Computation of binding free energy with molecular dynamics and grand canonical Monte Carlo simulations

Yuqing Deng[1] and Benoît Roux[1,2,a)]
[1]*Biosciences Division, Argonne National Laboratory, 9700 S. Cass Avenue, Argonne, Illinois 60439, USA*
[2]*Department of Biochemistry and Molecular Biology, Gordon Center for Integrative Science, University of Chicago, 929 57th Street, Chicago, Illinois 60637, USA*

(Received 13 December 2007; accepted 17 January 2008; published online 18 March 2008; publisher error corrected 26 March 2008)

The binding of a ligand to a receptor is often associated with the displacement of a number of bound water molecules. When the binding site is exposed to the bulk region, this process may be sampled adequately by standard unbiased molecular dynamics trajectories. However, when the binding site is deeply buried and the exchange of water molecules with the bulk region may be difficult to sample, the convergence and accuracy in free energy perturbation (FEP) calculations can be severely compromised. These problems are further compounded when a reduced system including only the region surrounding the binding site is simulated. To address these issues, we couple molecular dynamics (MD) with grand canonical Monte Carlo (GCMC) simulations to allow the number of water to fluctuate during an alchemical FEP calculation. The atoms in a spherical inner region around the binding pocket are treated explicitly while the influence of the outer region is approximated using the generalized solvent boundary potential (GSBP). At each step during thermodynamic integration, the number of water in the inner region is equilibrated with GCMC and energy data generated with MD is collected. Free energy calculations on camphor binding to a deeply buried pocket in cytochrome P450cam, which causes about seven water molecules to be expelled, are used to test the method. It concluded that solvation free energy calculations with the GCMC/MD method can greatly improve the accuracy of the computed binding free energy compared to simulations with fixed number of water. © *2008 American Institute of Physics*. [DOI: 10.1063/1.2842080]

I. INTRODUCTION

A number of recent studies have shown very encouraging results in calculations of standard (absolute) binding free energies using atomistic molecular dynamics (MD) simulations (see Ref. 1 for a review).[2–10] The progress is explained, in part, by several developments both formal and technical. Over the last years, many of the critical issues concerning the treatment of the standard state and the usage of restraint potentials to improve convergence have been resolved,[2,7,8,11–14] and strategies for the treatment of multiple ligand orientations and protein conformations have been elaborated.[6,8,9] In order to decrease the computational cost of free energy computations, methods have been designed to simulate reduced model systems in which the influence of the surrounding is incorporated implicitly via an effective "boundary potential."[15–20]

This progress signals that reliable free energy computations are increasingly becoming a reality. However, achieving a proper and complete equilibration of the solvent molecules during free energy computations can still represent a practical and important problem. While representative solvent configurations are generated spontaneously by unbiased MD trajectories when a binding site is exposed to the bulk phase, difficulties in sampling solvent configurations can become particularly acute when a binding site is deeply buried and inaccessible. In this case, the exchange of water molecules with the bulk region may be very slow and the convergence and accuracy in free energy perturbation (FEP) calculations based on unbiased molecular dynamics trajectories can be severely compromised. Difficulties in properly sampling the hydration states of a binding site are further compounded when a reduced system is simulated with an effective boundary potential.

A typical example is provided by the active site of cytochrome P450, a protein that oxidizes endogenous and xenobiotic substrates.[21] Previous simulation studies have shown that some conformational changes, including side chain rotameric states, are required to open up a channel allowing exchange between the cavity and the bulk phase;[21,22] about six water molecules are expelled from the buried cavity upon substrate binding. Such a change in hydration state must be captured in free energy calculations to yield accurate and meaningful results. Advance simulation methodologies to enhance sampling, such as replica exchange molecular dynamics[23] might be able to alleviate some of those problems. For example, recent applications[24,25] of Hamiltonian replica exchange[26] have shown the ability to accelerate protein folding by introducing hydrophobic replicas. However, constructing a novel replica exchange algorithm designed to enhance the sampling of water occupancy inside confined binding pockets is not straightforward.

To address this problem, Helms and Wade designed an alchemical transformation to include specific water molecules in a calculation of the binding free energy of camphor

a)Electronic mail: roux@chicago.edu.

to cytochrome P450.[27] Accordingly, the water molecules known to be displaced were "switched on" as the ligand camphor was "switched off" during the alchemical transformation. While this procedure takes into account the water displacement in the binding process, it implicitly assumes that one knows in advance the number of water molecules involved in the process. Systematic methods to account for the hydration level of internal cavities have been designed, though they require the specific evaluation of all probability of water occupancy.[3,12,27–31] Rigorously speaking, the number of water molecules in a reduced system should be in open thermodynamic equilibrium with the bulk phase. Thus, proper configurational averages are sampled only if the number of water molecules is allowed to fluctuate according to an effective grand canonical ensemble.[32] In this context, free energy computations with reduced systems should optimally be combined with a grand canonical Monte Carlo[32–36] (GCMC) to yield proper statistical averages. As the hydration state of a binding site is important,[31,37–39] a generally applicable simulation strategy able to naturally generate the proper hydration states would be highly desirable.

In this paper, we develop and test a free energy simulation framework that combines GCMC with MD generated with a reduced system. The method used in the present study is called the generalized solvent boundary potential (GSBP).[20] GSBP is a generalization of the spherical solvent boundary potential (SSBP), which was designed to simulate a solute in bulk water.[18] The reduction in system size afforded by GSBP permits inexpensive free energy computations quickly and with reasonable accuracy.[6,7,40] The present work is an extension of the developments in Ref. 32. It is our hope that the combination of GCMC and GSBP will provide a robust and efficient framework for the computations of binding free energies.

The rest of the paper is organized as follows. In the following section, we present the fundamental theory for free energy calculation with GCMC/MD. We then proceed to test calculations of solvation free energy and camphor/cytochrome P450 binding free energy. Last, we conclude the paper with a summary of the main results.

II. THEORY

A. GCMC and free energy calculation

Let us consider a solute inside a large solvent reservoir of N solvent molecules. Without loss of generality, we assume that the center of mass of the solute is fixed at the origin and that the internal degrees of freedom of the solute are represented by $\mathbf{X}$. The degrees of freedom of a solvent molecule, assumed here to be rigid for the sake of simplicity, are represented by $\mathbf{x}$, which includes the center-of-mass translation and rigid body rotation. The total potential energy of the full system is $U(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_N;\lambda)$, where λ is a thermodynamic switching parameter such that the solute is noninteracting (decoupled) when $\lambda=0$ and fully interacting when $\lambda=1$. By definition, the solvation free energy $\Delta G(\lambda)$ of the solute is

$$e^{-\beta\Delta G(\lambda)}=\frac{\int d\mathbf{X}\int d\mathbf{x}_1\cdots d\mathbf{x}_N e^{-\beta U(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_N;\lambda)}}{\int d\mathbf{X}\int d\mathbf{x}_1\cdots d\mathbf{x}_N e^{-\beta U(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_N;\lambda=0)}}. \quad (1)$$

Let us define a spherical inner region, centered on the solute in the solvent. The space outside the inner region is referred to as the outer region. For any instantaneous configuration of the system, it is possible to monitor the total number of solvent molecules inside the inner region. It is given by the discrete function $n'(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N)$, defined as

$$n'(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N)=\sum_{i=1}^{N}H(\mathbf{r}_i), \quad (2)$$

where $\mathbf{r}_i$ is the position of the center of the ith solvent molecule and H is a step-function equal to 1 with the molecule inside the inner region. The probability $\mathcal{P}_n(\lambda)$ of having exactly n solvent molecules inside the inner region is calculated from the average

$$\mathcal{P}_n(\lambda)=\langle\delta_{nn'}\rangle_{(\lambda)}=\frac{\int d\mathbf{X}\int d\mathbf{x}_1\cdots\int d\mathbf{x}_N\delta_{n,n'}e^{-\beta U(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_N;\lambda)}}{\int d\mathbf{X}\int d\mathbf{x}_1\cdots\int d\mathbf{x}_N e^{-\beta U(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_N;\lambda)}}, \quad (3)$$

where $\delta_{nn'}$ is a Kroenecker discrete delta function,

$$\delta_{nn'}=\begin{cases}1 & \text{if } n=n'(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N)\\ 0 & \text{otherwise.}\end{cases} \quad (4)$$

By construction, the probabilities $\mathcal{P}_n(\lambda)$ are normalized, i.e., $\Sigma_n\mathcal{P}_n(\lambda)=1$, via the completeness of the Kroenecker delta. Inserting the Kroenecker function in the expression for $\Delta G(\lambda)$ and summing over all possible number of solvent molecules yields,

$$e^{-\beta\Delta G(\lambda)}=\sum_n\frac{\int d\mathbf{X}\int d\mathbf{x}_1\cdots d\mathbf{x}_N\delta_{n,n'}e^{-\beta U(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_N;\lambda)}}{\int d\mathbf{X}\int d\mathbf{x}_1\cdots d\mathbf{x}_N e^{-\beta U(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_N;\lambda=0)}}. \quad (5)$$

It is useful to introduce the case $n=0$ and $\lambda=0$ as a reference state,

$$\begin{aligned}e^{-\beta\Delta G(\lambda)}&=\sum_n\frac{\int d\mathbf{X}\int d\mathbf{x}_1\cdots d\mathbf{x}_N\delta_{n,n'}e^{-\beta U(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_N;\lambda)}}{\int d\mathbf{X}\int d\mathbf{x}_1\cdots d\mathbf{x}_N e^{-\beta U(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_N;\lambda=0)}}\\&\quad\times\frac{\int d\mathbf{X}\int d\mathbf{x}_1\cdots d\mathbf{x}_N\delta_{0,n'}e^{-\beta U(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_N;\lambda=0)}}{\int d\mathbf{X}\int d\mathbf{x}_1\cdots d\mathbf{x}_N\delta_{0,n'}e^{-\beta U(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_N;\lambda=0)}}\\&=\mathcal{P}_0(\lambda=0)\\&\quad\times\sum_n\frac{\int d\mathbf{X}\int d\mathbf{x}_1\cdots d\mathbf{x}_N\delta_{n,n'}e^{-\beta U(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_N;\lambda)}}{\int d\mathbf{X}\int d\mathbf{x}_1\cdots d\mathbf{x}_N\delta_{0,n'}e^{-\beta U(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_N;\lambda=0)}}, \quad (6)\end{aligned}$$

where $\mathcal{P}_0(\lambda=0)$ is the probability of finding the inner region empty of any solvent molecules with a decoupled solute,

$$\mathcal{P}_0(\lambda=0)=\frac{\int d\mathbf{X}\int d\mathbf{x}_1\cdots d\mathbf{x}_N\delta_{0,n'}e^{-\beta U(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_N;\lambda=0)}}{\int d\mathbf{X}\int d\mathbf{x}_1\cdots d\mathbf{x}_N e^{-\beta U(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_N;\lambda=0)}}. \quad (7)$$

Since $\mathcal{P}_0(\lambda=0)$ is the probability of spontaneous occurrence for a configuration with zero solvent molecule in the inner region, it is associated with the reversible work $\Delta G_{\text{hs}}=-k_BT\ln[\mathcal{P}_0(\lambda=0)]$ for inserting a hard-sphere corresponding to radius of the inner region into the solvent. To compute

the configurational integral with exactly n solvent molecules in the inner region, one can arbitrarily pick the first n molecules and enforce that the remaining solvent molecules are restricted to the outer region. Since all the solvent molecules are identical, there are $N!/(N-n)!n! \approx N^n/n!$ equivalent choices, which gives the expression,

$$e^{-\beta\Delta G(\lambda)} = \mathcal{P}_0(\lambda=0)\sum_n \frac{N!}{n!(N-n)!}\frac{\int d\mathbf{X}\int_{\text{in}} d\mathbf{x}_1\cdots d\mathbf{x}_n\int_{\text{out}} d\mathbf{x}_{n+1}\cdots d\mathbf{x}_N e^{-\beta U(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_N;\lambda)}}{\int d\mathbf{X}\int_{\text{out}} d\mathbf{x}_1\cdots d\mathbf{x}_n\int_{\text{out}} d\mathbf{x}_{n+1}\cdots d\mathbf{x}_N e^{-\beta U(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_N;\lambda=0)}}$$
$$= \mathcal{P}_0(\lambda=0)\sum_n \frac{N^n}{n!}\frac{\int d\mathbf{X}\int_{\text{in}} d\mathbf{x}_1\cdots d\mathbf{x}_n\int_{\text{out}} d\mathbf{x}_{n+1}\cdots d\mathbf{x}_N e^{-\beta U(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_N;\lambda)}}{\int d\mathbf{X}\int_{\text{out}} d\mathbf{x}_1\cdots d\mathbf{x}_n\int_{\text{out}} d\mathbf{x}_{n+1}\cdots d\mathbf{x}_N e^{-\beta U(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_N;\lambda=0)}}$$
$$= \mathcal{P}_0(\lambda=0)\sum_n \frac{N^n}{n!}\frac{\int d\mathbf{X}\int_{\text{in}} d\mathbf{x}_1\cdots d\mathbf{x}_n e^{-\beta W(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_n;\lambda)}}{\int d\mathbf{X}\int_{\text{out}} d\mathbf{x}_1\cdots d\mathbf{x}_n e^{-\beta W(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_n;\lambda=0)}}, \tag{8}$$

where W is the potential of mean force (PMF) for the solute and the n tagged solvent molecules, defined as

$$e^{-\beta W(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_n;\lambda)} \propto \int_{\text{out}} d\mathbf{x}_{n+1}\cdots d\mathbf{x}_N e^{-\beta U(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_N;\lambda)}. \tag{9}$$

It is useful to express the total potential energy U in terms of contributions from the inner and outer regions,

$$U(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_N;\lambda) = U_{\text{ii}}(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_n;\lambda) + U_{\text{io}}(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_N;\lambda) + U_{\text{oo}}(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N), \tag{10}$$

where U_{ii} denotes the potential energy of the molecules within the inner region, U_{io} denotes the interaction energy between the molecules in the inner region and the ones in the outer region, and U_{oo} denotes the potential energy in the outer region only. From this decomposition, the PMF may be expressed as[20]

$$e^{-\beta W(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_n;\lambda)} = e^{-\beta U_{\text{ii}}(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_n;\lambda)}\frac{\int_{\text{out}} d\mathbf{x}_{n+1}\cdots d\mathbf{x}_N e^{-\beta[U_{\text{io}}(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_N;\lambda)+U_{\text{oo}}(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N)]}}{\int_{\text{out}} d\mathbf{x}_{n+1}\cdots d\mathbf{x}_N e^{-\beta U_{\text{oo}}(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N)}} = e^{-\beta[U_{\text{ii}}(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_n;\lambda)+\Delta W(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_n;\lambda)]}. \tag{11}$$

By definition, ΔW is equivalent to the solvation free energy of the solute and n solvent molecules in a fixed configuration embedded in its associated spherical exclusion region. It is equivalent to the reaction field solvent potential that is generated by GSBP.[20] Furthermore, one may note that

$$\lim_{\mathbf{x}_1,\ldots,\mathbf{x}_n\to\infty} W(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_n;\lambda=0) = n\Delta\mu_s + u_{\text{int}}(\mathbf{X}), \tag{12}$$

where $\Delta\mu_s$ is the excess solvation free energy of the solvent molecules in the bulk region (far away from the inner region) and $u_{\text{int}}(\mathbf{X})$ is the internal potential function of the solute (decoupled from the solvent). By virtue of the translational and rotational invariance of the bulk region and Eq. (12), the configurational integral in the denominator in Eq. (8) may be rewritten as

$$\int d\mathbf{X}\int_{\text{out}} d\mathbf{x}_1\cdots d\mathbf{x}_n e^{-\beta W(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_n;\lambda=0)}$$
$$= V^n e^{-n\beta\Delta\mu_s}(8\pi^2)^n\int d\mathbf{X} e^{-\beta u_{\text{int}}(\mathbf{X})}, \tag{13}$$

where V is the volume of the bulk region and $(8\pi^2)$ is the integral over the internal degrees of freedom of a solvent molecule, (N.B., $\int_{\text{out}} d\mathbf{x}_i = V8\pi^2$). If the solvent molecules were flexible, the above expressions would be modified to include configurational integral over internal degrees of freedom. It follows that the solvation free energy of the solute is

$$e^{-\beta\Delta G(\lambda)} = \mathcal{P}_0(\lambda=0) \times\sum_n \frac{\bar{\rho}^n}{n!}\frac{\int d\mathbf{X}\int_{\text{in}} d\mathbf{x}_1\cdots d\mathbf{x}_n e^{-\beta[W(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_n;\lambda)-n\Delta\mu_s]}}{(8\pi^2)^n\int d\mathbf{X} e^{-\beta u_{\text{int}}(\mathbf{X})}}, \tag{14}$$

where $\bar{\rho}=(N/V)$ is the average solvent density in the bulk. Equivalently, the solvation free energy of the solute may be expressed as

$$e^{-\beta\Delta G(\lambda)} = \mathcal{P}_0(\lambda=0)\sum_n (\bar{\rho})^n K_n(\lambda), \tag{15}$$

where $K_n(\lambda)$ are effective equilibrium association constants,

$$K_n(\lambda) = \frac{1}{n!}\frac{\int d\mathbf{X}\int_{\text{in}}\mathbf{x}_1\cdots d\mathbf{x}_n e^{-\beta[W(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_n;\lambda)-n\Delta\mu_s]}}{(8\pi^2)^n\int d\mathbf{X} e^{-\beta u_{\text{int}}(\mathbf{X})}}, \tag{16}$$

(note that K_0 is equal to 1). Equations (15) and (16) form the basis of the quasichemical free energy formalism that Grabowski *et al.* have used to discuss the solvation free energy of ions.[41]

One may also compute the solvation free energy using a thermodynamic integration,

$$\Delta G = -k_B T \int_0^1 d\lambda \frac{\partial}{\partial \lambda} \ln\left[\mathcal{P}_0(\lambda=0) \sum_n \frac{\bar{\rho}^n}{n!} \frac{\int d\mathbf{X} \int_{\text{in}} d\mathbf{x}_1 \cdots d\mathbf{x}_n e^{-\beta[W(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_n;\lambda)-n\Delta\mu_s]}}{(8\pi^2)^n \int d\mathbf{X} e^{-\beta u_{\text{int}}(\mathbf{X})}} \right]$$
$$= \int_0^1 d\lambda \sum_n \mathcal{P}_n(\lambda) \frac{\int d\mathbf{X} \int_{\text{in}} d\mathbf{x}_1 \cdots d\mathbf{x}_n (\partial W/\partial\lambda) e^{-\beta W(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_n;\lambda)}}{\int d\mathbf{X} \int_{\text{in}} d\mathbf{x}_1 \cdots d\mathbf{x}_n e^{-\beta W(\mathbf{X},\mathbf{x}_1,\ldots,\mathbf{x}_n;\lambda)}}$$
$$= \int_0^1 d\lambda \sum_n \mathcal{P}_n(\lambda) \left\langle \frac{\partial W}{\partial \lambda} \right\rangle_{\lambda,n}, \tag{17}$$

where the bracket $\langle \cdots \rangle_{\lambda,n}$ implies a constrained average with fixed coupling constant λ and fixed number of solvent molecules in the inner region, weighted by the probabilities $\mathcal{P}_n(\lambda)$ of n solvent molecules occupying the inner region,

$$\mathcal{P}_n(\lambda) = \frac{\bar{\rho}^n K_n(\lambda)}{\Sigma_m \bar{\rho}^m K_m(\lambda)}, \tag{18}$$

where Eqs. (4), (11), and (12) have been used.

Occupancy of an inner region by n solvent molecules distributed according to the probability $\mathcal{P}_n(\lambda)$ arises naturally in a GCMC simulation.[32] The reversible work expression can be written in terms of a grand canonical average of $\partial W/\partial\lambda$,

$$\Delta G = \int_0^1 d\lambda \left\langle\!\!\left\langle \frac{\partial W}{\partial \lambda} \right\rangle\!\!\right\rangle_\lambda. \tag{19}$$

This formulation shows how the fluctuations in the number of solvent molecules are naturally incorporated into the free energy calculation performed in a finite inner region. Equation (19) is the central result of this paper.

A formulation of free energy calculations based on Eq. (19) is essentially equivalent to familiar thermodynamic integration (TI),[42] where the averaging of $\partial W/\partial\lambda$ has been generalized to include the average over configuration and number of water molecules in the open ensemble. Interestingly, it is not possible to convert Eq. (19) into a standard FEP.[43] In the open system ensemble with fluctuating number of water molecules, FEP between two values of the coupling parameters λ_1 and λ_2 requires a reequilibration of the number of water molecules for each end-state λ_1 and λ_2. This highlights important formal differences between the two finite representations of an infinite thermodynamic system implemented respectively with SSBP (Ref. 18) and GSBP.[20] In GSBP, the radius of the inner region is fixed.[20] Accordingly, proper configurational averages can only be achieved if the number of water molecules is allowed to fluctuate according to the grand canonical ensemble.[32] In contrast, the radius of the inner region fluctuates dynamically in SSBP while the number of explicit water molecules within the inner region is constant.[18] For this reason, free energy calculations can rigorously be performed according to both the TI and FEP protocoles with the effective PMF in SSBP.[18] Although one should rigorously use only TI with GSBP, in practice, it is possible to use FEP if the perturbation proceeds by small steps for which $\mathcal{P}_n(\lambda)$ does not vary significantly.

B. Free energy decomposition

To enhance the sampling efficiency, we separate the binding free energy into the following components, defined as the progressive switching on of repulsive, dispersive, electrostatic (see Ref. 40 for details on the repulsive and dispersive contributions), and introduction/removal of restraint potentials,[6,7,40]

$$\Delta G_b^\circ = [\Delta G_{\text{rep}}^{\text{site}} - \Delta G_{\text{rep}}^{\text{bulk}}] + [\Delta G_{\text{disp}}^{\text{site}} - \Delta G_{\text{disp}}^{\text{bulk}}] + [\Delta G_{\text{elec}}^{\text{site}} - \Delta G_{\text{elec}}^{\text{bulk}}] + \Delta\Delta G_{\text{rstr}}^\circ, \tag{20}$$

where $\Delta G_{\text{rep}}^{\text{site}}$ and $\Delta G_{\text{rep}}^{\text{bulk}}$ are contributions from repulsion in binding site (restraints on) and bulk solvent, respectively, $\Delta G_{\text{disp}}^{\text{site}}$ and $\Delta G_{\text{disp}}^{\text{bulk}}$ are contributions from dispersion in binding site (restraints on) and bulk solvent, respectively, $\Delta G_{\text{elec}}^{\text{site}}$ and $\Delta G_{\text{elec}}^{\text{bulk}}$ are contributions from electrostatics in binding site (restraints on) and bulk solvent, respectively, and $\Delta\Delta G_{\text{rstr}}^\circ$ is the contribution from translational and rotational restraint potential with standard volume unit in liters. Here, we do not use restraint on the ligand conformation because the camphor does not have any rotable bonds. The solvation free energy is the sum of all bulk contributions,

$$\Delta G_{\text{solv}} = \Delta G_{\text{rep}}^{\text{bulk}} + \Delta G_{\text{disp}}^{\text{bulk}} + \Delta G_{\text{elec}}^{\text{bulk}}. \tag{21}$$

III. COMPUTATIONAL DETAILS

The CHARMM PARAM27 force field was used for the protein and for the heme in its reduced state.[44] The solvent water model was TIP3P.[45] The benzene model,[46] after which the CHARMM22 phenylalanine was parametrized,[44] was used in the benzene solvation free energy calculations. The generalized AMBER force field was used for camphor;[47] the ligand topology and parameters were generated with ANTECHAMBER 1.27[48] and with AM1-BCC partial charges.[49,50]

Crystal structures of the cytochrome P450/camphor complex in the Protein Data Bank (2CPP) were used to generate the all-atom structures. Hydrogen atoms were added using the HBUILD[51] facility in CHARMM.[52] The missing side chain coordinates of LYS205 were predicted from SCWRL.[53] The explicit simulation spheres were henceforth constructed from those all-atom structures. A 15 Å radius sphere was defined around the center of the binding pocket with 121 generalized multiple basis functions for GSBP.[20] The system was hydrated with 20 cycles of MC and MD (10 000 MC moves followed by 10 000 MD steps with 2 fs time step).[32] The MC consists of rigid body translation, rotation, and GCMC moves with equal probability. The anchor atoms for the restraint were selected at random on the protein and camphor ligand. The reference ligand center of mass and orientation angle values were calculated as the average of a 20 ps unrestrained MD trajectory.

All the free energy perturbation calculations were carried

out with the CHARMM program[52] version c34a2 modified for the present study. The simulations in the binding site consist of 18 windows of ligand repulsion, five windows of ligand dispersion, ten windows of ligand electrostatics, and 15 windows of restraint removal.[6] Two trajectories were run for each of the windows. Each of the trajectories consists of six cycles of GCMC/MD run (8000 MC steps followed by 16 ps MD) to equilibrate the system and 8 ps data collection. A further 16 ps MD and 8 ps data collection was done after the GCMC/MD cycles (112 ps total for each window). For the fixed water number simulation, each trajectory was 20 ps of equilibration and 40 ps of data collection. No GCMC was applied in the removal of the restraint potentials on the ligand.

In the bulk GCMC/MD, FEP calculations were done in a 10 Å sphere with nine basis functions of generalized multiples for GSBP. Ten cycles of GCMC/MD run (10 000 MC steps followed by 20 ps MD) were run to equilibrate the system and 8 ps data collection. A further 20 ps MD and 8 ps data collection was done after the GCMC/MD cycle (96 ps total for each trajectory). The fixed water number calculation in the bulk ($\Delta G_{\text{int}}^{\text{bulk}}$) calculation was done with SSBP.[18,40] A solute molecule and 400 water molecules were integrated with the 2 fs time step Langevin dynamics. The length of each simulation window was 40 ps equilibration followed by 40 ps of sampling. The relevant coupling parameters were the same for the bulk and site calculations. All bonds involving hydrogen atoms were fixed with the SHAKE algorithm.[54] To ensure convergence, we performed a series of free energy calculations for each system, starting from the configurations of each window saved at the end of the last run.

IV. RESULTS AND DISCUSSIONS

A. Solvation free energy

The excess chemical potential (hydration free energy) of water and benzene has been calculated first to test and validate the simulation strategy. The number of water is shown in Fig. 1 as the simulation progresses from a noninteracting water to a fully interacting one. As expected for a water molecule solute, the number of water solvent molecules fluctuates around the mean value, with no net change. The averages are reported in Table I. It is observed that about two solvent water molecules are driven out of the simulation sphere to accommodate the solute. Because of its size, benzene is expected to replace several water molecules when it is fully solvated. As the progression in Fig. 1 shows, the number of water molecules steadily decreases when the repulsive potential of benzene is gradually introduced (stages 1–10). The change is more obvious during the initial stages of the solute insertion, when the purely repulsive cavity is switched on. As the dispersive/attractive potential is added (stages 10–14), the number of water molecules increases slightly, consistent with a slight stabilization by the van der Waals attractive interaction of water around a nonpolar solute.[34] Because the polarity of benzene is almost negligible, the addition of electrostatic potential (stages 14–24) does not cause any further change in the number of water molecules.

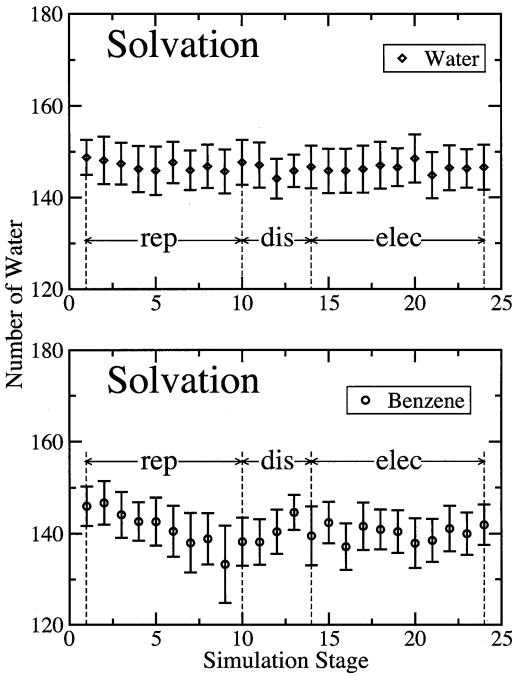

FIG. 1. Shown in the plot is the number of water molecules during solvation calculation. The upper panel is the solvation of water. The lower panel is the solvation of benzene. The simulation is divided into 24 stages, starting from noninteracting in stage one and progressing to full-interacting solute (water or benzene) in stage 24. In stages 1–10, the solute repulsion is turned on gradually. In stages 10–14, the solute attraction is added. At last, in stages 14–24, the solute particle charges are turned on. The errors are standard deviation of five complete runs of solvation free energy calculations.

Table I reports the free energy decomposition for the hydration free energies of water and benzene. For water, the free energy values from the decomposition match closely with the earlier calculation computed from periodic boundary conditions (PBC) and SSBP.[40] The chemical potential (−6.2 kcal/mol) used as input in the GCMC moves agrees quite well with the $-6.09\pm$ kcal/mol chemical potential computed from FEP. This is encouraging and reflects on the self-consistency of the method. The statistical errors were estimated from the last five runs of a series of 11 runs of simulations. For water, the precision of the free energy values is high, in spite of fluctuations in the number of water molecules in the GSBP sphere. One may note that in the GSBP sphere the average number of water molecules in the water calculation is slightly higher than 140, the number of water in a 10 Å sphere with bulk density of 0.0334 Å^{-3}. This overestimated number might be due to finite size effects; the surface tension causing an increase in the local pressure and density inside the droplet. The agreement of benzene hydration free energies with previous PBC and SSBP calculations[40] is not perfect but is quite reasonable. The nearly exact match of the current benzene solvation value to the experiment is partly fortuitous and we would not expect this small difference to hold for other molecules. Given the sensitivity of the repulsive free energy contribution to the details of the simulations,[40] such small differences are not surprising. These simulations show that the current GCMC/MD protocol can be used for calculating solvation free energies with finite simulation systems and GSBP.

TABLE I. Free energies of water and benzene hydration. GCMC/MD water simulation in a 10 Å radius sphere. $\mu=-6.2$ kcal/mol, $\bar{\rho}=0.0334$ Å^{-3}. Ten cycles of 10k-step MC and 10k-step MD for equilibration and 4k-step of MD for data collection. Free energy values in kcal/mol. The experimental values are taken from Ref. 56. The PBC and SSBP values are taken from Ref. 40.

Solute		ΔG_{rep}	ΔG_{disp}	ΔG_{elec}	ΔG	Expt.
Water	GCMC/MD	4.96 ± 0.06	-2.88 ± 0.01	-8.17 ± 0.06	-6.09 ± 0.04	−6.3
	PBC	4.98	−2.65	−8.26	−5.93	
	SSBP	5.26	−2.86	−8.12	−5.72	
Benzene	GCMC/MD	13.87 ± 0.13	-12.81 ± 0.08	-1.84 ± 0.02	-0.79 ± 0.21	−0.87
	PBC	13.30	−11.47	−1.83	0.00	
	SSBP	14.36	−12.66	−2.02	−0.31	

B. P450 bind camphor

The progression of the number of water molecules in the binding site during the free energy calculation is shown in Fig. 2. The trend in the change is similar to that of benzene hydration. About five water molecules get expelled from the binding site in the first stage when the repulsive potential of camphor is introduced. Except for the initial decrease of the number of water molecules in the initial stage, the number of water molecules fluctuates around an average value for the rest of the simulation. The larger fluctuations in the number of water molecules in the early stages might be caused by the confinement of the binding pocket. The large fluctuations observed during the free energy simulation justifies the approximation where the number of water molecules is not reequilibrated at the two end points of the FEP. Ultimately, six to seven water molecules are displaced from the simulation region when camphor binds. This agrees with the number used in the alchemical free energy simulation where water molecules are switched off while camphor is switched on.[27] In a sphere of 6.0 Å radius around camphor, the number of water changes from nine to one when camphor interaction is turned on. There are three more water molecules in this sphere than in the binding pocket defined in Ref. 27. Free energy calculations[27] have shown that six water molecules are most stable in the binding pocket. These differences could be a result of the different definitions of the binding pocket.

Table II lists the free energy values of P450-camphor binding. For the sake of comparison, two calculations with a fixed number of water molecules were also carried out, in analogy with the holo and apo states of P450. The first system, denoted as *Fix W. Holo*, contains 37 water molecules in the inner region, corresponding to the number of bound water during the simulation of P450 with bound camphor. The second system, denoted as *Fix W. Apo*, contains 45 water molecules in the inner region, corresponding to the number of bound water during the simulation without camphor. The difference of eight water molecules roughly matches the change of 6.7 water molecules observed in the course of the free energy calculations. The decomposition of the free energies shows the usual change from the bulk solution to the binding site—both the repulsive and dispersive contribution favors the binding and the electrostatics tends to be unfavorable. As shown in Table II, these calculations with fixed number of water molecules clearly miss the mark. They either predict a binding free energy value more favorable by 6 kcal/mol or more unfavorably by 20 kcal/mol. For the case where the binding site is not sufficiently hydrated enough (Fix W. Holo), the repulsive contribution differs the most. It accounts for the 6 kcal/mol over-favorable difference. The overhydrated binding site (Fix W. Apo), on the other hand, yields a total binding free energy of 12.60 kcal/mol. The free energy decomposition shows that the repulsive free energy in the binding site is highly unfavorable (40 kcal/mol). The dispersive free energy changes somewhat in favor of binding. The electrostatics turns around and becomes favorable to binding. It is apparent that camphor is pressed by the water and protein in that binding

TABLE II. Binding free energies of P450 to camphor are listed in the table. GCMC/MD stands for the calculation where the number of water molecules is allowed to fluctuate with GCMC moves. *Fix W. Apo* and *Fix W. Holo* stand for a fixed number of water molecules equal to 45 and 37 in the MD region, respectively. These are the number of water molecules with or without camphor bound. All free energies are in kcal/mol.

	Site			Bulk	
	GCMC/MD	Fix. W. Holo	Fix W. Apo	SSBP	Expt.
Δn_{water}	6.7	⋯	⋯	⋯	⋯
ΔG_{rep}	13.66 ± 0.30	7.99 ± 0.13	40.26 ± 1.39	21.15 ± 0.21	⋯
ΔG_{disp}	-35.40 ± 0.03	-35.50 ± 0.06	-36.99 ± 0.08	-21.15 ± 0.15	⋯
ΔG_{elec}	-1.37 ± 0.06	-1.52 ± 0.04	-5.08 ± 0.09	-4.80 ± 0.04	⋯
$\Delta\Delta G^{\circ}_{\text{rstr}}$	10.05 ± 0.03	9.98 ± 0.03	9.61 ± 0.06	⋯	⋯
Sum	-13.05 ± 0.34	-19.05 ± 0.14	7.80 ± 1.32	-4.80 ± 0.15	−3.5[a]
$\Delta G^{\circ}_{\text{b}}$	-8.25 ± 0.37	-14.25 ± 0.21	12.60 ± 1.33	⋯	−7.75[b]

[a]Experimental hydration free energy from Ref. 57.
[b]Experimental binding free energy from Ref. 27.

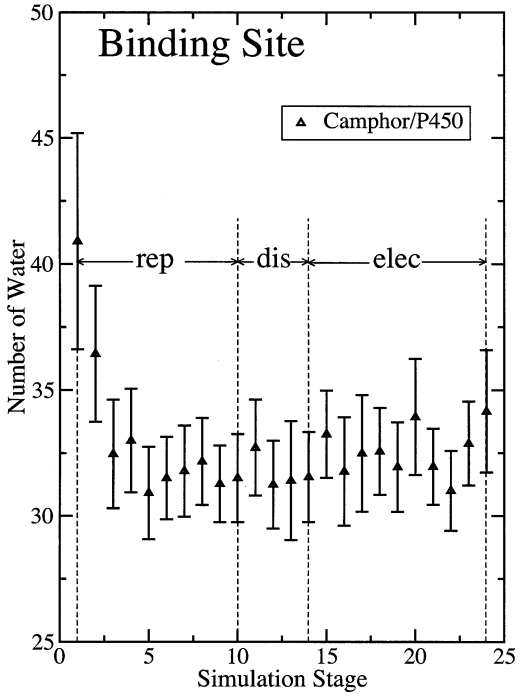

FIG. 2. Shown in the plot is the number of water molecules as camphor binds to cytochrome P450 alchemically. The simulation is divided into 24 stages, starting from noninteracting in stage one and progressing to full-interacting camphor in stage 24. In stages 1–10, the camphor repulsion is turned on gradually. In stages 10–14, the camphor attraction is added. At last, in stages 14–24, the camphor particle charges are turned on. The errors are standard deviation of five complete runs of binding free energy calculations.

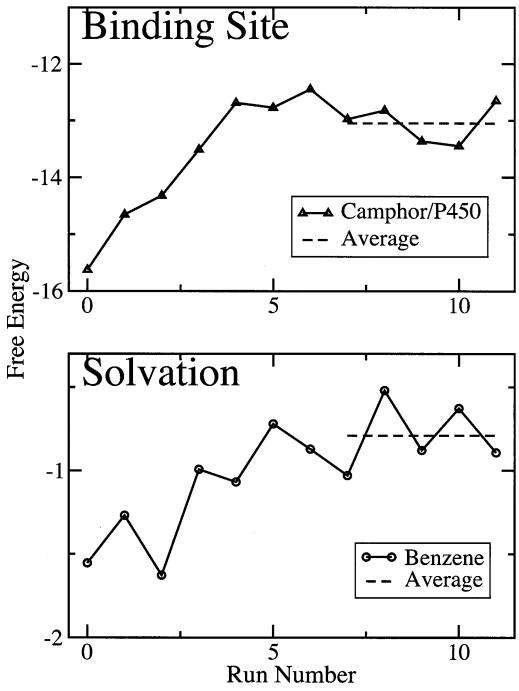

FIG. 3. Shown in the plot is the convergence of free energy calculations. Starting from run 1, each window of the free energy calculations is launched from the end configuration of the previous run. The upper panel shows the binding of camphor to P450. The lower panel shows the solvation of benzene.

pocket. Hence, the repulsive free energy becomes greater and dispersive and electrostatic free energies become more negative.

These results can be intuitively understood. The effect of water in the binding pocket manifests itself as a pressure on the ligand camphor. When the number is fixed, and too large, the pressure makes the binding highly unfavorable. Clearly, free energy simulations with a fixed number of water molecules can be extremely inaccurate in the case of a buried binding site. With GCMC/MD, the final binding free energy is quite close to both the experimental value (−7.75 kcal/mol), and with the results of previous FEP computations including an explicit alchemical transformation on water (−6.96 kcal/mol).[27] The difference between the present and previous values might be due to the force fields for the protein, water, or camphor. The solvation free energy of the camphor molecule indicates that the current model is not optimal. Compared to the experimental value (−3.5 kcal/mol), the calculated solvation free energy (−4.80 kcal/mol) is too favorable. In contrast, the model used in Ref. 27 yields a solvation free energy of −2.61 kcal/mol, which is too unfavorable. Furthermore, a better heme model[55] might improve the free energy values. Nonetheless, despite the differences in force field, both simulations agree that six water molecules are dispelled upon camphor binding and the computed binding free energies are close to the experimental value. It is likely that the good agreement is partly due to some cancellation of errors; the favorable dispersive interactions between the camphor and its surrounding being slightly overestimated both in the bulk solvent and in the binding pocket. In the GCMC/MD method, the number of water molecules in the buried pocket adjusts naturally as a result of the microscopic interactions. In contrast, the number of water to couple with the ligand in the alchemical transformation was required as an input in Ref. 27. Because it may be difficult to know the correct number of water molecules that will be displaced by a ligand *a priori*, the GCMC/MD method is more flexible and convenient.

Last, to address issues of convergence, we conduct a series of free energy runs, the end configurations (for each window) of a previous run serves as the start configurations of the next run. As shown in Fig. 3, for all the solvation and binding free energy calculation, the results approach a stable value after a few runs of calculations. We used the last five runs to calculate the values in all the tables. These results do not provide a formal proof of convergence, though it increases our confidence in the GCMC/MD simulation method.

V. CONCLUSION

In this paper, we have extended the GCMC within the framework of GSBP to compute free energies. We have tested the theory with solvation and binding free energy calculations. The chemical potential of water computed with the GCMC/MD simulations agrees well with the experimental value and previously published data. Furthermore, the water chemical potential computed with alchemical FEP recovers the chemical potential entered in the GCMC. Hydration free energy of benzene shows agreement with the experimental and computed values as well. Our simulation indicates that the hydration of benzene in a 10 Å sphere displaces about

four to five water molecules. From a practical point of view, it is worth noting that the SSBP method,[18,40] which represents the bulk system using a fixed number of solvent molecule and a flexible radius, is simpler in its usage than GCMC/MD. With SSBP, the standard free energy perturbation and thermodynamic integration expressions hold rigorously.[18] Free energy simulations in the bulk are possible with a fixed radius, but then the number of solvent molecule must be allowed to vary according to GCMC. Nonetheless, the GCMC/MD offers an alternative strategy that is advantageous for some circumstances such as poorly accessible binding pockets. We applied the GCMC/MD simulation to compute the binding free energy of camphor to a buried pocket of cytochrome P450. The results show that the GCMC captures the six to seven water molecules expelled upon camphor binding, and the binding free energy values agrees well with experimental data and a previous computational study. The results indicate that coupling MD with GCMC can provide an efficient treatment of hydration in binding processes in finite simulation systems. This treatment becomes particularly useful for MD/FEP performed with a buried binding pocket, where solvent diffusion could be too slow to allow the number of solvent molecules to equilibrate with the bulk region as the ligand is decoupled with its surrounding. Additional advantages might be gained by combining the GCMC/MD algorithm with Hamiltonian replica exchange methodologies.[26] Such methods should be of general use for free energy computations on biological systems.

ACKNOWLEDGMENTS

We thank Hyung-June Woo, Devleena Shivakumar, and Aaron Dinner for helpful discussions. This work was supported by the National Science Foundation through Grant No. MCB-0415784.

[1] N. Huang and M. P. Jacobson, Curr. Opin. Drug Discov. Devel. **10**, 325 (2007).
[2] S. Boresch, F. Tettinger, M. Leitgeb, and M. Karplus, J. Phys. Chem. B **107**, 9535 (2003).
[3] D. Hamelberg and J. A. McCammon, J. Am. Chem. Soc. **126**, 7683 (2004).
[4] H.-J. Woo and B. Roux, Proc. Natl. Acad. Sci. U.S.A. **102**, 6825 (2005).
[5] H. Fujitani, Y. Tanida, M. Ito, G. Jayachandran, C. D. Snow, M. R. Shirts, E. J. Sorin, and V. S. Pande, J. Chem. Phys. **123**, 084108 (2005).
[6] Y. Deng and B. Roux, J. Chem. Theory Comput. **2**, 1255 (2006).
[7] J. Wang, Y. Deng, and B. Roux, Biophys. J. **91**, 2798 (2006).
[8] D. L. Mobley, J. D. Chodera, and K. A. Dill, J. Chem. Phys. **125**, 084902 (2006).
[9] D. L. Mobley, J. D. Chodera, and K. A. Dill, J. Chem. Theory Comput. **3**, 1231 (2007).
[10] D. L. Mobley, A. P. Graves, J. D. Chodera, A. C. McReynolds, B. K. Shoicet, and K. A. Dill, J. Mol. Biol. **371**, 1118 (2007).
[11] J. Hermans and S. Subramaniam, Isr. J. Chem. **27**, 225 (1986).
[12] B. Roux, M. Nina, R. Pomès, and J. C. Smith, Biophys. J. **71**, 670 (1996).
[13] J. Hermans and L. Wang, J. Am. Chem. Soc. **119**, 2707 (1997).
[14] M. K. Gilson, J. A. Given, B. L. Bush, and J. Andrew McCammon, Biophys. J. **72**, 1047 (1997).
[15] M. Berkowitz and J. A. McCammon, Chem. Phys. Lett. **90**, 215 (1982).
[16] A. Warshel and G. King, Chem. Phys. Lett. **121**, 124 (1985).
[17] A. Brünger, C. L. Brooks III, and M. Karplus, Chem. Phys. Lett. **105**, 495 (1984).
[18] D. Beglov and B. Roux, J. Chem. Phys. **100**, 9050 (1994).
[19] J. W. Essex and W. L. Jorgensen, J. Comput. Chem. **16**, 951 (1995).
[20] W. Im, S. Bernèche, and B. Roux, J. Chem. Phys. **114**, 2924 (2000).
[21] T. I. Oprea, G. Hummer, and A. E. García, Proc. Natl. Acad. Sci. U.S.A. **94**, 2133 (1997).
[22] V. Cojocaru, P. J. Winn, and R. C. Wade, Biochim. Biophys. Acta **1770**, 390 (2007).
[23] K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).
[24] H. Fukunishi, O. Watanabe, and S. Takada, J. Chem. Phys. **116**, 9058 (2002).
[25] P. Liu, X. Huang, R. Zhou, and B. J. Berne, J. Phys. Chem. B **110**, 19018 (2006).
[26] Y. Sugita, A. Kitao, and Y. Okamoto, J. Chem. Phys. **113**, 6042 (2000).
[27] V. Helms and R. C. Wade, J. Am. Chem. Soc. **120**, 2710 (1998).
[28] Z. Li and T. Lazaridis, J. Phys. Chem. B **109**, 662 (2005).
[29] Y. Lu, C.-Y. Yang, and S. Wang, J. Am. Chem. Soc. **128**, 11830 (2006).
[30] Z. Li and T. Lazaridis, Phys. Chem. Chem. Phys. **9**, 573 (2007).
[31] C. Barillari, J. Taylor, R. Viner, and J. W. Essex, J. Am. Chem. Soc. **129**, 2577 (2007).
[32] H.-J. Woo, A. R. Dinner, and B. Roux, J. Chem. Phys. **121**, 6392 (2004).
[33] S. Vaitheeswaran, J. C. Rasaiah, and G. Hummer, J. Chem. Phys. **121**, 7955 (2004).
[34] S. Vaitheeswaran, H. Yin, J. C. Rasaiah, and G. Hummer, Proc. Natl. Acad. Sci. U.S.A. **101**, 17002 (2004).
[35] M. Clark, F. Guarnieri, I. Shkurko, and J. Wiseman, J. Chem. Inf. Model. **46**, 231 (2006).
[36] H. Yin, G. Hummer, and J. C. Rasaiah, J. Am. Chem. Soc. **129**, 7369 (2007).
[37] V. Helms and R. C. Wade, Proteins: Struct., Funct., Genet. **32**, 381 (1998).
[38] G. Hummer, Mol. Phys. **105**, 201 (2007).
[39] M. D. Collins, G. Hummer, M. L. Quillin, B. W. Matthews, and S. M. Gruner, Proc. Natl. Acad. Sci. U.S.A. **102**, 16668 (2005).
[40] Y. Deng and B. Roux, J. Phys. Chem. B **108**, 16567 (2004).
[41] P. Grabowski, D. Riccardi, M. Gomez, D. Asthagiri, and L. Pratt, J. Phys. Chem. A **106**, 9145 (2002).
[42] J. G. Kirkwood, J. Chem. Phys. **3**, 300 (1935).
[43] R. Zwanzig, J. Chem. Phys. **22**, 1420 (1954).
[44] A. D. MacKerell, Jr., D. Bashford, M. Bellot, R. L. Dumbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, J. Phys. Chem. B **102**, 3586 (1998).
[45] W. L. Jorgensen, J. Chandrasekhar, and J. D. Madura, J. Chem. Phys. **79**, 926 (1983).
[46] W. L. Jorgensen and D. L. Severance, J. Am. Chem. Soc. **112**, 4768 (1990).
[47] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, J. Comput. Chem. **25**, 1157 (2004).
[48] J. Wang, W. Wang, P. A. Kollman, and D. A. Case, J. Mol. Graphics Modell. **25**, 247 (2006).
[49] A. Jakalian, B. L. Bush, D. B. Jack, and C. I. Bayly, J. Comput. Chem. **21**, 132 (2000).
[50] A. Jakalian, D. B. Jack, and C. I. Bayly, J. Comput. Chem. **23**, 1623 (2002).
[51] A. T. Brunger and M. Karplus, Proteins: Struct., Funct., Genet. **4**, 148 (1988).
[52] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).
[53] A. A. Canutescu, A. A. Shelenkov, and R. L. Dunbrack Jr., Protein Sci. **12**, 2001 (2001).
[54] J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, J. Comput. Phys. **23**, 327 (1977).
[55] F. Autenrieth, E. Tajkhorshid, J. Baudry, and Z. Luthey-Schulten, J. Comput. Chem. **25**, 1613 (2004).
[56] C. C. Chambers, G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, J. Phys. Chem. **100**, 16385 (1996).
[57] R. Pomes, E. Eisenmesser, C. B. Post, and B. Roux, J. Chem. Phys. **111**, 3387 (1999).