Molecular simulation with variable protonation states at constant *p*H

Harry A. Stern
Department of Chemistry, University of Rochester, Rochester, New York 14627

(Received 13 February 2007; accepted 26 March 2007; published online 30 April 2007)

A new method is presented for performing molecular simulations at constant *p*H. The method is a Monte Carlo procedure where trial moves consist of relatively short molecular dynamics trajectories, using a time-dependent potential energy that interpolates between the old and new protonation states. Conformations and protonation states are sampled from the correct statistical ensemble independent of the trial-move trajectory length, which may be adjusted to optimize efficiency. Because moves are not instantaneous, the method does not require the use of a continuum solvation model. It should also be useful in describing buried titratable groups that are not directly exposed to solvent, but have strong interactions with neighboring hydrogen bond partners. The feasibility of the method is demonstrated for a simple test case: constant-*p*H simulations of acetic acid in aqueous solution with an explicit representation of water molecules. © *2007 American Institute of Physics*. [DOI: 10.1063/1.2731781]

I. INTRODUCTION

A fundamental property of many systems in chemistry and biology is the ability to exchange protons with the environment. In particular, the structure and function of many proteins depends strongly on the protonation state of titratable amino acid residues, as demonstrated by *p*H dependence of stability or activity.[1–13] Over the last decade or so several molecular simulation methods have been proposed in which protonation states are variable and the *p*H is a fixed parameter. These methods have recently been reviewed by Mongan and Case.[14]

The earliest approach is due to Mertz and Pettitt,[15] who treated the protonation state as an additional continuous degree of freedom, assigned it a fictitious kinetic energy, and incorporated it into an extended Lagrangian, as is done in Car-Parrinello[16] or Nosé-Hoover dynamics.[17–20] Similar methods have been proposed by Börjesson and Hünenberger[21,22] as well as Brooks and co-workers, who added a restraining potential to reduce simulation time spent in unphysical fractional protonation states.[23–25]

In a second category of methods, the system is restricted to physically meaningful, discrete protonation states. Ordinary molecular dynamics is performed; periodically, Monte Carlo moves between different protonation states are attempted. In the methods of Baptista and co-workers,[26–28] Mongan and Case,[14] and Antosiewicz and co-workers,[29–33] the trial Monte Carlo moves consist of an instantaneous switch between protonation states. Changing the protonation state of an acidic group without allowing the solvent to relax will lead to a large, unfavorable change in energy and thus a low probability for acceptance of Monte Carlo moves. Therefore, these methods must necessarily make use of a continuum solvation model, which can adjust to the new protonation state instantaneously. (The method of Baptista is a hybrid, in which ordinary molecular dynamics is run with explicit solvent, instantaneous protonation-state moves are made with continuum electrostatics, and after each move, the solute is frozen in the new protonation state to allow the solvent molecules to relax.) In contrast, the method of Bürgi *et al.*,[34] in which trial moves are free energy calculations, that does not require the use of implicit solvent. However, performing an entire free energy calculation for every trial move is prohibitively expensive, unless the free energy calculations are very approximate.

In this paper, a new method is presented for performing molecular simulations with variable protonation states. As with earlier methods, our approach is not intended to describe the dynamics of proton transfer to and from solution, rather, to visit conformations and protonation states with the correct statistical probability for a system in equilibrium with a bath at constant temperature and *p*H. The method falls in the second category described above: it alternates sampling over configurations for a given protonation state with Monte Carlo moves attempted between physically meaningful, discrete protonation states. Trial moves consist of relatively short molecular dynamics trajectories (not free energy calculations) using a time-dependent potential energy that interpolates between the old and new protonation states. In essence, this procedure is hybrid Monte Carlo[35] with a time-dependent Hamiltonian. It samples conformations and protonation states from the correct statistical ensemble, independent of the trial-move trajectory length, which may therefore be adjusted to optimize efficiency. Because moves are not instantaneous, the method does not require an impicit solvent model, and should also be useful in describing buried titratable groups that are not directly exposed to solvent, but have strong interactions with neighboring hydrogen bond partners. The feasibility of the method is demonstrated for a simple test case: simulations of acetic acid in aqueous solution at constant *p*H, with an explicit representation of water molecules.

II. THEORY

Consider a molecular system which may exist in a finite number of states Γ, each defined by the presence or absence

of various labile protons. The only states considered are those in which protons are covalently bound to particular acidic groups, or are entirely absent. Classical mechanics is used for simplicity, but extending the present treatment to quantum statistics is straightforward with the path-integral formulation.[36–39] Each atom $i \in \Gamma$ has mass m_i, position $\mathbf{r}_i$, and momentum $\mathbf{p}_i$. The system energy is given by a Hamiltonian,

$$H(\Gamma,\mathbf{r}_{i\in\Gamma},\mathbf{p}_{i\in\Gamma}) = \left[\sum_{i\in\Gamma}\frac{|\mathbf{p}_i|^2}{2m_i}\right] + U(\Gamma,\mathbf{r}_{i\in\Gamma}), \tag{1}$$

where U is a potential that depends explicitly on the protonation state as well as on the positions of the atoms. The system can exchange energy and labile protons (but not other atoms) with a bath at temperature T and constant pH. By definition,

$$p\mathrm{H} \equiv -\log_{10} a_{\mathrm{H}^+}, \tag{2}$$

where a_{H^+} is the proton activity. The proton chemical potential μ_{H^+} is given by

$$\beta\mu_{\mathrm{H}^+} = \beta\mu^0_{\mathrm{H}^+} + \ln a_{\mathrm{H}^+} \tag{3}$$

$$= \beta\mu^0_{\mathrm{H}^+} - p\mathrm{H}(\ln 10), \tag{4}$$

where $\beta = 1/k_BT$ and $\mu^0_{\mathrm{H}^+}$ is the standard-state proton chemical potential. The probability distribution for observing the system in a particular state Γ with positions $\mathbf{r}_{i\in\Gamma}$ and momenta $\mathbf{p}_{i\in\Gamma}$ is then given by a semigrand ensemble,[40–42]

$$\rho(\Gamma,\mathbf{r}_{i\in\Gamma},\mathbf{p}_{i\in\Gamma}) = \frac{1}{\Xi}\frac{1}{h^{N_\Gamma}\Omega_\Gamma}\exp(\beta\mu_{\mathrm{H}^+}n^\Gamma_{\mathrm{H}^+} - \beta H) \tag{5}$$

$$= \frac{1}{\Xi}\frac{1}{h^{N_\Gamma}\Omega_\Gamma}\exp([\beta\mu^0_{\mathrm{H}^+} - p\mathrm{H}(\ln 10)]n^\Gamma_{\mathrm{H}^+} - \beta H), \tag{6}$$

where

$$\Xi = \sum_\Gamma \int \frac{1}{h^{N_\Gamma}\Omega_\Gamma}\exp([\beta\mu^0_{\mathrm{H}^+} - p\mathrm{H}(\ln 10)]n^\Gamma_{\mathrm{H}^+} - \beta H) \times \prod_{i\in\Gamma} d^3r_i d^3p_i \tag{7}$$

is a semigrand partition function. Here h is Planck's constant, and N_Γ, Ω_Γ, and $n^\Gamma_{\mathrm{H}^+}$ are the number of degrees of freedom, the degeneracy (i.e., the product of factorials of numbers of indistinguishable atoms of each kind), and the number of labile protons, for each state Γ. Equation (6) provides a precise definition for the term "constant-pH simulation:" just as a constant-temperature simulation is one that visits points in phase space with the Boltzmann distribution, a constant-pH simulation visits protonation states and points in phase space for each state with probability distribution given by Eq. (6).

It is convenient to replace the "real" system described above by a fictitious system for which protons do not vanish in deprotonated states, but instead are replaced with "ghost" atoms, in a similar approach to that used in alchemical free energy calculations.[43] This fictitious system is to be defined in such a way that the marginal probability distribution for the system being in a given protonation state and with real atoms at given positions is the same for the real and fictitious systems. The ghost atoms have the same mass, holonomic constraints, and interactions with covalent neighbors. However, they do not interact with any other atoms. For the fictitious system, the number of atoms, constraints, and degrees of freedom are constants, equal to the numbers for the state(s) of the actual system in which all labile protons are present. At this point the potential U is assumed to have the particular form,

$$U(\Gamma,\mathbf{r}_{i\in\Gamma}) = U_{\mathrm{FF}}(\Gamma,\mathbf{r}_{i\in\Gamma}) + U^{\mathrm{valence,H+}}_{\mathrm{FF}}(\Gamma,\mathbf{r}_{i\in\Gamma}) + b_\Gamma. \tag{8}$$

The first two quantities are given by a molecular mechanics force field,[44–56] the parameters of which will depend on the protonation state. Here $U^{\mathrm{valence,H_+}}_{\mathrm{FF}}$ denotes intramolecular (valence) force field terms acting on labile protons. All other terms in the force field are included in U_{FF}. The quantities b_Γ depend only on the protonation state, not on the positions, and represent the energy of forming covalent bonds to labile protons. This energy is not taken into account by the force field itself. The Hamiltonian for the fictitious system is then defined to be

$$\tilde{H}(\Gamma,\mathbf{r}_i,\mathbf{p}_i) = \left[\sum_i \frac{|\mathbf{p}_i|^2}{2m_i}\right] + \tilde{U}(\Gamma,\mathbf{r}_i), \tag{9}$$

$$\tilde{U}(\Gamma,\mathbf{r}_i) = U_{\mathrm{FF}}(\Gamma,\mathbf{r}_{i\in\Gamma}) + U^{\mathrm{valence,H+}}_{\mathrm{FF}}(\Gamma,\mathbf{r}_i) + \tilde{b}_\Gamma, \tag{10}$$

which includes kinetic energies and force field valence terms for all labile protons whether they are present or absent (i.e., replaced by ghost atoms). The quantities $\tilde{b}_\Gamma$ are now defined so that they are the sum of the energy of forming covalent bonds to labile protons, the standard proton chemical potential, and a correction related to the different ideal gas free energies of the real and fictitious systems,

$$\tilde{b}_\Gamma = b_\Gamma - \mu^0_{\mathrm{H}^+}n^\Gamma_{\mathrm{H}^+} + k_BT\ln\frac{Q^{\mathrm{id}}(\Gamma)}{\tilde{Q}^{\mathrm{id}}(\Gamma)}, \tag{11}$$

where

$$Q^{\mathrm{id}}(\Gamma) = \frac{1}{h^{N_\Gamma}\Omega_\Gamma}\int \exp\left(-\beta\left[\sum_{i\in\Gamma}\frac{|\mathbf{p}_i|^2}{2m_i}\right]\right)\prod_{i\in\Gamma} d^3p_i, \tag{12}$$

$$\tilde{Q}^{\mathrm{id}}(\Gamma) = \int \exp\left(-\beta\left[\sum_i \frac{|\mathbf{p}_i|^2}{2m_i}\right] - \beta U^{\mathrm{valence,H+}}_{\mathrm{FF}}\right)\prod_{i\notin\Gamma} d^3r_i\prod_i d^3p_i. \tag{13}$$

In Eq. (12), the integral is over the momenta for all atoms in state Γ. In Eq. (13), the integral is over the positions for ghost atoms only and momenta for all atoms (real and ghost). These integrals will be independent of the positions of the real atoms $\mathbf{r}_{i\in\Gamma}$. The ratio $Q^{\mathrm{id}}/\tilde{Q}^{\mathrm{id}}$ is not dimensionless, but different choices of units will merely result in adding the same overall constant to each $\tilde{b}_\Gamma$.

The fictitious system will be sampled from the probability distribution,

$$\tilde{\rho}(\Gamma,\mathbf{r}_i,\mathbf{p}_i)=\frac{1}{\tilde{\Xi}}\exp[-p\mathrm{H}(\ln 10)n_{\mathrm{H}^+}^{\Gamma}-\beta\tilde{H}], \tag{14}$$

where

$$\tilde{\Xi}=\sum_{\Gamma}\int \exp[-p\mathrm{H}(\ln 10)n_{\mathrm{H}^+}^{\Gamma}-\beta\tilde{H}]\prod_i d^3r_i d^3p_i. \tag{15}$$

In that case, the marginal probability distribution that the system is in a state Γ and that the real atoms have positions $\mathbf{r}_{i\in\Gamma}$, obtained by integrating over positions of ghost atoms and all momenta, will be the same for the real and fictitious systems,

$$\int \rho(\Gamma,\mathbf{r}_{i\in\Gamma},\mathbf{p}_{i\in\Gamma})\prod_{i\in\Gamma} d^3p_i=\int \tilde{\rho}(\Gamma,\mathbf{r}_i,\mathbf{p}_i)\prod_{i\notin\Gamma} d^3r_i\prod_i d^3p_i, \tag{16}$$

as desired.

In principle, the quantities $\tilde{b}_{\Gamma}$ could be estimated from the dissociation energy of a labile proton, the standard proton chemical potential, and Eqs. (12) and (13). It is convenient, however, to simply treat them as adjustable parameters and fit them so as to reproduce experimental acid dissociation constants, thereby compensating for errors in the force field (as well as obviating the need for the ideal gas correction or the proton standard chemical potential). For example, consider an acid that exists in a protonated state HA and a deprotonated state A^- with a measured $p\mathrm{K}_a$. The difference in parameters $\Delta\tilde{b}\equiv\tilde{b}_{\mathrm{HA}}-\tilde{b}_{\mathrm{A}^-}$ is to be chosen such that the marginal probability of observing each protonation state at a specified $p\mathrm{H}$ is equal to the fraction which would be observed experimentally in dilute solution; that is,

$$p\mathrm{K}_a=p\mathrm{H}-\log_{10}\left[\frac{\int \tilde{\rho}(\mathrm{HA},\mathbf{r}_i,\mathbf{p}_i)\prod_i d^3r_i d^3p_i}{\int \tilde{\rho}(\mathrm{A}^-,\mathbf{r}_i,\mathbf{p}_i)\prod_i d^3r_i d^3p_i}\right] \tag{17}$$

$$=p\mathrm{H}-\log_{10}\frac{\tilde{\rho}(\mathrm{HA})}{\tilde{\rho}(\mathrm{A}^-)}, \tag{18}$$

or equivalently,

$$\tilde{\rho}(\mathrm{A}^-)=\frac{1}{1+10^{p\mathrm{H}-p\mathrm{K}_a}}. \tag{19}$$

This will be the case if $\Delta\tilde{b}$ is set to

$$\Delta\tilde{b}=-p\mathrm{K}_a(\ln 10)+k_BT\ln\left[\frac{\int \exp(-\beta[U_{\mathrm{FF}}(\mathrm{HA},\mathbf{r}_i)+U_{\mathrm{FF}}^{\mathrm{valence,H+}}(\mathrm{HA},\mathbf{r}_i)])\prod_i d^3r_i}{\int \exp(-\beta[U_{\mathrm{FF}}(\mathrm{A}^-,\mathbf{r}_i)+U_{\mathrm{FF}}^{\mathrm{valence,H+}}(\mathrm{A}^-,\mathbf{r}_i)])\prod_i d^3r_i}\right]. \tag{20}$$

The ratio of configuration integrals in Eq. (20) may be estimated by standard methods for free energy calculations, such as the Bennett acceptance ratio method.[57–59] The parameters thereby obtained might be expected to be fairly transferable among chemically similar functional groups, although this will depend on the particular force field used.

In order to sample protonation states, positions, and momenta of the fictitious system, a Markov chain is constructed, defined by a transition probability distribution $P(\Gamma,\mathbf{r}_i,\mathbf{p}_i\rightarrow\Gamma',\mathbf{r}_i',\mathbf{p}_i')$ such that

$$\sum_{\Gamma}\int \tilde{\rho}(\Gamma,\mathbf{r}_i,\mathbf{p}_i)P(\Gamma,\mathbf{r}_i,\mathbf{p}_i\rightarrow\Gamma',\mathbf{r}_i',\mathbf{p}_i')\prod_i d^3r_i d^3p_i$$
$$=\tilde{\rho}(\Gamma',\mathbf{r}_i',\mathbf{p}_i'), \tag{21}$$

where $\tilde{\rho}(\Gamma',\mathbf{r}_i',\mathbf{p}_i')$ is given by Eq. (14). Two kinds of transitions will be performed: transitions in which the positions and momenta are changed, but the protonation state is kept the same, and transitions in which the protonation state as well as the positions and momenta are changed.

The first kind may be performed with any of the usual means for visiting states according to the canonical distribution; for instance, molecular dynamics with periodic resampling of velocities from the Boltzmann distribution, ordinary constant-temperature Monte Carlo, Langevin, or Nosé-Hoover dynamics.[17–20,60] All of these methods generate transitions between points in phase space within the same protonation state Γ, such that the transition probability distribution $Q(\Gamma,\mathbf{r}_i,\mathbf{p}_i\rightarrow\Gamma,\mathbf{r}_i',\mathbf{p}_i')$ satisfies

$$\int \tilde{\rho}(\Gamma,\mathbf{r}_i,\mathbf{p}_i)Q(\Gamma,\mathbf{r}_i,\mathbf{p}_i\rightarrow\Gamma,\mathbf{r}_i',\mathbf{p}_i')\prod_i d^3r_i d^3p_i$$
$$=\tilde{\rho}(\Gamma,\mathbf{r}_i',\mathbf{p}_i'). \tag{22}$$

The second kind of move may be attempted with arbitrary probability $p_{\Gamma\rightarrow\Gamma'}$ as long as this probability is symmetric,

$$p_{\Gamma\rightarrow\Gamma'}=p_{\Gamma'\rightarrow\Gamma}. \tag{23}$$

A trajectory is run for a time t, during which the potential energy is switched between the two protonation states. That is, dynamics is run with the time-dependent Hamiltonian,

$$H_{\Gamma\to\Gamma'}(\tau,\mathbf{r}_i,\mathbf{p}_i)=\left[\sum_i \frac{|\mathbf{p}_i|^2}{2m_i}\right]+U_{\Gamma\to\Gamma'}(\tau,\mathbf{r}_i),$$

where

$$U_{\Gamma\to\Gamma'}(\tau=0,\mathbf{r}_i)=\tilde{U}(\Gamma,\mathbf{r}_i), \tag{24}$$

$$U_{\Gamma\to\Gamma'}(\tau=t,\mathbf{r}_i)=\tilde{U}(\Gamma',\mathbf{r}_i), \tag{25}$$

$$U_{\Gamma\to\Gamma'}(\tau,\mathbf{r}_i)=U_{\Gamma'\to\Gamma}(t-\tau,\mathbf{r}_i). \tag{26}$$

The potential may be switched from one protonation state to another in any manner, as long as forward and reverse switches are mirror images of each other under time reversal, i.e., Eq. (26) is satisfied. (Note the switching in one direction does not necessarily need to be symmetric in time.) The simplest possibility is linear interpolation

$$U_{\Gamma\to\Gamma'}(\tau,\mathbf{r}_i)=\left(1-\frac{\tau}{t}\right)\tilde{U}(\Gamma,\mathbf{r}_i)+\left(\frac{\tau}{t}\right)\tilde{U}(\Gamma',\mathbf{r}_i), \tag{27}$$

but more complex switching schemes could be be used.

Hamiltonian dynamics defines a reversible, volume-conserving map $\mathbf{x}\to\phi\mathbf{x}$ between points in phase space.[61–63] Here $\phi\mathbf{x}$ denotes the final point of a trajectory started from the initial point $\mathbf{x}=(\mathbf{r}_i,\mathbf{p}_i)$. Let σ denote momentum reversal; that is, if $\mathbf{x}=(\mathbf{r}_i,\mathbf{p}_i)$, then $\sigma\mathbf{x}=(\mathbf{r}_i,-\mathbf{p}_i)$. Then

$$\phi\sigma\phi\mathbf{x}=\sigma\mathbf{x}.$$

In the present case there are two time-dependent potentials $U_{\Gamma\to\Gamma'}$ and $U_{\Gamma'\to\Gamma}$ satisfying Eq. (26). Let ϕ and ψ be the maps generated by dynamics with $U_{\Gamma\to\Gamma'}$ and $U_{\Gamma'\to\Gamma}$, respectively. If dynamics with $U_{\Gamma\to\Gamma'}$ takes an initial point $\mathbf{x}$ to a final point $\phi\mathbf{x}$, then dynamics with $U_{\Gamma'\to\Gamma}$ will take $\sigma\phi\mathbf{x}$ back to $\sigma\mathbf{x}$. That is,

$$\psi\sigma\phi\mathbf{x}=\sigma\mathbf{x}.$$

In addition to being reversible, the map generated by Hamiltonian dynamics is volume conserving (whether or not the potential is time dependent). That is, the Jacobian of the variable transformation from initial to final phase points is unity,

$$\left|\frac{\partial\phi\mathbf{x}}{\partial\mathbf{x}}\right|=1, \tag{28}$$

and likewise for ψ. Equivalently, if δ is the Dirac delta function,

$$\delta(\mathbf{x}-\mathbf{x}')=\delta(\phi\mathbf{x}-\phi\mathbf{x}'),$$

and likewise for ψ. Momentum reversal is also volume conserving,

$$\delta(\mathbf{x}-\mathbf{x}')=\delta(\sigma\mathbf{x}-\sigma\mathbf{x}').$$

Therefore, the conditional probability distribution $p(\Gamma,\mathbf{x}\to\Gamma',\mathbf{x}')$ for attempting a trial move to a protonation state Γ' and phase point $\mathbf{x}'$, given the current protonation state Γ and phase point $\mathbf{x}$, is symmetric with momentum reversal,

$$p(\Gamma,\mathbf{x}\to\Gamma',\mathbf{x}')=p_{\Gamma\to\Gamma'}\delta(\mathbf{x}'-\phi\mathbf{x}) \tag{29}$$

$$=p_{\Gamma'\to\Gamma}\delta(\sigma\mathbf{x}'-\sigma\phi\mathbf{x}) \tag{30}$$

$$=p_{\Gamma'\to\Gamma}\delta(\psi\sigma\mathbf{x}'-\psi\sigma\phi\mathbf{x}) \tag{31}$$

$$=p_{\Gamma'\to\Gamma}\delta(\psi\sigma\mathbf{x}'-\sigma\mathbf{x}) \tag{32}$$

$$=p(\Gamma',\sigma\mathbf{x}'\to\Gamma,\sigma\mathbf{x}). \tag{33}$$

That is,

$$p(\Gamma,\mathbf{r}_i,\mathbf{p}_i\to\Gamma',\mathbf{r}_i',\mathbf{p}_i')=p(\Gamma',\mathbf{r}_i',-\mathbf{p}_i'\to\Gamma,\mathbf{r}_i,-\mathbf{p}_i).$$

It should be noted that a discretization of Hamilton's equations such as the velocity Verlet integrator will also give a reversible, volume-conserving map. This is shown explicitly in the appendix. Therefore, the trial-move probability distribution will also be symmetric for discrete, approximate molecular dynamics trajectories, independent of the time step.

Moves are accepted with probability given by the Metropolis criterion,

$$a(\Gamma,\mathbf{r}_i,\mathbf{p}_i\to\Gamma',\mathbf{r}_i',\mathbf{p}_i')=\min\left[1,\frac{\tilde{\rho}(\Gamma',\mathbf{r}_i',\mathbf{p}_i')}{\tilde{\rho}(\Gamma,\mathbf{r}_i,\mathbf{p}_i)}\right] \tag{34}$$

$$=\min[1,\exp(-p\mathrm{H}(\ln 10)\Delta n_{\mathrm{H}^+}-\beta\Delta\tilde{H})], \tag{35}$$

where

$$\Delta n_{\mathrm{H}^+}=n_{\mathrm{H}^+}^{\Gamma'}-n_{\mathrm{H}^+}^{\Gamma}, \tag{36}$$

$$\Delta\tilde{H}=\tilde{H}(\Gamma',\mathbf{r}',\mathbf{p}')-\tilde{H}(\Gamma,\mathbf{r},\mathbf{p}). \tag{37}$$

Note that $\Delta\tilde{H}$ includes the change in kinetic as well as potential energies. The transition probability distribution $R(\Gamma,\mathbf{r}_i,\mathbf{p}_i\to\Gamma',\mathbf{r}_i',\mathbf{p}_i')$ is the product of the trial-move distribution and the acceptance probability

$$R(\Gamma,\mathbf{r}_i,\mathbf{p}_i\to\Gamma',\mathbf{r}_i',\mathbf{p}_i')=p(\Gamma,\mathbf{r}_i,\mathbf{p}_i\to\Gamma',\mathbf{r}_i',\mathbf{p}_i')a(\Gamma,\mathbf{r}_i,\mathbf{p}_i\to\Gamma',\mathbf{r}_i',\mathbf{p}_i'), \tag{38}$$

which satisfies detailed balance

$$\tilde{\rho}(\Gamma,\mathbf{r}_i,\mathbf{p}_i)R(\Gamma,\mathbf{r}_i,\mathbf{p}_i\to\Gamma',\mathbf{r}_i',\mathbf{p}_i')=\tilde{\rho}(\Gamma',\mathbf{r}_i',\mathbf{p}_i')R(\Gamma',\mathbf{r}_i',\mathbf{p}_i'\to\Gamma,\mathbf{r}_i,\mathbf{p}_i), \tag{39}$$

since $\tilde{\rho}(\Gamma,\mathbf{r}_i,\mathbf{p}_i)=\tilde{\rho}(\Gamma,\mathbf{r}_i,-\mathbf{p}_i)$. The net transition probability distribution due to both kinds of moves is then

$$P(\Gamma,\mathbf{r}_i,\mathbf{p}_i \to \Gamma',\mathbf{r}_i',\mathbf{p}_i') = \begin{cases} R(\Gamma,\mathbf{r}_i,\mathbf{p}_i \to \Gamma',\mathbf{r}_i',\mathbf{p}_i'), & \Gamma \neq \Gamma' \\ \left[1 - \sum_{\Gamma'' \neq \Gamma} p_{\Gamma \to \Gamma''}\right] Q(\Gamma,\mathbf{r}_i,\mathbf{p}_i \to \Gamma,\mathbf{r}_i',\mathbf{p}_i') + \sum_{\Gamma'' \neq \Gamma} \left[p_{\Gamma \to \Gamma''} - \int R(\Gamma,\mathbf{r}_i,\mathbf{p}_i \to \Gamma'',\mathbf{r}_i'',\mathbf{p}_i'') \prod_i d^3r'' d^3p''\right] \\ \times \prod_i \delta(\mathbf{r}_i - \mathbf{r}_i')\delta(\mathbf{p}_i - \mathbf{p}_i'), & \Gamma = \Gamma' \end{cases} \tag{40}$$

where the second term for the case $\Gamma = \Gamma'$ is due to rejected protonation-state-change moves. The transition probability distribution given by Eq. (40) satisfies Eq. (21), as desired.

III. NUMERICAL RESULTS

To demonstrate the feasibility of the method, simulations of acetic acid were performed using the force field parameters of Jorgensen *et al.*[47] in a bath of 249 TIP4P water molecules.[64] The parameters for each protonation state are given in supplemental information.[65] Constraints were applied to bond lengths and angles for the water molecules and the labile proton in acetic acid. Simulations were performed in a cubic box of length 19.8 Å. The electrostatic energy and forces were computed using the Ewald sum.[60] Although there is still controversy in the literature,[66] some degree of consensus has emerged that Ewald summation is most likely the most reliable method for giving results that match as closely as possible those of an infinite aperiodic system.[67,68] The particular form of the Ewald sum used in this work is the inclusion of the mean of the Ewald potential, so that the sum remains independent of the choice of screening parameter, even for a system with net charge. Such a choice gives ionic solvation free energies that become independent of system size for relatively small solvent boxes.[69–73]

A parameter $\Delta\tilde{b}$ was determined for the OH covalent bond from Eq. (20). To compute the ratio of configuration integrals (i.e., the free energy change), the change in the force field parameters for the two protonation states was divided into 34 steps ("lambda values"). For each step, independent molecular dynamics simulations were run for 10 ns, at a constant temperature of 298.15 K, and resampling velocities from the Boltzmann distribution at 0.5 ps intervals. All molecular dynamics simulations were run with a time step of 2 fs and the velocity Verlet integrator.[60]

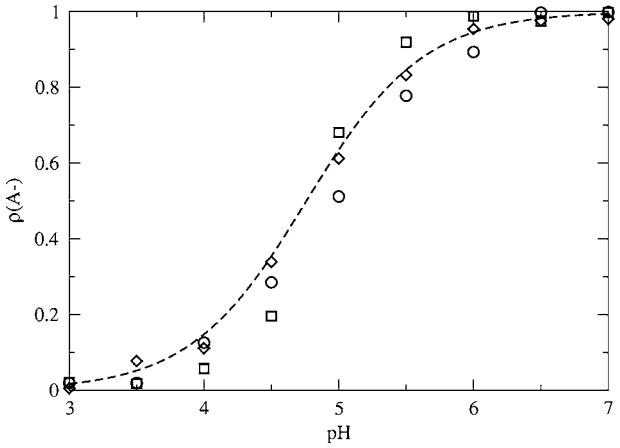

FIG. 1. Fraction of acetate ion, $\rho(A^-)$, as a function of pH. Squares, diamonds, circles are the fractions observed in three independent constant-pH simulations; dotted line is $1/(1+10^{pK_a-pH})$, with $pK_a=4.76$.

The free energy change for each step was estimated using the Bennett acceptance ratio method,[57] and added to obtain the total free energy change for switching force field parameters from the deprotonated state to the protonated state, 97.00 ± 0.03 kcal/mol. Using this value and the experimental pK_a for acetic acid (4.76), $\Delta\tilde{b}$ was set to -103.49 kcal/mol. That is, the potential energy for a conformation in the deprotonated state was just that given by the force field (with parameters for the deprotonated state) the potential energy for a conformation in the protonated state was that given by the force field (with parameters for the protonated state) minus 103.49 kcal/mol.

Constant-pH simulations were then run at a series of different pH values, ranging from 3.0 to 7.0. For each pH, three independent simulations were run for 10 ns of ordinary constant-temperature dynamics. Every 10 ps, a change in protonation state was attempted. This change was itself performed over 10 ps, during which the potential energy was changed by simple linear interpolation from one state to another (more complicated interpolation schemes did not significantly change acceptance probabilities). The overall time for each simulation was therefore 20 ns. The fraction of moves accepted was about 20%. Switching the protonation state over a time longer than 10 ps did not significantly improve acceptance probabilities per unit time, but switching it more quickly led to lower acceptance probabilities; a switching time of 10 ps seemed close to optimal.

The fraction of the deprotonated state observed in the simulations is shown as a function of pH in Fig. 1. There is a good agreement with the expected titration curve, Eq. (19). This is a demonstration of the consistency of the method, and that good sampling over protonation states can be achieved in explicit solvent with reasonable computational expense.

IV. CONCLUSIONS

A method has been presented for performing molecular simulations with variable protonation states, such that conformations and protonation states are visited with the correct statistical probability for a system in equilibrium with a bath at constant temperature and pH. The method relies on rela-

tively short (in this case, 10 ps) molecular dynamics trajectories used as trial Monte Carlo moves; in essence, hybrid Monte Carlo with a time-dependent Hamiltonian. Correct sampling is independent of the trajectory length, so that it may be adjusted to achieve optimal move acceptance per unit time.

The primary motivation for the current work was to be able to conduct simulations with variable protonation states using explicit solvent. The method should also be useful in treating groups not directly exposed to solvent, but making strong interactions with neighboring hydrogen bond partners; for instance, titratable residues located in the interior of a protein. The numerical simulations presented in this work were performed on a system with only one titratable group. For more complicated systems, it might be useful to perform moves in which several protonation states are changed at once.

The present work addresses only the problem of sampling over protonation states. Whether or not this or any other method can correctly predict experimental protonation states and *p*H-dependent conformational changes will depend on the ability of the force field and solvation model used to describe interactions of titratable groups.

ACKNOWLEDGMENTS

This work was supported in part by the Camille and Henry Dreyfus Foundation and the National Science Foundation (MCB-0501265). The author thanks Bruce Berne for helpful discussions.

APPENDIX: REVERSIBILITY AND PHASE-SPACE VOLUME CONSERVATION OF DISCRETE DYNAMICS

It is shown that the velocity Verlet integrator with a time-dependent potential energy that symmetrically switches between two states generates a reversible, volume-conserving map between phase points. Assume dynamics is run for N time steps, each of length Δt, such that the total trajectory length is $t=N\Delta t$. The Hamiltonian is given by

$$H_{\Gamma\to\Gamma'}(\tau,\mathbf{r},\mathbf{p})=\frac{|\mathbf{p}|^2}{2m}+U_{\Gamma\to\Gamma'}(\tau,\mathbf{r}),$$

such that

$$U_{\Gamma\to\Gamma'}(\tau,\mathbf{r})=U_{\Gamma'\to\Gamma}(t-\tau,\mathbf{r}).$$

The forces are

$$\mathbf{F}_{\Gamma\to\Gamma'}(\tau,\mathbf{r})=-\nabla U_{\Gamma\to\Gamma'}(\tau,\mathbf{r})$$

and also satisfy

$$\mathbf{F}_{\Gamma\to\Gamma'}(\tau,\mathbf{r})=-\mathbf{F}_{\Gamma'\to\Gamma}(t-\tau,\mathbf{r}).$$

Given a phase point at step k, the point at step $k+1$ is

$$\mathbf{r}_{k+1}=\mathbf{r}_k+\frac{\mathbf{p}_k}{m}\Delta t+\mathbf{F}_{\Gamma\to\Gamma'}(k\Delta t,\mathbf{r}_k)\frac{(\Delta t)^2}{2m}, \tag{A1}$$

$$\mathbf{p}_{k+1}=\mathbf{p}_k+[\mathbf{F}_{\Gamma\to\Gamma'}(k\Delta t,\mathbf{r}_k)+\mathbf{F}_{\Gamma\to\Gamma'}((k+1)\Delta t,\mathbf{r}_{k+1})]\frac{\Delta t}{2}. \tag{A2}$$

Consider a "forward" trajectory $\mathbf{r}_k$, $\mathbf{p}_k$ using the potential $U_{\Gamma\to\Gamma'}$, and a "reverse" trajectory $\mathbf{r}'_k$, $\mathbf{p}'_k$ using the potential $U_{\Gamma'\to\Gamma}$. If $\mathbf{r}'_k=\mathbf{r}_{N-k}$ and $\mathbf{p}'_k=-\mathbf{p}_{N-k}$, then

$$\begin{aligned}\mathbf{r}'_{k+1}&=\mathbf{r}'_k+\frac{\mathbf{p}'_k}{m}\Delta t+\mathbf{F}_{\Gamma'\to\Gamma}(k\Delta t,\mathbf{r}'_k)\frac{(\Delta t)^2}{2m}\\&=\mathbf{r}_{N-k}-\frac{\mathbf{p}_{N-k}}{m}+\mathbf{F}_{\Gamma\to\Gamma'}((N-k)\Delta t,\mathbf{r}_{N-k})\frac{(\Delta t)^2}{2m}\\&=\mathbf{r}_{N-k-1},\end{aligned} \tag{A3}$$

and

$$\begin{aligned}\mathbf{p}'_{k+1}&=\mathbf{p}'_k+[\mathbf{F}_{\Gamma'\to\Gamma}(k\Delta t,\mathbf{r}_k)+\mathbf{F}_{\Gamma'\to\Gamma}((k+1)\Delta t,\mathbf{r}'_{k+1})]\frac{\Delta t}{2}\\&=-\mathbf{p}_{N-k}+[\mathbf{F}_{\Gamma\to\Gamma'}((N-k)\Delta t,\mathbf{r}_{N-k})\\&\quad+\mathbf{F}_{\Gamma\to\Gamma'}((N-k-1)\Delta t,\mathbf{r}_{N-k-1})]\frac{\Delta t}{2}=\mathbf{p}_{N-k-1}.\end{aligned} \tag{A4}$$

It follows that if $\mathbf{r}'_0=\mathbf{r}_N$ and $\mathbf{p}'_0=-\mathbf{p}_N$, then $\mathbf{r}'_N=\mathbf{r}_0$ and $\mathbf{p}'_N=-\mathbf{p}_0$. That is, if the reverse trajectory starts at the final point of the forward trajectory (with momenta reversed), then it will terminate at the initial point of the forward trajectory (again, with momenta reversed).

Furthermore, each step of the dynamics conserves phase-space volume. This can be seen by calculating the Jacobian,

$$\begin{vmatrix}\frac{\partial\mathbf{r}_{k+1}}{\partial\mathbf{r}_k} & \frac{\partial\mathbf{r}_{k+1}}{\partial\mathbf{p}_k}\\ \frac{\partial\mathbf{p}_{k+1}}{\partial\mathbf{r}_k} & \frac{\partial\mathbf{p}_{k+1}}{\partial\mathbf{p}_k}\end{vmatrix}=\begin{vmatrix}1+\nabla\mathbf{F}_k\frac{(\Delta t)^2}{2m} & \frac{\Delta t}{m}\\ \left[\nabla\mathbf{F}_k+\nabla\mathbf{F}_{k+1}\left(1+\nabla\mathbf{F}_k\frac{(\Delta t)^2}{2m}\right)\right]\frac{\Delta t}{2} & 1+\nabla\mathbf{F}_{k+1}\frac{(\Delta t)^2}{2m}\end{vmatrix}=1. \tag{A5}$$

Since each time step conserves volume, so will the entire trajectory.

[1] S. Kuramitsu and K. Hamaguchi, J. Biochem. (Tokyo) **87**, 1215 (1980).
[2] C. M. Carr and P. S. Kim, Cell **73**, 823 (1993).
[3] D. R. Ripoll, Y. N. Vorobjev, A. Liwo, J. A. Vila, and H. A. Scheraga, J. Mol. Biol. **264**, 770 (1996).
[4] M. Stöckelhuber, A. A. Nögel, C. Eckerskorn, J. Kohler, D. Rieger, and M. Schleicher, J. Cell. Sci. **109**, 1825 (1996).
[5] M. Schaefer, M. Sommer, and M. Karplus, J. Phys. Chem. B **101**, 1663 (1997).
[6] M. L. F. Ruano, J. Perez-Gil, and C. Casals, J. Biol. Chem. **273**, 15183 (1998).

[7] J. B. Murray, C. M. Dunham, and W. G. Scott, J. Mol. Biol. **315**, 121 (2002).
[8] M.-R. Mihailescu and J. P. Marino, Proc. Natl. Acad. Sci. U.S.A. **101**, 1189 (2004).
[9] O. K. Gasymov, A. R. Abduragimov, T. N. Yusifov, and B. J. Glasgow, Biochemistry **43**, 12894 (2004).
[10] P. L. Sorgen, H. S. Duffy, D. C. Spray, and M. Delmar, Biophys. J. **87**, 574 (2004).
[11] V. Saksmerprome and D. H. Burke, J. Mol. Biol. **341**, 685 (2004).
[12] H. Sackin, M. Nanazashvili, L. G. Palmer, M. Krambis, and D. E. Walters, Biophys. J. **88**, 2597 (2005).
[13] K. Palczewski, Annu. Rev. Biochem. **75**, 743 (2006).
[14] J. Mongan and D. A. Case, Curr. Opin. Struct. Biol. **15**, 157 (2005).
[15] J. E. Mertz and B. M. Pettitt, Int. J. Supercomput. Appl. **8**, 47 (1994).
[16] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
[17] S. Nosé and M. L. Klein, Mol. Phys. **50**, 1055 (1983).
[18] S. Nosé, J. Chem. Phys. **81**, 511 (1984).
[19] S. Nosé, Mol. Phys. **52**, 255 (1984).
[20] W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).
[21] U. Borjesson and P. H. Hünenberger, J. Chem. Phys. **114**, 9706 (2001).
[22] U. Börjesson and P. H. Hünenberger, J. Phys. Chem. B **108**, 13551 (2004).
[23] M. S. Lee, F. R. Salsbury, Jr., and C. L. Brooks III, Proteins: Struct., Funct., Genet. **56**, 738 (2004).
[24] J. Khandogin and C. L. Brooks III, Biophys. J. **89**, 141 (2005).
[25] J. Khandogin and C. L. Brooks III, Biochemistry **45**, 9363 (2006).
[26] A. M. Baptista, P. J. Martel, and S. B. Petersen, Proteins: Struct., Funct., Genet. **27**, 523 (1997).
[27] A. M. Baptista, V. H. Teixeira, and C. M. Soares, J. Chem. Phys. **117**, 4184 (2002).
[28] M. Machuqueiro and A. M. Baptista, J. Phys. Chem. B **110**, 2927 (2006).
[29] A. M. Walczak and J. M. Antosiewicz, Phys. Rev. E **66**, 051911 (2002).
[30] M. Długosz and J. M. Antosiewicz, Chem. Phys. **302**, 161 (2004).
[31] M. Długosz, J. M. Antosiewicz, and A. D. Robertson, Phys. Rev. E **69**, 021915 (2004).
[32] M. Długosz and J. M. Antosiewicz, J. Phys. Chem. B **109**, 13777 (2005).
[33] M. Długosz and J. M. Antosiewicz, J. Phys.: Condens. Matter **17**, S1607 (2005).
[34] R. Bürgi, P. A. Kollman, and W. F. van Gunsteren, Proteins: Struct., Funct., Genet. **47**, 469 (2002).
[35] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, Phys. Lett. B **195**, 216 (1987).
[36] R. P. Feynman, *Statistical Mechanics* (Addison-Wesley, Reading, MA, 1972).
[37] B. J. Berne and D. Thirumalai, Annu. Rev. Phys. Chem. **37**, 401 (1986).
[38] Q. Wang and J. K. Johnson, J. Chem. Phys. **107**, 5108 (1997).
[39] H. A. Stern and B. J. Berne, J. Chem. Phys. **115**, 7622 (2001).
[40] R. A. Alberty and I. Oppenheim, J. Chem. Phys. **91**, 1824 (1989).
[41] R. A. Alberty, J. Chem. Phys. **114**, 8270 (2001).
[42] R. J. Silbey, R. A. Alberty, and M. G. Bawendi, *Physical Chemistry*, 4th ed. (Wiley, New York, 2004).
[43] P. Kollman, Chem. Rev. (Washington, D.C.) **93**, 2395 (1993).
[44] W. L. Jorgensen, J. D. Madura, and C. J. Swenson, J. Am. Chem. Soc. **106**, 6638 (1984).
[45] W. L. Jorgensen and J. Tirado-Rives, J. Am. Chem. Soc. **110**, 1657 (1988).
[46] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).
[47] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, J. Am. Chem. Soc. **118**, 11225 (1996).
[48] G. A. Kaminski and W. L. Jorgensen, J. Phys. Chem. **100**, 18010 (1996).
[49] T. A. Halgren, J. Comput. Chem. **17**, 490 (1996).
[50] A. D. MacKerell, Jr., D. Bashford, M. Bellot *et al.*, J. Phys. Chem. B **102**, 3586 (1998).
[51] N. Foloppe and A. D. MacKerell, Jr., J. Comput. Chem. **21**, 86 (2000).
[52] A. D. MacKerell, Jr., and N. K. Banavali, J. Comput. Chem. **21**, 105 (2000).
[53] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, J. Phys. Chem. B **105**, 6474 (2001).
[54] M. L. P. Price, D. Ostrovsky, and W. L. Jorgensen, J. Comput. Chem. **22**, 1340 (2001).
[55] J. W. Ponder and D. A. Case, Adv. Protein Chem. **66**, 27 (2003).
[56] A. D. MacKerell, J. Comput. Chem. **25**, 13 (2004).
[57] M. R. Shirts, J. W. Pitera, W. C. Swope, and V. S. Pande, J. Chem. Phys. **119**, 5740 (2003).
[58] M. R. Shirts, E. Bair, G. Hooker, and V. S. Pande, Phys. Rev. Lett. **91**, 140601 (2003).
[59] M. R. Shirts and V. S. Pande, J. Chem. Phys. **122**, 134508 (2005).
[60] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon, Oxford, 1987).
[61] J. M. Sanz-Serna and M. P. Calvo, *Numerical Hamiltonian Problems* (Chapman and Hall, London, 1994).
[62] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration* (Springer, Berlin, 2002).
[63] H. A. Stern, J. Comput. Chem. **25**, 749 (2004).
[64] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[65] See EPAPS Document No.E-JCPSA6-126-012718 for the force field parameters used for acetic acid and acetate ion. This document can be reached through a direct link in the online article's HTML reference section or via the EPAPS homepage (http://www.aip.org/pubservs/epaps.html).
[66] D. A. C. Beck, R. S. Armen, and V. Daggett, Biochemistry **44**, 609 (2005).
[67] T. E. Cheatham and B. R. Brooks, Theor. Chem. Acc. **99**, 279 (1998).
[68] T. E. Cheatham, Curr. Opin. Struct. Biol. **14**, 360 (2004).
[69] G. Hummer and D. M. Soumpasis, J. Chem. Phys. **98**, 581 (1992).
[70] G. Hummer, L. R. Pratt, and A. E. García, J. Phys. Chem. **100**, 1206 (1996).
[71] G. Hummer, L. R. Pratt, and A. E. García, J. Chem. Phys. **107**, 9275 (1997).
[72] G. Hummer, L. R. Pratt, A. E. García, B. J. Berne, and S. W. Rick, J. Phys. Chem. B **101**, 3017 (1997).
[73] G. Hummer, L. R. Pratt, and A. E. García, J. Phys. Chem. A **102**, 7885 (1998).