# Constant-pH Hybrid Nonequilibrium Molecular Dynamics—Monte Carlo Simulation Method

Yunjie Chen and Benoît Roux*

Department of Biochemistry and Molecular Biology, Department of Chemistry, University of Chicago, Chicago, Illinois 60637, United States

ABSTRACT: A computational method is developed to carry out explicit solvent simulations of complex molecular systems under conditions of constant pH. In constant-pH simulations, preidentified ionizable sites are allowed to spontaneously protonate and deprotonate as a function of time in response to the environment and the imposed pH. The method, based on a hybrid scheme originally proposed by H. A. Stern (*J. Chem. Phys.* **2007**, *126*, 164112), consists of carrying out short nonequilibrium molecular dynamics (neMD) switching trajectories to generate physically plausible configurations with changed protonation states that are subsequently accepted or rejected according to a Metropolis Monte Carlo (MC) criterion. To ensure microscopic detailed balance arising from such nonequilibrium switches, the atomic momenta are altered according to the symmetric two-ends momentum reversal prescription. To achieve higher efficiency, the original neMD—MC scheme is separated into two steps, reducing the need for generating a large number of unproductive and costly nonequilibrium trajectories. In the first step, the protonation state of a site is randomly attributed via a Metropolis MC process on the basis of an intrinsic pK_a; an attempted nonequilibrium switch is generated only if this change in protonation state is accepted. This hybrid two-step inherent pK_a neMD—MC simulation method is tested with single amino acids in solution (Asp, Glu, and His) and then applied to turkey ovomucoid third domain and hen egg-white lysozyme. Because of the simple linear increase in the computational cost relative to the number of titratable sites, the present method is naturally able to treat extremely large systems.
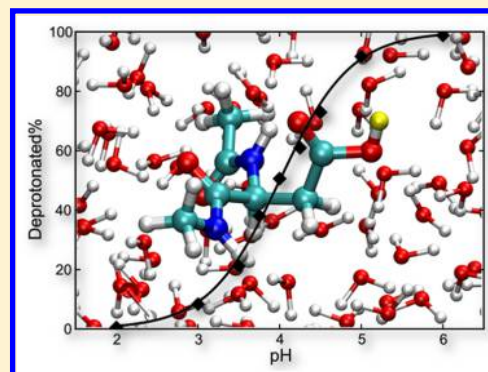
I. INTRODUCTION

In conventional MD simulations based on classical force fields, chemical bonds are commonly never broken and the protonation state of each ionizable residue is fixed and ascribed externally by the user. This is a considerable limitation when trying to study pH-dependent biological processes. These include, to name a few, membrane insertion,[1] fibril formation,[2] protein denaturation,[3] proton gradient-driven ATP synthesis,[4] and membrane transporters.[5,6] Typically, in the absence of specific information, the protonation state of the ionizable residues is assigned on the basis of their reference pK_a from the isolated amino acid in solution. However, the pK_a of specific residues in a protein is very sensitive to the environment, and actual pK_a values can deviate considerably from standard references. The issue is particularly important in the case of histidine because its pK_a is close to physiological pH. Furthermore, the dynamical movement of a protein can be coupled with an alternation of protonation state.

At the most fundamental level, protonation and deprotonation involve breaking or forming chemical bonds. For this reason, a calculation of absolute proton affinity can be carried out only from a quantum mechanical (QM) treatment. However, a pure QM approach is impractical for biomolecular systems, where there can be a very large number of ionizable sites. Slightly more tractable approaches are based on mixed quantum mechanical and molecular mechanical (QM/MM) treatments. While QM/MM methods can address issues of absolute proton affinity in the context of enzyme reactions,[7,8] they remain too computationally expensive to treat multiple ionizable sites in large bimolecular systems.

In the context of large biomolecular systems, an effective computational method to carry out simulations under conditions of constant pH must necessarily be built on the basis of empirical molecular mechanical (MM) models. In constant-pH simulations, preidentified titratable sites are allowed to spontaneously change their protonation state as a function of time in response to the environment and the imposed pH. Monte Carlo (MC) is a natural method for treating the abrupt transitions between discrete protonation states. For example, in a hybrid MD—MC simulation algorithm, one can imagine a cycle that consists of carrying out an equilibrium trajectory for some time, followed by attempts to make a transition to a different protonation state that is either accepted or rejected according to a Metropolis criterion. The first hybrid MD—MC methodology with discrete protonation states was introduced by Bürgi et al.[9] The method, which relied on a thermodynamic integration (TI) calculation carried out with explicit solvent to evaluate the free energy difference and the acceptance probability associated with the change in protonation state, was computationally expensive and suffered from convergence issues due to the short TI calculations. Soon

afterward, Mongan, McCammon, and colleagues[10,11] proposed a hybrid MD−MC methodology with sudden jumps between discrete protonation states in conjunction with an implicit solvent model based on the generalized Born (GB) approximation. The algorithm effectively achieves a correct Boltzmann sampling and avoids the issues arising from short TI simulations. However, it cannot be readily transposed to explicit solvent simulations. Sudden jumps between discrete protonation states unavoidably lead to large differences in electrostatic energy and to a rejection of nearly all proposed change. The issue stems from the lack of overlap between the explicit solvent configurations of the protonated and unprotonated states. Similar difficulties are encountered in trying to design an effective grand canonical MC algorithms for ionic solution in which ions are created and destroyed in explicit solvent.[12]

To circumvent the configurational overlap problem in constant-pH simulations, some treatments have resorted to an empirical combination of explicit/implicit representations: the equilibrium MD is carried out with explicit solvent, but the protonation/deprotonation MC processes are carried out with an implicit solvent model.[13−16] While such constant-pH simulations have the ability to explore the accessible ionization states of a system in the presence of explicit solvent, it is important to keep in mind that the actual pK_a values are derived from a CE or GB approximation. Although the usage of continuum solvent models in the treatment of ionizable sites in proteins has a long history,[17−19] they are necessarily approximate and there is a need for a simulation methodology that is based entirely on an explicit solvent representation. There are many complex situations where a titratable site may be located in a narrow aqueous crevice for which the usage of an implicit solvent approximation may be limited. Fundamentally, the statistical properties arising from such combined MD−MC methods do not stem from a single consistent Hamiltonian, which is problematical.

A different strategy to address the configurational overlap problem consists of considering an extended Lagrangian with a continuous $\lambda(t)$ that is dynamically propagated to control the protonation state of the site. This leads to the so-called λ-dynamics algorithm developed by Brooks and co-workers.[20−27] The original concept of expanding the degrees of freedom of a system in order to include a continuous thermodynamic coupling parameter λ as a dynamical variable in the sampling process goes back to early work by Jorgensen,[28] Tidor,[29] and Hünenberger[30] (see also ref 31). In this method, because λ is allowed to evolve as a continuous dynamical degree of freedom, there are no discrete protonation and deprotonation events and the configurational overlap problem is not encountered. However, carefully calibrated λ-dependent potentials must be incorporated into the Hamiltonian to increase computational efficiency. For example, an unfavorable energy barrier is often introduced between the physical end states to decrease the probability of the unphysical intermediate region. This barrier must not be too large in order to enable rapid transitions between end states $\lambda = 0$ and $\lambda = 1$. For this reason, there remains unavoidably a nonzero probability q of visiting the unphysical states ($\delta < \lambda < 1 - \delta$, with $\delta \approx 0.1$). Strictly speaking, only end states $\lambda = 0$ or $\lambda = 1$ are physical. Assuming independence of the sites for the sake of simplicity, the probability of a physically meaningful configuration of the entire system with n sites scales like $(1 - q)^n$, with $(1 - q) < 1$. For example, even if q is 0.1, the probability of generating a physical state is not larger than about 0.12 with 20 sites (in practice, much larger values of q are tolerated to increase

computational efficiency). As a result, an algorithm based on sampling a continuous λ does not scale very well for a bimolecular system with a large number n of ionizable sites. This remains the major limitation of an algorithm sampling the dynamics of a continuous coupling parameter $\lambda(t)$.

A different solution to the configurational overlap problem to construct an explicit solvent constant-pH simulation method is based on the enveloping distribution sampling (EDS) method.[32,33] EDS is a method used to generate a hybrid energy surface from a finite number of states such that the statistical mechanical distribution of the end states will overlap;[34] a smoothing parameter s controls the magnitude of the energy barrier and the extent of configurational overlap between the end states. In this constant-pH simulation method, EDS is used to mix the energy surface of the protonated and unprotonated states for a small set of N preselected ionizable sites. Systems with different values of the EDS smoothing parameter are then connected via a Hamiltonian replica-exchange to facilitate interconversion between the different protonation states during a hybrid MD−MC simulation. The EDS constant-pH simulation method can handle explicit solvent. However, it requires that all of the accessible protonated and unprotonated states be enumerated explicitly into the smoothed EDS Hamiltonian. If there are N sites, then 2^N energy surfaces need to be mixed into the smoothed EDS Hamiltonian. From a computational point of view, a simulation requires at least as many processors as ionization states. While this limitation can be mitigated to some extent by including only the relevant states and leaving out those anticipated to have small probabilities, the method remains fundamentally designed to treat systems comprising only a small number of sites.

Scalability to a large number of ionizable sites remains the most essential feature of a truly useful constant-pH simulation method for biomolecular systems. It is worth mentioning that several of the previously proposed constant-pH simulation methods have introduced replica-exchange between systems at different pH.[14−16,32,33] While usage of H-REMD may lead to enhanced sampling for specific systems, it does not alter the fundamental nature (and inherent limitations) of the currently available algorithms regarding the scaling properties. Constant-pH simulation methods that are inherently limited to a small number of sites are, ultimately, not able to address this fundamental issue. In this regard, it is important to recall that a constant-pH simulation is not really necessary for investigating the effect of pH when only a small number of ionizable sites groups are involved (e.g., less than 3 or 4) and that all of the possible states can be enumerated explicitly. All of the effects from changing the pH can be assessed rigorously by computing the relative free energies between all of the relevant protonation states. As long as all the relevant states can be enumerated and treated via free energy perturbation (FEP), such an exhaustive approach is viable. Constant-pH algorithms that are applicable to only a small number of ionizable sites offer no specific advantage over direct FEP calculations.

The true challenge facing a constant-pH simulation method for bimolecular systems appears when the number of ionizable sites becomes very large, for example, when simulating the pH dependence of a viral particle,[35] where a complete enumeration of all possible states is completely infeasible. The only approach that could be potentially scaled to a very large system is some type of hybrid MD−MC with discrete states.[10,36,37] However, the algorithm must be significantly enhanced to circumvent the configurational overlap problem in the presence of explicit

solvent. A naive MC algorithm relying on discrete (instantaneous) protonation/deprotonation attempts in the presence of explicit solvent fails because the coordinates of the surrounding solvent atoms are kept fixed while the state of the ionizable site is changed. The large energy variations essentially result in complete rejection of all attempts to change protonation states. What is needed is a scheme that would allow the solvent configuration of the system to dynamically adapt to the new state, at least partially. This occurs naturally if the coupling parameter λ is progressively switched between the end states during the MD trajectory. This is essentially the MC scheme with dynamical state switching that was previously formulated by Harry Stern.[38,39] It consists of carrying out short nonequilibrium molecular dynamics (neMD) switching trajectories to generate a physically plausible configuration with an altered protonation state, which is then subsequently accepted or rejected according to a Metropolis Monte Carlo (MC) criterion. While a continuous $\lambda(t)$ evolving according to a pre-established time-dependent schedule is also involved in the neMD trajectory, the outcome is a random walk in terms of the discrete end state only, thus avoiding the poor scaling properties of λ-dynamics because of the simple linear increase in the computational cost relative to the number of titratable sites. Consequently, the present hybrid neMD−MC method is naturally able to treat systems of arbitrary size.

In this article, we formulate and implement a hybrid neMD−MC method to carry out simulations under conditions of constant pH. To ensure microscopic detailed balance arising from the nonequilibrium switches, the atomic momenta are altered according to the symmetric two-ends momentum reversal prescription.[39−41] To achieve higher efficiency, the original neMD−MC scheme is separated into two steps, reducing the need for generating a large number of unproductive and costly nonequilibrium trajectories. In the first step, the protonation state of a site is randomly attributed via a Metropolis MC process on the basis of an intrinsic pK_a; an attempted nonequilibrium switch is generated only if this change in protonation state is accepted. The simulation method is tested with single amino acids in solution (Asp, Glu, and His) and is then applied to turkey ovomucoid third domain and hen egg-white lysozyme. The article is concluded with an outlook to further expand the present algorithm.

II. THEORETICAL BACKGROUND

For the sake of simplicity, we consider here a solvated molecule comprising a single preidentified ionizable site, as illustrated in Figure 1. The formalism can be easily generalized to multiple ionizable sites. It is assumed that, when the molecule is unprotonated, the proton is transferred to a reference ionizable group located far away in solution (i.e., the number of particles in the system is the same for the two states). Let $\mathbf{r}$ represent all of the coordinates within the system and $U_p(\mathbf{r})$, $U_u(\mathbf{r})$, correspond to the potential energy function for the protonated (p) and unprotonated (u) systems, respectively. In principle, the free energy difference ΔG

$$e^{-\beta \Delta G} = \frac{\int d\mathbf{r}\ e^{-\beta U_p(\mathbf{r})}}{\int d\mathbf{r}\ e^{-\beta U_u(\mathbf{r})}} \tag{1}$$

ought to be directly related to the experimental pK_a of the site in the molecule. However, to match the experimental pK_a value with a MM force field, it is necessary to introduce an adjustable offset constant C
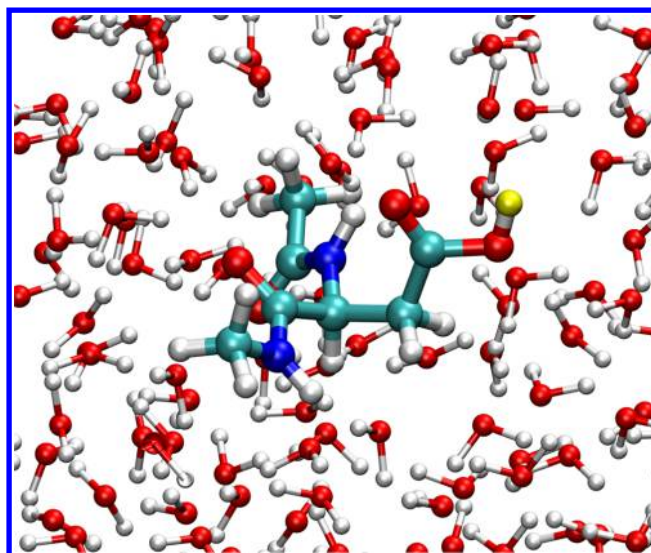

Figure 1. An aspartic dipeptide with blocked ends solvated in bulk water is used as an example of a molecule comprising an ionizable site. Within the framework of constant-pH simulations, a preidentified ionizable site is allowed to spontaneously protonate and deprotonate as a function of time in response to the environment and the imposed pH. In practice, the labile H atom (yellow sphere) is not annihilated in the unprotonated state. It is a fully interacting particle in the protonated state, and it is converted to a dummy noninteracting particle in the deprotonated state.

$$10^{pK_a} = \frac{\int d\mathbf{r}\ e^{-\beta[U_p(\mathbf{r})-C]}}{\int d\mathbf{r}\ e^{-\beta[U_u(\mathbf{r})]}} \tag{2}$$

given by

$$C \equiv k_B T\ \ln(10)\ pK_a + \Delta G \tag{3}$$

The use of an empirical offset constant C is necessary because the MM potential energy function is not designed to account for a number of critical factors, including the QM energy for binding a proton to an ionizable group, the standard state of the solvated proton, and effects associated with zero-point vibrations. A force field-dependent offset constant must be ascribed to each type of titratable site in a protein to calibrate the method. The value of these offset constants depends on the details of the MM force field. For each type of titratable site, an offset constant can be determined via an explicit solvent alchemical FEP calculation on a corresponding reference molecule. An amino acid dipeptide typically provides a reasonably good reference system for ionizable side chains (more details are provided in Section II.3). Once this calibration has been carried out, the resulting empirical offset constants are then assumed to be fully transferable. In other words, it is assumed that they do not vary with the sequence and conformation of the protein. The assumption of transferable empirical offset constants is a feature that is common to all MM treatments of ionizable groups in proteins.[9−11,13−17,20−27,32,33] This is also the reason why all MM treatments of ionizable groups in proteins, rather than absolute pK_a values, always focus on the issue of pK_a shifts relative to their corresponding reference compounds.

For the subsequent developments we construct an alchemical potential energy function $U(\mathbf{r},\lambda)$ using the continuous coupling parameter λ, which takes a value of 1 when the system is protonated and a value of 0 when the system is unprotonated

$$U(\mathbf{r},\lambda) = \lambda(U_p(\mathbf{r}) - C) + (1 - \lambda)U_u(\mathbf{r}) \tag{4}$$

The function $U(\mathbf{r},\lambda)$, defined in terms of a continuous coupling parameter λ, is equivalent to the alchemical potential function commonly defined for FEP/MD calculations[42] and λ-dynamics.[20−27]

The probability of finding the ionizable group protonated (P_1) or unprotonated (P_0) can then be expressed by the Henderson−Hasselbalch equations

$$P_1 = \frac{10^{(pK_a - pH)}}{1 + 10^{(pK_a - pH)}} \tag{5}$$

and

$$P_0 = \frac{1}{1 + 10^{(pK_a - pH)}} \tag{6}$$

respectively. To be consistent with the probability ratio $P_1/P_0 = 10^{(pK_a - pH)}$ from the Henderson−Hasselbalch equations, we construct an extended pH-dependent potential energy $\tilde{U}(\mathbf{r},\lambda)$, accounting for the protonated and unprotonated state as

$$\tilde{U}(\mathbf{r},\lambda) = \lambda(U_p(\mathbf{r}) - C + k_B T \ln(10)\, pH) + (1 - \lambda)U_u(\mathbf{r}) \tag{7}$$

and the corresponding Hamiltonian

$$\tilde{H}(\mathbf{r},\mathbf{p},\lambda) = K(\mathbf{p}) + \tilde{U}(\mathbf{r},\lambda) \tag{8}$$

where $\mathbf{p}$ represent all of the momenta within the system. For the sake of clarity, we will write $\mathbf{x}$ as a short-hand notation for $(\mathbf{r},\mathbf{p})$ in the following.

II.1. Constant-pH Simulation Algorithm. Our ultimate goal is to design a constant-pH simulation propagation algorithm $(\mathbf{x},\lambda \rightarrow \mathbf{x}',\lambda')$ such that it generates the proper Boltzmann equilibrium distribution for the system

$$\rho_{eq}(\mathbf{x},\lambda) = \frac{e^{-\beta\tilde{H}(\mathbf{x},\lambda)}}{\sum_{\lambda'} \int d\mathbf{x}'\, e^{-\beta\tilde{H}(\mathbf{x}',\lambda')}} \tag{9}$$

One possible avenue consists of generating a random walk in the space of $\mathbf{x}$ and λ that satisfies microscopic detailed balance

$$\rho_{eq}(\mathbf{x},\lambda)\, T(\mathbf{x},\lambda \rightarrow \mathbf{x}',\lambda') = \rho_{eq}(\mathbf{x}',\lambda')\, T(\mathbf{x}',\lambda' \rightarrow \mathbf{x},\lambda) \tag{10}$$

where $\lambda \rightarrow \lambda'$ represents any transition between the two end points (i.e., $0 \rightarrow 1$ or $1 \rightarrow 0$) and $\lambda' \rightarrow \lambda$ represents the reverse transition (i.e., $1 \rightarrow 0$ or $0 \rightarrow 1$). In practice, however, this is difficult to implement because of the strong coupling due to the large variations in electrostatic energy accompanying changes in the protonation state (λ). To circumvent this problem, we adopt a hybrid simulation method that combines the advantages of Monte Carlo (MC) with the strengths of classical molecular dynamics (MD). It consists of carrying out short nonequilibrium MD (neMD) trajectories upon a change in the protonation state of the site in order to generate a new configuration $\mathbf{x}'$ that is subsequently accepted or rejected via a Metropololis criterion. During the neMD trajectory, the proposed change in the discrete state variable λ is controlled externally while the MD propagation allows the continuous dynamical degrees of freedom $\mathbf{x}$ to dynamically adapt to the new state. This hybrid neMD−MC scheme is essentially the constant-pH simulation method that was initially elaborated by Harry Stern.[38] The scheme is designed to allow the system to evolve between the two states to generate the proposed move $(\mathbf{x},\lambda) \rightarrow (\mathbf{x}',\lambda')$. The hybrid neMD−MC scheme is composed of a time-dependent nonequilibrium MD switching process, followed by an MC step based on an acceptance criteria. Formally, the transition probability T is

expressed as the product of the probability of a *proposed* move T_p and the probability to *accept or reject* the proposed move T_a

$$T(\mathbf{x},\lambda \rightarrow \mathbf{x}',\lambda') = T_p(\mathbf{x},\lambda \rightarrow \mathbf{x}',\lambda')\, T_a(\mathbf{x},\lambda \rightarrow \mathbf{x}',\lambda') \tag{11}$$

During the dynamical switching process, the parameter λ follows a fixed time-dependent schedule over a time interval τ, whereas the remaining degrees of freedom are propagated dynamically according to the mixed-state potential energy function, $\tilde{U}(\mathbf{r},\lambda_{sw}(t))$ defined in eq 7. Here, the function $\lambda_{sw}(t)$ given by the linear form

$$\lambda_{sw}(t) = \lambda + (\lambda' - \lambda)\left(\frac{t - t_1}{\tau}\right) \tag{12}$$

represents the time-dependent continuous coupling parameter, which is externally controlled to drive the system between the two end points, λ and λ', during a neMD switching trajectory (note that λ can take the value 0 or 1 and λ' can take the value 1 or 0). For example, see the first switch in the time interval τ between t_1 and t_2 in Figure 2. The coupling parameter starts in the initial
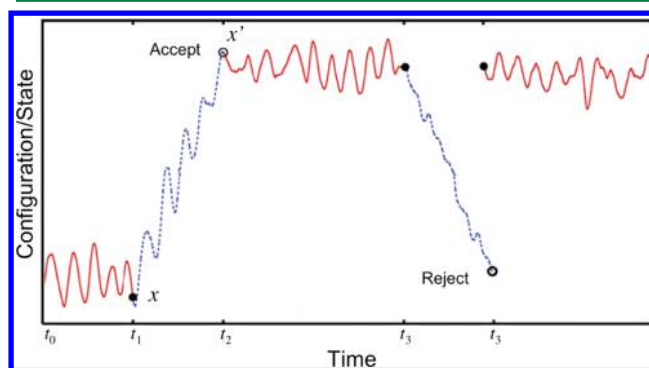

Figure 2. Schematic representation of the constant-pH hybrid neMD−MC simulation scheme. The red wiggly line represents the first stage of the simulation. The blue wiggly dashed line represents the second stage of the simulation. A Metropolis criterion is applied after the second stage (open black circle). If it is accepted, then the next MD simulation will start with the last configuration from the switching simulation; otherwise, the switching simulation will be discarded, and the next cycle of simulation will start with the last configuration from the MD stage.

state λ at time t_1 and is gradually altered in a time-dependent manner to reach the final state λ' at time $t_2 = t_1 + \tau$, where τ is the length of the switch.

The time-dependent $\lambda_{sw}(t)$ for the forward and backward switching, $\lambda \rightarrow \lambda'$ and $\lambda' \rightarrow \lambda$, must be consistent to guarantee detailed balance, as previously noted by Stern.[38] This is automatically satisfied if $\lambda_{sw}(t)$, for the sake of simplicity, is chosen to be symmetric with respect to the two end states over the interval τ. The linear form of eq 12 fulfills this condition. When a deterministic (reversible symplectic) propagator and a correct momentum reversal prescription are used,[40,41,43] the transition probability of the proposed move is inherently symmetric

$$T_p(\mathbf{x},\lambda \rightarrow \mathbf{x}',\lambda') = T_p(\mathbf{x}',\lambda' \rightarrow \mathbf{x},\lambda) \tag{13}$$

and the acceptance or rejection probability is given by the Metropolis criterion

$$T_a(\mathbf{x} \rightarrow \mathbf{x}') = \min\left[1, e^{-\beta[\tilde{H}(\mathbf{x}',\lambda') - \tilde{H}(\mathbf{x},\lambda)]}\right] \tag{14}$$

By substitution, it can be shown readily that

$$\frac{\rho_{eq}(\mathbf{x}',\lambda')}{\rho_{eq}(\mathbf{x},\lambda)} = \frac{T_p(\mathbf{x},\lambda \rightarrow \mathbf{x}',\lambda')}{T_p(\mathbf{x}',\lambda' \rightarrow \mathbf{x},\lambda)} \frac{T_a(\mathbf{x},\lambda \rightarrow \mathbf{x}',\lambda')}{T_a(\mathbf{x}',\lambda' \rightarrow \mathbf{x},\lambda)}$$

$$= \frac{\min[1, e^{-\beta[\tilde{H}(\mathbf{x},\lambda) - \tilde{H}(\mathbf{x}',\lambda')]}]}{\min[1, e^{-\beta[\tilde{H}(\mathbf{x}',\lambda') - \tilde{H}(\mathbf{x},\lambda)]}]}$$

$$= \frac{e^{-\beta \tilde{H}(\mathbf{x}',\lambda')}}{e^{-\beta \tilde{H}(\mathbf{x},\lambda)}} \tag{15}$$

Therefore, the Metropolis construct satisfies eq 10. It is also possible to use a stochastic propagator for the neMD trajectory and satisfy detailed balance by using a Metropolis acceptance criterion based on the external work.[41] The switching trajectory when attempting a change in protonation state has some analogies with the short TI calculations carried out to evaluate the free energy difference associated with the change in protonation state in the method proposed early on by Bürgi et al.[9] However, the Metropolis criterion applied to the discrete end state in the present hybrid neMD−MC algorithm formally resolve the issues arising from the poorly converged thermodynamic integration and guarantees that an equilibrium Boltzmann distribution will be achieved.

As illustrated in Figure 2, the constant-pH simulation scheme with dynamical state switching comprises multiple rounds, each of which is composed of two stages of MD simulation. The first stage is simply a standard MD simulation where the system is kept in a given protonation state λ (red line). The second stage is a neMD switching simulation where the state of the system is gradually switched from one to another (blue dashed line). At the end of the second stage, a Metropolis MC criterion is applied to determine if the switching will be accepted (open black circle). If it is accepted, then the next MD simulation will start with the last configuration from the switching simulation (red line again); otherwise, the switching simulation will be discarded, and the next round simulation will start with the last configuration from the MD stage.

II.2. Two-Step Inherent pK_a Algorithm. If the pH differs markedly from the inherent pK_a of an ionizable group, then the probability of the protonated state is expected to be either very close to 0 or 1. For example, if an aspartic residue is in an environment at a high pH, then the side chain should be essentially always unprotonated. As a consequence, almost all attempts to protonate the side chain ought to be rejected. Similarly, if an arginine residue is in an environment at a pH of 1, then the side chain should be predominantly protonated. Almost all attempts to deprotonate the side chain will be rejected. In such situations, generating a large number of computationally expensive neMD switching trajectories can become wasteful and inefficient. It is possible, however, to redesign the algorithm to take advantage of the information about the inherent pK_a of the site.

The idea is to separate the transition probability T into two distinct steps

$$T(\mathbf{x},\lambda \rightarrow \mathbf{x}',\lambda') = T^{(i)}(\lambda \rightarrow \lambda') \, T^{(s)}(\mathbf{x} \rightarrow \mathbf{x}'|\lambda \rightarrow \lambda') \tag{16}$$

where $T^{(i)}$ represents the transition probability for λ describing a protonation/deprotonation process according to some *inherent* p$K_a^{(i)}$ of the site and $T^{(s)}$ is the transition probability of an attempted $\mathbf{x} \rightarrow \mathbf{x}'$ neMD *switch*, conditional on the transition $\lambda \rightarrow \lambda'$ taking place. The transition probability of the first step, $T^{(i)}$, is

constructed such that it obeys the following microscopic detailed balance relation

$$\pi^{(i)}(\lambda) \, T^{(i)}(\lambda \rightarrow \lambda') = \pi^{(i)}(\lambda') \, T^{(i)}(\lambda' \rightarrow \lambda) \tag{17}$$

where $\pi^{(i)}(\lambda)$ is some arbitrarily assigned inherent equilibrium probability. Assuming an inherent p$K_a^{(i)}$ for the site, the equilibrium probability ratio for the first step is

$$\frac{\pi^{(i)}(\lambda')}{\pi^{(i)}(\lambda)} = 10^{(\lambda'-\lambda)(\mathrm{p}K_a^{(i)}-\mathrm{pH})} \tag{18}$$

The first step does not involve any changes in the atomic coordinates of the system; the process affects only the value of the ionization state variable λ. Complex changes in the atomic coordinates of the system are absorbed into the second step. To make progress, it is useful to rewrite the transition probability $T^{(s)}$ as the product of the probability of a *proposed* move $T_p^{(s)}$ and the probability to *accept or reject* the proposed move $T_a^{(s)}$

$$T^{(s)}(\mathbf{x} \rightarrow \mathbf{x}'|\lambda \rightarrow \lambda') = T_p^{(s)}(\mathbf{x} \rightarrow \mathbf{x}'|\lambda \rightarrow \lambda')$$
$$T_a^{(s)}(\mathbf{x} \rightarrow \mathbf{x}'|\lambda \rightarrow \lambda') \tag{19}$$

Imposing that the two-step scheme satisfies the global detailed balance relation

$$\frac{T_a^{(s)}(\mathbf{x} \rightarrow \mathbf{x}'|\lambda \rightarrow \lambda')}{T_a^{(s)}(\mathbf{x}' \rightarrow \mathbf{x}|\lambda' \rightarrow \lambda)}$$
$$= \frac{\rho_{eq}(\mathbf{x}',\lambda')}{\rho_{eq}(\mathbf{x},\lambda)} \frac{T_p^{(s)}(\mathbf{x}' \rightarrow \mathbf{x}|\lambda' \rightarrow \lambda)}{T_p^{(s)}(\mathbf{x} \rightarrow \mathbf{x}'|\lambda \rightarrow \lambda')} \frac{T^{(i)}(\lambda' \rightarrow \lambda)}{T^{(i)}(\lambda \rightarrow \lambda')} \tag{20}$$

and assuming that the proposed move in $\mathbf{x}$ are inherently symmetric and deterministic

$$\frac{T_p^{(s)}(\mathbf{x}' \rightarrow \mathbf{x}|\lambda' \rightarrow \lambda)}{T_p^{(s)}(\mathbf{x} \rightarrow \mathbf{x}'|\lambda \rightarrow \lambda')} = 1 \tag{21}$$

we obtain the relation

$$\frac{T_a^{(s)}(\mathbf{x} \rightarrow \mathbf{x}'|\lambda \rightarrow \lambda')}{T_a^{(s)}(\mathbf{x}' \rightarrow \mathbf{x}|\lambda' \rightarrow \lambda)} = \frac{\rho_{eq}(\mathbf{x}',\lambda')}{\rho_{eq}(\mathbf{x},\lambda)} \frac{\pi^{(i)}(\lambda)}{\pi^{(i)}(\lambda')}$$

$$= \frac{e^{-\beta \tilde{H}(\mathbf{x}',\lambda')}}{e^{-\beta \tilde{H}(\mathbf{x},\lambda)}} 10^{(\lambda-\lambda')(\mathrm{p}K_a^{(i)}-\mathrm{pH})}$$

$$= \frac{e^{-\beta[\tilde{H}(\mathbf{x}',\lambda') + \lambda' k_B T \ln(10)(\mathrm{p}K_a^{(i)}-\mathrm{pH})]}}{e^{-\beta[\tilde{H}(\mathbf{x},\lambda) + \lambda k_B T \ln(10)(\mathrm{p}K_a^{(i)}-\mathrm{pH})]}}$$

$$= \frac{e^{-\beta \tilde{H}^*(\mathbf{x}',\lambda')}}{e^{-\beta \tilde{H}^*(\mathbf{x},\lambda)}} \tag{22}$$

where the modified Hamiltonian $\tilde{H}^*(\mathbf{x},\lambda)$, defined as

$$\tilde{H}^*(\mathbf{x},\lambda) = K(\mathbf{p}) + \lambda U_p(\mathbf{r}) + (1-\lambda)U_u(\mathbf{r})$$
$$- \lambda(C - k_B T \ln(10)\mathrm{p}K_a^{(i)}) \tag{23}$$

is shifted by p$K_a^{(i)}$. In the two-step neMD−MC scheme, the functional form of the shifted Metropolis acceptance probability $T_a^{(s)}$, is

$$T_a^{(s)}(\mathbf{x}' \rightarrow \mathbf{x}|\lambda' \rightarrow \lambda) = \min[1, e^{-\beta[\tilde{H}^*(\mathbf{x}',\lambda') - \tilde{H}^*(\mathbf{x},\lambda)]}] \tag{24}$$

With respect to the first step, a number of methods could be used to generate the random walk in λ. As this is a simple two-state system, one might even generate the transition $\lambda \rightarrow \lambda'$ directly from the equilibrium probabilities. Interestingly, it can be shown that a simple Metropolis MC acceptance probability

$$T^{(i)}(\lambda \rightarrow \lambda') = \min[1, 10^{(\lambda'-\lambda)(pK_a^{(i)}-pH)}] \qquad (25)$$

provides the most effective method to simulate the first step (see Appendix I).

The overall efficiency of the two-step scheme reaches its highest level when the acceptance probability is equivalent for the protonation and deprotonation processes

$$\langle T^{(s)}(\mathbf{x},\lambda \rightarrow \mathbf{x}',\lambda')\rangle_\lambda = \langle T^{(s)}(\mathbf{x}',\lambda' \rightarrow \mathbf{x},\lambda)\rangle_{\lambda'} \qquad (26)$$

where $\langle...\rangle_\lambda$ denotes an average over all possible transition paths, given the initial configuration is sampled from Boltzmann distribution eq 9 with λ. Using the actual pK_a value of the site (if it were known) would always result in the highest efficiency from an importance sampling point of view. This is demonstrated in the following derivation.

$$
\begin{aligned}
\frac{\langle T^{(s)}(\mathbf{x},\lambda \rightarrow \mathbf{x}',\lambda')\rangle_\lambda}{\langle T^{(s)}(\mathbf{x}',\lambda' \rightarrow \mathbf{x},\lambda)\rangle_{\lambda'}} &= e^{-\beta[\tilde{G}^*(\lambda')-\tilde{G}^*(\lambda)]} \\
&= e^{-\beta[\Delta G_{\lambda\rightarrow\lambda'}+(\lambda'-\lambda)(C-k_BT\ln(10)pK_{a_{exp}})]} \\
&= e^{-\beta\cdot 0} \\
&= 1
\end{aligned}
$$
$$(27)$$

where the third equation results from the definition of C, eq 3.

In practice, the actual pK_a value of the site is not known, but it should be possible to choose some reasonable value capturing the dominant effects. For example, one could use the pK_a of the reference system $pK_a^{(ref)}$. In this case, the overall efficiency will be close to optimal if the pK_a of the site is shifted only by a small amount. Alternatively, it is possible to adaptively adjust the value of $pK_a^{(i)}$ during the simulation based on an earlier estimate (see Discussion). Ultimately, the two-step scheme is rigorously valid, although a well-chosen value for $pK_a^{(i)}$ can greatly increase the overall efficiency of the simulation.

For systems with multiple ionizable sites, the total cost of the constant-pH neMD−MC simulation scales linearly with the total number of sites if one attempts to change the ionization state of one site at a time. However, it might be sometimes advantageous to attempt changing the ionization state of multiple sites simultaneously. For example, when several titratable sites are in close proximity, attempts to change the ionization state of a single site at a time may be inefficient. In this case, attempts to protonate one site while deprotonating another site may be more advantageous. It is possible to include such two-sites processes in the MC move set to reach optimal efficiency.

II.3. Empirical Offset Constants. In practice, there are two stages for implementing a constant-pH hybrid simulation method. In the first stage, the free energy ΔG_{aa} of the protonated state relative to the unprotonated state, defined in eq 1, must be calculated for all the ionizable amino acids (aa) using alchemical FEP simulations with explicit solvent. Then, in a second stage, the empirical offset constant C_{aa}

$$C_{aa} \equiv k_BT \ln(10)pK_{a_{aa}} + \Delta G_{aa} \qquad (28)$$

must be used to shift the extended pH-dependent MM potential energy

$$
\begin{aligned}
\tilde{U}(\mathbf{r},\lambda) &= \lambda(U_p(\mathbf{r}) - C_{aa} + k_BT \ln(10)pH) \\
&\quad + (1-\lambda)U_u(\mathbf{r})
\end{aligned}
$$
$$(29)$$

The latter MM potential energy $\tilde{U}(\mathbf{r},\lambda)$ must then be used to carry out the constant-pH hybrid neMD−MC simulations of the protein with explicit solvent. In practice, the proton is treated as a dummy noninteracting particle for the unprotonated state (denoted H^0 in the following). That is, the dummy H atom is never annihilated, but its interaction with the environment is turned off (Figure 1). In this alchemical process, a number of internal MM covalent terms (bonds, angles, dihedrals) are kept to avoid the problem of a wandering noninteracting free proton in the simulation;[44−46] their influence cancels out in the treatment of the standard state.

The actual value of the empirical offset constant C_{aa} depends on the simulation protocol used to carry out the alchemical FEP calculation yielding the ΔG_{aa}. One possibility is to treat only the ionizable entity in the explicit solvent simulation box. This has the consequence that charge neutrality is not maintained during the protonation/deprotonation process. In this case, the empirical offset constant C_{aa} must be based on ΔG_{aa}, also calculated from an alchemical FEP simulation with just the amino acid in the explicit solvent box, with no constraint on charge neutrality. Alternatively, charge neutrality could be strictly enforced during all of the protonation/deprotonation processes by including some charge-canceling counterpart in the simulation. Maintaining charge neutrality is a useful way to avoid some important simulation artifacts associated with the free energy to insert or remove net charges in the system. With Ewald lattice summation and tinfoil conducting boundary conditions, such charging free energies necessarily include a spurious shift due to the Galvani potential of the bulk water phase in the finite simulation system.[47] The magnitude of such artifact depends on the details of the simulation system, such as the volume fraction occupied by the solvent.[47] While the issue may be safely ignored in the case of a solvated protein surrounded by a large number of water molecules, it can become much more severe when the volume fraction occupied by the solvent is much less than 85% of the entire system, a condition often encountered with simulation of membrane proteins.
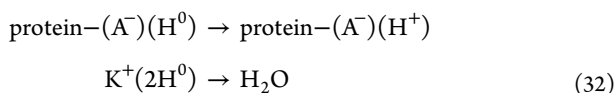
In the present study, the alchemical protocol that we use couples the following reactions

$$\text{protein}-(A^-)(H^0) \rightarrow \text{protein}-(A^-)(H^+)$$
$$H_2O \rightarrow Cl^-(2H^0) \qquad (30)$$
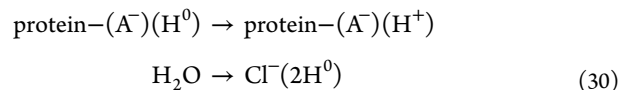
for the protonation of acidic residues (Asp, Glu, Tyr, COO-terminal) and

$$\text{protein}-B(H^0) \rightarrow \text{protein}-B(H^+)$$
$$K^+(2H^0) \rightarrow H_2O \qquad (31)$$
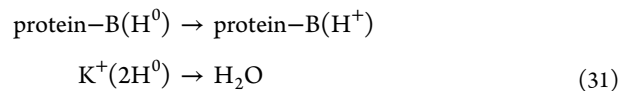
for the protonation of basic residues (His, Lys, Arg, N-terminal). Here, H^0 represents the dummy noninteracting hydrogen. In these alchemical transformations, the oxygen atom of the water molecule is converted into K^+ or a Cl^- ion while the two hydrogen atoms are transformed into dummy particles. It is worth noting that charge neutrality could also be maintained the simultaneous creation or destruction of a pair of opposite charges via

$$\text{protein}-(A^-)(H^0) \rightarrow \text{protein}-(A^-)(H^+)$$

$$K^+(2H^0) \rightarrow H_2O \tag{32}$$

for the protonation of acidic residues and

$$\text{protein}-(B)(H^0) \rightarrow \text{protein}-(B)(H^+)$$

$$H_2O \rightarrow Cl^-(2H^0) \tag{33}$$
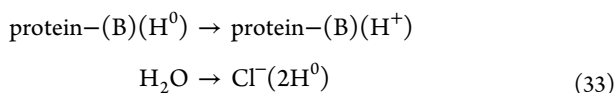
for the protonation of basic residues. However, this second protocol is less advantageous because it requires the creation of opposite charges during the alchemical switch process, yielding large energy differences that affect the overall Metropolis acceptance probability. The first protocol based on eqs 30 and 31 yields more efficient hybrid MD−MC simulations because the two charging free energies appear in opposite directions of the alchemical processes.

III. COMPUTATIONAL DETAILS

Three dipeptides were simulated with the hybrid algorithm: aspartic acid, glutamic acid, and histidine. In addition, two protein systems were simulated with the hybrid algorithm: the turkey ovomucoid third domain (OMTKY3, PDB ID 1OMU) and hen egg-white lysozyme (HEWL, PDB ID 1AKI). The amino acid dipeptides have acetylated N-terminus and *N*-methylamide C-terminus, whereas OMTKY3 and HEWL have a standard charged N- and C-terminus. All systems were simulated with explicit solvent with periodic boundary conditions. The CHARMM27 force field[48] with the TIP3 water potential[49] was used to model the microscopic interactions. The amino acid dipeptides were solvated in a 24 Å cubic box. OMTKY3 and HEWL were solvated in a 46 and 60 Å cubic boxes, respectively. To prevent any drift, the center-of-mass of the protein (or dipeptide) was weakly restrained to the center of the periodic box with a harmonic potential. Particle-mesh Ewald (PME) summation[50] was used to treat the electrostatic interactions, with a real-space cutoff set to 14 Å and grid spacing smaller than 0.5 Å. The LJ interactions were smoothly truncated with a switching function from 10 to 12 Å. The equations of motion were integrated with a 2 fs time step, and SHAKE[51] was used to constrain covalent bonds involving hydrogen atoms. The program CHARMM, version c36b1, was used to carry out all of the MD simulations.[52]

Special patches were created to handle the switch between the protonated and unprotonated states of the amino acids. Those allow alterations to the protein-structure-file (PSF) that represent the atomic model underlying a given simulated system.[52] In the case of aspartic and glutamic acids, two dummy hydrogen are introduced to allow the two carboxylate oxygen atoms to be protonated. In the unprotonated state, the acidic group has two dummy hydrogens, one on each carboxylate oxygen. When the protonated state is generated, one of the two dummy hydrogen particles is alchemically transformed into a fully interacting hydrogen atom. In the case of histidine, both nitrogen atoms δ and ϵ can be protonated. To determine the offset constant ΔG_{aa} of the ionizable sites, FEP/MD simulations of 10 ns were carried out for the dipeptide systems. WHAM was used to calculate the free energy difference between the two ionizable states.[53] The FEP simulation provides the reference values C_{aa} for aspartic acid, glutamic acid, and histidine that are then used for the OMTKY3 and HEWL simulations.

To set charge neutrality and maintain this neutrality during changes in ionization states, a number of special counterions able to alchemically interconvert into a water molecule were included in the simulation system. Similar procedures have been used previously in free energy perturbation (FEP) calculations involving charged species.[47,54] The number of these counterion−water "molecules" is adjusted to correspond to the number of ionizable sites. These counterion−water molecules were harmonically restrained in the bulk solvent away from the protein to ensure that this alchemical transformation occurs in the bulk region and never in close contact to the protein. Their restrained positions were generated by a script that sought to maximize these distances as much as possible. In the dipeptide systems, the distance of the counterion−water molecule to the ionizable site is 13 Å. In the OMTKY3 system, five counterion−water molecules were included; the shortest distance among them, and with ionizable protein sites, is 16.0 Å. In the HEWL system, 16 counterion−water molecules were included; the shortest distance among them, and with ionizable protein sites, is 17.5 Å.

In-house Python scripts were used for controlling the hybrid neMD−MC simulation scheme. In the hybrid algorithm, the equilibrium MD simulations are carried out in the NPT ensemble at a temperature of 300 K and 1 atm pressure, whereas the neMD switching simulations are carried out in the NVE ensemble with no constant temperature or pressure control. The PERT command of CHARMM is used for switching the Hamiltonian.[52] For peptides and proteins with multiple ionizable sites, a maximum of one protonation site is perturbed per cycle. For each cycle, one site is randomly picked. One may also assign weight for each site so that important sites can be perturbed more often. The expression $T^{(i)}$ in eq 25 is used to decide if a switch will be performed for this site. If it is not, then another site will be picked, until one acceptance is obtained or all sites have been visited. The neMD−MC switch is skipped for this cycle if none of the transition trials are accepted. After the neMD switching trajectory, the energy difference is calculated. The expression $T^{(s)}$ in eq 24 is used to accept or reject the new proposed state. If the state is accepted, then the new conformation and velocities are kept; otherwise, the configuration from the neMD switching simulation is discarded, and the simulation is continued from the position and momenta at the end of the MD stage. After the neMD−MC step, the configuration and state of the system is recorded. For the neMD switches, the momenta of all particles in the system are treated according to a symmetric two-ends momentum reversal prescription.[39,40] Accordingly, momenta have a probability of 0.5 to be kept unchanged and a probability of 0.5 to be reversed both before and after the switch. This prescription can greatly reduce the probability of different regions of configurational space being isolated from one another.[40] Protonation states were randomly picked for each residue to minimize the impact of the initial state to initiate the system. A first round of MD and neMD was then carried out, and the first generated configuration was accepted unconditionally.

During the hybrid neMD−MC simulation, a counterion−water molecule is randomly picked together with the ionizable state of a protein site for subsequent perturbation into K^+ or Cl^- ions according to eq 30 and eq 31. During the protonation process of the histidine, a $K−H_2O$ is perturbed into neutral H_2O according to eq 31. In case the system contains no $K−H_2O$, a neutral H_2O is perturbed into $Cl−H_2O$ according to eq 33. Similarly, during the deprotonation process of aspartic acid or glutamate, a $Cl−H_2O$ is perturbed into neutral H_2O according to eq 30. In case the system contains no $Cl−H_2O$, a neutral H_2O is perturbed into $K−H_2O$ according to eq 32. For the three dipeptide systems, the counterion−water molecule and the

ionizable site are coupled via eqs 32 and 33. In OMTKY3, the ionization of the five acidic residues (Asp7, Asp27, Glu10, Glu19, Glu43) is coupled to a counterion−water molecule via eq 30 (there are no basic sites). In HEWL, the ionization of nine acidic residues (Glu7, Asp18, Glu35, Asp48, Asp52, Asp66, Asp87, Asp101, and Asp119) is coupled to a counterion−water molecule via eq 30, and the ionization of His15 is coupled to a counterion−water molecule via eq 31.

In the following analysis, we report the net computational cost of the hybrid simulation by adding the number of time steps used for both the equilibrium MD and neMD simulations. For the three dipeptide systems, a total of 15 hybrid neMD−MC simulations was generated to calculate the statistical uncertainty of the results. The neMD switching trajectories were set to 10 ps, whereas the equilibrium MD trajectories were set to a very short length of only 0.2 ps. Such short equilibrium trajectories were used to reveal any possible sampling issues arising from the neMD switches. The total length of the hybrid simulations, calculated by accounting for the equilibrium MD and the neMD switching trajectory of each cycle, is 40 ns. In the case of the aspartic acid dipeptide, a scan over nine pH values was also carried out (with 0.2 ps of equilibrium MD and 10 ps of neMD for the switching per cycle). Again, a total of 15 trajectories was generated to calculate the statistical uncertainty of the results. For the OMTKY3 system, a total of seven sets of hybrid simulations with different parameters was carried out (each was carried out as 30 separate simulations with randomized initial conditions to reduce the computational time). Simulations 1, 5, and 6 comprised 15 000 cycles, and simulations 2, 3, 4, and 7 comprised 10 000 cycles. Each cycle of hybrid simulation included 10 ps of equilibrium MD and 20 ps of neMD for the switching. For simulation 7, the length of the neMD trajectory was 50 ps. Simulation 7 is the longest trajectory, for a total length of 600 ns (including equilibrium and switching contributions). Simulations 1, 2, 5, 6, and 7 were carried out at pH of 4, and simulations 3 and 4 were carried out at pH of 3. For simulation 1, 3, 5, and 7, the value of $pK_a^{(i)}$ was set to the experimental pK_a. For simulations 2, 4, and 6, the value of $pK_a^{(i)}$ was set to $pK_a^{(ref)}$. To test if this affects convergence, the proton could be added to a single oxygen of the carboxylate group only in simulations 5 and 6.

A single set of hybrid neMD−MC simulation was generated for the HEWL system (separated into 30 simulations with randomized initial conditions to reduce the computational time). Each cycle of hybrid simulation included 10 ps of equilibrium MD and 50 ps of neMD for the switching. The total computational cost of the hybrid simulation is 1.1 μs, including the equilibrium and switching contributions, corresponding to 18 838 cycles of neMD−MC.

IV. RESULTS AND DISCUSSION

IV.1. Free Energy Difference Calculations.
All of the calibration free energy difference ΔG_{aa} and constant C_{aa} values for the different reactions used in the present hybrid simulations are given in Table 1 (the dummy H^0 are not explicitly included for the sake of clarity). The upper half presents reactions for which charge neutrality are not maintained, whereas the lower half presents those for which charge neutrality is maintained. The difference comes from the ΔG of 81.2 kcal/mol for perturbing a K^+ ion into a neutral H_2O molecule, and 69.2 kcal/mol for perturbing a Cl^- ion into a neutral H_2O molecule. The quantities ΔG_{aa} and C_{aa} for the second protocol based on eqs 32 and 33 can

Table 1. Free Energy Difference ΔG_{aa} and Constant C_{aa} for Different Reactions[a]

reaction	ΔG_{aa} (kcal/mol)	C_{aa} (kcal/mol)
$Asp^- \rightarrow Asp^-(H^+)$	45.5	40.0
$Glu^- \rightarrow Glu^-(H^+)$	49.0	43.0
$His \rightarrow His(H_{\delta N}^+)$	−4.2	−12.4
$His \rightarrow His(H_{\epsilon N}^+)$	−18.8	−27.0
$Asp^- + H_2O \rightarrow Asp^-(H^+) + Cl^-$	−23.7	−29.2
$Glu^- + H_2O \rightarrow Glu^-(H^+) + Cl^-$	−20.2	−26.2
$His + K^+ \rightarrow His(H_{\delta N}^+) + H_2O$	77.0	68.8
$His + K^+ \rightarrow His(H_{\epsilon N}^+) + H_2O$	62.4	54.2

[a]Dummy H^0 particles are left out for the sake of clarity.

also be calculated in a similar manner (values not shown in Table 1).

IV.2. Validation of the Hybrid neMD−MC Scheme.
For each test case, 15 independent hybrid simulations were generated to estimate the average and error. One configuration and state are recorded per round after the neMD−MC switch. The pK_a of each protonation site is calculated as

$$pK_a = pH + \log\left[\frac{p_1}{p_0}\right] \tag{34}$$

where p_1 and p_0 are the probability of the protonated and deprotonated states being extracted from the simulation, respectively. To further validate the neMD−MC algorithm, the conditional density for a given λ was examined to see if it is internally consistent based on the Crooks identity.[55] The latter depicts the relationship of the work evaluated from forward and backward transitions.[55] Here, the forward (f) transition is defined as the protonation process, $\lambda = 0 \rightarrow 1$, and the backward (b) transition is defined as the deprotonation process, $\lambda = 1 \rightarrow 0$. When the initial configuration is sampled from an equilibrated ensemble, the Crooks equation states that

$$\frac{P_f(+W)}{P_b(-W)} = e^{-(\Delta G - W)/k_B T} \tag{35}$$

where P_f is the forward work probability distribution, P_b is the backward work probability distribution, and W is the external input work for the switch. W is equivalent to ΔH if a deterministic switch is used.[41] Equation 35 can be transformed into

$$k_B T \ln\left[\frac{P_f(+W)}{P_b(-W)}\right] = W - \Delta G \tag{36}$$

Let us define

$$Q(W) = k_B T \ln\left[\frac{P_f(+W)}{P_b(-W)}\right] \tag{37}$$

Then, eq 36 becomes

$$Q(W) = W - \Delta G \tag{38}$$

We fitted $Q(W)$ and W observed from simulations to eq 38 and calculated the coefficient of determination (R^2). If the distribution of starting configuration is the equilibrated ensemble, then R^2 should be close to 1. Therefore, this verifies that the conditional density for a given λ is correct by examining R^2.

IV.3. Isolated Dipeptides. The hybrid neMD−MC scheme was applied to isolated dipeptides in explicit water. The pK_a is calculated using eq 34. For all three dipeptides, the error between the experimental and calculated pK_a is less than 0.3. The evolution of the calculated pK_a and its standard deviation for an aspartic acid residue is plotted in Figure 3. The evolution for
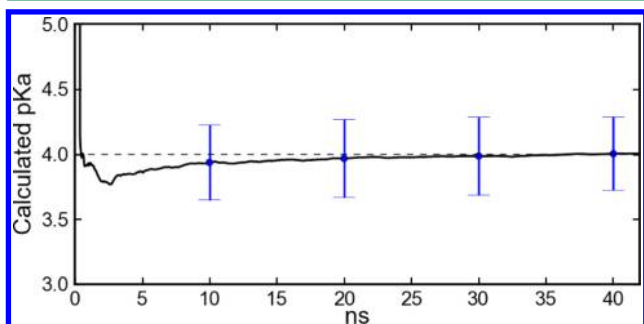

Figure 3. Evolution of the calculated pK_a for the aspartic dipeptide in water. The flat dashed line is the experimental pK_a of the system. The black line is calculated from the time average of all 15 trajectories, and the error bars show the standard deviation. Each round of hybrid neMD−MC simulation uses 0.2 ps for the equilibrium MD and 10 ps for the neMD switch. The time axis reflects the total computational cost of the hybrid simulation by including the number of time steps used for the equilibrium MD and for the neMD simulation.

glutamate acid and histidine is similar (not shown). The results show that the calculated pK_a converges toward the experimental pK_a. A scan over nine pH values was also carried out for an aspartic acid dipeptide in water. The results are shown in Figure 4. The percentage of protonated states calculated for pH varying from 2 to 6 agrees well with the theoretical curve.
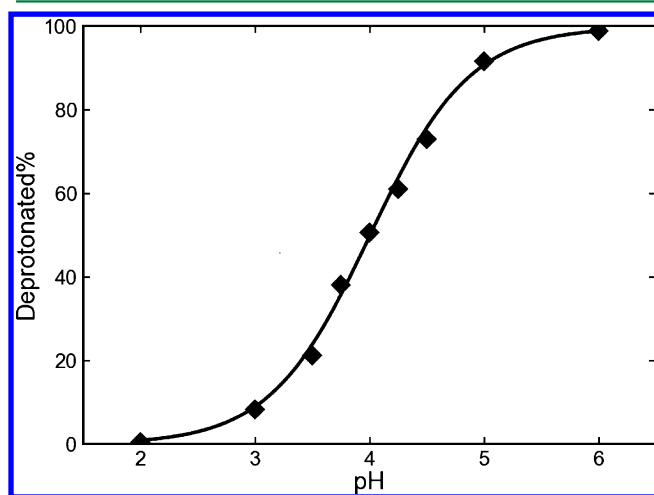

Figure 4. pH scan for the aspartic dipeptide in water. Constant-pH simulations are carried out at pH values of 2, 3, 3.5, 3.75, 4, 4.25, 4.5, 5, and 6. Each point is calculated from the time average of 15 trajectories to present the percentage of protonated states for that pH value. The solid curve presents the theoretical values.

The distribution of the energy difference of all switches for the aspartic acid dipeptide is plotted in Figure 5a. The relation between $Q(W)$ and W is plotted in Figure 5b. The coefficient of determination when fitting $Q(W)$ and W to eq 38 is 0.99. On the basis of this result, it is safe to conclude that the Crooks equation is satisfied here and that the initial configuration is drawn from an
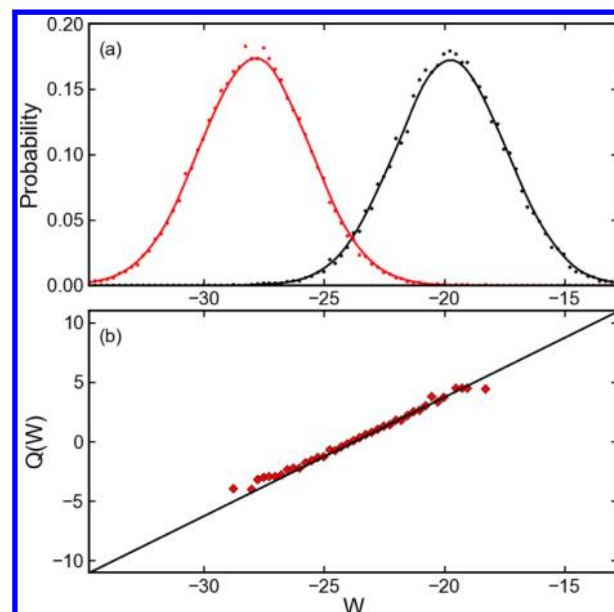

Figure 5. (a) Distribution of the energy difference for the protonation process (black) and deprotonation process (red) for the aspartic acid dipeptide. The sign is flipped for the deprotonation process. The distribution is plotted with dots using 60 000 data points. The solid curve presents the fitted distribution using the Nadaraya−Watson kernel regression estimate with bandwidth of 0.5. (b) The relationship between of $Q(W)$ and W is displayed. $Q(W)$ is defined as eq 38. The dots represent the simulation results, and the solid line presents the theoretical line. The coefficient of determination is 0.99 when fitting the data to the theoretical line. The dots represent all observable $Q(W)$; for other W, either $P_f(+W)$ or $P_b(-W)$ is zero.

equilibrated ensemble. The average acceptance ratio using different switching time for aspartic acid dipeptide is plotted in Figure 6. As expected, it is observed that the acceptance ratio
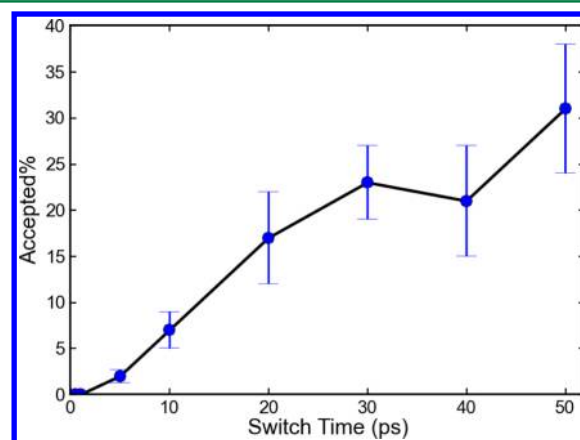
Figure 6. Acceptance ratio vs switching time for isolated aspartic acid residue. Switching time varies from 0.5 to 50 ps. Each point is calculated from the average of five trajectories, and the error bars show the standard deviation.

increases steadily when using longer switching time. Longer switching times increase the computational cost, but this is mitigated by a larger acceptance probability. In principle, it should be possible to optimize the switching time to maximize the efficiency of the algorithm.

IV.4. Multisite Systems. The present constant-pH hybrid neMD−MC scheme was applied to multisites protein systems

Table 2. Simulation Results for OMTKY3[a]

simulation[b]	Asp7	Asp27	Glu10	Glu19	Glu43	error[c]
1	3.5 (0.8)[d]	3.6 (1.3)	4.2 (0.1)	3.6 (0.4)	4.4 (−0.4)	0.73
2	3.5 (0.8)	4.3 (2.0)	4.2 (0.1)	3.5 (0.3)	4.2 (−0.6)	1.01
3	3.1 (0.4)	4.0 (1.7)	3.8 (−0.3)	3.4 (0.2)	4.7 (−0.1)	0.80
4	3.2 (0.5)	5.5 (3.2)	4.0 (−0.1)	3.4 (0.2)	4.1 (−0.7)	1.49
5	3.7 (1.0)	4.4 (2.1)	4.0 (−0.1)	3.5 (0.3)	4.4 (−0.4)	1.06
6	3.6 (0.9)	4.6 (2.3)	4.1 (0.0)	3.7 (0.5)	4.5 (−0.3)	1.13
7	3.4(0.7)	3.5 (1.2)	4.0 (−0.1)	3.6 (0.4)	4.4 (−0.4)	0.67
$pK_a^{(exp)}$	2.7	2.3	4.1	3.2	4.8	
$pK_a^{(ref)}$	4.0 (1.3)	4.0 (1.7)	4.4 (0.3)	4.4 (1.2)	4.4 (−0.4)	1.12

[a]Experimental values from Schaller and Robertson.[57] The solution contains 10 mM KCl. [b]Seven sets of simulations with different parameters: simulations 1, 5, and 6 comprised 15 000 cycles, and simulations 2, 3, 4, and 7 comprised 10 000 cycles; each cycle consisted of 10 ps MD simulation with 20 ps neMD switching simulation (except for simulation 7, for which the switching length is 50 ps); accordingly, the longest trajectory, simulation 7, has a total length of 600 ns. Simulations 1, 2, 5, 6, and 7 were carried out at pH 4, and simulations 3 and 4 were carried out at pH 3; for simulations 1, 3, 5, and 7, $pK_a^{(i)}$ was set to the experimental pK_a; for simulations 2, 4, and 6, $pK_a^{(i)}$ was set to $pK_a^{(ref)}$. The proton could be added to a single oxygen only in simulations 5 and 6. [c]The number in the parentheses shows the deviation from experimental value. [d]Root-mean-square deviation (RMSD) error relative to the experimental values.

Table 3. Simulation Results for HEWL

	experiment[a]	simulation[b]	error	trial	accepted	%
Glu7	2.85	2.86	0.01	2215	273	12.3
His15	5.36	3.85	1.51	4024	160	4.0
Asp18	2.66	2.74	0.08	1140	45	3.9
Glu35	6.20	3.98	2.22	641	29	4.5
Asp48	1.60	1.41	0.19	135	8	5.9
Asp52	3.68	3.99	0.31	3312	39	1.2
Asp66	0.90	0.83	0.07	39	2	5.1
Asp87	2.07	3.03	0.96	1276	126	9.9
Asp101	4.09	3.57	0.52	3468	111	3.2
Asp119	3.20	3.06	0.14	2588	134	5.2
Average			0.93[c]	1884[d]	93	5.5

[a]Experimental results from Bartik et al.[58] [b]Experimental pK_a is used as $pK_a^{(i)}$. [c]RMSD error from experimental value. [d]Each cycle of simulation consisted of 10 ps MD simulation with 50 ps neMD switching simulation. The trajectory has a total of 18 838 rounds of simulation. The total length is 1.1 μs.

with explicit solvent. The calculated pK_a's for OMTKY3 are shown in Table 2. The calculated pK_a's for HEWL are shown in Table 3. For OMTKY3, the RMSD error of calculated pK_a from experimental value can get to as low as 0.67 pH units (simulation 7 in Table 2) with a proper choice of the parameters of the algorithm. The error is significantly smaller than previous results from λ-dynamics[20] and is also smaller than the error assuming the reference pK_a. For HEWL, the RMSD error of calculated pK_a's compared to experiment is 0.93 (Table 3). These results match the best performance from other existing methods.[9−11,20,21,56]

The average acceptance ratio for each residue in the hybrid simulations of OMTKY3 is shown in Table 4. For simulation 6 in Table 2, the average acceptance ratio of Glu19 for either protonation or deprotonation process is calculated. Of around

Table 4. Acceptance Ratio for OMTKY3

simulation	Asp7	Asp27	Glu10	Glu19	Glu43	average
1	5.0	0.8	10.6	3.8	7.2	5.4
2	6.4	1.1	11.2	2.7	9.5	5.6
3	5.5	0.5	7.4	3.9	8.7	4.0
4	4.4	0.5	10.5	2.7	7.6	4.6
5	10.2	1.6	10.9	5.3	10.3	6.5
6	8.7	1.3	8.9	3.3	13.3	6.5
7	8.7	3.2	20.3	11.1	17.1	12.0

2500 trials to protonate the Glu19, only 2.1% are accepted, whereas 10% are accepted of around 500 trials to unprotonated it. This reveals the imbalance of the acceptance ratio between protonation and deprotonation processes. Simulation 5 in Table 2 is carried out in the attempt to reduce such imbalance. It uses the exact same settings as those in simulation 6 except that $pK_a^{(i)}$ is set as the experimental value, 3.2. Roughly 8% of 1000 protonation trials and 4% of 2000 deprotonation trials are accepted. The total number of successful transitions increases 1.6-fold compared with those for simulation 6. This provides a good example of how the overall efficiency of the algorithm can be increased by the choice of the inherent $pK_a^{(i)}$. The average acceptance ratio for each protonation site is calculated for simulation 1 in Table 2; the acceptance ratio can get as high as 10% for Glu10 or as low as 0.8% for Asp27. Especially for Asp27, of over 3000 attempted switches, less than 25 are accepted. Such a low acceptance probability is clearly a source of inefficiency. The calculated pK_a values converge much faster for Glu10 than for Asp27. The low acceptance ratio for Asp27 is also part of the reason that the calculated pK_a for it is not accurate.

In general, the total length needed to converge the hybrid simulation depends on the acceptance ratio of the residue with highest energy barrier opposing a change in the ionization state. Increasing the length of the neMD switching trajectory could help to increase the average acceptance ratio. However, it is not always desirable to use longer neMD switches for all sites. The

data from the hybrid simulation 7 of OMTKY3 in Table 2 was carried out to demonstrate this observation. The simulation uses exactly the same settings as those for hybrid simulation 1 except that neMD switching trajectories of 50 ps were used. As a result, the average acceptance ratio is increased by a factor of 3 for Asp27, a factor less than 2 for Glu10, and a factor less than 2.5 overall. Considering that the switching length is increased by a factor of 2.5, the overall efficiency is actually lower. However, this does not mean that the length of switch is an irrelevant parameter. Empirically, we find that it is preferable to choose a customized length for each site so that the average acceptance probability is close to about 10%. In a follow-up simulation, neMD switching trajectories of 50 ps were used for Asp7 and Glu19, 30 ps for Asp27, and 20 ps for Glu10 and Glu43. The average acceptance ratio versus switching time increased compared to both simulations 1 and 7 in Table 2 (results not shown).

In practice, it may be difficult to pick the best inherent value $pK_a^{(i)}$ or switching time for each residue. In the absence of additional information, the best choice is to use the reference pK_a and a fixed switching length in the early stage of the simulation. The results can then be used to adjust the parameters before pursuing the simulation further. This general idea could be developed into an adaptive version of the algorithm that could, on-the-fly, automatically adjust the $pK_a^{(i)}$ and the switching length. This adaptive program would adjust parameters periodically, but at a low frequency, to keep the average acceptance ratio around 10% for each site and for either process. Finally, we would like to emphasize that these simulation parameters affect only the efficiency of the algorithm, but they do not alter the expected outcome of the constant-pH simulations.

V. SUMMARY

A new computational method was developed to carry out explicit solvent simulations of complex molecular systems under conditions of constant pH. Preidentified ionizable sites are allowed to spontaneously protonate and deprotonate as a function of time in response to the environment and the imposed pH. To generate physically plausible configurations with altered protonation states that are subsequently accepted or rejected according to a Metropolis Monte Carlo (MC) criterion, the method consists of carrying out short nonequilibrium molecular dynamics (neMD) switching trajectories. To achieve higher efficiency, the random protonation/deprotonation processes are separated into two steps, reducing the need for generating a large number of unproductive and costly nonequilibrium trajectories. In the first step, the protonation state of a site is randomly attributed via a Metropolis MC process on the basis of an intrinsic pK_a value; a nonequilibrium switch is generated only when this change in protonation state is accepted. This hybrid two-step inherent pK_a neMD–MC simulation method was illustrated with application to turkey ovomucoid third domain and hen egg-white lysozyme. The illustrative results demonstrate that the present method is practical and able to treat multisites of proteins with explicit solvent. While biological systems were used here as a primary motivation, the present method offers a general framework to simulate the effect of pH in a wide range of nonbiological molecular systems and materials.

Charge neutrality of the simulated system is a feature of constant-pH simulation that deserves some special attention. In practice, it is possible to carry out constant-pH simulations without compensating for the change in total charge in the system during the protonation/deprotonation events. With

Ewald lattice sum and tinfoil boundary conditions, the total net charge is always accompanied by a uniform canceling background charge, which makes the simulation cell of the periodic system neutral.[47] However, because the spatial average over the entire simulation box is constrained to be exactly zero by the tinfoil boundary conditions, the Galvani potential of the bulk water phase floats with respect to the standard reference vacuum potential.[47] This phenomenon can give rise to a spurious shift in the free energy of charged moieties that is proportional to the volume fraction of the solvent in the simulation box, which would affect the apparent pK_a's of the sites attached to a solvated protein. This issue can be avoided by keeping the overall charge of the system neutral during the protonation/deprotonation events by introducing counter-reactions involving the transformation of an ion into a water molecules. However, the fluctuations in the energy become larger due to the counter-reactions, which decreases the acceptance probability and the efficiency of the algorithm. It is possible to compensate by adjusting the length of the neMD switching trajectories to restore the efficiency of the method (Figure 6).

Because the sampling time is expected to grow linearly with the number of titratable sites, a fundamental advantage of the present method is its natural ability to scale to extremely large systems. A number of enhancements of the method will be considered in the near future. For example, while the algorithm considered only independent protonation and deprotonation attempts in the applications presented here, it is possible that including correlated attempts would be useful, particularly when ionizable sites are strongly coupled. For example, transferring the proton from one site to another when several ionizable sites are in close proximity from one another may yield a higher acceptance probability and increase the overall efficiency of the algorithm. Such correlated attempts could be easily included in the MC move set on the basis of distance with no extra computational cost. Another feature of interest will be to enable a Hamiltonian replica-exchange treatment with multiple pH values to expand the scope of the method.

■ APPENDIX

Appendix: Relaxation Time Analysis for the Two-Step Scheme

Let us assume that the ratio of the two states in eq 18 is r

$$\frac{\pi^{(i)}(\lambda')}{\pi^{(i)}(\lambda)} = r \tag{39}$$

with $r > 1$, and average acceptance $T^{(s)}$ for both protonation and deprotonation process is p

$$T^{(s)}(\mathbf{x}' \rightarrow \mathbf{x}|\lambda' \rightarrow \lambda) = p \tag{40}$$

The total transition matrix is

$$M = \begin{pmatrix} 1-p & p \\ \dfrac{p}{r} & \dfrac{1-p}{r} \end{pmatrix} \tag{41}$$

The eigenvalue for this matrix is $\lambda_1 = 1$ and $\lambda_2 = (r - p - rp)/r$. Therefore, the relaxing time is

$$t_{relax} = \frac{1}{1 - \lambda_2}$$

$$= \frac{r}{(r + 1)p} \tag{42}$$

For example, if the two states are roughly equiprobable, $r \approx 1$ and $t_{relax} \approx p/2$. Alternatively, instead of using a Metropolis criterion to pick the next protonation state λ, one could pick λ with correct probability, regardless of the current protonation state

$$P(\lambda = 0) = \pi^{(i)}(0)$$

$$P(\lambda = 1) = \pi^{(i)}(1) \tag{43}$$

In this case, the transition matrix is

$$M' = \begin{pmatrix} \dfrac{1 - rp}{r + 1} & \dfrac{rp}{r + 1} \\[2mm] \dfrac{p}{r + 1} & \dfrac{1 - p}{r + 1} \end{pmatrix} \tag{44}$$

The eigenvalue for this matrix is $\lambda_1 = 1$ and $\lambda_2 = 1 - p$, and the relaxing time is

$$t'_{relax} = \frac{1}{p} \tag{45}$$

Because the relaxation time here is always larger than that of the previous scheme, picking the next round's protonation state using the Metropolis criterion is expected to be more efficient.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: roux@uchicago.edu.

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) O'Keefe, D. O.; Cabiaux, V.; Choe, S.; Eisenberg, D.; Collier, R. J. pH-dependent insertion of proteins into membranes: B-chain mutation of diphtheria toxin that inhibits membrane translocation, Glu-349→Lys. *Proc. Natl. Acad. Sci. U.S.A* **1992**, *89*, 6202−6206.

(2) Clippingdale, A. B.; Wade, J. D.; Barrow, C. J. The amyloid-β peptide and its role in Alzheimer's disease. *J. Pept. Sci.* **2001**, *7*, 227−249.

(3) Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*; W.H. Freeman and Company: New York, 1999.

(4) Rastogi, V. K.; Girvin, M. E. Structural changes linked to proton translocation by subunit c of the ATP synthase. *Nature* **1999**, *402*, 263−268.

(5) Feng, L.; Campbell, E. B.; Hsiung, Y.; MacKinnon, R. Structure of a eukaryotic CLC transporter defines an intermediate state in the transport cycle. *Science* **2010**, *330*, 635−641.

(6) Feng, L.; Campbell, E. B.; MacKinnon, R. Molecular mechanism of proton transport in CLC Cl$^-$/H$^+$ exchange transporters. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 11699−11704.

(7) Warshel, A. *Computer Modelling of Chemical Reactions in Enzymes and Solutions*; John Wiley & Sons: New York, 1991.

(8) Field, M. J. *A Practical Introduction to the Simulation of Molecular Systems*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2007.

(9) Bürgi, R.; Kollman, P. A.; van Gunsteren, W. F. Simulating proteins at constant pH: an approach combining molecular dynamics and Monte Carlo simulation. *Proteins: Struct., Funct., Bioinf.* **2002**, *47*, 469−480.

(10) Mongan, J.; Case, D. A.; McCammon, J. A. Constant pH molecular dynamics in generalized Born implicit solvent. *J. Comput. Chem.* **2004**, *25*, 2038−2048.

(11) Williams, S. L.; de Oliveira, C. A. F.; McCammon, J. A. Coupling constant pH molecular dynamics with accelerated molecular dynamics. *J. Chem. Theory Comput.* **2010**, *6*, 560−568.

(12) Shelley, J. C.; Patey, G. N. A configuration bias Monte-Carlo method for ionic-solutions. *J. Chem. Phys.* **1994**, *100*, 8265−8270.

(13) Baptista, A. M.; Teixeira, V. H.; Soares, C. M. Constant-pH molecular dynamics using stochastic titration. *J. Chem. Phys.* **2002**, *117*, 4184−4200.

(14) Meng, Y.; Roitberg, A. E. Constant pH replica exchange molecular dynamics in biomolecules using a discrete protonation model. *J. Chem. Theory Comput.* **2010**, *6*, 1401−1412.

(15) Sabri Dashti, D.; Meng, Y.; Roitberg, A. E. pH-replica exchange molecular dynamics in proteins using a discrete protonation method. *J. Phys. Chem. B* **2012**, *116*, 8805−8811.

(16) Swails, J. M.; York, D. M.; Roitberg, A. E. Constant pH replica exchange molecular dynamics in explicit solvent using discrete protonation states: implementation, testing, and validation. *J. Chem. Theory Comput.* **2014**, *10*, 1341−1352.

(17) Bashford, D.; Karplus, M. pKa's of ionizable groups in proteins: atomic detail from a continuum, electrostatic model. *Biochemistry (Moscow)* **1990**, *29*, 10219−10225.

(18) Alexov, E. G.; Gunner, M. R. Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. *Biophys. J.* **1997**, *72*, 2075.

(19) Georgescu, R. E.; Alexov, E. G.; Gunner, M. R. Combining conformational flexibility and continuum electrostatics for calculating pKa's in proteins. *Biophys. J.* **2002**, *83*, 1731−1748.

(20) Lee, M. S.; Salsbury, F. R.; Brooks, C. L. Constant-pH molecular dynamics using continuous titration coordinates. *Proteins: Struct., Funct., Bioinf.* **2004**, *56*, 738−752.

(21) Khandogin, J.; Brooks, C. L. Toward the accurate first-principles prediction of ionization equilibria in proteins. *Biochemistry (Moscow)* **2006**, *45*, 9363−9373.

(22) Arthur, E. J.; Yesselman, J. D.; Brooks, C. L. Predicting extreme pKa shifts in staphylococcal nuclease mutants with constant pH molecular dynamics. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 3276−3286.

(23) Nikolova, E. N.; Goh, G. B.; Brooks, C. L.; Al-Hashimi, H. M. Characterizing the protonation state of cytosine in transient G C Hoogsteen base pairs in duplex DNA. *J. Am. Chem. Soc.* **2013**, *135*, 6766−6769.

(24) Goh, G. B.; Knight, J. L.; Brooks, C. L. Constant pH molecular dynamics simulations of nucleic acids in explicit solvent. *J. Chem. Theory Comput.* **2012**, *8*, 36−46.

(25) Goh, G. B.; Knight, J. L.; Brooks, C. L. Towards accurate prediction of protonation equilibrium of nucleic acids. *J. Phys. Chem. Lett.* **2013**, *4*, 760−766.

(26) Goh, G. B.; Knight, J. L.; Brooks, C. L. pH-dependent dynamics of complex RNA macromolecules. *J. Chem. Theory Comput.* **2013**, *9*, 935−943.

(27) Goh, G. B.; Hulbert, B. S.; Zhou, H.; Brooks, C. L. Constant pH molecular dynamics of proteins in explicit solvent with proton tautomerism. *Proteins: Struct., Funct., Bioinf.* **2014**, *82*, 1319−1331.

(28) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(29) Tidor, B. Simulated annealing on free-energy surfaces by a combined molecular-dynamics and monte-carlo approach. *J. Phys. Chem.* **1993**, *97*, 1069−1073.

(30) Börjesson, U.; Hünenberger, P. H. Explicit-solvent molecular dynamics simulation at constant pH: methodology and application to small amines. *J. Chem. Phys.* **2001**, *114*, 9706−9719.

(31) Baptista, A. Comment on "Explicit-solvent molecular dynamics simulation at constant pH: methodology and application to small amines". *J. Chem. Phys.* **2002**, *116*, 7766−7768.

(32) Itoh, S. G.; Damjanović, A.; Brooks, B. R. pH replica-exchange method based on discrete protonation states. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 3420−3436.

(33) Lee, J.; Miller, B. T.; Damjanovic, A.; Brooks, B. R. Constant pH molecular dynamics in explicit solvent with enveloping distribution sampling and Hamiltonian exchange. *J. Chem. Theory Comput.* **2014**, *10*, 2738−2750.

(34) Lin, Z.; van Gunsteren, W. F. Enhanced conformational sampling using enveloping distribution sampling. *J. Chem. Phys.* **2013**, *139*, 144105.

(35) Zhao, G.; Perilla, J. R.; Yufenyuy, E. L.; Meng, X.; Chen, B.; Ning, J.; Ahn, J.; Gronenborn, A. M.; Schulten, K.; Aiken, C.; Zhang, P. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **2013**, *497*, 643−646.

(36) Duane, S.; Kennedy, A. D.; Pendleton, B. J.; Roweth, D. Hybrid Monte Carlo. *Phys. Lett. B* **1987**, *195*, 216−222.

(37) Brass, A.; Pendleton, B.; Chen, Y.; Robson, B. Hybrid Monte Carlo simulations theory and initial comparison with molecular dynamics. *Biopolymers* **1993**, *33*, 1307−1315.

(38) Stern, H. A. Molecular simulation with variable protonation states at constant pH. *J. Chem. Phys.* **2007**, *126*, 164112.

(39) Stern, H. A. Erratum: "Molecular simulation with variable protonation states at constant pH" [J. Chem. Phys.126, 164112 (2007)]. *J. Chem. Phys.* **2007**, *127*, 079901.

(40) Chen, Y.; Roux, B. Efficient hybrid non-equilibrium molecular dynamics-Monte Carlo simulations with symmetric momentum reversal. *J. Chem. Phys.* **2014**, *141*, 114107.

(41) Chen, Y.; Roux, B. Generalized Metropolis acceptance criterion for hybrid non-equilibrium molecular dynamics - Monte Carlo simulations. *J. Chem. Phys.* **2015**, *142*, 024101.

(42) Berneche, S.; Roux, B. The ionization state and the conformation of Glu-71 in the KcsA K$^+$ channel. *Biophys. J.* **2002**, *82*, 772−780.

(43) Nilmeier, J. P.; Crooks, G. E.; Minh, D. D. L.; Chodera, J. D. Nonequilibrium candidate Monte Carlo is an efficient tool for equilibrium simulation. *Proc. Natl. Acad. Sci. U.S.A* **2011**, *108*, E1009−E1018.

(44) Boresch, S.; Karplus, M. The role of bonded terms in free energy simulations: theoretical analysis. *J. Phys. Chem. A* **2000**, *103*, 103−118.

(45) Boresch, S.; Karplus, M. The role of bonded terms in free energy simulations. 2. Calculation of their influence on free energy differences of solvation. *J. Phys. Chem. A* **2000**, *103*, 119−136.

(46) Shobana, S.; Roux, B.; Andersen, O. Free energy simulations: thermodynamic resversibility and variability. *J. Phys. Chem. B* **2000**, *104*, 5179−5190.

(47) Aleksandrov, A.; Simonson, T.; Roux, B. An overview of electrostatic free energy computations for solutions and proteins. *J. Chem. Theory Comput.* **2014**, *10*, 2690−2709.

(48) MacKerell, A. D.; Feig, M.; Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400−1415.

(49) Jorgensen, W. L.; Madura, J. D.; Swenson, C. J. Optimized intermolecular potential functions for liquid hydrocarbons. *J. Am. Chem. Soc.* **1984**, *106*, 6638−6646.

(50) York, D. M.; Darden, T. A.; Pedersen, L. G. The effect of long-range electrostatic interactions in simulations of macromolecular crystals: a comparison of the Ewald and truncated list methods. *J. Chem. Phys.* **1993**, *99*, 8345.

(51) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.

(52) Brooks, B. R.; et al. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.

(53) Souaille, M.; Roux, B. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput. Phys. Commun.* **2001**, *135*, 40−57.

(54) Dixit, S. B.; Chipot, C. Can absolute free energies of association be estimated from molecular mechanical simulations? The biotin-streptavidin system revisited. *J. Phys. Chem. A* **2001**, *105*, 9795−9799.

(55) Crooks, G. E. Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems. *J. Stat. Phys.* **1998**, *90*, 1481−1487.

(56) Machuqueiro, M.; Baptista, A. M. Acidic range titration of HEWL using a constant-pH molecular dynamics method. *Proteins: Struct., Funct., Bioinf.* **2008**, *72*, 289−298.

(57) Schaller, W.; Robertson, A. D. pH, ionic strength, and temperature dependences of ionization equilibria for the carboxyl groups in turkey ovomucoid third domain. *Biochemistry (Moscow)* **1995**, *34*, 4714−4723.

(58) Bartik, K.; Redfield, C.; Dobson, C. M. Measurement of the individual pKa values of acidic residues of hen and turkey lysozymes by two-dimensional ^{1}H NMR. *Biophys. J.* **1994**, *66*, 1180−1184.