

Charge-Neutral Constant pH Molecular Dynamics Simulations Using a Parsimonious Proton Buffer

Serena Donnini,[†] R. Thomas Ullmann,[‡] Gerrit Groenhof,^{*,§} and Helmut Grubmüller^{*,‡}

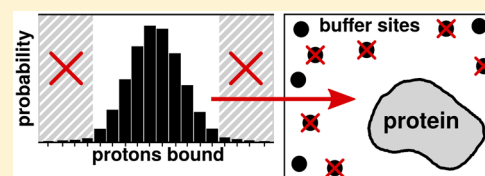
[†]Nanoscience Center and Department of Biological and Environmental Sciences, University of Jyväskylä, P. O. Box 35, 40014 Jyväskylä, Finland

[‡]Department of Theoretical and Computational Biophysics, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

[§]Nanoscience Center and Department of Chemistry, University of Jyväskylä, P. O. Box 35, 40014 Jyväskylä, Finland.

Supporting Information

ABSTRACT: In constant pH molecular dynamics simulations, the protonation states of titratable sites can respond to changes of the pH and of their electrostatic environment. Consequently, the number of protons bound to the biomolecule, and therefore the overall charge of the system, fluctuates during the simulation. To avoid artifacts associated with a non-neutral simulation system, we introduce an approach to maintain neutrality of the simulation box in constant pH molecular dynamics simulations, while maintaining an accurate description of all protonation fluctuations. Specifically, we introduce a proton buffer that, like a buffer in experiment, can exchange protons with the biomolecule enabling its charge to fluctuate. To keep the total charge of the system constant, the uptake and release of protons by the buffer are coupled to the titration of the biomolecule with a constraint. We find that, because the fluctuation of the total charge (number of protons) of a typical biomolecule is much smaller than the number of titratable sites of the biomolecule, the number of buffer sites required to maintain overall charge neutrality without compromising the charge fluctuations of the biomolecule, is typically much smaller than the number of titratable sites, implying markedly enhanced simulation and sampling efficiency.



INTRODUCTION

The structure of biomolecules in solution depend on pH.^{1–8} The pH determines the protonation state of titratable residues and therefore the charge distribution of a biomolecule. Because the charge distribution is critical to the structure and stability of biomolecules, their enzymatic activity in the case of enzymes, or their ability to interact with binding partners, e.g., in the case of proteins acting in signal transduction or as constituents of more complex biomolecular structures, can all be controlled through the pH. Consequently, the pH value in cells and cellular compartments is tightly regulated.^{9–11}

Despite its importance, molecular dynamics (MD) simulations of biomolecules typically do not include the effect of pH explicitly. Instead, the protonation states of protonatable groups of biomolecules are kept fixed during the simulation and the pH is accounted for implicitly by assigning a fixed protonation state to each protonatable residue, derived either from chemical intuition, heuristics,¹² or more elaborate estimates based on free energy calculations.¹³ Therefore, conformational changes or biological processes triggered by, or accompanied by, changes of protonation states cannot be modeled realistically. Moreover, assigning a realistic fixed protonation state is not trivial, as the proton affinities of residues in a biomolecule, measured by their pK_a values, are often unknown and may greatly differ from the pK_a values of isolated residues in solution. These differences arise because the proton affinity of a particular residue strongly depends on interactions with its environment in the

biomolecule, that is, on interactions with constituents of the environment, such as water molecules or lipids, and with other residues. Determining proton affinities of residues in large biomolecules is further complicated by (i) the often strong mutual coupling of the protonation states of interacting residues, which may in addition be coupled to their conformation,^{14–17} and (ii) the long-range nature of the electrostatic interaction.

Ideally, one would want to perform biomolecular simulations at a fixed pH value and allow the protonation of the titratable residues to adapt dynamically to changes in their environment during the simulation. Therefore, various approaches have been developed for including the effect of the pH in molecular dynamics simulations, such as Metropolis Monte Carlo,^{18–20} enveloping distribution sampling,²¹ or λ -dynamics.^{22–24}

In λ -dynamics based approaches, a continuous titration coordinate is introduced for each titratable residue, which defines the state of the titratable site. For example, in this work $\lambda = 0$ and $\lambda = 1$ correspond to the protonated and deprotonated states of a titratable group, respectively. The dynamics of λ evolves on a free energy surface that depends on the intrinsic proton affinity of the site, its interactions with the environment, and the pH of the solution. The free energy gradients are evaluated on-the-fly during the simulation. In such simulations,

Received: December 8, 2015

Published: February 16, 2016

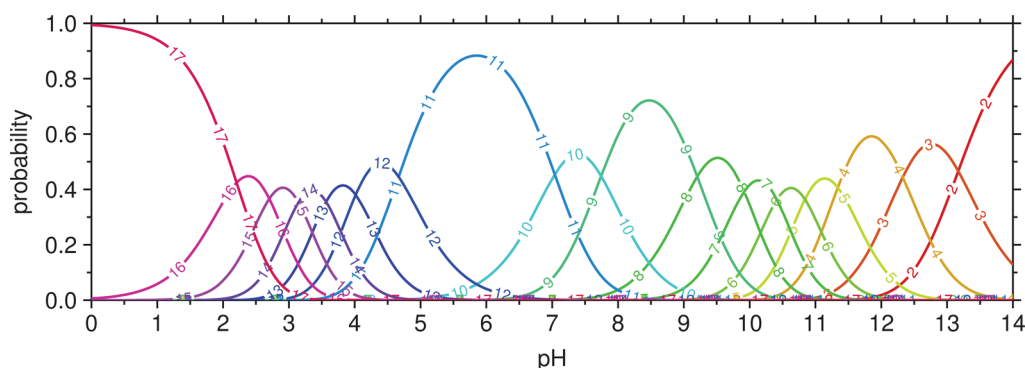


Figure 1. pH-dependent probability distributions of the protonation macrostates of the turkey ovomucoid inhibitor third domain (PDB 2GKR). The total number of protons bound is indicated on each curve. The number of titratable sites for this protein is 18.

protons are typically not transferred explicitly between the protein and solvent, but the effect of the proton activity of the surrounding medium is instead implicitly accounted for by coupling the system to a proton bath characterized by a constant chemical potential of the proton.

There are two major advantages of this implicit representation. The proton chemical potential accounts for the activity of the proton in the surrounding solution, which is determined by a complex equilibrium of differently protonated water species such as the hydroxyl ion, the hydronium ion, the Zundel ion,²⁵ and the Eigen ion.²⁶ In addition, the concentration of the relevant water species can be very low under physiological conditions, which would require model systems of computationally intractable size to include a large enough number of each relevant water species in an explicit representation. The implicit representation circumvents the necessity of including an accurate, explicit model for these complex coupled interconversion equilibria within the simulation and allows one to concentrate on the protonation equilibria of the biomolecule instead.

The disadvantage of the implicit representation is that the charge of the explicitly represented ionic species of the biomolecule is not balanced by the same amount of complementary charge carried by explicitly modeled counterions in the surrounding solution. The total charge of the system can therefore deviate from neutrality, which has been shown to cause severe artifacts if infinite periodic boundary conditions are applied in combination with the Ewald summation.^{27–29} Because Ewald summation techniques are widely used in the framework of molecular dynamics simulations in explicit solvent with periodic boundary conditions, the issue of neutrality in constant pH MD is particularly relevant.

Several approaches have been proposed for keeping the charge constant in constant pH MD simulations. In one approach, the protonation or deprotonation of each titratable site is coupled to the appearance or disappearance of an ion in solution to compensate for the difference in the net charge of the biomolecule. This approach is routinely used to avoid artifacts associated with a non-neutral simulation box in free energy calculations involving alchemical transformations that affect the total charge in the system.^{30,31} Recently, Wallace and Shen implemented this approach to prevent charge fluctuations in constant pH MD simulations.³² The precise chemical compound used to model the counterion is not important; it only has to carry the appropriate amount of charge. One could, for example, use hydronium ions as counterions.³³ Protonation of a site on the biomolecule causes deprotonation of the

coupled site and *vice versa*. Consequently, the sum of the charges on the biomolecule and the coupled sites is preserved, keeping the total charge of the simulation box constant. The disadvantage of coupling each titratable site of the biomolecule to a separate counterion is that the number of coupled sites increases with the number of titratable sites. Furthermore, to avoid biasing the pK_a estimate of the titratable sites, the local environment around each coupled site must remain identical during the simulation. In addition, direct interactions between coupled ions and the biomolecule must be avoided as well. In practice, therefore, the coupled sites are restrained to positions that maximize their separation. However, for larger systems with many titratable sites and thus many coupled sites, maintaining a sufficiently large separation between these ions requires large simulation boxes, which increases the computational effort significantly.

An alternative approach to avoid charge fluctuations of the system is to constrain the total charge of the biomolecule.^{34,35} In this way, only a single protonation macrostate, defined as the number of titratable protons bound to the biomolecule, can be simulated, but the distribution of these protons over the titratable sites is allowed to vary. However, whereas the average charge on the biomolecule is captured in this approach, fluctuations in the total number of protons bound are suppressed, which is unphysical. For example, Figure 1 shows that in the turkey ovomucoid inhibitor typically three or more protonation macrostates are populated significantly at all but the extreme pH values, rather than just one. The details of how this plot was obtained will be discussed later. This example demonstrates that a single protonation macrostate alone is not sufficient for a realistic simulation of a biomolecule at constant pH.

The aim of this work is to extend the approach by Wallace and Shen^{32,33} by combining the use of coupled sites^{32,33} with a constraint to overcome the disadvantages mentioned above, namely, the introduction of as many coupled sites as there are titratable sites in the system on the one hand and the artificial restriction to a single protonation macrostate on the other hand. The development of the combined approach was motivated by the expectation that the fluctuations of the overall charge of a biomolecule are typically much smaller than the total number of titratable sites of that biomolecule. Indeed, as shown in Figure 1, the number of macrostates populated at any given pH value is much smaller than the number of titratable sites in the ovomucoid inhibitor. The remaining protonation macrostates are rarely visited and can thus be

neglected. Therefore, a few coupled sites would be sufficient to compensate for realistic charge fluctuations of a biomolecule.

Accordingly, we introduce only a small number of additional titratable sites, which we term “buffer sites”, or buffers, in the simulation box and couple these to the overall protonation of the biomolecule with a charge constraint. To avoid affecting the pK_s of the titratable sites in the biomolecule, the free energy of deprotonation of the buffer is zero at all pH values. Because in this approach fluctuations of the total charge of the biomolecule are compensated for by opposite fluctuations of the total charge of the buffer sites in the solvent, the biomolecule’s macrostate can fluctuate. Introducing additional titratable sites to the system increases the number of ways protons can be distributed over these sites. Therefore, a single titration coordinate is used for all buffers simultaneously to avoid entropic artifacts. In our approach, the buffer sites effectively act as a proton buffer, without the necessity of modeling the complex coupled equilibria of the numerous protonated species in a real proton buffer.

THEORY

The constant pH MD approach³⁶ is first briefly summarized. Subsequently, the constraint equations, which are used to keep the total charge constant during the simulation, are presented. The proton buffer is then introduced, and the size of the proton buffer, i.e., the number of sites required to compensate for the fluctuations of the charge of the biomolecule, is discussed.

Constant pH Molecular Dynamics with the λ -Dynamics Approach. λ -Dynamics Approach briefly introduces the λ -dynamics method. Chemical Free Energy Contribution summarizes how the pH and the contributions to the deprotonation free energy due to the breakage and formation of chemical bonds, which are not described by classical force fields, are included in the Hamiltonian of the system. In Constraining the Interval of λ we present a biasing potential energy function for constant pH λ -dynamics that is better suited for simulations with a net charge constraint than the circular coordinate used in previous work.³⁶

λ -Dynamics Approach. In the λ -dynamics approach,^{37,38} the Hamiltonian of the system is expressed as

$$H(\lambda_i) = (1 - \lambda_i)H_{R_i} + \lambda_i H_{P_i} + \frac{1}{2}m_\lambda \dot{\lambda}_i^2 + U(\lambda_i) + \Delta G^{\text{chem}}(\lambda_i) \quad (1)$$

where λ_i is an additional degree of freedom for titrating site i , which interpolates the system between the reactant R_i ($\lambda_i = 0$) state with site i protonated and product P_i ($\lambda_i = 1$) state with site i deprotonated. The λ coordinate can be perceived as the coordinate of a particle moving between different chemical states. In eq 1, m_λ is the mass and $\frac{1}{2}m_\lambda \dot{\lambda}_i^2$ is the kinetic energy of this “ λ particle”. The λ -dependent potential term $U(\lambda)$ will serve as a biasing potential to limit the range of λ , and the term $\Delta G^{\text{chem}}(\lambda)$ is explained below.

Chemical Free Energy Contribution. To realistically describe protonation and deprotonation events at a given pH value, the effect of pH and the contributions from bond breaking and formation to the deprotonation free energy are described by an additional term, which shifts the protonation equilibrium by a certain free energy ΔG^{chem} .

To determine ΔG^{chem} , the equilibrium between a protonated and deprotonated acid in water is considered. ΔG^{chem} is then determined as

$$\Delta G^{\text{chem}}(\lambda) = (\ln 10)RT(pK_a - \text{pH}) - \Delta G^{\text{FF}}(\lambda) \quad (2)$$

where the first term on the right side of eq 2 is the usually experimentally determined free energy of deprotonation of a suitable model compound in solution, and the second term is the corresponding free energy difference obtained from a force field simulation. The pH-dependent term in eq 2 incorporates the dependency of the deprotonation free energy on the proton activity of the surrounding solution.

Constraining the Interval of λ . To restrict the λ space to the interval between the two physical states $\lambda = 0$ and $\lambda = 1$, two possibilities are considered, (i) the projection of an auxiliary, angular coordinate^{23,36,39,40} and (ii) a linear coordinate in combination with a biasing potential. The advantage of the angular coordinate is that it does not require any additional potential. The disadvantage is that it is difficult to control the height of the barrier between the two physical states due to the entropic contribution intrinsic to this projection.³⁶ The angular coordinate has the further disadvantage that it can lead to singularities in the constraint equations that are given below (eqs 6–8). The disadvantage of the linear coordinate is that it requires a stiff potential at the edge of the λ interval. The advantage is that the biasing potential can be used to control the barrier between the two physical states.

To avoid issues with singularities in the constraint equations, we have chosen to use a linear coordinate in combination with a biasing potential. The shape of the biasing potential is chosen such that (i) the λ values are restricted to the interval $[0,1]$, (ii) the average λ values in the protonated and deprotonated states are close to “0” and “1”, (iii) the ratio between time spent at intermediate states and time spent at the 0 and 1 states is small, and (iv) the transition rate between the 0 and 1 states is tunable.

The following functional form of the biasing potential, which we indicate as $U^{\text{dwp}}(\lambda)$ because of the double well shape (Figure 2),

$$U^{\text{dwp}}(\lambda) = -k \left[\exp\left(-\frac{(\lambda - 1 - b)^2}{2a^2}\right) + \exp\left(-\frac{(\lambda + b)^2}{2a^2}\right) \right] + d \left[\exp\left[-\frac{(\lambda - 0.5)^2}{2s^2}\right] \right] + 0.5w\{(1 - \text{erf}[r(\lambda + m)]) + (1 + \text{erf}[r(\lambda - 1 - m)])\} \quad (3)$$

fulfills requirements i–iv. Figure 2 illustrates the $U^{\text{dwp}}(\lambda)$ for various values of the barrier height. Because sampling of the λ coordinate can be controlled by adjusting the height of this

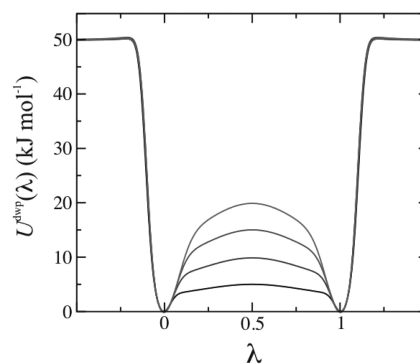


Figure 2. Double well potential with barrier heights 5, 10, 15, and 20 kJ mol^{-1} and outer walls 50 kJ mol^{-1} .

potential, we refer to this potential as adaptive biasing potential. In Appendix A, we describe in detail the motivation for the form of the potential in eq 3 and the procedure for optimizing the parameters.

Constraining the Total Charge of the System. The charge of a biomolecule with N^{sites} titratable sites depends on the total number of protons bound to those sites. The total number N^+ of protons bound is, using the convention for which a site is protonated at $\lambda = 0$ and deprotonated at $\lambda = 1$,

$$N^+ = \sum_i^{N^{\text{sites}}} (1 - \lambda_i) \quad (4)$$

Therefore, by constraining the sum over λ_i to a value c

$$\sum_i^{N^{\text{sites}}} \lambda_i = c \quad (5)$$

the overall charge of the biomolecule remains constant during the simulation. Note that c does not need to be equal to zero, because a net charge can also be compensated for by adding normal counterions. The constraint equation that must be fulfilled during the simulation is

$$\sigma(\lambda_1, \lambda_2, \dots, \lambda_{N^{\text{sites}}}) = \sum_i^{N^{\text{sites}}} \lambda_i - c \quad (6)$$

This requires a constraint force acting on the λ coordinates given by

$$G_i^\lambda = -\zeta \frac{\partial \sigma}{\partial \lambda_i} \quad (7)$$

where ζ is a Lagrange multiplier that needs to be determined. Here, we use the SHAKE algorithm to calculate ζ iteratively⁴¹

$$\zeta = \frac{\sigma(\lambda^u(t + \Delta t))}{\sum_i \Delta t^2 / m_i} \quad (8)$$

with $\lambda^u(t + \Delta t)$ the vector of λ_i 's at the unconstrained step. A detailed derivation of eq 8 is given in Appendix B. In our simulations the SHAKE algorithm converged within a few iterations.

Buffer Sites. The number of protons, and hence the charge, of a biomolecule with several titratable sites is typically not constant but fluctuates around an average protonation macrostate, as shown in Figure 3. A protonation *macrostate* is defined by the total number of titratable protons bound to a biomolecule, irrespective of the sites to which these protons are bound. Each macrostate is composed of several *microstates*, each of which has the same number of protons, but differs in the distribution of these protons over the titratable sites of the biomolecule.

Application of the constraint (eq 6) to the titratable sites of the biomolecule alone forces the biomolecule to sample a single macrostate only, which is not correct. To enable sampling all relevant macrostates, while keeping the total charge of the simulation box constant, we introduce additional titratable sites in the solvent, which we call buffer sites or buffers. These buffer sites are included in the constraint and can thus accept or donate protons to compensate for changes in the macroscopic protonation state of the protein. Thus, with the buffer sites, the charge on the biomolecule can fluctuate, while the charge of the total system, which now also includes these buffers, remains

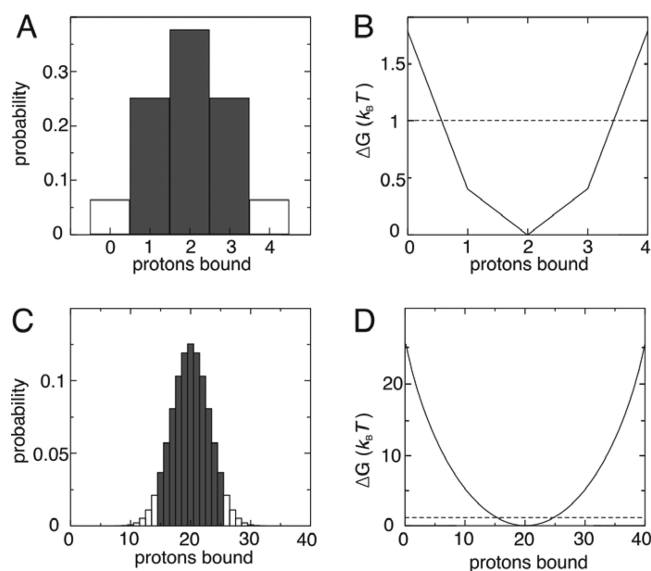


Figure 3. Protonation macrostate probabilities (A, C) and corresponding free energies (B, D) for a hypothetical biomolecule with $N^{\text{sites}} = 4$ (A, B) and $N^{\text{sites}} = 40$ (C, D) protonatable sites with identical $\text{p}K_a = \text{pH}$. Only the part of the distributions highlighted in gray is significantly populated. The dotted line indicates a free energy difference of $1 k_B T$.

constant. In analogy to a real buffer, our buffer sites consist of a mixture of a weak acid and its conjugate base. This is achieved by setting the free energy of deprotonation of the buffer sites to zero for all pH values; i.e., the free energies of the acid and its conjugate base are the same, independent of the pH ($\text{p}K_a = \text{pH}$ in eq 2). Thus, in the absence of a charge constraint, a 1:1 mixture of acidic and basic forms of the buffers would be obtained.

Simply including the buffers into the constraint equation, however, introduces an entropic bias in the macrostate probabilities of the biomolecule, because it changes the number of ways in which the macrostates can be realized. The probability p_j of protonation macrostate j of a biomolecule with N^{sites} protonatable sites is given by the sum of the Boltzmann weights of all microstates with j protons:

$$p_j = \frac{1}{\Xi} \sum' \exp\left(-\frac{E(\lambda_1, \lambda_2, \dots, \lambda_n)}{k_B T}\right) \quad (9)$$

where the prime indicates that the sum runs over all protonation microstates, which have the same number j of protons bound, i.e., $\sum_i^{N^{\text{sites}}} (1 - \lambda_i) = j$ (recall that $\lambda_i = 0$ corresponds to site i protonated), and Ξ is the semigrand canonical partition function for the titratable protons.

We define the symbol ω_j as the number of ways a macrostate with j protons and N^{sites} can be realized:

$$\omega_j = \frac{N^{\text{sites}}!}{j!(N^{\text{sites}} - j)!} \quad (10)$$

If we introduce N^{buffer} buffers and impose that the total number of protons present in the complete system is N^{sites} , there are $m = N^{\text{sites}} - j$ protons bound to the buffer. Because there are $N^{\text{buffer}}! / m!(N^{\text{buffer}} - m)!$ ways to distribute these m protons over the N^{buffer} buffers, the total number of ways in which macrostate j of the biomolecule can be realized has increased:

$$\omega_j^* = \frac{N^{\text{sites}}!}{j!(N^{\text{sites}} - j)!} \frac{N^{\text{buffer}}!}{m!(N^{\text{buffer}} - m)!} \quad (11)$$

where the asterisk (*) indicates that the system consist of the biomolecule and the buffers. Thus, the number of microstates that contribute to a protonation macrostate of the biomolecule is $N^{\text{buffer}}!/m!(N^{\text{buffer}} - m)!$ times larger and adds an artificial contribution (ΔG^*) to the true free energy. Furthermore, since this contribution is different for each macrostate, straightforward application of the charge constraint in combination with N^{buffer} buffer sites introduces an entropic contribution to the free energy of a macrostate j of

$$\Delta G_j^* = -RT \ln \left(\frac{N^{\text{buffer}}!}{m!(N^{\text{buffer}} - m)!} \right) \quad (12)$$

despite the zero free energy of protonating the buffers. Here, $m = N^{\text{buffer}} - j$, as before.

To avoid this bias in the macrostate distribution, we introduce a *single* protonation coordinate for all buffer sites simultaneously. With this approach, there is only *one* way of distributing the m protons over the N^{buffer} buffers and eq 11 reduces to eq 10. The m protons are “smeared out” evenly over all buffers. Although it may appear unphysical to have particles with noninteger charges in the simulation box, we consider the potential effects on the sampling of the biomolecule less severe than (i) the aforementioned entropic artifacts, (ii) suppressing macrostate fluctuations, or (iii) simulating with a net charge.²⁷ Typically, to minimize any bias, the buffers are restrained to positions far from the system of interest.

With a single titration coordinate (λ^{buffer}) for the N^{buffer} buffer sites, the constraint equation becomes

$$\sigma(\lambda_1, \lambda_2, \dots, \lambda_{N^{\text{sites}}}, \lambda^{\text{buffer}}) = \sum_i \lambda_i + N^{\text{buffer}} \lambda^{\text{buffer}} - c \quad (13)$$

Accordingly, the expression for ζ in eq 7 becomes

$$\zeta = \frac{\sigma(\lambda^u(t + \Delta t))}{\sum_i^{N^{\text{sites}}} \Delta t^2/m_i + N^{\text{buffer}} \Delta t^2/m_{\text{buffer}}} \quad (14)$$

with m_{buffer} the fictitious mass of λ^{buffer} .

Buffer Size. With N^{sites} titratable sites, a biomolecule can in principle sample $N^{\text{sites}} + 1$ protonation macrostates, each with a different total charge. Therefore, to sample the complete range of macrostates, also N^{sites} buffers would be required. At this point it may seem therefore that our buffer approach offers no major advantages with respect to the co-ions approach of Shen and co-worker,³² which introduces N^{sites} additional sites. However, as we will show below, the number of macrostates available to a biomolecule at any pH is typically much lower than the number of titratable sites, even if their pK_a values are identical. Therefore, in contrast to the co-ions approach, the number of buffers (N^{buffer}) can be chosen much smaller than the number of titratable sites on the biomolecule (N^{sites}), without compromising accuracy.

To illustrate this point, Figure 3 shows the macrostate distribution of a hypothetical biomolecule with N^{sites} non-interacting titratable sites, all with the same pK_a . The distribution of macrostates for such system is determined by the number of microstates ω_j that contribute to a given macrostate j (eq 10). Figure 3 shows two examples with 4 and

40 titratable sites, respectively. In both systems, the distributions peak at the center because the number of realizations is the highest if the number of protons is half the number of sites. The relative width of the distributions decreases with increasing N^{sites} . Assuming that protonation macrostates with a small population can be neglected (white area of the distributions in Figure 3), we conclude that, in our hypothetical system, the number of buffer sites required for keeping the charge constant while sampling the most relevant macrostates of the biomolecule is only a fraction of N^{sites} . We note furthermore that because the pK_a values of titratable sites in real biomolecules are usually not identical, the number of buffers required can be reduced even further, as we will show later for a representative set of proteins.

In the simple example above, we saw that the number of buffers required is related to the width of the macrostate distribution. A robust measure for the width of an arbitrarily shaped probability distribution can be derived from the Rényi–Süßmann entropy.^{42–44} This measure can be addressed as an effective configuration space volume denoted by $\tilde{\Omega}$.⁴⁴ Here, $\tilde{\Omega}$ quantifies the number of protonation macrostates effectively sampled by a protein

$$\tilde{\Omega} = \frac{1}{\sum_j p_j^2} \quad (15)$$

where the sum runs over the probabilities p of all protonation macrostates j .

An upper limit for the necessary number of buffer sites of a system is obtained by considering the hypothetical system of N^{sites} protonatable sites each of which is equally likely to be protonated or deprotonated. The probability distribution of protonation macrostates for this hypothetical system is given by a binomial distribution with success probability 0.5 and number of trials N^{sites} , which approaches⁴⁵

$$\tilde{\Omega}^{\text{max}} = \lim_{N^{\text{sites}} \rightarrow \infty} \tilde{\Omega} = \sqrt{\pi N^{\text{sites}}} \quad (16)$$

for large N^{sites} . Thus, the higher the number of titratable sites of the biomolecule, the smaller the number of buffer sites per titratable site needed to properly describe all charge and protonation fluctuations.

METHODS

The charge constraint with proton buffer approach for preserving charge neutrality in constant pH MD was implemented in the GROMACS molecular dynamics package (version 3.3).^{46–48} This code is available for download from <http://www.mpibpc.mpg.de/grubmueller/constpH>.

To test the validity of our approach, we used a model system consisting of four noninteracting acetic acid molecules in explicit water. Although this system is simpler than typical biomolecules, we preferred it as a suitable test system, because (a) the results can be compared to results obtained analytically and (b) any potential artifacts of our method can be well separated from sampling or force field issues that other systems might have. To that end, all interactions between the acetic acids and the buffer sites, either directly via the Coulomb potential or indirectly via water polarization, were excluded. This was achieved by truncating the Coulomb interactions beyond a cutoff and by freezing the water molecules that are outside the cutoff spheres of the acetic acids and the buffers.

Acetic acid was described with the GROMOS96 53A6⁴⁹ force field parameters for the carboxylic group of glutamic acid. The buffer sites were described as hydronium ions with parameters listed in Table S1 of the [Supporting Information](#). The acetates and buffer sites were placed in a rectangular box of dimensions $8 \times 8 \times 4 \text{ nm}^3$, which was filled with 8,698 SPC⁵⁰ water molecules. The acetic acid molecules and buffer sites were restrained at the heavy atom positions with a force constant of $1,000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ and were positioned in the simulation box at a distance of more than 2.7 nm from each other. Water molecules at a distance larger than 1.0 nm from the titratable sites were frozen. To validate the biasing potential (eq 3), constant pH MD simulations of a single buffer site were performed in a cubic box of $4 \times 4 \times 4 \text{ nm}^3$. In all simulations, the Coulomb and Lennard-Jones cutoffs were set to 1.0 nm, and interactions between atoms within the cutoff were evaluated at every step of the simulation. To correct for the neglect of electrostatic interactions beyond the cutoff, a reaction field correction with infinite dielectric constant was employed.⁵¹ Periodic boundary conditions were applied. Constant temperature was maintained by using the Berendsen thermostat⁵² at 300 K with coupling time of 0.5 ps. A leapfrog integrator was used with an integration time step of 2 fs. The bond distances and bond angles of water were constrained using the SETTLE algorithm.⁵³ All other bond distances were constrained using the LINCS algorithm.⁵⁴ To equilibrate the system prior to the constant pH MD simulations, the potential energy of the system was first minimized using a steepest descent approach, followed by a 500 ps MD simulation.

The mass and temperature of all λ particles were set to 20 u and 300 K, respectively. Each λ particle was coupled separately to a heat bath via the Andersen⁵⁵ thermostat with a coupling parameter of 6 ps^{-1} . The barrier height of the biasing potential U^{dwp} (eq 3) was chosen—as typical for the application of our method—as a trade-off between “localization” of λ at 0 and 1, respectively, and sufficient sampling. Here, a height of 7 kJ mol^{-1} was chosen, unless stated otherwise. The barrier height of the biasing potential describing the single titration coordinate λ^{buffer} , which deprotonates the buffer sites collectively, however, was set to zero. This is because fractional values of λ^{buffer} should not be penalized by the potential, as they compensate for protonation macrostates of the acetic acids. To prevent sampling beyond the $[0,1]$ interval, the walls of the adaptive biasing potential U^{dwp} were set to 200 kJ mol^{-1} . The complete set of parameters for the biasing potentials are listed in Table S2 of the [Supporting Information](#).

Prior to the constant pH MD simulations, the free energies of deprotonation of each acetic acid molecule and of the buffer sites were calculated by means of thermodynamic integration (TI). These free energies are required for the force field correction ($\Delta G^{\text{FF}}(\lambda)$ in eq 2). For each acetic acid a separate TI simulation was performed, in which the topology was linearly interpolated between the protonated and deprotonated state.

To obtain the correction potential (eq 2) of the buffers, we interpolated the topology linearly from the state with all buffers protonated to the state with all buffers deprotonated. Because the buffers are essentially identical, we could have computed the free energy of deprotonation of a single buffer as well. However, due to small variations in the local environment of the buffers in our test system, the $\Delta G^{\text{FF}}(\lambda_i)$ are not exactly identical either. To avoid that these differences obscure our test calculations, we computed the correction potential for the collective deprotonation of all four buffers instead. For a real

application both approaches for obtaining ΔG^{FF} should yield similar results.

Each TI run consisted of 11 simulations of 10 ns each, at fixed values of λ along the interval $[0,1]$. During the simulations the $\partial H/\partial \lambda$ was recorded, and the resulting $\partial H/\partial \lambda$ curve was integrated to yield a free energy profile of the deprotonation reaction ($\Delta G^{\text{FF}}(\lambda)$). Fourth order polynomials were fitted to the free energy profiles and used as $\Delta G^{\text{FF}}(\lambda_i)$ in eq 2. These fits were refined further by means of Boltzmann inversion. The details of these fits and their refinement are included as [Supporting Information](#) (Table S3 and Table S4). The reference pK_a of acetic acid was 4, whereas for the buffer sites $\text{pK}_a = \text{pH}$ by construction. In total 100 constant pH MD simulations of 5 ns each were performed for our test system at $\text{pH} = 4$. The first 0.2 ns of these runs were used for equilibration and discarded from data analysis.

A titration curve of the four acetic acids was obtained by performing constant pH MD simulations at different pH values ($\text{pH} = 1, 2, 3, 4, 5, 6, \text{ and } 7$). At each pH value 20 runs of 5 ns each were performed. The first 0.2 ns of each run was omitted from the data analysis. The average λ at each pH value was then plotted as a function of pH. Because the acetic acids are equivalent, the average was computed as

$$\langle \lambda \rangle = \frac{1}{N_{\text{sites}}} \sum_i^{N_{\text{sites}}} \langle \lambda_i \rangle \quad (17)$$

The pK_a value was estimated from a fit of the calculated titration curve to the Henderson–Hasselbalch equation.³⁶

The statistical error of the average λ was calculated as described previously.³⁶ The error bars in the titration curve plot were obtained by propagating the error of the four averages of the acetic acid λ values at each pH point.

The protonation macrostate probability distributions of a representative set of proteins were computed with Wang–Landau Monte Carlo simulations^{56,57} using a continuum electrostatics model for reasons of computational efficiency. These simulations were performed with a modified version of the MEAD software package⁵⁸ and the GMCT simulation software⁵⁹ as described previously.^{17,60}

RESULTS AND DISCUSSION

First, to verify our theoretical result that the number of buffer sites required to compensate the charge fluctuations of the biomolecule is much smaller than the number of titratable sites, we calculated the number of accessible macrostates for a series of proteins at $\text{pH} = 7$. Next, we tested if the adaptive biasing potential $U^{\text{dwp}}(\lambda)$ is capable of constraining the protonation ensemble mostly at $\lambda = 0$ (protonated) and $\lambda = 1$ (deprotonated), respectively, while at the same time permitting sufficiently many transitions for accurate sampling. Third, we verified for a system with four noninteracting acetic acids in explicit water that our approach of using a charge constraint in combination with a variable number of buffers preserves not only the total charge of the system but also the relative free energies of the accessible macrostates. Finally, we varied the pK_a values of the four acetic acids to mimic more closely a real biomolecule.

Macroscopic Protonation States in Proteins. To estimate how many macroscopic states are actually relevant at a given pH value, we computed the macrostate distributions for a representative set of 17 protein systems by means of Poisson–Boltzmann based Monte Carlo simulations. [Figure 4](#)

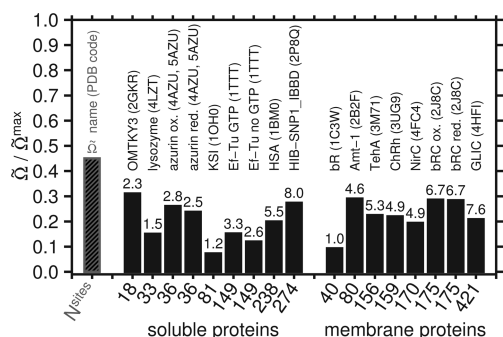


Figure 4. Thermally accessible macroscopic protonation states for 17 soluble and membrane proteins with different numbers N_{sites} of protonatable sites at pH 7.0. For each protein, the bar shows the ratio $\tilde{\Omega} / \tilde{\Omega}^{\max}$, while the value of $\tilde{\Omega}$ is reported on top of the bar. $\tilde{\Omega}$ is the number of thermally accessible macroscopic protonation states at pH 7.0, and $\tilde{\Omega}^{\max}$ is an upper limit for the number of thermally accessible protonation macrostates. $\tilde{\Omega}^{\max}$ is realized in a hypothetical system of N_{sites} protonatable sites equally likely to be protonated or deprotonated. The abbreviated protein name and the PDB code of the protein structure used in the simulations are stated above each bar.

shows for each protein the number of protonation macrostates effectively sampled by the protein ($\tilde{\Omega}$, eq 15) at pH = 7, the number of titratable sites (N_{sites}), and the ratio $\tilde{\Omega} / \tilde{\Omega}^{\max}$, where $\tilde{\Omega}^{\max}$ is the theoretical upper limit for the number of thermally accessible protonation macrostates (eq 16). For each system considered, the number $\tilde{\Omega}$ of macrostates populated at pH = 7 is much smaller than the N_{sites} number of titratable sites. Furthermore, the ratio $\tilde{\Omega} / \tilde{\Omega}^{\max}$ is significantly smaller than one for all proteins investigated, suggesting that our approximation for $\tilde{\Omega}^{\max}$ (eq 16) provides an excellent conservative estimate for the maximum number of buffers required in a constant pH MD simulation of a biomolecule with N_{sites} titratable sites. Because the ratio $\tilde{\Omega} / \tilde{\Omega}^{\max}$ never exceeds 1/3, the number of sites may be further reduced in practice without compromising the macrostate sampling.

For the ovomucoid inhibitor system, we also investigated the distribution of macrostates as a function of pH. Figure 1 shows these distributions. For all pH values, the number of macrostates populated is much smaller than the number of titratable sites ($N_{\text{sites}} = 18$), again confirming the theoretical upper limit, eq 16.

Constraining the Interval of λ with an Adaptive Biasing Potential. The adaptive biasing potential (Figure 2, eq 3) was tested in simulations of a single titratable site (hydronium) in explicit solvent, with which we also investigated the effect of the barrier height of this potential on the transition rate and distribution probabilities between the protonation states of the site.

Figure 5 shows the effect of the barrier height (4, 7, and 10 kJ mol⁻¹) on the sampling of λ . Increasing the barrier height increases the residence time at the end states. Accordingly, the transition rate decreases with a factor of about 2 when the barrier is increased from 4 to 7 kJ mol⁻¹ and of about 2.5 when the barrier is increased from 7 to 10 kJ mol⁻¹. Note that the relation between the rate and the barrier is not perfectly exponential ($\exp\left(-\frac{\Delta E^{\text{TST}}}{k_B T}\right) = 3.3$), as predicted by Arrhenius or Eyring theory,^{61,62} because of the curvature of the biasing potential. Indeed, taking into account the biasing potential explicitly in an analytical evaluation of the partition functions at the transition state ($\lambda = 0.5$) and in the reactant state ($\lambda <$

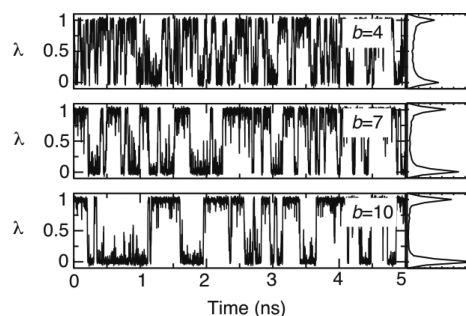


Figure 5. Dynamics of the deprotonation of a titratable site in explicit solvent with different barrier heights of the biasing potential. Trajectories (left) and distributions (right) of the deprotonation coordinate λ from 5 ns constant pH MD simulations with barrier heights 4 kJ mol⁻¹ ($b = 4$), 7 kJ mol⁻¹ ($b = 7$), and 10 kJ mol⁻¹ ($b = 10$). The titratable site is protonated at $\lambda = 0$ and deprotonated at $\lambda = 1$.

0.5),⁶¹ yields rate ratios (1.8 and 2.6, respectively) that are in very good agreement with our simulations. This aspect of the biasing potential should be taken into consideration when optimizing the parameters for a desired transition rate.

The λ distributions (Figure 5) peak at 0 and 1, respectively. Although the biasing potential (Figure 2) is strictly symmetric, the distributions are not. This asymmetry increases with barrier height, suggesting insufficient sampling in the 5 ns trajectories with the higher barriers. The height of the walls that prevent sampling beyond the physically relevant λ interval only has a minor effect on the sampling as shown in Figure S1 of the Supporting Information.

Constant pH MD with a Proton Buffer. We verified that our approach of combining a constraint on the charge with buffer sites preserves not only the charge but also the relative free energies of the protonation macrostates for a model system containing four acetic acids and four buffers. Three sets of constant pH MD simulations were performed. In the first set, no constraint was applied to the sum of the λ deprotonation coordinates, and the total charge of the simulation box was allowed to fluctuate. In the second and third sets of simulations, the charge constraint was applied to the acetic acids and the buffer sites. The difference between the second and third sets of constant pH MD simulations is that in the second set the protonation state of each buffer site was described by an independent λ coordinate (in total eight λ coordinates were propagated), whereas in the third set of simulations, a single collective λ^{buffer} coordinate was used to describe the extent of deprotonation for the four buffers simultaneously (in total five λ coordinates were propagated). In all simulations with the charge constraint, the sum of λ was fixed to four ($c = 4$, eq 5 and eq 13), which is the same as fixing the sum of protons to four. This, in combination with simulating at pH = pK_a, implies that there are four protons in the system that can bind to any of the eight titratable sites with equal probability. We verified that in all simulations with the charge constraint the total charge was conserved (Table S5 and Figure S2 in the Supporting Information).

To test whether the constraint biases the distribution of protons over the titratable sites of the four acetic acids, the protonation macrostate free energy profiles of the four acetic acids from the three sets of simulations were calculated and compared with the analytical solution (Figure 6). The free energy profiles as a function of protonation macrostate, which

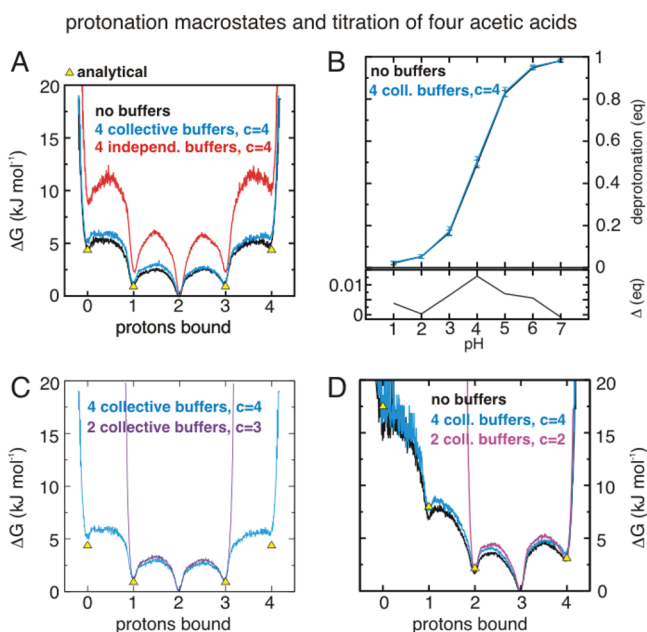


Figure 6. Protonation macrostates and titration of four acetic acids in constant pH MD simulations with explicit water. Simulations were performed without constraint (black), with charge constraint $c = 4$ and four independent buffer sites (red), with $c = 4$ and four collective buffer sites (i.e., a single titration coordinate for the four buffer sites) (blue), with $c = 3$ and two collective buffer sites (magenta), and with $c = 2$ and two collective buffer sites (purple). The free energy difference obtained analytically for four independent protonatable sites is also reported (yellow triangles). (A and C) Protonation macrostate free energy profiles at $\text{pH} = \text{p}K_a$. (B) Averaged titration curves of four acetic acids. The deprotonation (in equivalents, eq) is plotted as a function of pH. The lower graph shows the difference between the two titration curves. (D) Protonation macrostate free energy profiles at $\text{p}K_a$ values of 2.5, 3.0, 3.5, 4.5, respectively, and $\text{pH} = 4$.

is defined here as the total number of protons bound to the acetic acids (i.e., $n_p(t) = \sum_{i=1}^4 (1 - \lambda_i(t))$), were obtained directly from the probability distribution $p(n_p)$ of $n_p(t)$ over the three sets of simulations (i.e., $G = -RT \ln[p(n_p)]$). Convergence of the free energies for the simulations without charge constraint and with charge constraint and collective buffers is shown in Figure S3 and Figure S4 of Supporting Information, respectively.

Figure 6A shows that, in all three simulations, macrostates were sampled in which the number of protons bound to the acetic acid subsystem varied from 0 to 4, but their relative free energies were different. Without the charge constraint, the protonation macrostate free energies agree well with the reference free energies obtained analytically for four independent sites (eq 9; yellow symbols in Figure 6). Although in this simulation the overall charge of the simulation box was not conserved, the truncation of Coulomb interactions beyond a cutoff avoided the artifacts associated with a non-neutral simulation box.²⁷

When the four buffer sites are forced to titrate concertedly by using a single λ^{buffer} coordinate for all buffer sites simultaneously (blue curve in Figure 6A), the protonation macrostate free energies of the acetic acid subsystem agree well with both the analytical free energies, as well as the free energies obtained without constraints. Such agreement demonstrates that our approach for charge conservation in constant pH MD simulations reproduces the correct macrostate probabilities.

In contrast, when the constraint was applied on eight otherwise independent titratable sites (red curve in Figure 6A), the free energies of the macrostates deviated from the analytical solution. The central macrostate with two protons bound is overstabilized, while the macrostates with no protons bound and with four protons are much less populated. As was discussed in Theory, the use of a constraint in combination with independent buffer sites introduces an entropic bias in the relative free energies of the macrostates due to the increase in the number of ways protons can be distributed over the buffers. Because this number is the highest for the case in which there are two protons present on both the acetic acid subsystem and the buffer sites, the macrostate with two protons on the acetic acid subsystem is affected most (eq 11; red curve in Figure 6A).

Next, we performed simulations at various pH values and computed the titration curve of the four acetic acids with and without a constraint. In the simulations with constraint, the buffers were modeled with a single collective λ^{buffer} coordinate. Figure 6B shows that the averaged titration curves for simulations with and without the constraint are identical within statistical error, indicating that the proton buffer also does not bias the system at pH values different from $\text{pH} = \text{p}K_a$.

To investigate the effect of reducing the number of buffer sites, we also ran simulations with two instead of four buffers. In these simulations, the number of protons was constrained to three, eliminating the least populated macrostates of the acetic acid subsystem with zero and four protons, respectively. Although all five possible macrostates were significantly populated in our model system (Figure 6A), our aim was to verify whether the relative free energies of the remaining macrostates with one, two, or three protons remain the same if fewer buffers are used. Panel C of Figure 6 shows the protonation macrostate free energies of the acetic acid subsystem at $\text{pH} = \text{p}K_a$ with four and two buffer sites. Whereas the macrostates with all acetic acids protonated or deprotonated are now excluded, the remaining three macrostates are sampled with the same weights in both cases. This example demonstrates that our approach preserves the free energy differences between the macrostates, despite neglecting the two least populated macrostates.

Next, to study the performance of our method for the more realistic case of different $\text{p}K_a$ values, the reference $\text{p}K_a$ values of the four acetic acids were set to 2.5, 3.0, 3.5, and 4.5, respectively. Three sets of constant pH MD simulations at $\text{pH} = 4.0$ were performed: without charge constraint; with charge constraint $c = 4$ and four buffer sites; with $c = 2$ and two buffer sites. Panel D of Figure 6 shows the free energy profiles of the protonation macrostates in these simulations, along with the analytical solution. With four buffers, the macrostates are sampled with the same weights as without constraint. In both cases, the macrostate without protons bound is not populated. Decreasing the number of buffers to two, and setting the charge constraint $c = 2$, eliminates the two least populated macrostates (zero and one protons bound) but enables sampling of the remaining three macrostates with the correct thermodynamic weights. Note that also in this example our choice for the constraint and number of buffers caused the elimination of a macrostate that should not be neglected. Nevertheless, the agreement with the reference simulations without constraints shows once more that our approach not only conserves the charge but also maintains the relative free energies of the accessible protonation macrostates.

In our simple model with only four identical titratable sites the choice of which macrostates to neglect is obvious. For realistic biomolecules with more than four titratable sites, this choice may be more challenging. We showed that at any pH value most protonation macrostates are not significantly populated in proteins (Figure 1). To achieve a selective elimination of the nonrelevant macrostates, suitable values for the constraint and the number of buffers have to be determined. Since these parameters are not independent, some adjustments may be required prior to production runs. We suggest the following procedure.

First, to determine the number of buffer sites, we suggest to compute $\tilde{\Omega}^{\max}$ (eq 16), which is an upper limit for the necessary number of buffer sites. As an initial guess for the constraint, which determines the number of titratable protons in the system, we suggest to use the maximum number of protons bound to the biomolecule at the desired pH. This number can be estimated from the reference pK_a 's of the titratable sites, or from measured titration curves, if available. If optimal values for the number of buffer sites and the constraint are chosen, $\langle \lambda^{\text{buffer}} \rangle \approx 0.5$ in the simulation.

If during the simulation, the buffers are mostly protonated (i.e., $\lambda^{\text{buffer}} \approx 0$), there are either too few buffer sites or too many active protons in the system. In this case, one can therefore either increase the number of buffer sites, decrease the number of protons by *increasing* the constraint value, or both. The reason why in our implementation one needs to increase the constraint value in order to decrease the number of protons is that sites are protonated at $\lambda_i = 0$ and deprotonated at $\lambda_i = 1$, and the constraint is defined as the sum over λ_i (eq 13). We suggest to adjust the number of buffer sites and the constraint value until $\langle \lambda^{\text{buffer}} \rangle$ is about half on average in a test simulation.

In contrast, if the buffer sites are deprotonated for most of the simulation (i.e., $\lambda^{\text{buffer}} \approx 1$), there are either too many buffers or too few active protons in the system. The remedy in this situation is therefore to either decrease the number of buffers, or increase the number of protons by *decreasing* the constraint, or both. Again, as explained before, a decrease of the constraint corresponds in our implementation to an increase in the number of protons present in the system.

Optimizing the number of buffer sites and the constraint prior to the production runs may require several iterations, in particular if the titration curve or the pK_a s of the biomolecule are not known in advance. However, this initial effort is paid off quickly, because of a higher computational efficiency with a minimal number of buffer sites. Since the ratio between the number of relevant macrostates and the number of titratable sites decreases rapidly with the total number of titratable sites, this advantage will even be more pronounced when simulating large biomolecular systems, for which coupling each site to a separate and mutually independent buffer particle becomes computationally intractable.

Because we use a collective λ^{buffer} coordinate for all buffers, the buffers carry fractional charges. Although clearly unphysical, these fractional charges are strongly screened in the high dielectric water and therefore will only have minor effects on both the protonation and conformational sampling of a biomolecule. These effects can be further reduced by restraining each buffer at the largest possible distance from the biomolecule and the other buffers. In contrast, the artifacts caused by a net charge in the simulation box would be much more severe, in particular for membrane systems, or deeply

buried enzyme active sites, in which the net charges can significantly alter the pK_a of residues in contact with the hydrophobic interior.²⁷

CONCLUSIONS

At thermodynamic equilibrium, the total charge of a biomolecule fluctuates due to continuous protonation changes of titratable sites with pK_a s near the solution pH. Constant pH MD methods were introduced to model the changes in protonation, but the charge fluctuations may introduce artifacts that can affect the accuracy of the pK_a s.²⁷ To avoid these artifacts in constant pH MD simulations while maintaining an accurate statistical ensemble of all protonation fluctuations, we have developed an approach that uses a charge constraint in combination with a minimal proton buffer. While the constraint prevents charge fluctuations of the complete system, which contains both the biomolecule and the buffer sites, the buffer sites compensate for charge fluctuations of the biomolecule. We used a simple test system, consisting of four noninteracting titratable sites, to demonstrate that with our approach the thermodynamically accessible protonation macrostates of the biomolecule are sampled, while the simulation box remains neutral. A major advantage of our approach is that the number of buffer sites is adjustable. This, in combination with the notion that in realistic biomolecules only a small fraction of all possible protonation macrostates is populated at any pH value, implies that it is possible to perform realistic charge-neutral constant pH MD simulations with only a few buffer sites.

APPENDIX A

The biasing potential for λ -dynamics simulations must fulfill the following criteria: (1) λ values in the interval $[0,1]$, (2) average values of λ in protonated and deprotonated states close to "0" and "1", respectively, (3) small ratio between time spent at intermediate states and time spent at the 0 and 1 states, and (4) tunable transition rate.

To address condition 1, we have chosen to use a function U^{wall} which rises steeply beyond the λ interval, mimicking two walls

$$U^{\text{wall}}(\lambda) = 0.5w\{(1 - \text{erf}[r(\lambda + m)]) + (1 + \text{erf}[r(\lambda - 1 - m)])\} \quad (18)$$

with erf being the error function. The parameter w determines the height of the walls. As will become clear later, the effective height of the walls of the biasing potential is $w = w_0 - h/2$, where w_0 is the desired height of the walls and h is the height of the barrier potential. A high value of w has the purpose of mimicking an infinitely high wall (in this work $w = 200$ was used for most of the simulations).

The parameters r and m in eq 18 are chosen to obtain a steep rise in the potential beyond the $[0, 1]$ interval. To avoid compromising the molecular dynamics integrator by using too steep potentials, the steepness is decreased, with increasing height w of the walls. In practice, we choose a width $\sigma_0 = 0.02$ and define

$$e_1 = \text{erf}^{-1}[1 - 2/w] \quad (19)$$

$$e_{10} = \text{erf}^{-1}[1 - 20/w] \quad (20)$$

with erf^{-1} being the inverse error function, and

$$r = \frac{e_1 - e_{10}}{2\sigma_0} \quad (21)$$

$$m = 2\sigma_0 \frac{2e_1 - e_{10}}{e_1 - e_{10}} \quad (22)$$

To address condition 2, namely, enhance the sampling of λ at the physical states close to 0 and 1, two negative Gaussian functions are used. Each Gaussian mimics a minimum of the biasing potential at the desired values of $\lambda_0 = 0$ and $\lambda_1 = 1$, respectively:

$$U^{\min}(\lambda) = -k \left(\exp \left[-\frac{(\lambda - 1 - b)^2}{2a^2} \right] + \exp \left[-\frac{(\lambda + b)^2}{2a^2} \right] \right) \quad (23)$$

The depth k of the minima is set initially to half the height of the desired barrier h , and the parameters a and b are set initially to $a = 0.05$ and $b = -0.1$. The three parameters k , a , and b are adjusted iteratively until the desired shape of the potential is reached, as explained below.

The last conditions 3 and 4 are addressed by a third Gaussian, which mimics a barrier centered at $\lambda = 0.5$:

$$U^{\text{barrier}}(\lambda) = d \left(\exp \left[-\frac{(\lambda - 0.5)^2}{2s^2} \right] \right) \quad (24)$$

The parameter d defines the height of the Gaussian function. This is set to half of the desired barrier height h , because the overall height of the barrier will depend also on the depth of the minima. The height h can be changed to control the transition rates between the protonated and deprotonated states of λ , as required in condition 4. The width s of $U^{\text{barrier}}(\lambda)$ is chosen to obtain a relatively flat potential at the center of the λ interval. Here, we have used $s = 0.3$.

The final expression for the biasing potential, which we indicate as $U^{\text{dwp}}(\lambda)$ because of the double well shape (Figure 2), is

$$U^{\text{dwp}}(\lambda) = U^{\text{wall}}(\lambda) + U^{\min}(\lambda) + U^{\text{barrier}}(\lambda) \quad (25)$$

To avoid that, due to overlap between the three functions $U^{\text{wall}}(\lambda)$, $U^{\min}(\lambda)$, and $U^{\text{barrier}}(\lambda)$, the potential $U^{\text{dwp}}(\lambda)$ deviates from the desired shape, the parameters k , a , and b in eq 23 are adjusted iteratively until convergence is reached. The criteria for convergence are described in the following.

The parameter b , which shifts the position of the minima in $U^{\min}(\lambda)$, is adjusted iteratively:

$$b = b + 0.01x_0 \quad (26)$$

with

$$x_0 = \frac{\sum^{\lambda < 0.5} \lambda \exp[-U^{\text{dwp}}(\lambda)]}{\sum^{\lambda < 0.5} \exp[-U^{\text{dwp}}(\lambda)]} \quad (27)$$

until

$$\text{abs}[x_0] < \epsilon \quad (28)$$

with ϵ arbitrarily small. Here, we used $\epsilon = 0.005$. Note that in eq 27 it suffices to consider only half of the λ interval ($\lambda < 0.5$) due to the symmetry of the biasing potential.

The parameter a , which determines the width of the Gaussian in eq 23, is adjusted iteratively as well:

$$a = \frac{a}{1 + 0.01 \frac{\sigma - \sigma_0}{\sigma_0}} \quad (29)$$

with

$$\sigma = \sqrt{\frac{\sum^{\lambda < 0.5} (\lambda - x_0)^2 \exp[-U^{\text{dwp}}(\lambda)]}{\sum^{\lambda < 0.5} \exp[-U^{\text{dwp}}(\lambda)]}} \quad (30)$$

until

$$\text{abs} \left[\frac{\sigma - \sigma_0}{\sigma_0} \right] < \epsilon \quad (31)$$

with $\sigma_0 = 0.02$, as mentioned earlier.

The parameter k at every iteration is

$$k = k + (h/2 + \min[U^{\text{dwp}}(\lambda)]) \quad (32)$$

where min is the minimum of the considered function.

Setting the initial values of k and d to half the desired barrier height is a convenient choice for finding the parameters that describe the desired potential most closely. As a consequence of this choice, the effective walls of the biasing potential shift by half the barrier height from their value w in eq 18.

■ APPENDIX B

To fix the total number of protons of a system during a constant pH molecular dynamics simulation, a constraint is applied. The constraint equation for a system with N^{sites} protonatable sites, each described by a titration coordinate λ , is defined as

$$\sigma(\lambda_1, \lambda_2, \dots, \lambda_N) = \sum_i^{N^{\text{sites}}} \lambda_i - c \quad (33)$$

with constraint force

$$G_i^\lambda = -\zeta \frac{\partial \sigma}{\partial \lambda_i} \quad (34)$$

where ζ is the undetermined Lagrange multiplier.

The constraint force G_i^λ is determined using the SHAKE algorithm.⁴¹ In the leapfrog algorithm, the constrained equations of motion for λ_i are integrated as

$$\begin{aligned} \lambda_i^c(t + \Delta t) &= \lambda_i^c(t) + v_i^{\lambda_i^c} \left(t + \frac{\Delta t}{2} \right) \Delta t \\ &= \lambda_i^c(t) + \left[v_i^{\lambda_i^c} \left(t - \frac{\Delta t}{2} \right) + \{ F_i^\lambda(t) + G_i^\lambda(t) \} \frac{\Delta t}{m_i} \right] \Delta t \end{aligned} \quad (35)$$

where the superscript c stands for constrained, $v_i^{\lambda_i^c}$ is the velocity of λ_i and F_i^λ the force acting on λ_i .

Under the assumption that the constraint is fulfilled at the previous time step t , an unconstrained update to $t + \Delta t$ is performed, using only the forces F_i^λ . To find the Lagrange multiplier ζ , the constraint eq 33 is expanded around the unconstrained $\lambda_i^u(t + \Delta t)$'s in a Taylor series and truncated after the first order:

$$\begin{aligned} \sigma(\lambda^c(t + \Delta t)) &= \sigma(\lambda^u(t + \Delta t)) \\ &+ \sum_i \frac{\partial \sigma}{\partial \lambda_i} \Big|_{\lambda_i^u(t + \Delta t)} \{ \lambda_i^c(t + \Delta t) - \lambda_i^u(t + \Delta t) \} \end{aligned} \quad (36)$$

Upon substituting eq 35, we arrive at

$$\begin{aligned}\sigma(\lambda^c(t + \Delta t)) &= \sigma(\lambda^u(t + \Delta t)) + \sum_i \left. \frac{\partial \sigma}{\partial \lambda_i} \right|_{\lambda_i^u(t + \Delta t)} G_i^\lambda(t) \frac{\Delta t^2}{m_i} \\ &= \sigma(\lambda^u(t + \Delta t)) - \sum_i \zeta \frac{\Delta t^2}{m_i}\end{aligned}\quad (37)$$

The condition $\sigma(\lambda^c) = 0$ requires that

$$\zeta = \frac{\sigma(\lambda^u(t + \Delta t))}{\sum_i \Delta t^2 / m_i} \quad (38)$$

from which the constraint force $G_i^\lambda(t)$ is readily obtained for computing the constrained $\lambda_i^c(t + \Delta t)$ values (eq 35).

Because the Taylor series is truncated, the expression for ζ is not exact. Therefore, this procedure is repeated, starting from the new λ_i^c 's, until the error in the constraint equation is below a certain threshold. Typically, ζ converges to machine precision within a few iterations.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b01160.

Force field parameters for the hydronium ions, the parameters of the biasing potential U^{dwp} , the coefficients of the polynomial fits to ΔG^{FF} , dynamics of deprotonation of a single titratable site with different heights of the outer walls of the potential U^{dwp} , charge conservation of the constraint approach, and the convergence of the protonation macrostate free energies in our simulations at pH = pK_a (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*(G.G.) E-mail: geritt.xgroenhof@jyu.fi.

*(H.G.) E-mail: hgrubmu@gwdg.de.

Funding

This work was supported by the Academy of Finland, the Deutsche Forschungs Gemeinschaft (DFG Priority Program 1648 Software for Exascale Computing), and the Max Plank Society.

Notes

The authors declare no competing financial interest.

■ REFERENCES

- Creighton, T. E. *Proteins. Structures and Molecular Properties*; W. H. Freeman: New York, NY, USA, 1993.
- Dobson, C. M. *Nature* **2003**, *426*, 884–890.
- Haas, J.; Vöhringer-Martinez, E.; Bögehold, A.; Matthes, D.; Hensen, U.; Pelah, A.; Abel, B.; Grubmüller, H. *ChemBioChem* **2009**, *10*, 1816–1822.
- Stouffer, A. L.; Acharya, R.; Salom, D.; Levine, A. S.; Di Costanzo, L.; Soto, C. S.; Tereshko, V.; Nanda, V.; Stayrook, S.; DeGrado, W. F. *Nature* **2008**, *451*, 596–599.
- Schnell, J. R.; Chou, J. J. *Nature* **2008**, *451*, 591–595.
- Tournaire-Roux, C.; Sutka, M.; Javot, H.; Gout, E.; Gerbeau, P.; Luu, D.-T.; Bligny, R.; Maurel, C. *Nature* **2003**, *425*, 393–397.
- Wiley, D. C.; Skehel, J. J. *Annu. Rev. Biochem.* **1987**, *56*, 365–394.
- Nehrke, K. *Wormbook: the Online Review of C. Elegans Biology* **2014**, 1–22.
- Boron, W. F. *Adv. Physiol. Ed.* **2004**, *28*, 160–179.
- Occhipinti, R.; Boron, W. F. *Prog. Biophys. Mol. Biol.* **2015**, *117*, 43–58.
- Krulwich, T. A.; Sachs, G.; Padan, E. *Nat. Rev. Microbiol.* **2011**, *9*, 330–343.
- Sondergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, *7*, 2284–2295.
- Bürgi, R.; Kollman, P. A.; van Gunsteren, W. F. *Proteins: Struct., Funct., Genet.* **2002**, *47*, 469–480.
- Bashford, D. *Front. Biosci., Landmark Ed.* **2004**, *9*, 1082–1099.
- Isom, D. G.; Cannon, B. R.; Castaneda, C. A.; Robinson, D.; García-Moreno E, B. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 17784–17788.
- Damjanović, A.; Brooks, B. R.; García-Moreno E, E. B. *J. Phys. Chem. A* **2011**, *115*, 4042–4053.
- Ullmann, R. T.; Ullmann, G. M. *J. Phys. Chem. B* **2011**, *115*, 10346–10359.
- Baptista, A. M.; Teixeira, V. H.; Soares, C. M. *J. Chem. Phys.* **2002**, *117*, 4184–4200.
- Dlugosz, M.; Antosiewicz, M. *Chem. Phys.* **2004**, *302*, 161–170.
- Mongan, J.; Case, D. A.; McCammon, J. A. *J. Comput. Chem.* **2004**, *25*, 2038–2048.
- Lee, J.; Miller, B. T.; Damjanovic, A.; Brooks, B. R. *J. Chem. Theory Comput.* **2014**, *10*, 2738–2750.
- Mongan, J.; Case, D. A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 157–163.
- Lee, M. S.; Salsbury, F. R., Jr.; Brooks, C. L., III. *Proteins: Struct., Funct., Genet.* **2004**, *56*, 738–752.
- Khandogin, J.; Brooks, C. L. *Biophys. J.* **2005**, *89*, 141–157.
- Zundel, G. *Angew. Chem., Int. Ed. Engl.* **1969**, *8*, 499–509.
- Wicke, E.; Eigen, M.; Ackermann, T. *Z. Phys. Chem.* **1954**, *1*, 340–364.
- Hub, J.; de Groot, B. L.; Grubmüller, H.; Groenhof, G. *J. Chem. Theory Comput.* **2014**, *10*, 381–390.
- Hünenberger, P. H.; McCammon, J. A. *J. Chem. Phys.* **1999**, *110*, 1856–1872.
- Hummer, G.; Pratt, L. R.; Garcia, A. E. *J. Phys. Chem.* **1996**, *100*, 1206–1215.
- Dixit, S. B.; Chipot, C. *J. Phys. Chem. A* **2001**, *105*, 9795–9799.
- Donnini, S.; Juffer, A. H. *J. Comput. Chem.* **2004**, *25*, 393–411.
- Wallace, J. A.; Shen, J. K. *J. Chem. Phys.* **2012**, *137*, 184105.
- Chen, W.; Wallace, J. A.; Yue, Z.; Shen, J. K. *Biophys. J.* **2013**, *105*, L15–L17.
- Knight, J. L.; Brooks, C. L., III. *J. Comput. Chem.* **2011**, *32*, 3423–3432.
- Knight, J. L.; Brooks, C. L., III. *J. Chem. Theory Comput.* **2011**, *7*, 2728–2739.
- Donnini, S.; Tegeler, F.; Groenhof, G.; Grubmüller, H. *J. Chem. Theory Comput.* **2011**, *7*, 1962–1978.
- Kong, X.; Brooks, C. L., III. *J. Chem. Phys.* **1996**, *105*, 2414–2423.
- Tidor, B. *J. Phys. Chem.* **1993**, *97*, 1069–1073.
- van der Vegt, N. F. A.; Briels, W. J. *J. Chem. Phys.* **1998**, *109*, 7578–7582.
- Abrams, J. B.; Rosso, L.; Tuckerman, M. E. *J. Chem. Phys.* **2006**, *125*, 074115.
- Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
- Süßmann, G. Z. *Z. Naturforsch., A: Phys. Sci.* **1997**, *52a*, 49–52.
- Schleich, W. P. Chapter Appendix C: Süßmann Measure. In *Quantum Optics in Phase Space*; Wiley-VCH: Weinheim, Germany, 2005; pp 611–613.
- Ullmann, R. T.; Ullmann, G. M. *J. Phys. Chem. B* **2011**, *115*, 507–521.
- Jacquet, P.; Szpankowski, W. *IEEE Trans. Inf. Theory* **1999**, *45*, 1072–1081.
- Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43–56.
- Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306–317.
- van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.; Berendsen, H. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

- (49) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–1676.
- (50) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction models for water in relation to protein hydration. In *Intermolecular Forces*, Vol. 14; Pullman, B., Ed.; Reidel: Dordrecht, The Netherlands, 1981; pp 331–342; DOI: [10.1007/978-94-015-7658-1_21](https://doi.org/10.1007/978-94-015-7658-1_21)
- (51) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1995**, *102*, 5451–5459.
- (52) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A. D.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (53) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952–962.
- (54) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (55) Andersen, H. C. *J. Chem. Phys.* **1980**, *72*, 2384–2393.
- (56) Wang, F.; Landau, D. P. *Phys. Rev. Lett.* **2001**, *86*, 2050–2053.
- (57) Wang, F.; Landau, D. P. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **2001**, *64*, 056101.
- (58) Bashford, D. An object-oriented programming suite for electrostatic effects in biological molecules An experience report on the MEAD project. In *Scientific Computing in Object-Oriented Parallel Environments*; Ishikawa, Y., Oldehoeft, R., Reynders, J., Tholburn, M., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 1997; Vol. 1343.
- (59) Ullmann, R. T.; Ullmann, G. M. *J. Comput. Chem.* **2012**, *33*, 887–900.
- (60) Ullmann, R. T.; Andrade, S. L. A.; Ullmann, G. M. *J. Phys. Chem. B* **2012**, *116*, 9690–9703.
- (61) Eyring, H. *J. Chem. Phys.* **1935**, *3*, 107–115.
- (62) Evans, M. G.; Polanyi, M. *Trans. Faraday Soc.* **1935**, *31*, 875–894.