\left[-\beta\left(U_{\mathrm{ii}}\left(\mathbf{X}_U, \mathbf{X}_n\right) + \Delta W\left(\mathbf{X}_U, \mathbf{X}_n\right)\right)\right]}{Z \Lambda_U^3 \Lambda^{3N} n!\,(N-n)!} \times \int_{\Omega} d(\mathbf{Y}_{N-n}) \exp[-\beta U(\mathbf{Y}_{N-n})]. \tag{9}$$

Noting that, in the limit of an infinitely large domain, we can take

$$\lim_{N\to\infty, \frac{N}{V}\to\rho} \frac{1}{\Lambda^{3N-3n}\,(N-n)!} \int_{\Omega} d\left(\mathbf{Y}_{N-n}\right) \times \exp[-\beta U\left(\mathbf{Y}_{N-n}\right)] \propto \exp\left[-\beta(N-n)\mu\right], \tag{10}$$

where V is the volume of the domain, ρ is the bulk density, and μ is the chemical potential of one of the solvent molecules. We now have

$$P\left(\mathbf{X}_U, \mathbf{X}_n, n\right) \propto \frac{1}{\Lambda_U^3 \Lambda^{3n} n!} \exp[-\beta(U_{\mathrm{ii}}(\mathbf{X}_U, \mathbf{X}_n) + \Delta W(\mathbf{X}_U, \mathbf{X}_n) + (N-n)\mu)]. \tag{11}$$

C. Transition probabilities

Define the ratio of equilibrium probability densities for state a, with n_a solvent molecules, and state b, with n_b solvent molecules, as

$$\frac{P\left(\mathbf{X}_{U_b}, \mathbf{X}_{n_b}, n_b\right)}{P\left(\mathbf{X}_{U_a}, \mathbf{X}_{n_a}, n_a\right)} = \frac{n_a! \Lambda^{3n_a}}{n_b! \Lambda^{3n_b}} \exp[-\beta(W(\mathbf{X}_{U_b}, \mathbf{X}_{n_b}) - W(\mathbf{X}_{U_a}, \mathbf{X}_{n_a}) - (n_b - n_a)\mu)]. \tag{12}$$

We now wish to substitute in the *excess* chemical potential μ', noting that

$$\mu = \mu^0 + \mu' \tag{13}$$

and

$$\mu^0 = \frac{1}{\beta} \ln[\rho \Lambda^3] \tag{14}$$

is the chemical potential of a monatomic ideal gas that has the same density and particle mass of our solvent. Making these substitutions,

$$\frac{P\left(\mathbf{X}_{U_b}, \mathbf{X}_{n_b}, n_b\right)}{P\left(\mathbf{X}_{U_a}, \mathbf{X}_{n_a}, n_a\right)} = \frac{n_a!}{n_b!} \rho^{n_b - n_a} \exp[-\beta(W(\mathbf{X}_{U_b}, \mathbf{X}_{n_b}) - W(\mathbf{X}_{U_a}, \mathbf{X}_{n_a}) - (n_b - n_a)\mu')]. \tag{15}$$

Equation (15) represents the relative probabilities for two states with a different solute configuration and a different number of interior solvent molecules within our hybrid simulation.

1. Insertion/deletion attempts

We wish to perform grand canonical Monte Carlo (GCMC) insertion and deletion moves over some subdomain $\Omega_{\mathrm{id}} \subseteq \Omega_{\mathrm{i}}$. Since the entire domain Ω_{i} is not subject to GCMC control, Eq. (15) must be modified for the criteria governing insertion or deletion attempts. If we label the number of molecules in state a within Ω_{id} as n_a^{id}, the configurational probability distribution requires the additional counting factor $\frac{n_a!}{n_a^{\mathrm{id}}!\left(n_a - n_a^{\mathrm{id}}\right)!}$. For a single insertion or deletion attempt, we note that $n_b - n_b^{\mathrm{id}} = n_a - n_a^{\mathrm{id}}$ and Eq. (15) becomes

$$\frac{P\left(\mathbf{X}_{U_b}, \mathbf{X}_{n_b}, n_b\right)}{P\left(\mathbf{X}_{U_a}, \mathbf{X}_{n_a}, n_a\right)} = \frac{n_a^{\mathrm{id}}!}{n_b^{\mathrm{id}}!} \rho^{n_b - n_a} \exp[-\beta(W(\mathbf{X}_{U_b}, \mathbf{X}_{n_b}) - W(\mathbf{X}_{U_a}, \mathbf{X}_{n_a}) - (n_b - n_a)\mu')]. \tag{16}$$

Assuming that we do not attempt to insert or delete more than one molecule at a time and that the probabilities of attempting an insertion or deletion are both equal to p_{id}, the total selection probability for inserting or deleting a particular molecule is

$$S_{\mathrm{ins}} = \frac{p_{\mathrm{id}}}{\left(n^{\mathrm{id}} + 1\right) V_{\Omega_{\mathrm{id}}}}, \qquad S_{\mathrm{del}} = \frac{p_{\mathrm{id}}}{n^{\mathrm{id}}}, \tag{17}$$

where $V_{\Omega_{\mathrm{id}}}$ is the volume of Ω_{id}. The factor of $\left(n^{\mathrm{id}} + 1\right)$ in the denominator of the insertion selection is consistent with the presented formulation and corresponds to giving an inserted molecule a random label.

The detailed balance condition imposes that, for a transition between initial i and final f states, the forward and reverse acceptance probabilities (A) must satisfy

$$P(\mathbf{X}_i) S\left(\mathbf{X}_i \to \mathbf{X}_f\right) A\left(\mathbf{X}_i \to \mathbf{X}_f\right) = P(\mathbf{X}_f) S(\mathbf{X}_f \to \mathbf{X}_i) A(\mathbf{X}_f \to \mathbf{X}_i). \tag{18}$$

The Metropolis criterion satisfies this requirement:

$$A(\mathbf{X}_i \to \mathbf{X}_f) = \min\left(1, \frac{S(\mathbf{X}_f \to \mathbf{X}_i)}{S(\mathbf{X}_i \to \mathbf{X}_f)} \frac{P(\mathbf{X}_f)}{P(\mathbf{X}_i)}\right). \tag{19}$$

Combining equations (16), (17), and (18) into the Metropolis algorithm gives the following acceptance probabilities for

insertion and deletion attempts:

$$A\left[(\mathbf{X}_U, \mathbf{X}_n, n) \to (\mathbf{X}_U, \mathbf{X}_{n+1}, n+1)\right] = \min\left(1, \frac{\langle n\rangle \exp[-\beta(W(\mathbf{X}_U, \mathbf{X}_{n+1}) - W(\mathbf{X}_U, \mathbf{X}_n) - \mu')]}{n+1}\right),$$

$$A\left[(\mathbf{X}_U, \mathbf{X}_n, n) \to (\mathbf{X}_U, \mathbf{X}_{n-1}, n-1)\right] = \min\left(1, \frac{n \exp[-\beta(W(\mathbf{X}_U, \mathbf{X}_{n-1}) - W(\mathbf{X}_U, \mathbf{X}_n) + \mu')]}{\langle n\rangle}\right), \tag{20}$$

where $\langle n\rangle = \rho V_{\Omega_{\mathrm{id}}}$. We see that these acceptance probabilities reduce to the form familiar to GCMC movesets.[74–79]

D. Implicit solvent model

The computational benefit of a hybrid solvent model arises from using simple, inexpensive, implicit solvent models to reproduce the bulk solvent PMF defined by Eq. (8). For this work, which includes no charged interactions, we model the solvent PMF as a combination of repulsive and attractive components, $\Delta W(\mathbf{X}_U, \mathbf{X}_n) = \Delta w^{(\mathrm{rep})}(\mathbf{X}_U, \mathbf{X}_n) + \Delta w^{(\mathrm{att})}(\mathbf{X}_U, \mathbf{X}_n)$. This separation closely mirrors the Weeks-Chandler-Andersen (WCA) decomposition of the LJ interaction, given by

$$U^{\mathrm{LJ}}(r_{ij}) = 4\epsilon_{ij}\left[\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}}\right], \tag{21}$$

where r_{ij} is the distance between atoms i and j, ϵ_{ij} gives the minimum value of this energy, and $2^{1/6}\sigma_{ij}$ is the distance at which that minimum occurs. The WCA decomposition[80] separates this potential, $U^{\mathrm{LJ}}(r_{ij}) = u^{(\mathrm{rep})}(r_{ij}) + u^{(\mathrm{att})}(r_{ij})$, where the components are defined as follows:

$$u^{(\mathrm{rep})}(r_{ij})(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} U^{(\mathrm{LJ})}(r_{ij}) + \epsilon_{ij}, & r_{ij} < 2^{1/6}\sigma_{ij} \\ 0, & r_{ij} \geq 2^{1/6}\sigma_{ij} \end{cases}, \tag{22}$$

$$u^{(\mathrm{att})}(r_{ij}) = \begin{cases} -\epsilon_{ij}, & r_{ij} < 2^{1/6}\sigma_{ij} \\ U^{(\mathrm{LJ})}(r_{ij}), & r_{ij} \geq 2^{1/6}\sigma_{ij} \end{cases}. \tag{23}$$

The attractive PMF contribution is modeled as a sum of atomic components, each integrating the attractive component of the WCA potential over the system domain:

$$\Delta w^{(\mathrm{att})}(\mathbf{X}_U, \mathbf{X}_n) = \rho \sum_i \int_\Omega u^{(\mathrm{att})}(\|\mathbf{x}_i - \mathbf{y}\|)\, h_i(\mathbf{x}_i, \mathbf{y})\, d\mathbf{y}, \tag{24}$$

where $\mathbf{y} \in \Omega$, ρ is the average solvent particle density, $h_i(\mathbf{x}_i, \mathbf{y})$ is the reference distribution used to describe the mean normalized solvent density surrounding atom i at position $\mathbf{x}_i$, and the summation is performed over all atoms.

An appropriate approximation for the distribution function would be a set of overlapping Heaviside functions H for the explicit region (which is centered at the origin) and each particle of the system:

$$h_i(\mathbf{x}_i, \mathbf{y}) = h(\mathbf{X}_u, \mathbf{X}_n, \mathbf{y}) = H(\|\mathbf{y}\| - R_o - \sigma_e) \times \prod_i H(\|\mathbf{y} - \mathbf{x}_i\| - \sigma_i), \tag{25}$$

$$H(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}, \tag{26}$$

where σ_i is the radius applied to particle i and is calculated as $\sigma_i = \frac{1}{2}2^{1/6}\sigma_{ii} + \sigma_s$, with σ_s representing a parameterized probe radius. R_o corresponds to the radius of the spherical explicit region. The parameter σ_e is added to augment this surface definition and adds significant accuracy to this simple approximation of the distribution function. This distribution function excludes bulk solvent density from the inner domain and regions of overlap between explicit particles and the outer domain. Unfortunately, the use of Eq. (25) necessitates an expensive numerical calculation for the evaluation of the presented implicit solvent model. For this reason, such a distribution function is used for analysis of the implicit solvent model but is not implemented for hybrid simulations.

Instead, we will approximate the distribution function for atom i with a simple two body overlap between this atom and the explicit domain:

$$h_i(\mathbf{x}_i, \mathbf{y}) = H(\|\mathbf{y}\| - R_o - \sigma_e)\, H(\|\mathbf{y} - \mathbf{x}_i\| - \sigma_i). \tag{27}$$

This representation of the distribution function allows for a decomposition of the overall PMF into atomic components that can be calculated analytically.

The repulsive component of the implicit solvent model is proportional to the volume of the explicitly defined region, $\Delta w^{(\mathrm{rep})}(\mathbf{X}_U, \mathbf{X}_n) = pV(\mathbf{X}_U, \mathbf{X}_n)$, where p is a parameterizable constant used to represent the bulk solvent pressure and V is the system volume. Though the choice of using volume over surface-area for this PMF component may seem curious, such volume terms were present in early developments of scaled particle theory.[81] The volume dependence of cavity formation has indeed been demonstrated for sub-nanometer length scales,[82–87] expected and shown[88] to be the relevant length scale for *atomic* solvation forces. A surface area term is still expected to contribute to the *absolute* cost of cavity formation for the explicit region, but is omitted from our calculations since there are no large-scale geometry changes within the simulations. The volume of our system is calculated as

$$V(\mathbf{X}_U, \mathbf{X}_n) = \int_\Omega \left(1 - \sum_i h_i(\mathbf{x}_i, \mathbf{y})\right) d\mathbf{y}. \tag{28}$$

This is not equivalent to the volume of the inner domain Ω_{i} and approximates the effect of explicit particles that overlap with the outer domain.

It is important to note that when used in these hybrid solvent models, the above implicit solvent PMF requires no solute parameterization as long as the solute atoms do not overlap with the boundary, and thus requires a total of three parameters only. This significantly simplifies the parameterization procedure when compared to implicit solvent models that often require new radii or other parameters for each atom present within the simulation.[31,32,34]

E. Mean force analysis

The implicit solvent model will be compared to explicit solvent results using mean-force analysis on a system of bulk solvent. Following Eq. (8), we can define the mean force for atom i given the static configuration of *explicit* solvent molecules ($\mathbf{X}_n$) as[73,88–90]

$$\begin{aligned} F_i(\mathbf{X}_n) &= -\frac{\delta \Delta W(\mathbf{X}_n)}{\delta \mathbf{x}_i} \\ &= \frac{\int_{\Omega_o} d(\mathbf{Y}_{N-n}) \left(-\frac{\delta U_{io}(\mathbf{X}_n, \mathbf{Y}_{N-n})}{\delta \mathbf{x}_i}\right) \exp\left[-\beta\left(U_{io}(\mathbf{X}_n, \mathbf{Y}_{N-n}) + U_{oo}(\mathbf{Y}_{N-n})\right)\right]}{\int_{\Omega_o} d(\mathbf{Y}_{N-n}) \exp\left[-\beta\left(U_{io}(\mathbf{X}_n, \mathbf{Y}_{N-n}) + U_{oo}(\mathbf{Y}_{N-n})\right)\right]} \\ &= -\left\langle \frac{\delta U_{io}(\mathbf{X}_n, \mathbf{Y}_{N-n})}{\delta \mathbf{x}_i} \right\rangle_{\mathbf{Y}_{N-n}}, \end{aligned} \tag{29}$$

where $\langle \cdots \rangle_{\mathbf{Y}_{N-n}}$ indicates an ensemble average over $\mathbf{Y}_{N-n}$, the configurations of solvent in the outer region. This mean force can be calculated from explicit solvent simulations by averaging the force on atom i from solvent molecules in the exterior domain. This quantity is then compared to the forces obtained with the implicit solvent model using the *full* distribution function (Eq. (25)).

When using the approximate, two-body distribution function given in Eq. (27), we are assuming that the mean forces on atom i are independent of all other explicit particles in the inner domain. Therefore, it is useful to define the analogous force from the full system as

$$F_i(\mathbf{x}_i) = -\left\langle \frac{\delta U_{io}(\mathbf{X}_n, \mathbf{Y}_{N-n})}{\delta \mathbf{x}_i} \right\rangle_{\mathbf{x}_{j\neq i}, \mathbf{Y}_{N-n}} = \langle F_i(\mathbf{X}_n) \rangle_{\mathbf{x}_{j\neq i}}, \tag{30}$$

where, in addition to all outer solvent molecules, we are now also averaging over all molecules $\mathbf{x_{j\neq i}}$ in the *inner* region. Note that this corresponds to simplifying the total PMF to a single-body term, a common approximation for simple implicit solvent models.[51,53,62,63] For this study on spherical systems, these forces must be spherically symmetric. Equation (30) then calculates a radial mean force that is used to produce the optimal parameter set for the implicit solvent model.

III. THE SMOOTHLY DECOUPLED PARTICLE INTERFACE

A. Addition of an intermediate domain

Equation (11) presents a target configurational probability density from a hybrid simulation and *exactly* reproduces any thermodynamic property that is a function of ($\mathbf{X}_U, \mathbf{X}_n$), as long as the function $\Delta W(\mathbf{X}_U, \mathbf{X}_n)$ is accurately calculated. In practice, however, developing accurate approximations to $\Delta W(\mathbf{X}_U, \mathbf{X}_n)$ in the form of implicit solvent models is a formidable challenge when attempting to reduce computational expense. For the SDPI model, we will derive an alternate formalism that provides the same theoretical result as Eq. (11) but in practice provides a much more straightforward and accurate approximation to the solvent PMF.

Consider partitioning the total system domain into three regions ($\Omega = \Omega_i + \Omega_s + \Omega_o$). The interior and exterior regions (Ω_i and Ω_o, respectively) are now separated by an intermediate, or switching, region Ω_s and are not in contact. The vectors $\mathbf{X}_n$, $\mathbf{Y}^s_m$, $\mathbf{Y}^o_{N-n-m}$ correspond to the configurations of the subscripted molecules within the interior, intermediate, and exterior regions, respectively. Following the reasoning outlined above, Eq. (6) can be extended to include this additional subregion:

$$\begin{aligned} Z = \frac{1}{\Lambda_U^3 \Lambda^{3N}} \sum_n \frac{1}{n!} \int d(\mathbf{X}_U) \int_{\Omega_i} d(\mathbf{X}_n) \sum_m \frac{1}{m!\,(N-n-m)!} \\ \times \int_{\Omega_s} d\left(\mathbf{Y}^s_m\right) \int_{\Omega_o} d\left(\mathbf{Y}^o_{N-n}\right) \exp\left[-\beta U\right]. \end{aligned} \tag{31}$$

Using the development of Sec. II B, we can integrate over the coordinates of the outermost region to obtain a joint probability of finding n solvent molecules within the interior region, m molecules within the intermediate region, and the configuration $\left(\mathbf{X}_U, \mathbf{X}_n, \mathbf{Y}^s_m\right)$:

$$\begin{aligned} P\left(\mathbf{X}_U, \mathbf{X}_n, n, \mathbf{Y}^s_m, m\right) \propto \frac{1}{n!m!\Lambda_U^3 \Lambda^{3(n+m)}} \exp\Big[-\beta\Big(U_{ii}\left(\mathbf{X}_U, \mathbf{X}_n\right) \\ + U_{is}\left(\mathbf{X}_U, \mathbf{X}_n, \mathbf{Y}^s_m\right) + U_{ss}\left(\mathbf{Y}^s_m\right)\Big)\Big] \\ \times \exp\Big[-\beta\Big(\Delta W\left(\mathbf{X}_U, \mathbf{X}_n, \mathbf{Y}^s_m\right) \\ + (N-n-m)\,\mu\Big)\Big], \end{aligned} \tag{32}$$

where U_{is} is the interaction potential between particles in Ω_i and those in Ω_s, etc.

We will now introduce an unphysical potential U^λ for all interactions with the m molecules in the intermediate region. The nature of this potential will be described in more detail below. The configurational probability of this new system is

$$\begin{aligned} P\left(\mathbf{X}_U, \mathbf{X}_n, n, \mathbf{Y}^s_m, m\right) \propto \frac{1}{n!m!\Lambda_U^3 \Lambda^{3(n+m)}} \exp\Big[-\beta\Big(U_{ii}(\mathbf{X}_U, \mathbf{X}_n) \\ + U^\lambda_{is}\left(\mathbf{X}_U, \mathbf{X}_n, \mathbf{Y}^s_m\right) + U^\lambda_{ss}\left(\mathbf{Y}^s_m\right)\Big)\Big] \\ \times \exp\Big[-\beta\Big(\Delta W^\lambda\left(\mathbf{X}_U, \mathbf{X}_n, \mathbf{Y}^s_m\right) \\ + (N-n-m)\,\mu\Big)\Big]. \end{aligned} \tag{33}$$

For nontrivial cases, the introduction of this non-physical potential will alter the configurations of molecules within the intermediate region so that they no longer match the configurations expected from the original supersystem. However, we note that if we can ensure that

$$\frac{1}{\Lambda^{3m}}\int_{\Omega_s} d\mathbf{Y}^s_m \exp\Big[-\beta\Big(U^\lambda_{is}\left(\mathbf{X}_U,\mathbf{X}_n,\mathbf{Y}^s_m\right)+U^\lambda_{ss}\left(\mathbf{Y}^s_m\right) + \Delta W^\lambda\left(\mathbf{X}_U,\mathbf{X}_n,\mathbf{Y}^s_m\right)\Big)\Big] = \exp\Big[-\beta\Delta W\left(\mathbf{X}_U,\mathbf{X}_n\right)\Big], \tag{34}$$

where the right side represents the solvent PMF developed in Sec. II B, then the configurations sampled from the innermost region will have the same equilibrium probability as those sampled from the original supersystem. This equation demonstrates that the overall goal of the SDPI model is to reproduce the bulk solvent PMF using an implicit solvent model in conjunction with sampling over the degrees of freedom located within the switching region.

It is straightforward to extend the results of Sec. II to include this intermediate region and potential. Most notably, the acceptance criteria for insertion and deletion attempts (Eq. (20)) become

$$\begin{aligned} &A\left[\left(\mathbf{X}_U,\mathbf{X}_{n_a},n_a,\mathbf{Y}^s_{m_a},m_a\right)\rightarrow\left(\mathbf{X}_U,\mathbf{X}_{n_b},n_b,\mathbf{Y}^s_{m_b},m_b\right)\right] \\ &\quad=\min\left(1,\frac{\langle n\rangle\exp\left[-\beta\left(W^\lambda\left(\mathbf{X}_U,\mathbf{X}_{n_b},\mathbf{Y}^s_{m_b}\right)-W^\lambda\left(\mathbf{X}_U,\mathbf{X}_{n_a},\mathbf{Y}^s_{m_a}\right)-\mu'\right)\right]}{n^{id}_b+m^{id}_b}\right), \\ &A\left[\left(\mathbf{X}_U,\mathbf{X}_{n_b},n_b,\mathbf{Y}^s_{m_b},m_b\right)\rightarrow\left(\mathbf{X}_U,\mathbf{X}_{n_a},n_a,\mathbf{Y}^s_{m_a},m_a\right)\right] \\ &\quad=\min\left(1,\frac{\left(n^{id}_b+m^{id}_b\right)\exp\left[-\beta\left(W^\lambda\left(\mathbf{X}_U,\mathbf{X}_{n_a},\mathbf{Y}^s_{m_a}\right)-W^\lambda\left(\mathbf{X}_U,\mathbf{X}_{n_b},\mathbf{Y}^s_{m_b}\right)+\mu'\right)\right]}{\langle n\rangle}\right), \end{aligned} \tag{35}$$

where the id superscript once again notates molecules located within $\Omega_{id}\subseteq(\Omega_i\cup\Omega_s)$, the subdomain subject to GCMC control. Here, $\langle n\rangle=\rho V_{\Omega_{id}}$ and it is assumed that $n^{id}_b+m^{id}_b=n^{id}_a+m^{id}_a+1$. In the limit of instantaneous switching and an infinitely thin intermediate region Ω_s, these results reduce to those of the traditional hybrid model presented in Sec. II.

B. Introduction of a decoupling potential

The purpose of an intermediate region that interacts with an unphysical potential is to smoothly decouple the particles between the explicit and implicit regions. For the spherical systems of this study, we define the decoupling parameter as a simple polynomial:

$$\lambda(r)=\begin{cases}1, & r\le R_i\\ 1+\dfrac{2(r-R_i)^3}{(R_o-R_i)^3}-\dfrac{3(r-R_i)^2}{(R_o-R_i)^2}, & R_i<r<R_o\\ 0, & r\ge R_o\end{cases}, \tag{36}$$

where the intermediate domain Ω_s corresponds to the spherical shell bound by the radii R_i and R_o. This parameter smoothly transitions particles between the fully interacting ($\lambda=1$) and the fully decoupled ($\lambda=0$) states, as demonstrated in Fig. 1. Figure 2 illustrates the spatial arrangement of molecules within the separate regions of both the traditional and SDPI hybrid models.

A SDPI simulation will contain explicit particles from the innermost and intermediate domains. The total energy of this system is

$$W^\lambda\left(\mathbf{X}_U,\mathbf{X}_n,\mathbf{Y}^s_{n_s}\right)=\sum_i W_i\left(\mathbf{x}_i\right)=\sum_i\Bigg[\lambda\left(\|\mathbf{x}_i\|\right)\Delta w_i\left(\mathbf{x}_i\right)+\left(1-\lambda\left(\|\mathbf{x}_i\|\right)\right)K_i+G\left(\mathbf{x}_i\right)+\frac{1}{2}\sum_{j\neq i}U^\lambda\left(\mathbf{x}_i,\mathbf{x}_j\right)\Bigg], \tag{37}$$

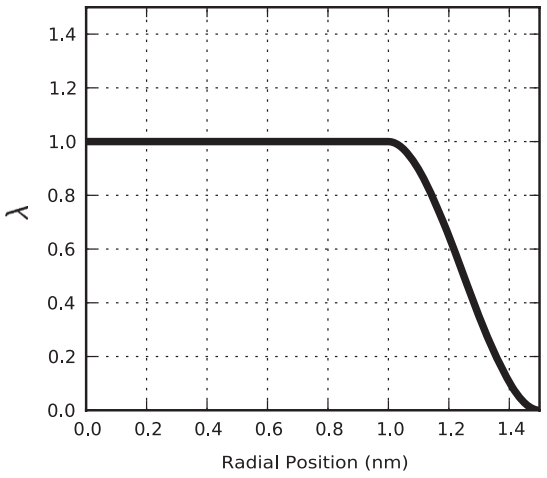

FIG. 1. Behavior of the coupling parameter for a SDPI model in which R_i = 1.0 nm and R_o = 1.5 nm. Equation (38) utilizes this parameter to determine the extent to which a particle at a given radial position interacts with the rest of the system.

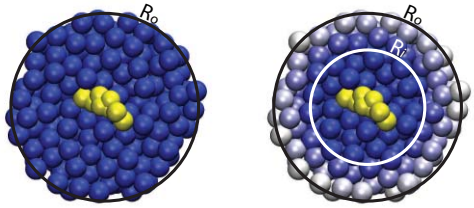

FIG. 2. Composition of spherical hybrid models ($R_o = 2.5$ nm) centered on a cholesterol molecule (yellow) and containing MARTINI CG water particles. (a) Traditional hybrid model. All solvent molecules within R_o are explicit. Contributions to the PMF from the region outside of R_o are approximated using implicit solvent methods. (b) SDPI model. The innermost region (Ω_i) contains explicit particles and is bound by R_i (1.5 nm for this example). The exterior region (Ω_o, outside of R_o) is accounted for using implicit solvent methods. The intermediate, or switching, region (Ω_s) is bound by R_i and R_o. The particles in this region are colored by the strength of their interactions with the rest of the system. The particles at R_i are fully interacting (blue) and become decoupled (white) as they approach R_o.

where, assuming that the center of the spherical domain is placed at the origin, $\|\mathbf{x}_i\|$ represents the radial position of atom i. The terms $G(\mathbf{x}_i)$ and K_i, described in more detail below, are used to counteract the gradient in chemical potential experienced by particles in the switching region. The implicit solvent component of the total potential, ΔW, is our approximation to the bulk solvent PMF and, for this study, is decomposable into atomic components Δw_i. Since the solvent used here interacts only via the LJ potential, we decouple this interaction in a manner directly analogous to that used in free energy studies:[91]

$$U^{\lambda}\left(\mathbf{x}_i, \mathbf{x}_j\right) = \lambda\left(\|\mathbf{x}_i\|\right)\lambda\left(\left\|\mathbf{x}_j\right\|\right) U^{LJ}\left(r_{\mathrm{sc}}\left(\mathbf{x}_i, \mathbf{x}_j\right)\right), \quad (38)$$

where U^{LJ} is given by Eq. (21) and

$$r_{\mathrm{sc}}\left(\mathbf{x}_i, \mathbf{x}_j\right) = \left(\alpha\sigma_{ij}^6\left(1-\lambda\left(\|\mathbf{x}_i\|\right)\lambda\left(\left\|\mathbf{x}_j\right\|\right)\right)^p + \left\|\mathbf{x}_i - \mathbf{x}_j\right\|^6\right)^{\frac{1}{6}}. \quad (39)$$

The terms α and p modulate the scaling of interactions and are set to 0.5 and 1, respectively.

Thus, a particle within this system interacts both with other explicit particles (via $U^{\lambda}(\mathbf{x}_i, \mathbf{x}_j)$) and with the bulk region (via $\Delta w_i(\mathbf{x}_i)$). As can be seen by the above equation, a particle close to the outer boundary R_o is essentially noninteracting. The result of such a decoupling will be a gradient in chemical potential across the switching region that will lead to large differences in the solvent density. In order to ensure Eq. (34), we will attempt to correct for this gradient so that the solvent will maintain the appropriate bulk density everywhere (note that this is not absolutely necessary and that there are other methods that could satisfy Eq. (34)). We introduce the constant factor K_i, corresponding to the implicit solvent PMF for atom i isolated from any other explicit regions or particles. This term is approximately equal to the excess chemical potential of our solvent. Scaling this term by $(1-\lambda(\|\mathbf{x}_i\|))$ mostly accounts for the drop in chemical potential across the boundary. However, it is still necessary to introduce the factor

$$G(\mathbf{x}_i) \approx -(1-\lambda(\|\mathbf{x}_i\|))K_i - \int_0^{\|\mathbf{x}_i\|} dr\left\langle\frac{dW_i(r)}{dr}\right\rangle, \quad (40)$$

equivalent to the difference between the expected and *actual* excess chemical potential associated with inserting a solvent particle at position $\mathbf{x}_i$, since a simple scaling by $(1-\lambda(\|\mathbf{x}_i\|))$ does not appropriately account for any path-dependent effects. The addition of these terms ensures that the solvent maintains bulk density across the domain. Of course, higher-order structural metrics cannot be expected to be recovered within the intermediate region.

We approximate $G(\mathbf{x}_i)$ using a set of overlapping Gaussians,[62,63]

$$G(\mathbf{x}_i) = \sum_{j=0}^{N_j} g_j(\|\mathbf{x}_i\| - r_j) = \sum_j a_j \exp\left[-\frac{(\|\mathbf{x}_i\| - r_j)^2}{2\sigma_g^2}\right], \quad (41)$$

where the Gaussian centers are equally spaced at $r_j = (j + 0.5)\sigma_g$ and $N_j = (\frac{R_o}{\sigma_g} + 4)$. The Gaussians centered outside of R_o are added to ensure appropriate behavior of $G(\mathbf{x}_i)$ and its derivatives at the boundary. The coupling of Eqs. (38) and (40) warrants a self-consistent procedure for the determination of the coefficients a_j, described in more detail in Sec. IV.

For the calculation of $\Delta w_i(\mathbf{x}_i)$, we will use the same implicit solvent model presented in Sec. II D (Eqs. (28) and (24)) with an analogous two-body distribution function:

$$h_i^{(\mathrm{rep})}(\mathbf{x}_i, \mathbf{y}) = (1-\lambda(\|\mathbf{y}\|))H(\|\mathbf{y} - \mathbf{x}_i\| - \sigma_i), \quad (42)$$

where $\lambda(\|\mathbf{y}\|)$ is given by Eq. (36).

Therefore, the SDPI hybrid model assumes that single-body potentials can effectively correct the chemical potential gradient and approximate the solvent PMF. This is analogous to the use of a single-body potential term used in the traditional hybrid model, and in Sec. V we attempt to demonstrate that such an assumption is much more valid when coupled with the intermediate switching domain present in the SDPI model.

IV. METHODS

A. Simulation methodology

Simulations were conducted using a monatomic water model from the MARTINI coarse-grained force field[22,23] and a modified version of GROMACS 3.3.1.[92] All simulations were performed using Langevin dynamics integration with a 10 fs time step (a conservative value for this forcefield[22]) coupled to a heat bath at 400 K with a friction coefficient of 91 ps^{-1}. Explicit solvent trajectories were maintained at constant pressure (1 atm) using the Berendsen coupling algorithm.[93] Explicit solvent and traditional hybrid models used a LJ potential that is switched at 0.9 nm and cut off at 1.2 nm in conjunction with a dispersive tail correction.[94] Because this work attempts to demonstrate accuracy and not computational efficiency, the SDPI hybrid models employ a full LJ interaction that is not cut off. This allowed for a more straightforward implementation of the implicit solvent methods for this model. More optimized methods are currently being implemented for future analyses.

All thermodynamic properties presented in this study were calculated at a frequency not higher than once every picosecond, or 100 MD steps. Simulation lengths for all analyses were typically between 10 and 100 ns (10^6 to 10^7 MD steps), depending on the convergence of results, which were determined using a bootstrapping analysis.

B. Hybrid model parameters

All results were compiled with both the traditional and SDPI hybrid models at three different sizes (spheres of radii $R_o = 2.5$, 2.0, and1.5 nm). The inner radii for these three systems in the SDPI hybrid simulations were $R_i = 1.5$, 1.5, and 1.0 nm, respectively. The switching region length of 0.5 nm for the two smaller hybrid models corresponds to roughly one solvation layer for this coarse-grained (CG) system.

Traditional hybrid models were simulated using 250 insertion/deletion GCMC attempts for every 100 steps of MD. Such moves were only attempted over a "buffer region" covering the outer 0.5 nm of the radial domain. SDPI hybrid models, which were subject to much higher acceptance ratios, included 100 insertion/deletion attempts at an interval of every 100 MD steps over a buffer region covering the outermost 0.3 nm of the radial domain.

C. Mean force analysis

Bulk explicit solvent mean forces were calculated on a system of pure solvent via Eq. (30) from a full explicit simulation from which radial forces on all particles within a 2.0 nm spherical domain were averaged at 0.01 nm intervals. These forces were then compared to the analogous implicit solvent forces (for a traditional hybrid model) over a range of values for the parameters defined in Sec. II D. Specifically, $\sigma_s \in [0.05, 0.39]$ nm and $\sigma_e \in [0.00, 0.06]$ nm, tested in intervals of 0.01 nm. The optimal value for pressure (p) for each parameter set was found via least-squares fitting. The overall best-fit parameters were chosen based on the minimum value of mean-squared error,

$$\chi^2 = \frac{1}{N_r}\sum_{i=0}^{N}\left(F_i(\mathbf{x}_i) - \frac{d\Delta w_i(\mathbf{x}_i)}{d\mathbf{x}_i}\right)^2, \tag{43}$$

where N_r is the number of radial data points (201).

In addition, *true* mean forces (via Eq. (29)) were calculated by choosing an equilibrium configuration from a bulk explicit solvent simulation, freezing all water molecules within a 1.5 nm spherical domain, and averaging forces on these molecules while sampling over all solvent exterior to this region. These were then compared to the implicit solvent forces calculated using the traditional hybrid model with the two-body approximation (Eq. (27)) as well as the many-body distribution function (Eq. (25)). For the calculations using the many-body distribution function, $\Delta w^{(\mathrm{att})}$ was calculated via numerical integration on a three-dimensional grid of 0.0012 nm spacing. The volume derivatives of the system were calculated using the Shrake-Rupley quadrature method[95] with 996 grid points on the surface of the sphere. Finally, the explicit mean forces were compared to mean forces calculated using the SDPI hybrid model with R_i =1.5 nm and $R_o = 2.5$ nm. SDPI mean forces were calculating by freezing all molecules within R_i and averaging forces on these molecules while sampling over all solvent within the switching region.

D. Parameterization of $G(r)$

The effective chemical potential term, $G(r)$, was found using an adaptive method implemented in a pure MC simulation for which the insertion/deletion attempts described above were augmented with translation moves for SDPI models of unrestrained solvent. The simulations used $\sigma_g = 0.0125$ nm and began with all Gaussian coefficients a_i set to zero. As the simulation progressed, if the difference between the bulk density (8.28 particles/nm^3 for the CG water studied at these thermodynamic conditions) and the calculated density within a radial shell of the hybrid simulation was, within statistical error, higher or lower than a prescribed tolerance, the coefficients were raised or lowered accordingly. This procedure is similar to that used in the development of other effective potentials for hybrid models.[62,63] As the radial bulk density profile approached the correct value, the tolerance in error was lowered along with the amount by which the coefficients were altered. All Gaussian coefficients a_i centered at radii $r_j < 0.5$ nm were held at 0. The optimal $G(r)$ curve was chosen once the *maximum* relative error in density along the radial coordinate was less than 1%. To ensure smooth forces, a double-exponential smoothing method[96] was then applied to $G(r)$. The resulting potentials demonstrated larger deviations (less than 2% maximum relative error in density) but were more stable for dynamics simulations.

E. Density profiles and free energy calculations

To determine the extent to which boundary artifacts are present in the hybrid models studied, density profiles were calculated for systems containing no solute and in which all solvent molecules were unrestrained. In the absence of any boundary artifacts, one would expect to find the average density to be equivalent to that of the bulk value ρ everywhere within the spherical domain.

Solvation free energies for water and cholesterol molecules[23] were calculated using the hybrid models and compared to the analogous values obtained from pure explicit solvent simulations. All free energies were calculated using the multistate Bennet acceptance ratio estimator (MBAR).[97] The soft-core LJ potential[91] was used to decouple the solute from the simulation (analogous to that of Eq. (39)). This potential was decoupled according to the parameter λ_{MBAR}, which for a particular simulation was held fixed at a prescribed value between 1.0 (fully interacting) and 0.0 (fully decoupled) at equally spaced intervals of 0.05. Thus, separate simulations for 21 values of λ_{MBAR} were performed for each free energy calculation. Intramolecular interactions of cholesterol were not decoupled. For the SDPI simulations, the $\lambda(r)$ parameter for the molecule being decoupled (given by Eq. (36)) was simply scaled by λ_{MBAR}.

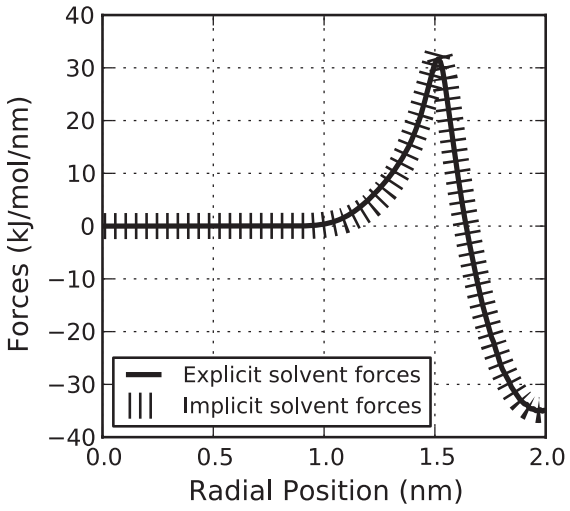

FIG. 3. Comparison of implicit solvent forces to the analogous radial forces obtained from explicit solvent simulations. This optimal fit is associated with an average mean-squared error value of $\chi^2 = 0.30\,(\mathrm{kJ/mol/nm})^2$.

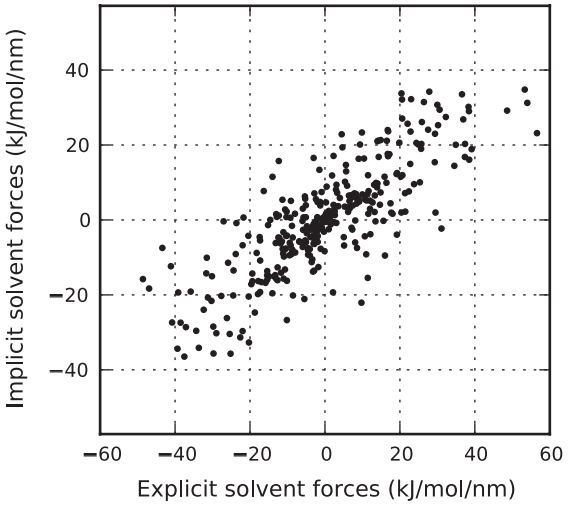

FIG. 4. For a static configuration of solvent particles within the interior domain, comparison of implicit solvent forces obtained using the two-body distribution function to the true mean forces obtained from explicit solvent simulations. This optimal fit is associated with an average mean-squared error of $\chi^2 = 105.25\,(\mathrm{kJ/mol/nm})^2$.

For water free energies calculated using the hybrid models, the water molecule being annihilated was constrained to various positions within the hybrid model to test the uniformity of results across the domain. The cholesterol molecule was restrained to the center of the hybrid regions using a simple harmonic potential applied to the center of mass with a spring constant of 75 $\mathrm{kJ/mol/nm^2}$.

V. RESULTS AND DISCUSSION

A. Mean force analysis and parameterization

Figure 3 compares the explicit solvent radial mean forces (given by Eq. (30)) and the best-fit implicit solvent forces. These forces were calculated on solvent molecules within Ω_i by averaging the force contribution from all molecules outside of this region. The parameters for the optimal implicit solvent forces are $\sigma_e = 0.03$nm, $\sigma_s = 0.22$nm, and $p = 86.76\ \mathrm{kJ/mol/nm^3}$, giving an average mean-squared error value of $\chi^2 = 0.30\,(\mathrm{kJ/mol/nm})^2$. These parameters are used in all hybrid calculations.

Despite the excellent fit, it is important to note that these are not *true* mean forces, but the projection of such forces onto a single-body radial term. True explicit mean force components (given by Eq. (29)) and the corresponding optimal implicit solvent forces are shown in Fig. 4. These optimal implicit solvent forces give an average mean-squared error of $\chi^2 = 105.25\,(\mathrm{kJ/mol/nm})^2$. Figure 5 compares these same explicit forces to implicit forces obtained using the full geometrical distribution function (Eq. (25)) rather than the two-body approximation, resulting in an average mean-squared error of $\chi^2 = 120.70\,(\mathrm{kJ/mol/nm})^2$. These results demonstrate that this implicit solvent model is not capable of reproducing the true mean forces and that even the use of the full geometrical distribution function does not provide any benefit. Nevertheless, the extremely good fit shown in Fig. 3 is noteworthy and demonstrates that any single-body potential capable of reproducing explicit solvent configurations must match the behavior predicted by the simple theory used in this study and outlined in Sec. II D. Additionally, any model that would hope to improve upon these results would require a more detailed and computationally expensive representation of the solvent density than that which is used here, most likely through heuristic many-body terms[61] or integral equation and other more complex theories.[70–72]

Figure 6 illustrates the ability of the SDPI model to reproduce mean forces within the explicit region. These forces give an average mean-squared error of $\chi^2 = 20.24\,(\mathrm{kJ/mol/nm})^2$, a significantly more accurate result than that obtained with the implicit solvent model used in the traditional hybrid methods.

Figure 7 displays the converged curves of $G(r)$ for the three hybrid models tested. These curves were parameterized

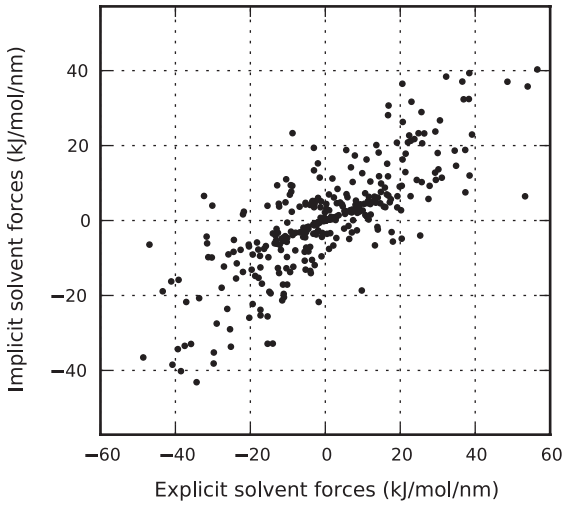

FIG. 5. For a static configuration of solvent particles within the interior domain, comparison of implicit solvent forces obtained using the full many-body distribution function to the true mean forces obtained from explicit solvent simulations. This optimal fit is associated with an average mean-squared error of $\chi^2 = 120.70\,(\mathrm{kJ/mol/nm})^2$.

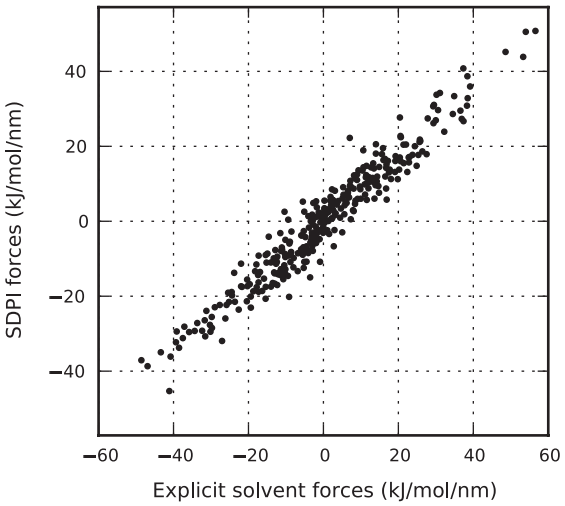

FIG. 6. For a static configuration of solvent particles within the interior domain, this plot illustrates SDPI mean solvent forces calculated from the implicit solvent model as well as by averaging over configurations of solvent in the switching region. These forces are compared to those obtained from explicit solvent simulations and are associated with an average mean-squared error of $\chi^2 = 20.24\,(\mathrm{kJ/mol/nm})^2$.

separately, but the level of quantitative similarity suggests that it may be possible to develop a simple transformation of a particular $G(r)$ potential for use in hybrid models of other dimensions. We are currently attempting to develop such a transformation, which would simplify the parameterization of these models significantly.

B. Density profiles

Figure 8 displays the density profiles for solutions of bulk CG water and lists the maximum error in density found for each model. These systems contain only unrestrained solvent, so a value of the bulk density ρ is desired at all positions within the spherical domain. Using the optimal parameter

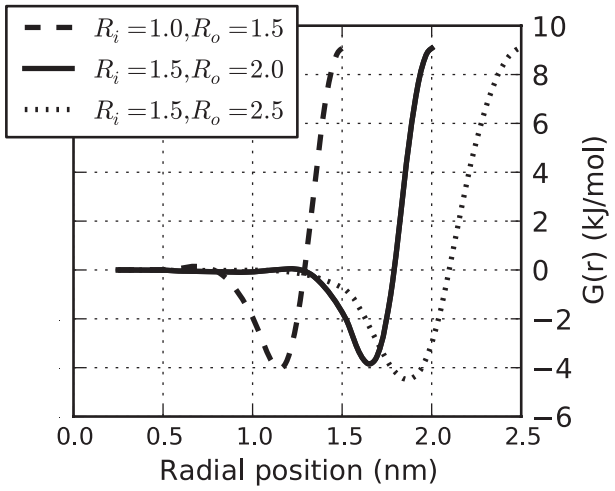

FIG. 7. The fit potential $G(r)$, given by Eq. (40), for the three hybrid models tested. This term counteracts the variations in chemical potential experienced by decoupled particles and maintains the correct bulk density throughout the switching region.

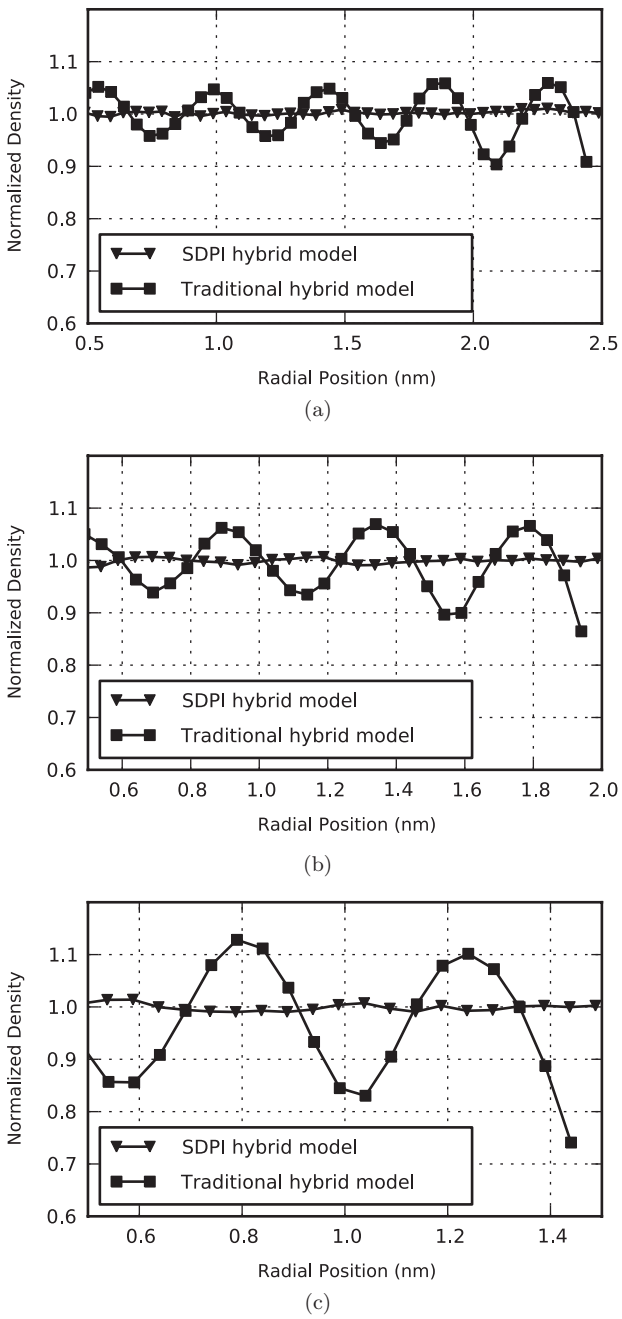

FIG. 8. Density profiles for the three hybrid model sizes tested, for which the maximum relative error excludes radial positions below 0.5 nm. (a) 2.5 nm hybrid models. The maximum relative error is 9.6% for the traditional model and 1.1% for the SDPI model. (b) 2.0 nm hybrid models. The maximum relative error is 13.5% for the traditional model and 1.2% for the SDPI model. (c) 1.5 nm hybrid models. The maximum relative error is 25.9% for the traditional model and 1.4% for the SDPI model.

set, very good results are obtained with the traditional hybrid model. Nevertheless, local deviations from bulk density are observed, particularly for the models with smaller explicit regions. SDPI hybrid models demonstrate nearly perfect results at all positions within the explicit domain with a *maximum* relative error of 1.4% among all models tested, compared to maximum errors in the 10%-26% range for traditional models. Density profiles taken from the traditional hybrid model also exhibit solvent layering, which may be indicative of a

TABLE I. Comparison of free energy values obtained from traditional and SDPI hybrid models. Free energies of CG water were obtained at variable radial positions within the spherical explicit domain. The values in parentheses represent the absolute error with respect to the value calculated from a full explicit solvent simulation (−21.72 ± .04 kJ/mol). Values of R_i are only relevant for SDPI models.

Domain radii R_o / R_i (nm)	Radial Position (nm)	Traditional	SDPI
2.5 / 1.5	0.0	−23.03 (−1.31) ± 0.06	−21.66 (0.06) ± 0.06
	0.5	−22.07 (−0.35) ± 0.06	−21.56 (0.16) ± 0.06
	1.0	−22.16 (−0.44) ± 0.06	−21.73 (−0.01) ± 0.06
	1.5	−21.96 (−0.24) ± 0.06	−21.84 (−0.12) ± 0.06
	2.0	−21.66 (0.06) ± 0.06	−21.65 (0.07) ± 0.02
	2.4	−21.66 (0.06) ± 0.06	−21.72 (0.00) ± 0.00
2.0 / 1.5	0.0	−22.00 (−0.28) ± 0.07	−21.47 (0.25) ± 0.06
	0.5	−22.25 (−0.53) ± 0.07	−21.70 (0.02) ± 0.06
	1.0	−21.88 (−0.16) ± 0.07	−21.61 (0.11) ± 0.06
	1.5	−21.50 (0.22) ± 0.10	−21.81 (−0.09) ± 0.05
	1.9	−21.58 (0.14) ± 0.06	−21.72 (0.00) ± 0.00
1.5 / 1.0	0.0	−18.48 (3.24) ± 0.06	−21.88 (−0.16) ± 0.06
	0.5	−21.65 (0.07) ± 0.06	−21.76 (−0.04) ± 0.06
	1.0	−21.43 (0.29) ± 0.06	−21.71 (0.01) ± 0.05
	1.4	−21.46 (0.26) ± 0.05	−21.72 (0.00) ± 0.00

higher-order effect resulting from errors in the solvent PMF near the explicit/implicit boundary.

C. Free energies

Table I lists the solvation free energies for CG water molecules at various radial positions within the hybrid models. The one-body implicit solvent potential used in this study performs very well in the traditional hybrid model, and the free energy values obtained from these simulations are quite accurate at a variety of positions within the explicit domain. Still, there are regions within these domains that produce free energies of relatively large error. It is interesting to note that the free energies are most erroneous near the center of the explicit region, a result that again suggests that the errors associated with such traditional hybrid models are a higher-order effect in which large density fluctuations throughout the explicit region result from small errors near the boundary. The free energies obtained using SDPI hybrid methods exhibit nearly perfect quantitative results everywhere, with the largest error being 0.25 ± 0.06 kJ/mol (at the center of the region for which $R_i = 1.5$, $R_o = 2.0$ nm).

Table II lists the solvation free energies for cholesterol obtained from hybrid simulations. The traditional hybrid model performs well but is again subject to significant error. The SDPI hybrid model is consistently more accurate, even when comparing results from the smallest SDPI model (absolute error of 0.22 ± 0.12 kJ/mol for $R_o = 1.5$ nm) and the largest traditional model (absolute error of −1.81 ± 0.10 kJ/mol for $R_o = 2.5$ nm). The absolute error obtained in the 1.5 nm SDPI model is particularly relevant to note: given that the switching region for these simulations was bound by the radii $R_i = 1.0$ nm and $R_o = 1.5$ nm, and that the end-to-end distance of cholesterol is occasionally more than 2.0 nm, these results demonstrate that a high level of accuracy is still obtained from simulations in which the explicit domain is tightly fit around the solute.

For the traditional model, one would expect to obtain the best results for larger systems or near the center of the explicit region. It is interesting to note that this is not the case. We suspect the reason for this to be a higher-order mediation caused by small artifacts near the explicit-implicit boundary, not unlike the role that solvent has been shown to play under confinement.[64–66]

TABLE II. Comparison of free energy values for cholesterol obtained from traditional and SDPI hybrid models of varying sized explicit regions. The values in parentheses represent the absolute error with respect to the value calculated from a full explicit solvent simulation (−32.70 ± .06 kJ/mol). Values of R_i are only relevant for SDPI models.

Domain radii R_o/R_i (nm)	Traditional	SDPI
2.5 / 1.5	−34.51 (−1.81) ± 0.10	−33.10 (−0.40) ± 0.09
2.0 / 1.5	−33.64 (−0.94) ± 0.09	−32.38 (0.32) ± 0.12
1.5 / 1.0	−34.21 (−1.51) ± 0.10	−32.48 (0.22) ± 0.12

VI. CONCLUSIONS

This work demonstrates that the accurate calculation of the bulk solvent PMF for hybrid models with an instantaneous switching at the explicit-implicit boundary is a formidable challenge, and that even well-parameterized models can result in boundary artifacts that lead to less accurate (and somewhat unpredictable) results. We present a new method, termed the SDPI hybrid model, in which explicit particles are slowly decoupled from the system across a switching region, allowing for a smooth removal of local interactions with the bulk solvent. The theoretical development of this model presents the conditions that are necessary to reproduce equilibrium properties from the original supersystem. This work does not present an analogous condition for the reproduction of kinetic properties from any particular thermostat. Given the finite domain and inclusion of Monte Carlo moves, it is expected that certain

kinetic properties (such as detailed water dynamics near the explicit/implicit boundary) will deviate from those calculated using full explicit methods. However, we expect any such artifacts to play a minor role for molecules within the explicit region and for timescales longer than the relaxation of water. Future work on more complex solutes will be used to determine which kinetic quantities can be faithfully reproduced.

The SDPI model presents nearly perfect thermodynamic results with respect to full explicit simulations at all regions within the explicit domain. It is important to note that the addition of this switching region does not add to the computational complexity of traditional hybrid models and actually provides a significant computational advantage by allowing one to simulate a much smaller number of particles while obtaining more accurate results. This advantage may be further highlighted when one considers that simpler implicit solvent models (such as the radial potential of this work) can now be used without penalizing model's accuracy. The model used in this work requires no additional parameters for any solute and three additional parameters for the solvent (σ_e, σ_s, and p, defined in Sec. II D). This is a substantial benefit over implicit solvent models that require new parameters for each atom type.

The SDPI model necessitates the use of a parameterized potential to counteract the gradient in chemical potential experienced by decoupled particles. Current work is aimed at extending the transferability of SDPI methods by developing a simple method of transforming this potential for use with explicit domains of varying shapes and sizes. Additionally, we are currently adapting these methods for use in more complex solvents that can be used for a wider range of biomolecular applications. While the theoretical development for such applications of the SDPI method will remain unchanged, the incorporation of polar solvents will likely require an orientational dependance of the chemical potential correction term, $G(r)$, and that the electrostatic implicit solvent model appropriately accounts for the change in solvent dielectric across the switching region. This could be accomplished by models utilizing simple geometrical overlaps similar to the nonpolar model used in this work[35] or by solutions to Poisson's equation that incorporate a smoothly varying dielectric.[98,99] We are particularly excited to see the potential applications of the SDPI model with popular three-point water models.

ACKNOWLEDGMENTS

We are very grateful to Nathan Baker, John Chodera, Lutz Maibaum, and Hans Andersen for enlightening discussions on the development of this model, as well as to Jeffrey Weber for the careful reading of this manuscript. This work was funded by the National Science Foundation awards DMS-0900700 and MCB-0954714 and the National Institutes of Health Grant No. R01-GM062868. Computer resources were provided by the National Science Foundation awards CNS-0619926. J.A.W was supported by the National Science Foundation Graduate Research Fellowship Program.

[1] S. Patel and C. Brooks, J. Comput. Chem. **25**, 1 (2004).
[2] E. Harder, B. Kim, R. Friesner, and B. Berne, J. Chem. Theory Comput. **1**, 169 (2005).
[3] S. W. Rick, S. J. Stuart, and B. J. Berne, J. Chem. Phys. **101**, 1 (1994).
[4] J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson, and T. Head-Gordon, J. Phys. Chem. B **114**, 2549 (2010).
[5] V. Voelz, G. Bowman, K. Beauchamp, and V. Pande, J. Am. Chem. Soc. **132**, 1526 (2010).
[6] G. Bowman and V. Pande, Proc. Natl. Acad. Sci. U.S.A. **107**, 10890 (2010).
[7] D. Shaw, R. Dror, J. Salmon, J. Grossman, K. Mackenzie, J. Bank, C. Young, M. Deneroff, B. Batson, and K. Bowers, in *Proceedings of the ACM/IEEE Conference on Supercomputing (SC09)* (IEEE Computer Society Press, Washington, DC 2009).
[8] J. Zhou, I. F. Thorpe, S. Izvekov, and G. A. Voth, Biophys. J. **92**, 4289 (2007).
[9] S. Izvekov and G. Voth, J. Phys. Chem. B **109**, 2469 (2005).
[10] S. Izvekov and G. A. Voth, J. Chem. Phys. **123**, 134105 (2005).
[11] W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen, J. Chem. Phys. **128**, 244114 (2008).
[12] W. G. Noid, P. Liu, Y. Wang, J.-W. Chu, G. S. Ayton, S. Izvekov, H. C. Andersen, and G. A. Voth, J. Chem. Phys. **128**, 244115 (2008).
[13] A. Das and H. C. Andersen, J. Chem. Phys. **131**, 034102 (2009).
[14] V. Krishna, W. G. Noid, and G. A. Voth, J. Chem. Phys. **131**, 024103 (2009).
[15] A. Das and H. C. Andersen, J. Chem. Phys. **132**, 164106 (2010).
[16] L. Lu, S. Izvekov, A. Das, H. C. Andersen, and G. A. Voth, J. Chem. Theory Comput. **6**, 954 (2010).
[17] V. Molinero and W. Goddard, J. Phys. Chem. B **108**, 1414 (2004).
[18] M. S. Shell, J. Chem. Phys. **129**, 144108 (2008).
[19] A. Chaimovich and M. S. Shell, Phys. Chem. Chem. Phys. **11**, 1901 (2009).
[20] A. Chaimovich and M. Shell, Phys. Rev. E **81**, 060104 (2010).
[21] L. Monticelli, S. Kandasamy, X. Periole, R. Larson, D. Tieleman, and S. Marrink, J. Chem. Theory Comput. **4**, 819 (2008).
[22] S. Marrink, H. Risselada, S. Yefimov, D. Tieleman, and A. de Vries, J. Phys. Chem. B **111**, 7812 (2007).
[23] H. J. Risselada, A. E. Mark, and S. J. Marrink, J. Phys. Chem. B **112**, 7438 (2008).
[24] S. C. L. Kamerlin, M. Haranczyk, and A. Warshel, ChemPhysChem **10**, 1125 (2009).
[25] V. Barone and M. Cossi, J. Phys. Chem. A **102**, 1995 (1998).
[26] A. Klamt and G. Schüürmann, J. Chem. Soc. Perkin Trans. **2**, 799 (1993).
[27] Q. Cui, J. Chem. Phys. **117**, 4720 (2002).
[28] M. J. Schnieders and J. W. Ponder, J. Chem. Theory Comput. **3**, 2083 (2007).
[29] M. J. Schnieders, N. A. Baker, P. Ren, and J. W. Ponder, J. Chem. Phys. **126**, 124114 (2007).
[30] J. Maple, Y. Cao, W. Damm, T. Halgren, G. Kaminski, L. Zhang, and R. Friesner, J. Chem. Theory Comput. **1**, 694 (2005).
[31] A. Onufriev, D. Bashford, and D. A. Case, Proteins **55**, 383 (2004).
[32] E. Gallicchio and R. Levy, J. Comput. Chem. **25**, 479 (2004).
[33] A. Vitalis and R. V. Pappu, J. Comput. Chem. **30**, 673 (2009).
[34] P. Labute, J. Comput. Chem. **29**, 1693 (2008).
[35] T. Grycuk, J. Chem. Phys. **119**, 4817 (2003).
[36] J. Swanson, J. A. Wagoner, N. Baker, and J. McCammon, J. Chem. Theory Comput. **3**, 170 (2007).
[37] A. Warshel, Annu. Rev. Biophys. Biomol. Struct. **32**, 425 (2003).
[38] J. Gao, P. Amara, C. Alhambra, and M. Field, J. Phys. Chem. A **102**, 4714 (1998).
[39] S. Dapprich, I. Komáromi, K. Byun, K. Morokuma, and M. Frisch, J. Mol. Struct.: THEOCHEM **461**, 1 (1999).
[40] B. Ensing, S. O. Nielsen, P. B. Moore, M. L. Klein, and M. Parrinello, J. Chem. Theory Comput. **3**, 1100 (2007).
[41] Q. Shi, S. Izvekov, and G. A. Voth, J. Phys. Chem. B **110**, 15045 (2006).
[42] M. Praprotnik, L. D. Site, and K. Kremer, J. Chem. Phys. **123**, 224106 (2005).
[43] M. Praprotnik, L. D. Site, and K. Kremer, Phys. Rev. E **73**, 066701 (2006).
[44] P. H. Hunenberger and J. A. McCammon, Biophys. Chem. **78**, 69 (1999).
[45] D. Beglov and B. Roux, J. Chem. Phys. **100**, 9050 (1994).
[46] W. Im, S. Berneche, and B. Roux, J. Chem. Phys. **114**, 2924 (2001).
[47] N. K. Banavali, W. Im, and B. Roux, J. Chem. Phys. **117**, 7381 (2002).
[48] D. Beglov and B. Roux, Biopolymers **35**, 171 (1995).
[49] Y. Deng and B. Roux, J. Chem. Phys. **128**, 115103 (2008).

[50]D. Shivakumar, Y. Deng, and B. Roux, J. Chem. Theory Comput. **5**, 919 (2009).
[51]C. Brooks and M. Karplus, J. Chem. Phys. **79**, 6312 (1983).
[52]M. Berkowitz and J. McCammon, Chem. Phys. Lett. **90**, 215 (1982).
[53]A. Brunger, C. Brooks, and M. Karplus, Chem. Phys. Lett. **105**, 495 (1984).
[54]M. S. Lee, F. R. Salsbury, and M. A. Olson, J. Comput. Chem. **25**, 1967 (2004).
[55]M. Lee and M. Olson, J. Phys. Chem. B **109**, 5223 (2005).
[56]V. Lounnas, S. Ludemann, and R. Wade, Biophys. Chem. **78**, 157 (1999).
[57]J. Essex and W. Jorgensen, J. Comput. Chem. **16**, 951 (1995).
[58]J. Rullmann and P. Vanduijnen, Mol. Phys. **61**, 293 (1987).
[59]A. Warshel and G. King, Chem. Phys. Lett. **121**, 124 (1985).
[60]Y. H. Li, G. Krilov, and B. J. Berne, J. Phys. Chem. B **109**, 463 (2005).
[61]P. Attard, Mol. Phys. **104**, 1951 (2006).
[62]G. Brancato, A. D. Nola, V. Barone, and A. Amadei, J. Chem. Phys. **122**, 154109 (2005).
[63]G. Brancato, N. Rega, and V. Barone, J. Chem. Phys. **128**, 144501 (2008).
[64]J. L. England and V. S. Pande, Biophys. J. **95**, 3391 (2008).
[65]J. England, D. Lucent, and V. Pande, Curr. Opin. Struct. Biol. **18**, 163 (2008).
[66]J. C. Rasaiah, S. Garde, and G. Hummer, Ann. Rev. Phys. Chem. **59**, 713 (2008).
[67]D. Chandler and H. C. Andersen, J. Chem. Phys. **57**, 1930 (1972).
[68]T. Yamazaki and A. Kovalenko, J. Phys. Chem. B **115**, 310 (2011).
[69]F. Hirata and P. Rossky, Chem. Phys. Lett. **83**, 329 (1981).
[70]D. Beglov and B. Roux, J. Chem. Phys. **104**, 8678 (1996).
[71]D. Beglov and B. Roux, J. Phys. Chem. B **101**, 7821 (1997).
[72]P. Koehl, H. Orland, and M. Delarue, Phys. Rev. Lett. **102**, 87801 (2009).
[73]B. Roux and T. Simonson, Biophys. Chem. **78**, 1 (1999).
[74]W. Im, S. Seefeld, and B. Roux, Biophys. J. **79**, 788 (2000).
[75]H. J. Woo, A. R. Dinner, and B. Roux, J. Chem. Phys. **121**, 6392 (2004).
[76]D. Adams, Mol. Phys. **28**, 1241 (1974).
[77]D. Adams, Mol. Phys. **29**, 307 (1975).
[78]D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed. (Academic, New York, 2002).
[79]B. Smit, Mol. Phys. **85**, 153 (1995).
[80]J. D. Weeks, D. Chandler, and H. C. Andersen, J. Chem. Phys. **54**, 5237 (1971).
[81]F. Stillinger, J. Solution Chem. **2**, 141 (1973).
[82]K. Lum, D. Chandler, and J. Weeks, J. Phys. Chem. B **103**, 4570 (1999).
[83]D. Chandler, Nature (London) **417**, 491 (2002).
[84]D. Huang, P. Geissler, and D. Chandler, J. Phys. Chem. B. **105**, 6704 (2001).
[85]D. Huang and D. Chandler, Phys. Rev. E **61**, 1501 (2000).
[86]D. Huang and D. Chandler, Proc. Natl. Acad. Sci. U.S.A. **97**, 8324 (2000).
[87]G. Graziano, J. Phys. Chem. B **110**, 11421 (2006).
[88]J. A. Wagoner and N. A. Baker, Proc. Natl. Acad. Sci. U.S.A. **103**, 8331 (2006).
[89]J. A. Wagoner and N. A. Baker, J. Comput. Chem. **25**, 1623 (2004).
[90]A. Ben-Naim, *Solvation Thermodynamics*, 2nd ed. (Plenum, New York, 1987).
[91]T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, and W. F. van Gunsteren, Chem. Phys. Lett. **222**, 529 (1994).
[92]H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, Comput. Phys. Commun. **91**, 43 (1995).
[93]H. Berendsen, J. Postma, and W. van Gunsteren, J. Chem. Phys. **81**, 3684 (1984).
[94]M. Shirts, D. Mobley, J. Chodera, and V. Pande, J. Phys. Chem. B **111**, 13052 (2007).
[95]A. Shrake and J. Rupley, J. Mol. Biol. **79**, 351 (1973).
[96]See http://www.itl.nist.gov/div898/handbook/ for NIST/SEMATECH e-Handbook of Statistical Methods.
[97]M. R. Shirts and J. D. Chodera, J. Chem. Phys. **129**, 124105 (2008).
[98]P. Qin, Z. Xu, W. Cai, and D. Jacobs, Commun. Comput. Phys. **6**, 955 (2009).
[99]C. Xue and S. Deng, Phys. Rev. E **81**, 016701 (2010).