

Charge-leveling and proper treatment of long-range electrostatics in all-atom molecular dynamics at constant pH

Jason A. Wallace and Jana K. Shen^{a)}

Department of Chemistry and Biochemistry, University of Oklahoma, Norman, Oklahoma 73019, USA

(Received 19 May 2012; accepted 21 October 2012; published online 13 November 2012)

Recent development of constant pH molecular dynamics (CpHMD) methods has offered promise for adding pH-stat in molecular dynamics simulations. However, until now the working pH molecular dynamics (pHMD) implementations are dependent in part or whole on implicit-solvent models. Here we show that proper treatment of long-range electrostatics and maintaining charge neutrality of the system are critical for extending the continuous pHMD framework to the all-atom representation. The former is achieved here by adding forces to titration coordinates due to long-range electrostatics based on the generalized reaction field method, while the latter is made possible by a charge-leveling technique that couples proton titration with simultaneous ionization or neutralization of a co-ion in solution. We test the new method using the pH-replica-exchange CpHMD simulations of a series of aliphatic dicarboxylic acids with varying carbon chain length. The average absolute deviation from the experimental pK_a values is merely 0.18 units. The results show that accounting for the forces due to extended electrostatics removes the large random noise in propagating titration coordinates, while maintaining charge neutrality of the system improves the accuracy in the calculated electrostatic interaction between ionizable sites. Thus, we believe that the way is paved for realizing pH-controlled all-atom molecular dynamics in the near future. © 2012 American Institute of Physics. [<http://dx.doi.org/10.1063/1.4766352>]

I. INTRODUCTION

Solution pH is an important environmental factor in biological and chemical processes. It affects the stability and function of proteins and nucleic acids by shifting the protonation/deprotonation equilibria of titratable (ionizable) groups. To enable molecular simulations of such proton-coupled phenomena, significant progress has been made in recent years in the development of the so-called constant pH molecular dynamics (CpHMD) technique, which allows the protonation states to fluctuate in response to changes in local chemical environment and solution pH. These methods fall into two main categories based on whether the protonation states are represented by discrete or continuous coordinates. In the discrete approach, Monte-Carlo steps are periodically performed to sample protonation states during molecular dynamics.¹⁻⁵ In the continuous approach, an additional titration degree of freedom is added for every titratable site and the titration degrees of freedom are propagated alongside conformational dynamics.⁶⁻⁸

Among these methods, the continuous pH molecular dynamics (pHMD) method^{7,8} based on λ dynamics,⁹ which will be referred to as CpHMD hereafter, has been successfully applied to pK_a predictions¹⁰ and pH-dependent protein folding¹¹ and conformational dynamics.^{12,13} The first version of the CpHMD technique makes use of the generalized Born (GB) implicit-solvent model for propagating both conformational and titration coordinates and the temperature-based

replica-exchange protocol for accelerated sampling in the coupled conformational and protonation-state space.¹⁴ Recently, the method was extended to conformational sampling in explicit solvent while propagating the titration coordinates via GB calculations, in an attempt to circumvent the conformational bias of implicit-solvent models while retaining the ability to quickly evaluate solvation free energies.¹⁵ This hybrid-solvent scheme, combined with a pH-based replica-exchange protocol,¹⁵ has been shown to offer more realistic pH-dependent conformational sampling and more accurate pK_a prediction for proteins¹⁵ as well as surfactant micelles.¹⁶

Most recently, two groups have reported on the development of CpHMD techniques based solely on explicit-solvent models and testing on small model systems.^{17,18} In these implementations the GB-based single-site CpHMD method⁷ was extended to explicit-solvent simulations by replacing the calculation of solvation forces due to the GB model with deriving forces on real and λ particles from interactions involving explicit water molecules. A noted key ingredient is the alternative functional form (other than $\sin^2\theta$) for the λ variable. However, two major issues have not been addressed. The long-range electrostatics was calculated using truncation with force shift¹⁸ or Ewald summation without accounting for the additional forces on λ particles.¹⁷ Potential artifacts due to the fluctuation in overall net charge were also neglected.

In this work we show that leveling the net charge of the system and properly treating long-range electrostatics are two critical factors in developing all-atom constant pH molecular dynamics techniques. While it has been a standard practice to neutralize the simulation system in explicit-solvent simulations under periodic conditions with particle mesh

^{a)}Present address: Department of Pharmaceutical Sciences, University of Maryland, Baltimore, Maryland 21201, USA. E-mail: jshen@rx.umaryland.edu.

Ewald (PME) treatment of extended electrostatics,¹⁹ it has not been attempted in constant pH simulations.²⁰ This is because in such simulations the net charge of the solute may fluctuate as a result of exchanging proton(s) with solvent. Residual net charge is not a concern for methods that use GB implicit-solvent model for computing forces on λ particles (titration coordinates).^{14,15} However, it needs to be compensated in fully explicit-solvent simulations. Here we present a technique which couples proton titration with simultaneous ionization or neutralization of a co-ion in solution. This approach allows the net charge of the system to remain constant during protonation or deprotonation of the solute. For treatment of long-range electrostatics we apply the generalized reaction field (GRF) method.²¹ Incorporation of the GRF method in the CpHMD framework is straightforward because electrostatic force is calculated from strictly pairwise interactions, unlike the smooth Ewald methods where the force is calculated in part from convolution over the charge interpolation grid.²² Although Ewald-based methods have been traditionally believed to be more accurate, evidence of artifacts due to imposed periodicity emerged.²³ Subsequently, numerous papers have been published comparing the two methods. GRF has been shown to give results comparable to the more computationally expensive Ewald methods in simulations of highly charged proteins,²⁴ RNA,²⁵ small peptide folding,²⁶ small peptides over long timescales,²⁷ and the calculation of protein-folding kinetics from MD simulations.²⁸ Additionally, in a recent pHMD study based on the discrete approach and a mixed-solvent scheme, the GRF method gave more accurate pK_a values than the particle mesh Ewald method.²⁹

We implemented the new method in the existing pHMD module of the CHARMM program and tested it by running titration simulations and obtaining pK_a 's for a series of aliphatic dicarboxylic acids, $\text{HO}_2\text{C}-(\text{CH}_2)_n-\text{CO}_2\text{H}$, where n varies from 3 to 7. This series of molecules has attractive qualities as a test case. There are two interacting titratable sites, so the accurate calculation of the two macroscopic pK_a values resulting from stepwise protonation/deprotonation is non-trivial and offers a stringent test for computational methods. Understanding the pK_a difference between the two is a classical problem dated back to the 1930s when Kirkwood developed analytical theories to describe the electrostatic influence of substituents on the dissociation constants of organic acids.^{30,31} It has continued to fascinate and challenge the theoretical and computational communities until now.^{1,4,32,33} Another reason for choosing the dicarboxylic acids is the small molecular size. Adequate sampling of all relevant conformations should not be an issue. Furthermore, since the experimental data are extrapolated to zero ionic strength,³⁴ potential complexity in calculating the deprotonation energies due to salt screening is absent. Finally, we note that while CpHMD methods are rooted on the λ -dynamics technique⁹ for free energy calculations, classical³⁵ or quantum mechanics/molecular mechanics free energy simulations³⁶ can also be applied to obtain pK_a 's. A major strength of the CpHMD methods is that the protonation states, or more precisely, the fraction of the unprotonated form, of all titratable sites are simultaneously determined in a "regular" molecular dynamics at a specified pH condition. The pK_a value of each group

can be then calculated in the same fashion as in experiment, i.e., by fitting the fraction of the unprotonated form for multiple pH conditions to the Henderson-Hasselbalch equation. In an upcoming work, we will demonstrate the capability of the fully explicit-solvent CpHMD method in determination of pK_a values for proteins.

The rest of the paper is organized as follows. We first present the new method and discuss the implementation. We then test the all-atom CpHMD simulation with the pH-based replica-exchange sampling protocol on the series of dicarboxylic acids. We calculate the pK_a values of the dicarboxylic acids and compare the results with experimental data as well as with two other methods, the GB-based CpHMD and the explicit-solvent CpHMD without co-ions. The new CpHMD method gives the closest agreement with experiment. The average absolute error of the calculated pK_a values is about 0.18 units. We also examine the pH-dependent conformational dynamics and changes in solvent distributions around the dicarboxylic acid. These results suggest that a fully atomistic "pH-stat" will soon become available for molecular dynamics simulations.

II. METHODS

A. Formalism of continuous constant pH molecular dynamics

Based on the λ -dynamics technique,⁹ continuous constant pH molecular dynamics utilizes an extended Hamiltonian to simultaneously propagate spatial (real) and titration (virtual) coordinates. Thus, the total Hamiltonian of the system can be written as

$$\mathcal{H}(\{\mathbf{r}_a\}, \{\boldsymbol{\theta}_i\}) = \sum_a \frac{1}{2} m_a \dot{\mathbf{r}}_a^2 + U^{\text{int}}(\{\mathbf{r}_a\}) + U^{\text{hybr}}(\{\mathbf{r}_a\}, \{\boldsymbol{\theta}_i\}) + \sum_i \frac{1}{2} m_i \dot{\boldsymbol{\theta}}_i^2 + U^*(\{\boldsymbol{\theta}_i\}), \quad (1)$$

where $a = 1, N_{\text{atom}}$ is the atomic index for spatial coordinates, and $i = 1, N_{\text{titr}}$ is the index for the continuous variables $\boldsymbol{\theta}_i$, which is related to the titration coordinate λ_i by $\lambda_i = \sin^2(\boldsymbol{\theta}_i)$. Boundaries are naturally imposed on the titration coordinates through the sine function, where $\lambda_i = 0$ corresponds to the protonated state and $\lambda_i = 1$ corresponds to the deprotonated state. For groups with two competing titration sites such as histidine or those containing a carboxyl group, a second continuous variable is included to facilitate the interconversion between the two tautomers. This is indicated in Eq. (1) by bold $\boldsymbol{\theta}$.

In Eq. (1) the first term is the kinetic energy of the real system (atoms), U^{int} is the internal potential energy, which is independent of titration, and U^* is a biasing potential, which contains three terms. U^{barr} is a harmonic potential to suppress the intermediate values of λ , U^{mod} is a potential of mean force (PMF) function for titrating the model compound, and U^{pH} is the free energy imposed on the protonation equilibrium due to the change in solution pH.

$$U^*(\{\boldsymbol{\theta}_i\}) = \sum_i -U^{\text{mod}}(\boldsymbol{\theta}_i) + U^{\text{barr}}(\boldsymbol{\theta}_i) + U^{\text{pH}}(\boldsymbol{\theta}_i), \quad (2)$$

where U^{pH} is given by

$$U^{\text{pH}}(\lambda_i) = \ln(10)k_B T (\text{p}K_a^{\text{mod}} - \text{pH})\lambda_i, \quad (3)$$

and $\text{p}K_a^{\text{mod}}$ is the experimental $\text{p}K_a$ of a model compound, also known as the reference $\text{p}K_a$. In GB simulations, due to the pairwise form of the solvation energy, U^{mod} for a single titratable site can be described by a quadratic function,

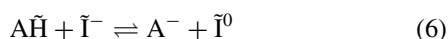
$$U^{\text{mod}}(\theta_i) = A(\sin^2(\theta_i) - B)^2, \quad (4)$$

where A and B are the fitting parameters. For groups with two competing titratable sites, U^{mod} is second order in both λ and x , where the latter represents the tautomeric degree of freedom.⁸ Finally, we consider U^{hybr} , which is a hybrid energy term that couples the dynamics of conformational and titration degrees of freedom. In GB simulations, it is a sum of van der Waals, Coulombic, and the GB electrostatic solvation energies:

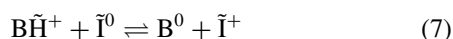
$$U^{\text{hybr}}(\{\mathbf{r}_a\}, \{\theta_i\}) = U^{\text{vdW}}(\{\mathbf{r}_a\}, \{\theta_i\}) + U^{\text{Coul}}(\{\mathbf{r}_a\}, \{\theta_i\}) + U^{\text{GB}}(\{\mathbf{r}_a\}, \{\theta_i\}). \quad (5)$$

B. Titratable co-ions for charge leveling

During proton titration the net charge of the solute fluctuates. This does not present a problem in GB-based CpHMD method. We also demonstrated that the potential artifacts are negligible in the hybrid-solvent scheme where GB model is applied to propagating the titration coordinates.¹⁵ However, a varying net charge is a major issue that needs to be addressed for developing an accurate CpHMD method based on the explicit-solvent model. If neglected, overestimation of electrostatic coupling between ionizable sites may occur, as will be shown in Sec. IV. Here we describe a charge-leveling scheme, which couples proton titration with charging or neutralization of a titratable dummy ion. Specifically, the deprotonation of each acidic site is coupled to the neutralization of a dummy anion and the deprotonation of each basic site is coupled to the ionization of a dummy cation. The dummy ion absorbs the change in charge during titration, thus mimicking the effect of gaining or releasing a proton to solution. Except for being titratable, these dummy ions are treated on the same footing as regular ions and randomly distributed in the simulation box at the beginning of a simulation. Chloride and sodium ions are used in the current implementation for no specific reason other than the availability of the force field parameters and the fact that they present the simplest chemical species that can be used as a charge reservoir. Since there is one dummy ion for each titratable site, we will refer to it as co-ion in the remainder of the paper. Thus, the net reaction under consideration is



for each acidic site, and



for each basic site. Here $\tilde{\text{H}}$ represents the dummy hydrogen atom attached to the titratable site (as CpHMD does not model bond breaking or formation). $\tilde{\text{I}}$ represents the corresponding

co-ion. Since on each side of the reactions only one of the species carries a charge, formation of an artificial contact pair is avoided.

In the CpHMD simulation, we aim to calculate the deprotonation free energy of a titration site in a specific environment (e.g., protein) by making use of a reference free energy, i.e., the free energy for deprotonating a reference (or model) compound in solution. The difference between the two can be obtained by considering classical interactions only, because the quantum mechanical energy of breaking or forming the bond with a proton for a titratable moiety in solution ($\Delta G_{\text{mod}}^{\text{quant}}$) and that embedded in the solute molecule (ΔG^{quant}) are approximately identical.

$$\begin{aligned} \Delta G^{\text{dprot}} - \Delta G_{\text{mod}}^{\text{dprot}} &= \Delta G^{\text{class}} + \Delta G^{\text{quant}} - \Delta G_{\text{mod}}^{\text{class}} - \Delta G_{\text{mod}}^{\text{quant}} \\ &\approx \Delta G^{\text{class}} - \Delta G_{\text{mod}}^{\text{class}}. \end{aligned} \quad (8)$$

With the incorporation of a titratable co-ion, the above free energy difference includes the contribution from the difference between neutralizing the co-ion in the solute environment and in solution. However, since the co-ion remains in solution, the free energy of neutralization will not be significantly affected by the presence of the titrating solute. Therefore,

$$\begin{aligned} \Delta G^{\text{dprot}}(\text{titr}; \text{ion}) - \Delta G_{\text{mod}}^{\text{dprot}}(\text{titr}; \text{ion}) \\ \approx \Delta G^{\text{dprot}}(\text{titr}) - \Delta G_{\text{mod}}^{\text{dprot}}(\text{titr}). \end{aligned} \quad (9)$$

Finally, the above free energy difference is related to the $\text{p}K_a$ shift in reference to the model compound $\text{p}K_a$ value by

$$\Delta G^{\text{dprot}} - \Delta G_{\text{mod}}^{\text{dprot}} = \ln(10)k_B T (\text{p}K_a - \text{p}K_a^{\text{mod}}). \quad (10)$$

The $\text{p}K_a$ can be obtained by running CpHMD titration at several pH conditions, calculating the fraction of the time spent in each protonation state, and fitting the unprotonated fraction to an appropriate titration model (see Sec. III).

C. Generalized reaction field for long-range electrostatics

Now we come back to the hybrid-energy term in the Hamiltonian (Eq. (1)). In the explicit-solvent simulations, $U^{\text{hybr}}(\{\mathbf{r}_a\}, \{\theta_i\})$ describes the van der Waals and Coulomb energies for solute-solute and solute-solvent interactions. In our previous work with the hybrid-solvent scheme, we mentioned large fluctuation in titration coordinates when using interactions with explicit solvent under periodic boundary conditions to calculate forces on λ particles.¹⁵ Detailed analysis revealed that this is a result of neglecting the effects due to long-range electrostatics (data not shown). In the present implementation, we employ the generalized reaction field method²¹ to account for the missing forces resulting from extended electrostatics. Thus, the hybrid energy becomes

$$U^{\text{hybr}}(\{\mathbf{r}_a\}, \{\theta_i\}) = U^{\text{vdW}}(\{\mathbf{r}_a\}, \{\theta_i\}) + U^{\text{Coul}}(\{\mathbf{r}_a\}, \{\theta_i\}) + U^{\text{GRF}}(\{\mathbf{r}_a\}, \{\theta_i\}), \quad (11)$$

where the GRF term is given by^{21,24}

$$U^{\text{GRF}} = -\frac{q_a q_b}{4\pi\epsilon_0\epsilon_{\text{in}}} \left(\frac{0.5C_{\text{RF}}r_{ab}^2}{R_c^3} + \frac{1 - 0.5C_{\text{RF}}}{R_c} \right). \quad (12)$$

Here r_{ab} is the distance between two atoms, q_a and q_b are the respective partial charges, ϵ_0 is the vacuum permittivity, ϵ_{in} is the dielectric constant (typically set to one) within R_c , the cut-off radius. C_{RF} , which governs the magnitude of the reaction field, carries the dependence on ionic strength and is given by

$$C_{\text{RF}} = -\frac{(2\epsilon_{\text{out}} - 2\epsilon_{\text{in}})(1 + \kappa R_c) + \epsilon_{\text{out}}(\kappa R_c)^2}{(\epsilon_{\text{in}} + 2\epsilon_{\text{out}})(1 + \kappa R_c) + \epsilon_{\text{out}}(\kappa R_c)^2}, \quad (13)$$

where ϵ_{out} is the dielectric constant (typically the value for water) of the surrounding medium outside of the cut-off sphere, and κ is the inverse Debye screening length, $\kappa^2 = 2N_A I e^2 / \epsilon_0 \epsilon_{\text{out}} k_B T$, where I is the ionic strength.

III. COMPUTATIONAL DETAILS

A. Simulation protocol

All simulations were carried out using an in-house version of the CHARMM program (version c35b3).³⁷ The explicit-solvent CpHMD method was implemented in the pHMD module. The pH-based replica-exchange protocol¹⁵ was added to the REPDSTR module.¹⁵

The CHARMM22 force field³⁸ was used. The parameters for the dicarboxylic acids were adapted from glu residue and the methylene unit of alkanes. For the explicit-solvent simulations we utilized the modified CHARMM TIP3P water model,³⁹ an updated van der Waals radius for the sodium ion,⁴⁰ and a modified sodium-chloride van der Waals interaction distance to reduce sodium chloride contact-ion pair formation.⁴¹

In all simulations, the SHAKE algorithm was applied to the bonds and angles involving hydrogen atoms to allow a 2 fs time step. The titration degrees of freedom were propagated using the Langevin algorithm with a collision frequency of 5 ps^{-1} . The simulation temperature was 300 K. The mass of the fictitious θ particles was set to 10 amu consistent with previous work.^{7,8} To accelerate sampling convergence the pH-based replica-exchange protocol was applied.¹⁵ In this protocol, the pH condition ranges from 3 to 7 with an interval of 0.5 units. Exchanges between adjacent replicas were attempted every 500 molecular dynamics steps (or 1 ps). Simulation of each replica lasted 5000 exchange steps or 5 ns.

For the GB-based simulations, the GBSW (generalized-Born model with simple switching) model⁴² was used with the atomic input radii of Nina *et al.*⁴³ and a surface tension coefficient of $5 \text{ cal mol}^{-1} \text{ \AA}^{-2}$. In the GB-based simulations, conformational dynamics was propagated via the Langevin algorithm with a collision frequency of 5 ps^{-1} . Non-bonded interactions were truncated at a cut-off radius of 20 \AA using a switching function. The ionic strength was set to zero.

For the explicit-solvent simulations, a cubic water box with 30 \AA edges was used. Any water molecule within 2.4 \AA of the solute was deleted. Two titratable chloride ions were added as co-ions to maintain a total net charge of -2 . Two sodium ions were added as counterions to achieve charge neutrality in the simulation system. In the simulation of the model compound, only one co-ion and one counterion were present. The initial locations of these ions were set randomly. The coordinates of the water molecules and ions were sub-

jected to energy minimization with the solute restrained prior to starting the CpHMD simulation. Simulations were carried out with periodic boundary conditions at ambient temperature and pressure using the Hoover thermostat⁴⁴ and Langevin piston pressure-coupling algorithm⁴⁵ as in the previous hybrid-solvent CpHMD simulations.¹⁵ The non-bonded interactions were truncated at 14 \AA , beyond which the electrostatic effects were treated by the GRF method. In the GRF term, ϵ_{in} was set to 1.0 and ϵ_{out} was set to 80.0. The ionic strength in C^{RF} was set to zero consistent with experimental data.

B. Model compound

For no particular reason other than convenience, we used azelaic acid, which has the longest carbon chain, as the model compound. Only one carboxylic site was allowed to titrate and the other site was kept protonated. The model $\text{p}K_a$, which is the microscopic value $\text{p}K_{1a}/\text{p}K_{1b}$, was set to 4.85, based on the experimental macroscopic $\text{p}K_1$ of 4.55³⁴ and the relationship in Eq. (18). Because of the fast exchange of the two carboxylate oxygens, the single-site titratable model was used.⁷ To determine the potential of mean force for model titration, thermodynamic integration $\Delta G = \int \langle dU/d\theta \rangle_\theta$ was applied, where the average force, $\langle dU/d\theta \rangle$, was obtained at θ values of 0.2, 0.4, 0.6, 0.7854, 1.0, 1.2, and 1.4 from 500 ps simulations in explicit solvent or 100 ps simulations with the GB model. Convergence was verified by examining the cumulative values of the average forces. The average force was then fit to the following equation:

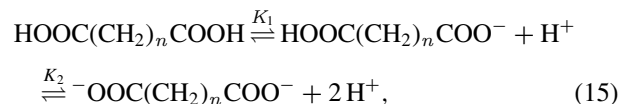
$$\langle dU/d\theta \rangle = 2A(\sin^2\theta - B)\sin(2\theta), \quad (14)$$

to obtain the A and B parameters (see Sec. IV). The same procedure was used to derive the PMF for the co-ion in explicit solvent.

C. Data analysis

1. Calculation of macroscopic $\text{p}K_a$ values for dicarboxylic acids

The deprotonation/protonation of a dicarboxylic acid involves four microscopic equilibria. Since the two carboxylic sites are chemically equivalent, $k_{1a} = k_{1b}$ and $k_{2a} = k_{2b}$. Thus, the titration of dicarboxylic acids can be described by two macroscopic equilibria,



where K_1 and K_2 are the macroscopic constants. By expressing K_1 and K_2 in terms of the population of the four microstates (Figure 1), one arrives at the following relationships:⁴⁶

$$K_1 = 2k_{1a} = 2k_{1b}, \quad (16)$$

$$K_2 = k_{2a}/2 = k_{2b}/2. \quad (17)$$

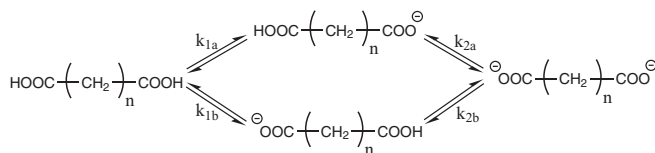


FIG. 1. Microscopic titration of dicarboxylic acids. k_{1a} , k_{1b} , k_{2a} , and k_{2b} are the equilibrium constants.

Thus, the macro- and microscopic $\text{p}K_a$ values are linked to each other by a simple additive shift:

$$\text{p}K_1 = \text{p}K_{1a} - \log 2, \quad (18)$$

$$\text{p}K_2 = \text{p}K_{2a} + \log 2. \quad (19)$$

To facilitate comparison with experiment, we will calculate and discuss the macroscopic $\text{p}K_a$'s in the remainder of the article. Following Ullmann,⁴⁶ we can write down the fraction of deprotonated as well as singly- and doubly protonated states as

$$P_0 = 1/Z, \quad (20)$$

$$P_1 = 10^{\text{p}K_2 - \text{pH}}/Z, \quad (21)$$

$$P_2 = 10^{\text{p}K_1 + \text{p}K_2 - 2\text{pH}}/Z, \quad (22)$$

where Z is the partition function,

$$Z = 1 + 10^{\text{p}K_2 - \text{pH}} + 10^{\text{p}K_1 + \text{p}K_2 - 2\text{pH}}. \quad (23)$$

Consequently, the average number of bound protons becomes^{4,46}

$$\langle P \rangle = P_1 + 2P_2 = \frac{10^{\text{p}K_2 - \text{pH}} + 2 \times 10^{\text{p}K_1 + \text{p}K_2 - 2\text{pH}}}{Z}. \quad (24)$$

Consistent with experiment, the two macroscopic $\text{p}K_a$'s are determined by fitting $\langle P \rangle$ obtained from the simulation to the equation above.

2. Cross-correlation analysis

The correlation between the two protonation events in a dicarboxylic acid can be quantified using the cross-correlation function of the λ values. The normalized cross-correlation function between two time series of properties x and y , shifted by an offset Δr , is defined by

$$R_{xy}(\Delta r) = \frac{1}{N} \sum_i \frac{[x(i) - \bar{x}][y(i + \Delta r) - \bar{y}]}{\sigma_x \sigma_y}, \quad (25)$$

where \bar{x} and \bar{y} are the population means and σ is the standard deviation. R_{xy} ranges from -1 (completely anti-correlated) to $+1$ (completely correlated). N is the total number of data points.

IV. RESULTS AND DISCUSSION

A. Potential of mean force for model compound titration

We found that the PMF for titrating the model compound and the co-ion in explicit solvent can be accurately described

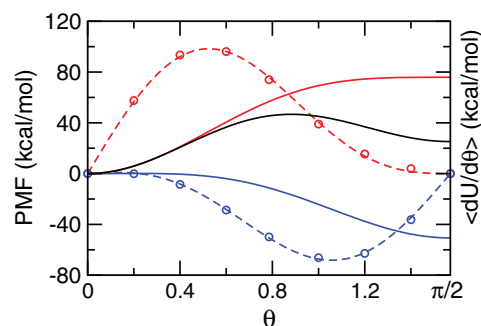


FIG. 2. Potential of mean force (PMF) for titrating the model compound and co-ion in explicit solvent. Solid curves are the PMFs for the deprotonation reaction of the model azelaic acid (blue), neutralization reaction of the chloride co-ion (red), and the sum of the two (black). The average force at each θ value is shown as circle and the fitting function is shown as dashed curve. Integration of the fitting function gives the PMF.

by a quadratic function as in GB simulations. Figure 2 shows a perfect fitting of the average force along the θ coordinate to Eq. (14) for the first deprotonation reaction of azelaic acid and the chloride co-ion. Table I summarizes the resulting parameters as well as the corresponding free energies in comparison to those from the GB-based simulation. It can be seen that the deprotonation free energy with the GB model is about 10 kcal/mol (or 20%) smaller than with the explicit-solvent model, indicating that the charged form is less favored by the GB model, consistent with our observation that solvation energies in GB simulations are underestimated.⁴⁷

B. Efficiency of replica exchange and sampling convergence

With the model PMF in place, we simulated the titration of azelaic acid allowing both carboxyl sites to protonate or deprotonate. The simulation continued for 5000 replica-exchange steps (or 5 ns) with pH conditions ranging from 3 to 7 with an interval of 0.5 unit. The flow of replicas in the control parameter space, which in this case is the pH, is critical for the efficiency of a replica-exchange simulation as it is directly linked to the mean first passage time for crossing the energy barriers in the state space.⁴⁸ We picked three replicas and examined the pH conditions they were subjected to during the simulation. Figure 3(a) shows that all three replicas had a full walk-through in the pH ladder. The acceptance probability for the pH exchange is above 50% for all neighboring pairs of pH conditions (Figure 3(b)). It is also worthwhile to notice that the exchange ratio is smaller in the middle of the

TABLE I. Parameters in the model PMF and the corresponding free energy in the explicit-solvent and GB-based CpHMD simulations.^a

Solvent	Parameters		ΔG (kcal/mol)
	A	B	
Explicit	-56.16	0.05	-50.54
GB	-62.39	0.18	-39.93

^aModel PMF corresponds to the blue curve in Figure 2.

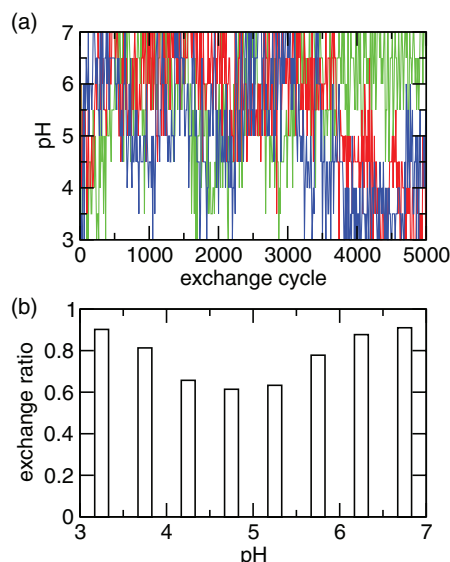


FIG. 3. Efficiency of replica exchange in the simulation of azelaic acid. (a) The pH condition visited by the replica after each exchange cycle. Three randomly chosen replicas are shown in blue, red, and green. (b) The exchange probability between adjacent pH replicas.

pH range. This is where doubly, singly protonated, and fully unprotonated states coexist (see Sec. IV C 1).

To examine the convergence of the protonation-state sampling, we plotted the unprotonated fraction of the two carboxylic groups as a function of replica-exchange steps (Figure 4). At the beginning the two S values deviate from each other significantly. However, as the simulation proceeds, they move closer and become nearly identical at the end of 5000 exchange steps (or 5 ns). This is required because the two carboxylic groups have identical microscopic pK_a values due to the equivalent chemical environment in solution.

C. Titration simulation of azelaic acid

1. Macroscopic pK_a values of azelaic acid

We first calculated the fractions of the population having two, one, and zero protons at each pH (Figure 5(b)). At pH 3 the azelaic acid is doubly protonated, while at pH 7 it is fully deprotonated. In the intermediate pH range between 4 and 6, these two states coexist with the singly protonated state,

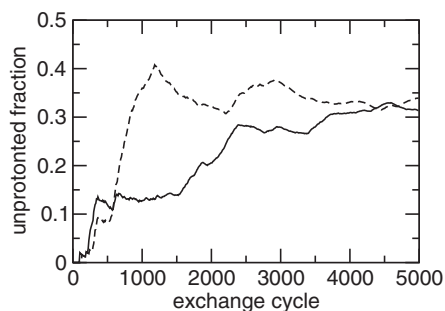


FIG. 4. Convergence of the CpHMD simulation of azelaic acid in explicit solvent with charge leveling. The unprotonated fraction for the two carboxylic groups at pH 4.5 as a function of the replica-exchange steps.

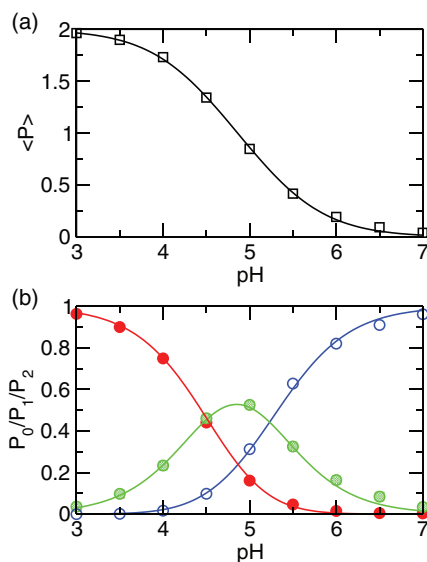


FIG. 5. Simulated titration of azelaic acid. (a) The average number of bound protons at each pH. Best fit to the data using Eq. (24) is shown as solid curve. The resulting pK_a 's are 4.4 and 5.3. (b) Fraction of the population having two (P_2 , red), one (P_1 , green), and zero (P_0 , blue) bound protons at each pH. Best fits to Eqs. (20)–(22) are shown as solid curves. The two pK_a 's obtained in all three fittings are within 0.2 units from those obtained in (a).

indicating that the two carboxylic groups are simultaneously titrating. From these fractions, the total average protonation, $\langle P \rangle$, can be calculated using Eq. (24). Figure 5(a) shows that fitting of $\langle P \rangle$ at each pH to Eq. (24) is nearly perfect (correlation coefficient, >0.999). The resulting macroscopic pK_a 's are 4.4 and 5.3, with a difference of 0.8 unit between them (Table II). These numbers are within 0.15 unit from the respective experimental values of 4.55, 5.41, and 0.86. Note that while the agreement for pK_1 is expected because the first ionization is described by the model PMF, it is very encouraging to see that the predicted pK_a difference is only 0.06 lower than experiment, suggesting that the new method is able to quantitatively capture the electrostatic influence on the protonation of the two carboxylic groups. For comparison we also calculated the pK_a 's using GB- and explicit-solvent simulations without charge leveling. While the pK_a difference in the GB simulation is 0.98, 0.18 unit larger than experiment, the simulation without charge leveling overestimates the pK_a difference by 2 units.

2. pH-dependence of the conformational states, solvent and counterion distributions

Examination of the conformational states at different pH conditions reveals that azelaic acid prefers states with larger end-to-end distance when it is doubly deprotonated (at pH 7) as compared to the doubly protonated form (at pH 3). At pH 7 the distribution of the distance between the two carboxyl carbons has greater intensity at 9 and 10 Å but reduced intensity at distances below 8.5 Å (Figure 6(a)). This is expected since there is electrostatic repulsion between the charged carboxylate groups at pH 7.

To investigate the response of solvent and counterions to different pH conditions, we calculated the radial distribution

TABLE II. Experimental and calculated pK_a 's of a series of dicarboxylic acids.

Acid		Expt. ^a	E+CL ^b	GB ^c	E-CL ^d
	n^e				
			pK_1		
Azelaic	7	4.55	4.4	4.20	4.2
Suberic	6	4.52	4.6	4.19	4.3
Pimelic	5	4.48	4.5	4.17	4.2
Adipic	4	4.42	4.7	4.04	4.1
Glutaric	3	4.34	4.0	3.57	3.7
AAE			0.18	0.43	0.38
			pK_2		
Azelaic	7	5.41	5.3	5.18	7.1
Suberic	6	5.40	5.2	5.25	7.1
Pimelic	5	5.42	5.3	5.25	7.0
Adipic	4	5.41	5.4	5.42	7.1
Glutaric	3	5.42	5.0	5.05	6.6
AAE			0.18	0.19	1.55
			ΔpK_a^f		
Azelaic	7	0.86	0.8	0.98	2.9
Suberic	6	0.88	0.6	1.06	2.8
Pimelic	5	0.94	0.8	1.08	2.8
Adipic	4	0.99	0.7	1.38	2.9
Glutaric	3	1.08	1.0	1.48	2.9
AAE			0.19	0.24	1.9

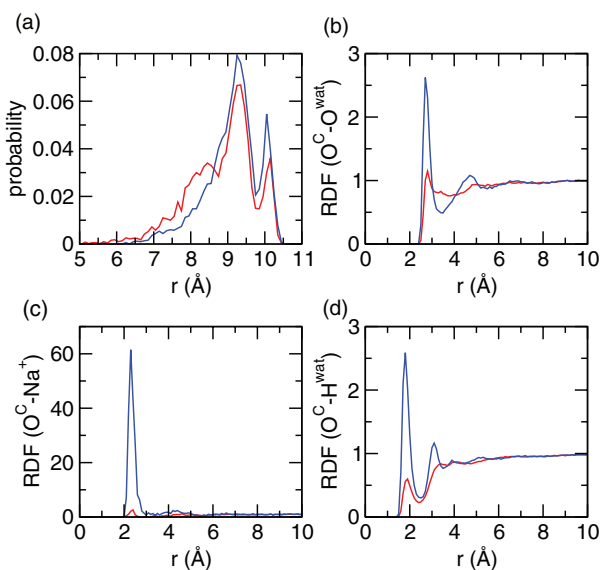
^aExperimental data are taken from Ref. 34.^bE+CL refers to the simulations based on explicit solvent with charge leveling.^cGB refers to the simulations based on the GB implicit-solvent model.^dE-CL refers to the simulations based on explicit solvent without charge leveling.^e n refers to the number of methylene groups in the dicarboxylic acid $\text{HO}_2\text{C}-(\text{CH}_2)_n-\text{CO}_2\text{H}$.^fIn all simulations, the microscopic pK_{1a} of azelaic acid (4.85) was used as the model pK_a (see Sec. III).

FIG. 6. pH-dependent conformation, solvent, and counterion distribution around azelaic acid. (a) Probability distribution of the distance between the two carboxylate carbons at pH 3 and 7. (b) Radial distribution function (RDF) between the carboxylate oxygen and water oxygen. (c) RDF between the carboxylate oxygen and sodium counterion. (d) RDF between the carboxylate oxygen and water hydrogen. Data obtained at pH 3 and 7 are colored red and blue, respectively.

function (RDF) involving the carboxylate oxygens of azelaic acid. At pH 7 there is a dramatic increase in the intensity of the first peak in the RDF between the carboxylate-oxygen and water-oxygen at 3 Å (Figure 6(b)), and in the RDF between carboxylate-oxygen and water-hydrogen at 2 Å (Figure 6(c)). This is due to the formation of hydrogen bond between the carboxylate group and water, whereby water acts as a hydrogen-bond donor, to solvate the charged azelaic acid. At pH 7 a strong peak appears in the RDF between the carboxylate-oxygen and sodium ion near 2.5 Å, suggesting that the counterion is attracted to the carboxylate group upon deprotonation (Figure 6(d)), consistent with the observation of ionization-induced condensation of counterions around the titratable site made in the CpHMD simulation of micelles.¹⁶

3. Distribution of charge-leveling co-ions around solute

A major assumption we made with the charge-leveling co-ions is that they remain in solution such that only the titration free energy of the solute is of concern (see Eq. (9)). To verify that it is the case, we calculated the minimum distance between the chloride co-ions and azelaic acid (Figure 7(a)). The probability distribution is relatively flat, indicating the absence of accumulation of co-ions around the titration sites. The probability elevates around 10 Å, which is a distance where the effect of electrostatic interactions on pK_a values becomes negligible.⁸ Furthermore, the distribution remains virtually the same at pH conditions of 3, 5, and 7, where the azelaic acid is neutral (pH 3) or ionized (charge of -1 at pH 5 and -2 at pH 7), suggesting that the co-ions do not respond to the change in the ionization states. For comparison, we also examined the minimum distance between the sodium counterions and azelaic acid. As shown in Figure 7(b), there is a significant population of counterions in the vicinity of azelaic acid as it becomes ionized at pH 5 and 7. This is expected because of the electrostatic attraction.

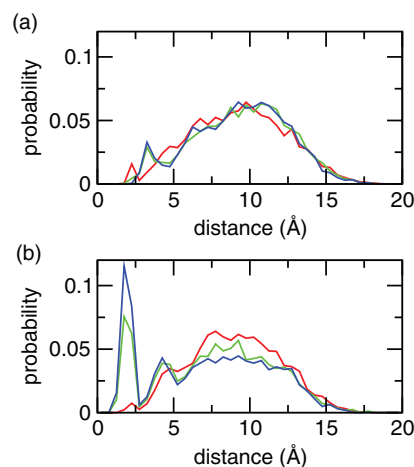


FIG. 7. Probability distribution of the minimum distance from the azelaic acid to (a) any co-ion and (b) any counterion. Three pH conditions are shown with the azelaic acid being doubly protonated (pH 3, red), singly protonated (pH 5, green), or fully deprotonated (pH 7, blue).

D. Titration simulations of a series of dicarboxylic acids

1. Accuracy of the calculated pK_a values

Table II summarizes the calculated and experimental pK_a values for dicarboxylic acids with 3–7 intervening methylene groups. For comparison, the values obtained with the GB-based method and the method without charge leveling (E-CL) are also listed. Examining the first macroscopic pK_a 's, we notice that the change in the experimental pK_a value as n decreases from 7 to 3 is very small (0.2 units), suggesting that the strength of electrostatic interaction does not significantly change. This can be attributed to the relatively large separation distance between the two carboxylic groups. Nevertheless, the trend is decreasing, which is captured by all three methods. We also notice that the error slightly increases with the decreasing number of intervening methylene groups (n) and becomes the largest for glutaric acid in all three methods. With the method E+CL, the absolute error increases from 0.15 for azelaic acid, to 0.34 for glutaric acid. With the GB method, the error is all negative and the absolute value increases from 0.35 for azelaic acid, to 0.77 for glutaric acid. Thus, the method E+CL performs slightly better. Interestingly, with the method E-CL, the error is also negative and the absolute value increases from 0.35 for azelaic acid, to 0.64 for glutaric acid, which is comparable to the GB method. We attribute the most pronounced error for glutaric acid ($n = 3$) in all three methods to the neglect of quantum effects. As the number of intervening methylene units decreases, the electronic-structure environment of the carboxylic groups starts to deviate from those in the model compound ($n = 7$). The classical treatment (Eq. (8)) is no longer very accurate. Similar observation was made in an early work by Potter, Gilson, and McCammon.³² Interestingly, the quantum effects tend to stabilize the doubly charged form, thereby reducing the magnitude of the electrostatic coupling due to charge-charge repulsion.

Calculation of the second pK_a of dicarboxylic acids provides a stringent test for the accuracy of modeling the electrostatic interaction between the two ionizable sites. Both E+CL and GB methods show only slight underestimation with the largest absolute errors of 0.42 and 0.37, respectively. By contrast, the method E-CL results in large overestimation with errors of 1.2–1.7 units, indicating that the electrostatic repulsion between the two charged carboxylic groups is grossly overestimated due to the uncompensated charges. This comparison provides strong evidence for the necessity of maintaining charge neutrality in the simulation system.

The next quantity to be compared is the difference between the first and second pK_a values, ΔpK_a . The trend follows as before. Both the E+CL and GB methods give very accurate predictions with the average absolute errors of 0.19 and 0.24, respectively. By contrast, the average absolute error is 1.9 for the E-CL method. It is worth noting that the difference is slightly underestimated in all simulations with the E+CL method, while it is slightly overestimated in all the GB simulations. There are two sources for the difference between the two pK_a 's. The first source arises from the so-called statistical factor due to the fact that the first deprotonation re-

action involves one of the two protons, while the second reaction involves only one proton (Eq. (15)).³⁰ It follows that $K_1 = 4K_2$ in the absence of interactions between the two titratable sites.³⁰ Thus, the statistical factor results in 0.6 units for ΔpK_a . The second source for ΔpK_a , which is more difficult to accurately account for, is electrostatic in nature. Deprotonation of the first carboxylic group imposes an electrostatic penalty for the deprotonation of the second group, and causes ΔpK_a to be greater than 0.6 units. Thus, it is also reasonable to expect that the instantaneous protonation states of the two equivalent carboxylic groups should be anti-correlated with one another. The ionization of one carboxylic group should favor the protonation of the other at an intermediate pH that allows both groups to be protonated. We will discuss this topic next.

2. Coupling between titration events

The difference between the two pK_a values for dicarboxylic acids is due to the anticorrelation between the two protonation events, since deprotonation of the first carboxylic group imposes an electrostatic penalty for the deprotonation of the second group. To understand the extent of the coupling between the two protonation events, we calculated the normalized cross-correlation function based on the λ values of two titrating sites in glutaric acid (Eq. (25)). Glutaric acid is a good example because it has the strongest coupling among the series of dicarboxylic acids and ΔpK_a was calculated with good accuracy using the E+CL method. For comparison we also calculated the cross correlation using the data from the GB and the explicit-solvent simulation without co-ion. The cross-correlation at zero delay time correlates with the accuracy of the calculated pK_a difference (Figure 8). There is a strong anti-correlation between the titration events in the simulation without charge leveling, which severely overestimates ΔpK_a . The degree of anti-correlation is reduced and the accuracy of the calculated ΔpK_a is improved with the GB model. In the simulation with the E+CL method the cross-correlation is virtually nonexistent while ΔpK_a is the most accurate. Thus, our data indicate that overestimation of the electrostatic repulsion between the ionized carboxylate groups in the GB and the explicit-solvent simulation without

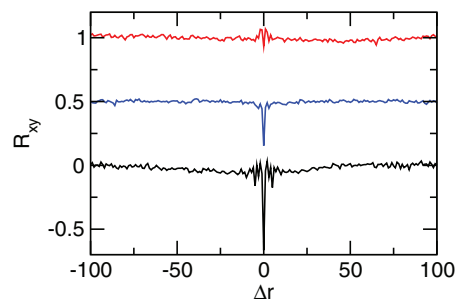


FIG. 8. Correlated titration of the two carboxylic groups in glutaric acid. R_{xy} is the normalized cross-correlation function based on the λ values of two titrating sites at pH 4. Δr is the offset in the exchange cycles. Results based on the GB simulation, explicit-solvent simulation with and without charge leveling are shown in blue, red, and black, respectively. For clarity the baseline for the first two sets of data is shifted.

co-ions exaggerates the coupling between the two titration events leading to an overestimation in the pK_a difference.

V. CONCLUSION

We have developed an all-atom CpHMD technique that makes use of the generalized reaction field for treatment of long-range electrostatics and a charge-leveling scheme that maintains the solute net charge constant by coupling titration to the ionization or neutralization of co-ions serving as a charge reservoir. We tested the new method on a series of aliphatic dicarboxylic acids with the number of intervening methylene units ranging from 3 to 7. We compared the calculated macroscopic pK_a values with those resulting from the GB-based simulation as well as the simulation without charge-leveling. The all-atom method with charge leveling delivers the most accurate pK_a values with the average absolute error being only 0.18 units for both macroscopic pK_a values. This performance is slightly better than the GB-based method, which gives the average absolute errors of 0.43 and 0.19 units for the respective first and second pK_a values. In contrast, the simulation without charge leveling overestimates the second pK_a value by an average of 1.55 units. Analysis reveals that the electrostatic repulsion between the two ionized carboxyl groups is too large, leading to exaggerated coupling between the two titration events and consequently an overestimation of the second pK_a . Our results demonstrate that proper treatment of long-range electrostatics and maintaining charge neutrality of the system are both critical. While the former removes large random noise in propagating titration coordinates, the latter corrects for systematic deviations from experiment, suggesting that the way is paved for realizing pH-controlled molecular dynamics in the near future. Nevertheless, further testing and improvement are needed. One major issue is related to potential artifacts of co-ions. Although in the current simulations, clustering of co-ions and interactions with the solute and counterions are not observed, such artifacts may arise as the number of titratable sites and hence co-ions becomes very large. Another issue is related to sampling convergence. The additional ions may further slow down convergence for explicit-solvent simulations. Work is underway to explore these topics using protein systems.

ACKNOWLEDGMENTS

J.K.S. is indebted to Dr. Charles L. Brooks III for support during the development of the GB-based CpHMD method which laid a foundation for this work. We acknowledge National Science Foundation (MCB-1054547) and National Institutes of Health (RO1GM098818) for financial support.

¹A. M. Baptista, V. H. Teixeira, and C. M. Soares, *J. Chem. Phys.* **117**, 4184 (2002).

²R. Bürgi, P. A. Kollman, and W. F. van Gunsteren, *Proteins* **47**, 469 (2002).

³J. Mongan, D. A. Case, and J. A. McCammon, *J. Comput. Chem.* **25**, 2038 (2004).

⁴M. Dlugosz and J. M. Antosiewicz, *Chem. Phys.* **302**, 161 (2004).

⁵Y. Meng and A. E. Roitberg, *J. Chem. Theory Comput.* **6**, 1401 (2010).

⁶U. Börjesson and P. H. Hünenberger, *J. Chem. Phys.* **114**, 9706 (2001).

⁷M. S. Lee, F. R. Salsbury, Jr., and C. L. Brooks III, *Proteins* **56**, 738 (2004).

⁸J. Khandogin and C. L. Brooks III, *Biophys. J.* **89**, 141 (2005).

⁹X. Kong and C. L. Brooks III, *J. Chem. Phys.* **105**, 2414 (1996).

¹⁰J. A. Wallace, Y. Wang, C. Shi, K. J. Pastoor, B.-L. Nguyen, K. Xia, and J. K. Shen, *Proteins* **79**, 3364 (2011).

¹¹J. Khandogin, J. Chen, and C. L. Brooks III, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18546 (2006).

¹²J. Khandogin and C. L. Brooks III, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 16880 (2007).

¹³J. Khandogin, D. P. Raleigh, and C. L. Brooks III, *J. Am. Chem. Soc.* **129**, 3056 (2007).

¹⁴J. Khandogin and C. L. Brooks III, *Biochemistry* **45**, 9363 (2006).

¹⁵J. A. Wallace and J. K. Shen, *J. Chem. Theory Comput.* **7**, 2617 (2011).

¹⁶B. H. Morrow, Y. Wang, J. A. Wallace, P. H. Koenig, and J. K. Shen, *J. Phys. Chem. B* **115**, 14980 (2011).

¹⁷S. Donnini, F. Tegeler, G. Groenhof, and H. Grubmüller, *J. Chem. Theory Comput.* **7**, 1962 (2011).

¹⁸G. B. Goh, J. L. Knight, and C. L. Brooks III, *J. Chem. Theory Comput.* **8**, 36 (2012).

¹⁹S. B. Dixit and C. Chipot, *J. Phys. Chem. A* **105**, 9795 (2001).

²⁰J. A. Wallace and J. K. Shen, *Methods Enzymol.* **466**, 455 (2009).

²¹I. G. Tironi, R. Sperb, P. E. Smith, and W. F. van Gunsteren, *J. Chem. Phys.* **102**, 5451 (1995).

²²U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, L. Hsing, and L. G. Pedersen, *J. Chem. Phys.* **103**, 8577 (1995).

²³P. H. Hünenberger and J. A. McCammon, *Biophys. Chem.* **78**, 69 (1999).

²⁴R. Gargallo, P. H. Hünenberger, F. X. Avilés, and B. Oliva, *Protein Sci.* **12**, 2161 (2003).

²⁵M. Nina and T. Simonson, *J. Phys. Chem. B* **106**, 3696 (2002).

²⁶A. Baumketner and J.-E. Shea, *J. Phys. Chem. B* **109**, 21322 (2005).

²⁷L. Monticelli, C. Simões, L. Belvisi, and G. Colombo, *J. Phys. Condens. Matter* **18**, S329 (2006).

²⁸A. Robertson, E. Luttmann, and V. S. Pande, *J. Comput. Chem.* **29**, 694 (2008).

²⁹M. Machuqueiro and A. M. Baptista, *Proteins* **72**, 289 (2008).

³⁰J. G. Kirkwood and F. H. Westheimer, *J. Chem. Phys.* **6**, 506 (1938).

³¹F. H. Westheimer and J. G. Kirkwood, *J. Chem. Phys.* **6**, 513 (1938).

³²M. J. Potter, M. K. Gilson, and J. A. McCammon, *J. Am. Chem. Soc.* **116**, 10298 (1994).

³³D. J. Price, J. D. Roberts, and W. L. Jorgensen, *J. Am. Chem. Soc.* **120**, 9672 (1998).

³⁴R. Gane and C. K. Ingold, *J. Chem. Soc.* **1931**, 2153.

³⁵T. Simonson, J. Carlsson, and D. A. Case, *J. Am. Chem. Soc.* **126**, 4167 (2004).

³⁶S. C. L. Kamerlin, M. Haranczyk, and A. Warshel, *J. Phys. Chem. B* **113**, 1253 (2009).

³⁷B. R. Brooks, C. L. Brooks III, A. D. Mackerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartles, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. K. T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus, *J. Comput. Chem.* **30**, 1545 (2009).

³⁸A. D. Mackerell, Jr., M. Feig, and C. L. Brooks III, *J. Comput. Chem.* **25**, 1400 (2004).

³⁹W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.* **79**, 926 (1983).

⁴⁰S. Y. Noskov and B. Roux, *J. Mol. Biol.* **377**, 804 (2008).

⁴¹Y. Luo and B. Roux, *J. Phys. Chem. Lett.* **1**, 183 (2010).

⁴²W. Im, M. S. Lee, and C. L. Brooks III, *J. Comput. Chem.* **24**, 1691 (2003).

⁴³M. Nina, D. Beglov, and B. Roux, *J. Phys. Chem. B* **101**, 5239 (1997).

⁴⁴W. G. Hoover, *Phys. Rev. A* **31**, 1695 (1985).

⁴⁵S. E. Feller, Y. Zhang, R. W. Pastor, and B. R. Brooks, *J. Chem. Phys.* **103**, 4613 (1995).

⁴⁶G. M. Ullmann, *J. Phys. Chem. B* **107**, 1263 (2003).

⁴⁷Y. Wang, J. A. Wallace, P. H. Koenig, and J. K. Shen, *J. Comput. Chem.* **32**, 2348 (2011).

⁴⁸W. Nadler and U. H. E. Hansmann, *Phys. Rev. E* **75**, 026109 (2007).