Published in final edited form as:
J Chem Theory Comput. 2017 December 12; 13(12): 5933–5944. doi:10.1021/acs.jctc.7b00875.

Constant-pH Molecular Dynamics Simulations for Large Biomolecular Systems

Brian K. Radak[†], **Christophe Chipot**[‡,¶], **Donghyuk Suh**[§], **Sunhwan Jo**[†,@], **Wei Jiang**[†], **James C. Phillips**[‖], **Klaus Schulten**[‖,¶], and **Benoît Roux**[⊥,§,#]

[†]Leadership Computing Facility, Argonne National Laboratory, Argonne, IL 60439-8643, USA

[‡]Laboratoire International Associé Centre National de la Recherche Scientifique et University of Illinois at Urbana-Champaign, Unité Mixte de Recherche n°7565, Université de Lorraine, Université de Lorraine, B.P. 70239, 54506 Vandoeuvre-lès-Nancy cedex France

[¶]Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL, 61801-2325, USA

[§]Department of Chemistry, University of Chicago, Chicago, IL, 60637-1454, USA

[‖]Theoretical and Computational Biophysics Group, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, 61801-2325, USA

[⊥]Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL, 60637-1454, USA

[#]Center for Nanoscale Materials, Argonne National Laboratory, Argonne, IL 60439-8643, USA

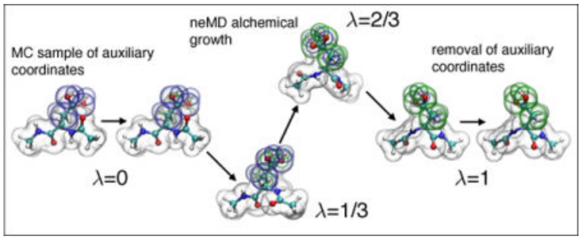
Abstract

An increasingly important endeavor is to develop computational strategies that enable molecular dynamics (MD) simulations of biomolecular systems with spontaneous changes in protonation states under conditions of constant pH. The present work describes our efforts to implement the powerful constant-pH MD simulation method based on a hybrid nonequilibrium MD/Monte Carlo (neMD/MC) technique within the highly scalable program NAMD. The constant-pH hybrid neMD/MC method has a number of appealing features; it samples the correct semi-grand canonical ensemble rigorously, the computational cost increases linearly with the number of titratable sites, and it is applicable to explicit solvent simulations. The present implementation of the constant-pH hybrid neMD/MC in NAMD is designed to handle a wide range of biomolecular systems with no constraints on the choice of force field. Furthermore, the sampling efficiency can be adaptively improved on-the-fly by adjusting algorithmic parameters during the simulation. Illustrative examples emphasizing medium and large scale applications on next-generation supercomputing architectures are provided.

Correspondence to: Benoît Roux.
[@]Current Address: Silcs Bio, LLC, Baltimore, MD, 21201-1193, USA

Supporting Information Available
Provided are detailed mathematical derivations in support of assertions made in the text, a complete listing of new and modified force field parameters, and a complete listing of fitting data from SNase simulations. This material is available free of charge via the Internet at http://pubs.acs.org/.

Graphical TOC Entry



1 Introduction

Most conventional molecular dynamics (MD) simulations of biomolecular systems aim to sample statistical mechanical ensembles with a fixed composition. This largely stems from the fact that many processes of interest are well described by classical, fixed valence force-field models (e.g., protein folding, conformational changes, ligand binding)[1–3]. Nonetheless, a vast amount of biochemistry is regulated by carefully buffered solutions and a complex interplay between multiple protonation states. This is clearly illustrated, for example, by the sensitivity of enzymes to pH (e.g., pH-rate studies) and the presence of distorted pH gradients in cancerous cells.[4,5] In many cases, the number of relevant states is relatively small (perhaps two to four) and can be studied by brute force enumeration. However, this approach quickly becomes untenable for larger systems or even simple solutions of modest concentration. Even if the number of truly relevant states is manageably small compared to the total number of possible states, it may still not be readily obvious which of the states is in fact important. If the feasibility of the calculation demands such an insight the investigator risks heavily biasing the results. Such systems require a constant-pH simulation approach that naturally accounts for variation of protonation states without a priori enumeration of the relevant states.

A classical MD simulation in the canonical ensemble typically samples according to a Hamiltonian $H(\boldsymbol{x})$, where $\boldsymbol{x}$ represents both the coordinates and momenta. Assuming that the system comprises m titratable sites, the Hamiltonian must be generalized to control the microscopic potential function upon changes in protonation states. For this purpose, we define a vector of coupling parameters, $\boldsymbol{\lambda} \equiv \{\lambda_1, \lambda_2, \ldots, \lambda_m\}$, where each element is a zero or one to indicate the absence or presence of a proton at a given site, respectively. The sum over all elements, $n_{\boldsymbol{\lambda}} \equiv \sum_{s=1}^{m} \lambda_s$ simply counts the total number of protons in the system that are coupled to the pH bath. It follows that the simulation samples from the probability distribution with partition function

$$Q_{\boldsymbol{\lambda}} = \int d\boldsymbol{x} e^{-\beta H(\boldsymbol{x};\boldsymbol{\lambda})}, \quad (1)$$

where $\beta \equiv 1/(k_B T)$, k_B is Boltzmann's constant, T is the absolute temperature, and the integral is over the system (or periodic cell) volume. The vector of coupling parameters is explicitly kept in the notation $H(\boldsymbol{x}; \boldsymbol{\lambda})$ as a reminder that the model is expected to represent a *family* of protonation possibilities. A constant-pH simulation samples according to this family of Hamiltonians by combining them into a single semi-grand canonical partition function:

$$\Xi(\mathrm{pH}) = \sum_{\boldsymbol{\lambda} \in \mathscr{S}} Q_{\boldsymbol{\lambda}} 10^{-n_{\boldsymbol{\lambda}} \mathrm{pH}}. \quad (2)$$

The summation extends over the complete set of possible protonation states $\mathscr{S}$, which has at most 2^m elements, although many of these states may be forbidden. It is also possible for $\boldsymbol{\lambda}$ to have elements representing protonation sites in solution to maintain charge neutrality, although this is not necessary.[6] Equation (2) has a form similar to that of an expanded ensemble[7] with the difference that each state has a weight that is explicitly pH-dependent and, thus, has a physical meaning. This differs from the conventional approach in which the weights are just arbitrary sampling devices. A Gibbs-sampling view of this problem[8] suggests that exploration of the space defined by $\boldsymbol{x}$ and $\boldsymbol{\lambda}$ can be accomplished by simply alternating sampling between the two. This is essentially the "stochastic titration" method first suggested by Baptista et al. [9]. Such an approach hinges on either the strict use of an implicit-solvent model[10] or else a sampling of the state space on an implicit representation followed by a period of solvent "relaxation".[9,11] These approximations are used in order to avoid very low efficiency due to steric clashes in explicit solvent.

More recently, Gibbs sampling methods have been generalized into a broad class of nonequilibrium MD/Monte Carlo (neMD/MC) schemes[6,12–15] and this is adopted here. In this scheme each Monte Carlo (MC) move consists of a short MD trajectory in which the system is driven from its current configuration and protonation state $(\boldsymbol{x}, \boldsymbol{\lambda})$ into a new candidate state $(\boldsymbol{x}', \boldsymbol{\lambda}')$. A "pure" Gibbs sampling scheme is recovered when the length of the trajectory is zero. The advantage of finite, non-zero length trajectories is that there is no need to rely on an auxiliary implicit-solvent model, which might otherwise limit either the transferability of the method or its extension beyond fixed-charge force fields. The disadvantage is that rejecting neMD/MC moves is generally expensive, since generating the candidate configuration requires a short MD simulation. However, it is difficult to compare this expense against other methods utilizing relatively expensive implicit-solvent models such as non-linear Poisson-Boltzmann. The latter models can be quite demanding for large systems and do not necessarily have cost scaling that coincides with explicit models, nor the same memory requirements.

Other constant-pH approaches are also possible, which do not sample the semi-grand canonical ensemble directly. For example, Lee et al. [16] developed a family of Hamiltonians based on enveloping distribution sampling (EDS), which can be reweighted to produce the desired statistics. Several research groups have also proposed variations based on continuous titration coordinates using an extended Lagrangian, whereby the elements of λ take fractional values and carry fictitious masses and momenta (so-called "λ-dynamics").[17,18] However, because protonation states fractionally coexist, some implementations do not appear to have included rigorous long-range electrostatics until recently;[19] this seems especially problematic for simulations of highly charged systems such as RNA.[20] Some implementations also require spurious modification of the bonded terms in the underlying force field model, for example, superposition of protonated and deprotonated carboxylate geometries.[21]

The motivation behind the current work is to address problem spaces that may not be appropriate for the other approaches described above. For example, methods based on an auxiliary implicit-solvent model are prone to fail when titrations do not occur in aqueous solvent, such as for membrane-bound proteins or otherwise buried sites.[11,22] Another concern is scalability to large systems with many titratable sites. The EDS-based approach in particular requires concurrent simulations (replicas) on the order of 2^m, such that even a system of modest size with 10 protonation sites would require a rather unwieldy 1,024 replicas. A similar argument can be made for continuous titration methods, which many research groups analyze by creating ad hoc bins for the fractional occupations observed near zero and one.[19] In this binning approach, the data outside of the bins is completely discarded in some cases. Clearly, the amount of time spent completely outside of the bins must be directly proportional to the number of fractional sites, thereby rendering less and less useable data as the system size increases. The neMD/MC approach addresses these shortcomings. As shown here, the new MC procedure naturally accounts for all types of environmental responses, including those found in crowded spaces such as lipid membranes. The cost of sampling also does not increase with the number of sites, although the overall sampling requirement obviously increases (the curse of dimensionality). The method also strictly respects the underlying model (e.g., no auxiliary implicit-solvent description is needed); the present study utilizes a fixed charge force field representation, but this is not algorithmically necessary. It is also noteworthy that an internally consistent description of tautomeric states is a natural part of the algorithm.[21,23] Lastly, the neMD/MC procedure can be extended to permit meaningful optimization based on the simulation history,[24] not least because of an iterative procedure for pK_a estimation.[6] All of these merits are seamlessly combined with the portability, scalability, and flexibility of the NAMD[25] simulation engine in order to permit constant-pH simulations on both commodity and capability computing resources.

2 Theory

The core theoretical arguments for the neMD/MC constant-pH approach have already been presented by Chen and Roux [6] based on earlier ideas due to Stern [12]. Some of these developments have also been known to the wider constant-pH community for some time (see, for example, a review by Mongan and Case [26] from over a decade ago). For clarity, the ideas needed to understand the new implementation are presented here.

2.1 Empirical Model Corrections

Consider the equilibrium of an arbitrary titratable system, A, interconverting between its protonated (HA) and deprotonated (A$^-$) forms:

$$\mathrm{HA} \overset{\Delta F_\mathrm{a}}{\rightleftharpoons} \mathrm{A}^- + \mathrm{H}^+,$$

where

$$10^{\mathrm{p}K_{\mathrm{a}}} \equiv e^{-\beta \Delta F_{\mathrm{a}}} = \frac{Q_{\mathrm{A}^-} Q_{\mathrm{H}^+}}{Q_{\mathrm{HA}}}. \qquad (3)$$

Most classical models do not provide a realistic dissociative model for protons because they neglect a physical description of covalent bond energies, nuclear quantum effects, and/or proton solvation. As such, direct evaluation of the pK_{a} is generally difficult and/or inconvenient. It is instead common to redefine the partition function ratio such that the model for a particular system exactly matches a known reference value $\mathrm{p}K_{\mathrm{a}}^{\mathrm{ref}}$:

$$10^{-\mathrm{p}K_{\mathrm{a}}^{\mathrm{ref}}} = \frac{Q_{\mathrm{A}^-}}{Q_{\mathrm{HA}}} e^{\beta \Delta E}, \qquad (4)$$

where

$$\Delta E \equiv (\Delta F_a - F_{\mathrm{H}^+}) - \frac{\ln 10}{\beta} \mathrm{p}K_{\mathrm{a}}^{\mathrm{ref}} \qquad (5)$$

and $F_{\mathrm{H}^+} \equiv -\beta^{-1} \ln Q_{\mathrm{H}^+}$ defines the absolute free energy of a proton in solution. For many force fields it is easiest to compute $\Delta F \equiv \Delta F_a - F_{\mathrm{H}^+}$ directly. However, this definition is slightly misleading since it implies that F_{H^+} is always a global constant. Clearly the actual physical quantity for F_{H^+} must be a constant, but this is only true for the model if it includes a meaningful description of bond breakage and formation. Otherwise, there are other additive errors in ΔF_a, which are not easily separated in the definition of ΔF. In practice, suffice to treat ΔF as a single free energy term even though its physical meaning is rather complex. From here on, all factors containing ΔE are assumed to be implicitly absorbed into the relevant partition function ratio.

It is also worth stating that ΔE is technically ensemble dependent. That is, ΔE formally depends on the system composition, volume, and temperature, and this dependence is necessarily inherited by the redefined systems. For systems in aqueous solution the dependence of $\mathrm{p}K_{\mathrm{a}}^{\mathrm{ref}}$ on volume can be safely ignored provided that the constant-pH ensemble is simulated at a density reasonably similar to that at which the reference data is generated. However, the effect of temperatures far from that at which $\mathrm{p}K_{\mathrm{a}}^{\mathrm{ref}}$ is measured may be non-negligible and therefore requires a correction.

2.2 Statistical Mechanical Connections

The above formalism can also be understood as a statistical mechanical form of the Henderson-Hasselbalch equation by identifying the protonated fraction as

$$P_{\mathrm{HA}}(\mathrm{pH})=\frac{Q_{\mathrm{HA}}10^{-\mathrm{pH}}}{Q_{\mathrm{A}^-}+Q_{\mathrm{HA}}10^{-\mathrm{pH}}}=\frac{1}{1+10^{\mathrm{pH}-\mathrm{p}K_{\mathrm{a}}}}, \quad (6)$$

where, by construction, $\mathrm{p}K_{\mathrm{a}}=\mathrm{p}K_{\mathrm{a}}^{\mathrm{ref}}$ for the reference system. The simplest variation is take the same species A within some other composition but no additional protonation sites. In this case different partition functions Q'_{A^-} and Q'_{HA} can be constructed, but their ratio is no longer directly equal to $\mathrm{p}K_{\mathrm{a}}^{\mathrm{ref}}$. Instead one finds that

$$\mathrm{p}K_{\mathrm{a}}=-\log\frac{Q'_{\mathrm{A}^-}}{Q'_{\mathrm{HA}}}=\mathrm{p}K_{\mathrm{a}}^{\mathrm{ref}}+\frac{\beta(\Delta F'_{\mathrm{a}}-\Delta F_{\mathrm{a}})}{\ln 10}.$$

The $\mathrm{p}K_a$ of other systems are thus seen to be shifted with respect to $\mathrm{p}K_{\mathrm{a}}^{\mathrm{ref}}$ by an amount that can be computed as a difference of relative free energies. In this two-state case the difference is exclusively a function of the original Hamiltonian definitions since the $F_{\mathrm{H}+}$ terms cancel; it does not depend on ΔE, except through the choice of reference state.

The construction of computing shifted $\mathrm{p}K_a$ values is not as straightforward when dealing with systems that possess more than two states. As per the general case described by Eq. (2), the state of a system with m protonation sites is completely defined by its occupancy vector $\boldsymbol{\lambda}$ (Figure 1). The number of states described by different permutations of $\boldsymbol{\lambda}$ will, in general, be considerably greater than two and a different shift value will be needed for each pair of states. Accordingly, the shift must instead be written as $\Delta E(\boldsymbol{\lambda}, \boldsymbol{\lambda}')$, where $\boldsymbol{\lambda}'$ is the occupancy vector for some other state. Subsets of the elements of $\boldsymbol{\lambda}$ can be organized into residues and these are the basic units used to define different values of $\Delta E(\boldsymbol{\lambda}, \boldsymbol{\lambda}')$. In practice, a change in $\boldsymbol{\lambda}$ usually only involves a few residues, and the change is computed by summing over the per residue shifts.

For any protonation site, the terms in Eq. (2) can be separated into two groups – those with and those without the proton present. This partitioning corresponds to separating the $Q_{\boldsymbol{\lambda}}$ into two groups based on whether a particular element is one (Ξ_1) or zero (Ξ_0). This splitting of the summation can be done for any site and defines two nonoverlapping summations over the set:

$$\Xi(\mathrm{pH})=\Xi_0(\mathrm{pH})+\Xi_1(\mathrm{pH})10^{-\mathrm{pH}}. \quad (7)$$

Note also that an extra pH-dependent factor has been factored out of Ξ_1 since each term in the summation has at least one more proton than those in Ξ_0. More complicated partitioning schemes with more than two groups can also be performed using sets of sites. For example, three groups can be used to enumerate the states of histidine based on its two sites [see Figure (1)]. Following the same procedure as for the two-state case yields

$$\mathrm{p}K_{\mathrm{a}}(\mathrm{pH}) = -\log\frac{\Xi_0(\mathrm{pH})}{\Xi_1(\mathrm{pH})}, \quad (8)$$

which is no longer a simple difference in relative free energies due to the pH-dependence. It is worth noting that this equation only contains the ratio Ξ_0/Ξ_1 such that the empirical corrections defined by $\Delta E(\boldsymbol{\lambda}, \boldsymbol{\lambda}')$ can be applied.

The dependence of the pK_a value on pH is frequently added into Eq. (6) by defining the Hill coefficient, n:

$$\mathrm{p}K_{\mathrm{a}}(\mathrm{pH}) \approx \mathrm{p}K_{\mathrm{a}}^{(\mathrm{a})} + (1-n)\left(\mathrm{pH} - \mathrm{p}K_{\mathrm{a}}^{(\mathrm{a})}\right). \quad (9)$$

This asserts that the deviation is no longer with respect to $\mathrm{p}K_{\mathrm{a}}^{\mathrm{ref}}$, but some other “apparent” pK_a, $\mathrm{p}K_{\mathrm{a}}^{(\mathrm{a})}$ (in most cases this is the pH value at which the occupied and unoccupied fractions are equal to 1/2). The pH dependence vanishes for $n = 1$ and the two-state case is recovered. This approximation might be considered as a first order series expansion about $\mathrm{p}K_{\mathrm{a}}^{(\mathrm{a})}$, although this viewpoint is quite different from the usual physical motivation for the Hill coefficient.[27]

2.3 neMD/MC Sampling

As in previous work[6,12] the neMD/MC scheme is composed of alternating sampling in $\boldsymbol{x}$ at a fixed protonation state $\boldsymbol{\lambda}$ using standard MD and neMD/MC moves sampling in both $\boldsymbol{x}$ and $\boldsymbol{\lambda}$ (Figure 2). Only the latter warrants additional comment. The neMD/MC detailed-balance equation has the form:

$$\frac{\rho(\boldsymbol{x}', \boldsymbol{\lambda}')}{\rho(\boldsymbol{x}, \boldsymbol{\lambda})} = \frac{T(\boldsymbol{x}, \boldsymbol{\lambda} \rightarrow \boldsymbol{x}', \boldsymbol{\lambda}')}{T(\boldsymbol{x}', \boldsymbol{\lambda}' \rightarrow \boldsymbol{x}, \boldsymbol{\lambda})}, \quad (10)$$

where T is the probability of the given transition and the ratio of equilibrium distribution functions is

$$\frac{\rho(\boldsymbol{x}', \boldsymbol{\lambda}')}{\rho(\boldsymbol{x}, \boldsymbol{\lambda})} = e^{-\Delta\varepsilon(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{\lambda}, \boldsymbol{\lambda}')} 10^{-\Delta n \mathrm{pH}}, \quad (11)$$

where

$$\Delta\varepsilon(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{\lambda}, \boldsymbol{\lambda}') \equiv \beta[H(\boldsymbol{x}'; \boldsymbol{\lambda}') - H(\boldsymbol{x}; \boldsymbol{\lambda}) + \Delta E(\boldsymbol{\lambda}, \boldsymbol{\lambda}')], \quad (12)$$

and $\Delta n \equiv n_{\boldsymbol{\lambda}'} - n_{\boldsymbol{\lambda}}$. Following Chen and Roux[6], the transition probability is split into two parts:

$$T(\boldsymbol{x},\boldsymbol{\lambda} \to \boldsymbol{x}',\boldsymbol{\lambda}') = T^{(\mathrm{i})}(\boldsymbol{\lambda} \to \boldsymbol{\lambda}')T^{(\mathrm{s})}(\boldsymbol{x} \to \boldsymbol{x}'|\boldsymbol{\lambda} \to \boldsymbol{\lambda}'), \quad (13)$$

where $T^{(\mathrm{i})}$ represents an "inherent" probability for the transition $\boldsymbol{\lambda} \to \boldsymbol{\lambda}'$, and $T^{(\mathrm{s})}$ is the transition probability of an attempted $\boldsymbol{x} \to \boldsymbol{x}'$ neMD "switch", conditional on the $\boldsymbol{\lambda} \to \boldsymbol{\lambda}'$ transition. The inherent transition probability is defined as

$$\frac{T^{(\mathrm{i})}(\boldsymbol{\lambda} \to \boldsymbol{\lambda}')}{T^{(\mathrm{i})}(\boldsymbol{\lambda}' \to \boldsymbol{\lambda})} = 10^{[\mathrm{p}K_{\mathrm{a}}^{(\mathrm{i})}(\boldsymbol{\lambda},\boldsymbol{\lambda}') - \Delta n\mathrm{pH}]} \quad (14)$$

where the quantity $\mathrm{p}K_{\mathrm{a}}^{(\mathrm{i})}(\boldsymbol{\lambda},\boldsymbol{\lambda}')$ is referred to as the pairwise inherent $\mathrm{p}K_{\mathrm{a}}$ for the transition between $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}'$. From this definition it follows that the conditional transition probability for a neMD switch is

$$\frac{T^{(\mathrm{s})}(\boldsymbol{x} \to \boldsymbol{x}'|\boldsymbol{\lambda} \to \boldsymbol{\lambda}')}{T^{(\mathrm{s})}(\boldsymbol{x}' \to \boldsymbol{x}|\boldsymbol{\lambda}' \to \boldsymbol{\lambda})} = e^{-\Delta\varepsilon(\boldsymbol{x},\boldsymbol{x}';\boldsymbol{\lambda},\boldsymbol{\lambda}')}10^{-\mathrm{p}K_{\mathrm{a}}^{(\mathrm{i})}(\boldsymbol{\lambda},\boldsymbol{\lambda}')}. \quad (15)$$

Adding and subtracting $\mathrm{p}K_{\mathrm{a}}^{(\mathrm{i})}(\boldsymbol{\lambda},\boldsymbol{\lambda}')$ in the exponent preserve the equilibrium distribution. These detailed-balance conditions can be satisfied by simple Metropolis criteria for both $T^{(\mathrm{i})}$ and $T^{(\mathrm{s})}$. In the latter case, one can also use a generalized neMD/MC criterion by replacing the energy difference $H(\boldsymbol{x}';\boldsymbol{\lambda}) - H(\boldsymbol{x};\boldsymbol{\lambda})$ with the nonequilibrium work applied during the switch (see Computational Methods).[6,13,15] The splitting of Eq. (13) into the transition probabilities $T^{(\mathrm{i})}$ and $T^{(\mathrm{s})}$ given by Eqs. (14) and (15) is a generalization of the method previously introduced by Chen and Roux[6].

Because it cancels exactly upon multiplication, the choice of $\mathrm{p}K_{\mathrm{a}}^{(\mathrm{i})}(\boldsymbol{\lambda},\boldsymbol{\lambda}')$ is completely arbitrary and does not affect detailed balance. However, it clearly affects sampling efficiency by partitioning effort between the two steps. It has been previously shown that choosing $\mathrm{p}K_{\mathrm{a}}^{(\mathrm{i})}(\boldsymbol{\lambda},\boldsymbol{\lambda}')$ to be the true $\mathrm{p}K_{\mathrm{a}}$ maximizes the efficiency because the free energy of the switching transformation effectively becomes zero.[6,24]

There are a few modifications to the inherent transition step that make the algorithm more useful for systems that contain multiple residues and/or residues with more than two states. First, there is an implicit proposal component in $T^{(\mathrm{i})}$ that is fixed with respect to $\boldsymbol{\lambda}$ and thus immediately falls out of the detailed-balance condition. That is, each residue (or group of residues) that can be titrated is assigned a fixed weight during the simulation. At the beginning of the neMD/MC step, the complete set of states permitted within this group is selected directly according to the probability mass function defined by the (normalized) weights. Two such choices are illustrated in Figure 1, whereby a carboxylate moiety and/or methyl imidazole group might be chosen.

Once the residue selections have been made, the remainder of the transition probability is split yet further into another proposal and a preliminary acceptance step:[8,28]

$$T^{(\mathrm{i})}(\boldsymbol{\lambda} \rightarrow \boldsymbol{\lambda}^{'}) = T_{\mathrm{p}}^{(\mathrm{i})}(\boldsymbol{\lambda} \rightarrow \boldsymbol{\lambda}^{'}) T_{\mathrm{a}}^{(\mathrm{i})}(\boldsymbol{\lambda} \rightarrow \boldsymbol{\lambda}^{'}) \qquad (16)$$

$$T_{\mathrm{p}}^{(\mathrm{i})}(\boldsymbol{\lambda} \rightarrow \boldsymbol{\lambda}^{'}) = \begin{cases} 0 & \boldsymbol{\lambda} = \boldsymbol{\lambda}^{'} \\ \frac{p(\boldsymbol{\lambda}, \boldsymbol{\lambda}^{'})}{1 - p(\boldsymbol{\lambda}, \boldsymbol{\lambda})} & \boldsymbol{\lambda} \neq \boldsymbol{\lambda}^{'} \end{cases} \qquad (17)$$

$$T_{\mathrm{a}}^{(\mathrm{i})}(\boldsymbol{\lambda} \rightarrow \boldsymbol{\lambda}^{'}) = \min\left[1, \frac{1 - p(\boldsymbol{\lambda}, \boldsymbol{\lambda})}{1 - p(\boldsymbol{\lambda}, \boldsymbol{\lambda}^{'})}\right], \qquad (18)$$

where

$$p(\boldsymbol{\lambda}, \boldsymbol{\lambda}^{'}) \equiv \frac{10^{\mathrm{p}K_{\mathrm{a}}^{\mathrm{i}}(\boldsymbol{\lambda}, \boldsymbol{\lambda}^{'}) - (n_{\boldsymbol{\lambda}^{'}} - n_{\boldsymbol{\lambda}})\mathrm{pH}}}{\sum_{\boldsymbol{\lambda}^{''}} 10^{\mathrm{p}K_{\mathrm{a}}^{\mathrm{i}}(\boldsymbol{\lambda}^{'}, \boldsymbol{\lambda}^{''}) - (n_{\boldsymbol{\lambda}^{''}} - n_{\boldsymbol{\lambda}^{'}})\mathrm{pH}}}. \qquad (19)$$

The first step directly samples all states that are not the current state, while the second step accepts or rejects this proposal using a Metropolis criterion and appropriate renormalization. Chodera and Shirts [8] refer to this approach as "Metropolized indepence sampling." This is because, for two states, $T_{\mathrm{p}}^{(\mathrm{i})}$ chooses the only possible candidate state 100% of the time and $T_{\mathrm{a}}^{(\mathrm{i})}$ reduces to a simple Metropolis criterion with the conventional exponential form. Conversely, when $p(\boldsymbol{\lambda}, \boldsymbol{\lambda})$ is very small, $T_{\mathrm{a}}^{(\mathrm{i})}$ essentially evaluates to one and $T_{\mathrm{p}}^{(\mathrm{i})}$ becomes an independence sampling amongst all possible trial states.

In practice, this algorithm will tend to propose the most probable state that is not the current state, unless the current state is strongly favored by the system pH. For example, at low pH a histidine residue is most probably in its protonated state. Since the only other states are neutral tautomers, one of these *must* be proposed and with their fixed tautomeric ratio (near 2 : 1 in single peptides[29]). Nonetheless, such a proposal will probably be rejected. Conversely, at high pH the histidine is likely to be in one of its neutral forms and the probability of proposing the protonated state is low. However, it is important to keep in mind that even if the inherent transition step is accepted, this only means that the algorithm then proceeds to the switch step, which itself can be accepted or rejected. Importantly, since all of the evaluations needed to compute $T_{\mathrm{p}}^{(\mathrm{i})}$ and $T_{\mathrm{a}}^{(\mathrm{i})}$ are exceedingly inexpensive compared to the cost of the full switch trajectory, it is worth repeating the process several times in order to choose a good candidate. The present algorithm chooses a maximum number of attempts (by default, the number of titratable residues in the system) in which to propose and preliminarily accept a switch move, which is then performed. If the maximum number of attempts is reached, then no switch is performed and neMD/MC cycle is considered as completed.

3 Computational Methods

The constant-pH implementation described here is available as a Tcl plugin, `namdcph`, for use in conjunction with NAMD 2.12 and later versions.[25] The focus here is on proteins, but the implementation is flexible enough to permit new residues (or molecules) with arbitrary numbers of sites and states, provided that force field definitions and reference pK_a values are available. All simulations here were carried out with the CHARMM36 force field[30] and thus the titratable amino acid selection is limited to only those definitions (serine is defined in CHARMM36, but not used here). Due to subtleties of the CHARMM residue-topology file format, terminal groups are not yet titratable and are instead fixed in their zwitterionic forms.

All MD simulations, constant-pH or otherwise, utilized the same simulation settings. Periodic boundary conditions were employed using particle mesh Ewald electrostatics[31] and smooth switching of the Lennard-Jones forces[32] between 10 and 12 Å, after which an isotropic long-range approximation was used. The exception to this is membrane simulations during which switching was performed between 8 and 10 Å and the isotropic correction was neglected. Langevin dynamics was employed at 298 K with a friction coefficient of 1 ps^{-1} coupled to heavy atoms only. Unless otherwise specified, integration was performed with an r-RESPA[33] multiple time-stepping scheme with an effective time step of 2 and 4 fs for short- and long-range interactions, respectively. The RATTLE[34] and SETTLE[35] algorithms were used to constrain covalent bonds involving hydrogen to their equilibrium lengths.

3.1 Alchemical Coupling and Force Field Modifications

When needed, alchemical coupling between protonation states was accomplished via linear coupling using a “dual-topology” paradigm with additional zero-length bonds between equivalent but otherwise non-interacting atoms (force constant 100 kcal/mol-Å^2).[36,37] The key advantage of this approach is that it resolves topological conflicts between different protonation states, especially when rigid bonds are used. The result of the isotropic harmonic bonds is that, when one set of alchemical atoms is completely uncoupled, the additional Boltzmann factors for the kinetic and potential energy of each atom have a Gaussian form in the Cartesian basis.[38] This is exploited during constant-pH MD by deleting the noninteracting atoms during “normal” MD and then resampling them when a neMD/MC switching trajectory is initialized (Figure 3). After a neMD/MC move in which the candidate state is accepted, the newly uncoupled atoms are again deleted, otherwise the initial coordinates before resampling are used. As a requirement for this procedure to work, the number of atoms must be constant before and after the switch. This means that the coordinates for a constant-pH trajectory can be analyzed and visualized as if they belonged to a conventional simulation with fixed composition.

Clearly, the above scheme necessitates the introduction of "dummy protons" for all deprotonated states. These phantom particles only interact with the system via a small number of bonded force-field terms. If made appropriately, these adjustments to the model produce the same ensemble averages, but do affect the free energy and dynamics of the system (see Supporting Information). Since the free energies are themselves part of the constant-pH simulation (via the reference free energy computations) this strategy poses no problem. The definition of the potential experienced by dummy atoms also affects the sampling efficiency and can be optimized, for example, to produce rapid transitions between configurations that might otherwise be separated by a barrier when the proton is coupled (see Supporting Information).

3.2 Switching Protocols and Alchemical Work

The neMD/MC switching trajectories additionally require a nonequilibrium protocol for uncoupling the initial protonation state and coupling the candidate state. Here a linear switching protocol is used such that the coupling constant changes at the beginning of each step (an asymmetric splitting of the Liouvillian[39]). This scheme is in line with the "BBK" leap-frog integrator used in NAMD.[25,40] A linear switch appears to be the most efficient scheme (in the sense of maximizing the mean acceptance probability) when the endpoints are linearly coupled and the transformation is antisymmetric in time.[24] With both these assumptions the work exerted during an N-step nonequilibrium protocol is (excluding integrator error[15,39,41])

$$W_p = \frac{1}{N}\sum_{t=0}^{N-1}[U_1(\boldsymbol{x}_t) - U_0(\boldsymbol{x}_t)], \qquad (20)$$

where U_0 and U_1 are the potential energy functions of the system before and after the switch, respectively. This form for the work is not necessarily ideal, since it assumes that the integrator error is completely negligible with respect to the free energy of the transformation. Other integrators and/or expressions for the work may improve upon this scheme,[39,41,42] but our preliminary experience is that any such errors are negligible relative to normal statistical uncertainty.

3.3 Titration Curve and pK_a Estimation

Sampling from a constant-pH, semi-grand canonical ensemble provides a direct estimate of any number of protonation-state populations by taking simple averages of the elements of the occupation vector, $\boldsymbol{\lambda}$. Specifically, the fractional population expressions defined in terms of pK_a and pH [Eq. (6)] can be expressed as ensemble averages of site occupancies. For example, if a residue has two states that differ only in the presence and absence of proton s, then the observed protonated fraction is simply $P_s(\text{pH}) = \langle \lambda_s \rangle$, where λ_s is the s-th element of $\boldsymbol{\lambda}$ and $\langle \cdot \rangle$ indicates an average at fixed pH. By taking many observations at different pH values these averages can be interpreted as a titration curve. The most straightforward analysis is to perform nonlinear regression using Eqs. (6) and (9) to determine both an apparent pK_a and a Hill coefficient.

3.3.1 Macro-/Microscopic Titrations—The simple formalism of averaging elements of $\boldsymbol{\lambda}$ only describes a method for considering *microscopic titrations* – movements of individual protons. However, most residues are best characterized by their unique *macroscopic titration*, which often includes multiple sites grouped together. For the imidazole example in Figure 1, this means grouping together *both* sites, since the neutral states are distinct, but convert into the same protonated state. The protonated fraction for this case is thus $\langle \lambda_s \lambda_{s+1} \rangle$ and this is the titration curve that can generally be observed in a laboratory titration. However, the two deprotonated fractions must be computed separately as $\langle \lambda_s (1 - \lambda_{s+1}) \rangle$ and $\langle (1 - \lambda_s) \lambda_{s+1} \rangle$ and correspond to microscopic titrations.

If sites are equivalent, the macro-/microscopic distinction can still be made, but might be less useful. Perhaps the most obvious example is a carboxylate, for which the macroscopic protonated fraction is the aggregate of two equivalent sites (i.e., $\langle \lambda_1 (1-\lambda_2) + (1-\lambda_1)\lambda_2 \rangle$, Figure 1). In principle the two components can be computed separately to yield two identical pK_a values. In general, a residue with q equivalent sites and p protons in the protonated state will have macroscopic (M) and microscopic (μ) values that differ as:

$$\mathrm{p}K_{\mathrm{a}}^{\mu} - \mathrm{p}K_{\mathrm{a}}^{\mathrm{M}} = \log\left(\frac{p}{q-p+1}\right). \quad (21)$$

Since this difference is a straightforward constant, we choose the macroscopic value as more intuitive in nearly all instances. However, it may still be useful to monitor the microscopic values separately, since agreement between equivalent sites is a necessary (but not sufficient) condition for statistical convergence.

3.3.2 Accurate Estimation with WHAM—Here we note a simple and straightforward use of the unbinned weighted histogram analysis method (UWHAM),[43] which has not, to our knowledge, been reported in the literature before. It is appropriate when data has been collected at multiple pH values and can be extended to variation of other parameters, such as temperature or additional bias potentials for enhanced sampling. The UWHAM equations only involve energy *differences* as a function of the parameter that is being varied (i.e., pH). Since the Hamiltonian of the system does not depend on the pH, all terms involving $\boldsymbol{x}$ cancel and only terms containing $\boldsymbol{\lambda}$ remain. If occupation vectors are tracked during simulations at $k = 1, \ldots, M$ pH values, then the protonated fraction of some state defined by the indicator function $\chi(\boldsymbol{\lambda}_t)$ is (see the discussion above):

$$P_{\chi}(\mathrm{pH}) = \frac{1}{N} \sum_{t=1}^{N} w_t(\mathrm{pH}) \chi(\boldsymbol{\lambda}_t), \quad (22)$$

where

$$w_t(\mathrm{pH}) \equiv \left[\sum_{k=1}^{M} \frac{N_k}{N} e^{f(\mathrm{pH}_k)-f(\mathrm{pH})} 10^{-(\mathrm{pH}_k-\mathrm{pH})_{n_t}}\right]^{-1} \qquad (23)$$

is the effective weight of sample t ($0 \le w_t(\mathrm{pH}) \le 1$), n_t is the number of protons observed at each of N_k samples observed at pH_k, and $N \equiv \sum_{k=1}^{M} N_k$. The summation over t thus includes all observed occupancy vectors from all pH values at which data were gathered. The function $f(\mathrm{pH})$ is the semi-grand potential with respect to the pH, which must first be determined at the M values where samples were accrued.[43–46] However, after this has been done *any* pH value may be inserted into Eq. (22) and so it is effectively an analytic estimate of the titration curve, albeit containing N parameters. This is to be contrasted with the two-parameter Hill coefficient approach.

Although the above is a strikingly simple special case of the traditional Ferrenberg-Swendsen, WHAM equations,[43–46] it has some potentially unexpected consequences. Specifically, for simple systems where only one proton is titrating Eq. (22) is *exactly* a sigmoid for each state (see Supporting Information). As such, all reweighted populations will fall on exactly the same curve without any fitting. This means that Eq. (6) can be inverted at any point to give the same value for the pK_a, as it should. This is clearly not the case when populations are counted separately at each pH, as is usually done. In practice one can still perform fitting on any selection of points, but the asymptotic standard error of the parameters will be numerically zero. This procedure obviates the need for Hill coefficients, since these would be exactly unity within numerical error.

In the present work, the reported titration curves are computed using UWHAM and all reported pK_a values reflect a nonlinear regression utilizing a Hill coefficient, unless it is rigorously unity. The observed populations in each fit are taken only at those pH values where data was collected and the population is neither exactly one nor zero (i.e., when the pH is very different from the pK_a). When applicable, the reported values are for the aggregate data over multiple runs and the error is the standard deviation of fits to the individual data sets; this quantity is generally larger and more realistic than the fitting error. If a replicate did not provide a meaningful estimate of a pK_a then the appropriate extremal pH value was assigned instead.

3.4 Reference Energy Shifts

Reference energy shifts were computed using a set of terminally blocked dipeptides solvated in a pre-equilibrated 39 Å cube of water (1981 molecules). After minimization (500 steps) and equilibration (1 ns) in the protonated state the system was then converted to a dual topology and the same process was repeated. The free energies between pairs of states at 298 K were first computed using a two-dimensional expanded ensemble scheme[47] in which the alchemical coupling constant (six values linearly spaced between zero and one) and thermostat temperature (seven values exponentially spaced between 290 and 325 K) were varied. Each simulation was 200 ns long with 10 ps between proposed state changes and data collection. In order to make conservative estimates of the free energies, the r-RESPA scheme was *not* used for these simulations and the first 500 ps were discarded as equilibration before analysis with UWHAM.[43]

A second round of energy shifts were computed for each amino acid by running constant-pH simulations at six pH values (eight for histidine) spaced at 0.2 unit intervals about the desired reference $\mathrm{p}K_\mathrm{a}$. The reported titration curves and $\mathrm{p}K_\mathrm{a}$ values reflect the pooled data from eight trials, while the error bars are the scaled standard error of the individual $\mathrm{p}K_\mathrm{a}$ estimates. All simulations attempted protonation moves every 10 ps with switch times of 15 ps (this seems to be near optimal for a solvent exposed carboxylate[24]) over the course of at least 10 ns and the first 1 ns was discarded as equilibration. In each case the computed/observed $\mathrm{p}K_\mathrm{a}$ was then compared with $\mathrm{p}K_\mathrm{a}^\mathrm{ref}$ and the free energy value was corrected if necessary. After a correction the complete assay was repeated until the observed and reference values agreed within reasonable certainty.

3.5 Membrane Translocation of a Titratable Peptide

A 1–palmitoyl–2–oleoyl–phosphatidylcholine (POPC) bilayer was constructed with approximate dimensions 57 × 57 × 126 Å (100 lipid units, 9282 water molecules) using the CHARMM-GUI membrane builder.[48] A terminally blocked pentapeptide, AADAA, was then immersed in the aqueous phase, roughly 50 Å away from the barycenter of the membrane. For comparison purposes, a reference assay of the pentapeptide immersed in a bulk aqueous environment was also prepared (39 Å cube, 1936 water molecules). The initial configurations, wherein the pentapeptide was located at distinct altitudes (between 0 and 50 Å at 10 Å intervals) along the normal to the lipid bilayer (i.e., the z coordinate) were generated by a 200 ps steered MD simulation.[49] Each of the six resulting structures were then equilibrated (50 ns) while the barycenter of the pentapeptide was held near a constant value of z by a positional harmonic restraint with a force constant of 100 kcal/mol-Å^2.

Constant-pH MD assays of the titration curves at each z-value were determined from up to eleven individual simulations, representing an aggregate time ranging from about 180 to 680 ns per z-value (2.4 μs total for the full translocation). In addition, the reference $\mathrm{p}K_\mathrm{a}$, determined using the assay in bulk water, was obtained from up to seven individual simulations, amounting to an aggregate time of 310 ns. All simulations attempted protonation moves every 10 ps with switch times of 10 ps. In accordance with the expectation that the $\mathrm{p}K_\mathrm{a}$ value of the aspartate would increase near the membrane, the estimated inherent $\mathrm{p}K_\mathrm{a}$ value was increased (as high as six units) for smaller values of z to achieve a more efficient sampling.

Lastly, in order to obtain the correct baseline in water of the $\mathrm{p}K_\mathrm{a}$ profile as a function of z and, hence, account for the fact that an appreciable fraction of the simulation assay is occupied by the lipid bilayer, causing a shift in the electrostatic potential, a separate 10 ns simulation was performed in the absence of the pentapeptide. Based on this additional simulation, the average electrostatic potential along the z-axis was computed, from whence a $\mathrm{p}K_\mathrm{a}$ shift of 2.4 units was inferred due to a difference in the Galvani potential.[50]

3.6 Simulation of a Globular Protein

Staphylococcal nuclease (PDB: 3BDC)[51] was solvated in a 86 Å truncated octahedron (15557 water molecules) with a NaCl concentration near 100 mM after neutralization at pH 7 (26 Na^+, 31 Cl^-) using CHARMM-GUI.[48] This was intended to reproduce as closely as possible the setup used by Huang et al. [19] in their constant-pH simulations of the same system, except omitting their use of a hydroxide force field. All simulations of this system also utilized hydrogen mass repartitioning of the protein.[52] The system was first minimized (500 steps) and then equilibrated (6 ns) with harmonic restraints (10 kcal/mol-$Å^2$) on the heavy atoms set against the crystal structure reference. These were decreased by half at 200 ps intervals and then removed completely after 1 ns. Equilibration also utilized pressure coupling at 1 bar with a Langevin-piston barostat (piston period of 50 fs and decay time of 25 fs). Constant-pH MD assays of the titration curves were performed on 12 pH values between 2.0 and 7.5 at 0.5 unit intervals and repeated four times. All simulations attempted protonation moves every 10 ps over 34 ns with switch times of 20 ps (i.e., 3400 neMD/MC cycles) and the first 1 ns was discarded as equilibration. Inherent pK_a values were assigned using the experimental values of Castañeda et al. [51] and fixed throughout the simulation (n.b., this does not affect the outcome of the simulation but only the efficiency of the sampling). If experimental values were not available, the reference pK_a was used instead.

4 Results and Discussion

4.1 Reference Simulations

A core component of calibrating the constant-pH approach described here is the computation of the reference energy shifts defined by Eq. (5) These are tabulated in Table 1 for the CHARMM36 protein force field. It is important to note that both components of these shifts may display some temperature dependence, although this is difficult to gauge unless the corresponding experimental data is available. These limitations are not specific to the present constant-pH treatment but are expected to arise with all simulations based on force fields.

After employing the results of Table 1 in constant-pH simulations, it can be seen that the reference pK_a construction was successful as all values are reproduced within 0.2 units. (Figure 4). It is worth noting that the simulations here are extremely conservative in length and large error bars are assumed (two and a half standard deviations of the mean). Combined with the unavoidable error bars from the reference free energy simulations (Table 1) it would seem that, in practice, pK_a values estimated from constant-pH simulations are only likely to be systematically accurate within ~0.3 units. This is because any pK_a calculation based on constant-pH is intrinsically a relative pK_a with respect to these reference quantities. These must always carry some statistical uncertainty into the simulation and this cannot be removed by additional sampling, hence it is effectively systematic. However, these errors could cancel considerably when examining pK_a values between different residues in the same system. The real strength of the method should be in determination of *correlations* between titratable sites. Conventional free energy simulations will likely be superior in strict quantitative estimation, but would require a great deal of manual intervention for determining which groups meaningfully interact.

4.2 pK_a Shifts From Peptide Translocation Across a Membrane

A key motivation of the neMD/MC constant-pH approach is to enable efficient changes of protonation states in crowded environments, such as lipid membranes. Existing methods based on implicit solvation models, for example, are unlikely, to be efficient in this regime. As a cogent example we demonstrate a titratable pentapeptide at various levels of insertion above a POPC lipid bilayer. The evolution of the pK_a of the central aspartate residue as a function of the POPC bilayer normal is shown in Figure 5. As a basis of comparison, the pK_a of the same pentapeptide in a bulk aqueous environment was determined to be 3.9. This result is in excellent agreement with the potentiometric titration of 3.9 in synthetic, uncharged alanine–based pentapeptides,[55] and consonant with the average measurement of 3.5 in a series of folded proteins.[56] From the onset, a shift of the pK_a can be observed as the permittivity of the environment progressively changes from that of water to the interior of the bilayer. While the pK_a remains nearly that in the bulk aqueous medium starting roughly 15 Å away from the head–group region, it increases almost linearly as the pentapeptide translocates towards the center of the membrane hydrophobic core. At $z = 0$, the pK_a peaks at 7.3, which corresponds to a shift of 3.4 units with respect to the bulk region, far from the interface. Obtaining a converged value of the pK_a when the peptide is buried deep in the interior of the POPC bilayer constitutes a daunting task, requiring substantial sampling, owing to the partial hydration of the titratable amino acid. As the pentapeptide partitions into the membrane, it is accompanied by a retinue of water molecules trapped amidst the lipid chains and preserving, at least in part, the hydration state of the carboxylic-acid moiety. As a basis of comparison, although their constant-pH simulations do not tackle the more difficult scenario wherein the pentapeptide lies in the middle of the bilayer, Teixeira et al. [57] predict a similar trend in the shift of the pK_a.

4.3 Virtual Titration of Staphylococcal Nuclease

A second motivation for the neMD/MC constant-pH approach is to enable efficient sampling of large numbers of protonation states. Scaling in this manner may be a limitation of methods that utilize intermediate states where protons are only partially interacting. As a demonstration of this ability we present simulations of a medium-sized globular protein, Staphylococcal nuclease (SNase, 143 residues), over a broad range of pH values. Representative titration curves (Figure 6) show that multiple titratable side chains are well described with a diverse set of responses to the protein environment (e.g., GLU10 and GLU52 differ by ~2 units in their apparent pK_a). Quantitative determinations of the apparent pK_a values (Table 2) show excellent agreement with both experimental and theoretical determinations of the carboxylate groups (to our knowledge, the only groups for which data is available). Complete fitting results, including Hill coefficient comparisons, are available in the Supporting Information.

Although Huang et al. [19] also used the CHARMM36 force field, it is unclear exactly how much agreement one can expect between the two sets of simulations. In many cases the values are extremely similar (as few as 0.1–0.2 units). Others differ by as much as 0.8 units, but these cases also have large relative statistical uncertainty. If perfect agreement is assumed to be possible, then our previous speculation that absolute pK_a values (regardless of statistical uncertainties) can only be trusted within 0.5 units seems to be reasonable.

Although SNase contains several lysine residues, nearly all of them have pK_a values outside the pH range used here and therefore show zero protonation events. This is *not* because these residues were not permitted to titrate, but is instead an intrinsic feature of the two-step inherent pK_a algorithm.[6] Accordingly, these residues are not listed in Table 2 and can only be said to have pK_a values greater than 7.5 based on the data here. Interstingly, LYS24 has an apparent pK_a of 8.4 and spent as much as 10% of the simulation at pH 7.5 in its neutral form (see Figure 6). Our simulation does not necessarily render a physically accurate description of SNase, but it highlights the fact that the method used here automatically captures unexpected behavior without any input from the user.

4.4 Practical Considerations

The examples above are intended to be representative of both typical and challenging cases amenable to constant-pH simulations. It is worth discussing possible limitations and shortcomings of the method, specifically as they could have been encountered in these demonstrations. Most glaringly, there are two adjustable parameters in the method, the switch time and the inherent pK_a, which strongly affect efficiency and, if chosen improperly, could have given rise to severely disappointing results. These parameters deserve closer individual discussion.

In previous work we analyzed the efficiency of a simple carboxylate system in explicit solvent and did a systematic test of short and long switch times.[24] In that work it was shown that an optimal switch (in the sense of maximizing the transition rate between states) should exist and depends on both the magnitude and intrinsic time scale of equilibrium fluctuations in the “force” along the interaction coupling. It was found that the optimal switch time was roughly an order of magnitude greater than the (apparent) time scale – a strikingly reasonable 11 ps. The present work seems to confirm that this estimate is transferable to titratable groups exposed to the solvent, even non-carboxylate moieties; we therefore recommend an initial value of 10–20 ps for essentially all residue types presented here. A modest extension of the same theoretical analysis also indicates that the optimal switch time almost always corresponds to an optimal mean acceptance probability of 20–25% (see Supporting Information). Although no adjustments seemed to be necessary in this work, a simple and reasonable adaptive scheme would be to track the acceptance rate (this is a standard output in the current code) and then increase (or possibly decrease) the switch time based on this simple criterion.

The two-step inherent pK_a algorithm is a critical component of the overall performance when simulating many residues across a broad range of pH values. A given simulation will naturally spend more time sampling residues with pK_a values close to the pH and therefore most physically relevant. This is clearly demonstrated by the SNase example above, wherein several lysine residues were permitted to titrate throughout the simulation (and occasionally did), but nearly all protonation state changes at pH ≤ 7 were by aspartate and glutamate, which had predicted pK_a values between two and six. In other words, the imposed pH and predicted pK_a values must closely coincide, otherwise titration of the site will be essentially ignored. This is also appealing from the standpoint that setting the pK_a of a residue to plus or minus infinity (or any large number in fact) effectively assigns a fixed protonation state.

This is a much more explicit practice than simply assuming a fixed valence when constructing the system topology.

The main disadvantage of the inherent pK_a algorithm is that residues for which titration is desired must have estimated pK_a values that are fairly accurate (or at least in the pH range being studied). Otherwise, efficiency will be severely impacted. If one assumes that most residues are only weakly shifted with respect to their reference value, then this simple estimate should be adequate in most cases. However, for larger shifts this can be problematic.

For example, consider a system with two aspartate residues, one of which is expected to be shifted towards neutrality by about two units, while the other is assumed to be near its reference value. In most biological applications, the shifted residue is of more probable importance and so it would seem reasonable to focus the majority of simulations near a pH of six. Imagine instead that these residues were misidentified and their behavior is reversed or, at the very least, that the residue assumed to be unshifted is also in fact shifted. In the former example the inherent pK_a algorithm will fail almost completely in the sense that very few state changes are likely to be successful (much less attempted). In the latter case, the results may still be highly biased, since the true, shifted behavior of the aspartate may be hidden by the narrow range of pH values. This scenario, although contrived, is a strong argument in favor of using a wide range of pH values (an extent of at least four units seems reasonable) or even using an expanded ensemble in which the pH is able to vary.[11,22,58] A complementary and/or alternative adjustment consists in selectively deactivating the inherent pK_a algorithm for a small subset of residues that are either suspected of being important or have otherwise uncharacterized behavior. This can be done by trivially setting the inherent pK_a equal to the pH. Since the particular value of the inherent pK_a only impacts efficiency, these residues could even be “re-activated” at a later time if data collection indicates that the behavior is not of interest.

5 Conclusion

This work introduces yet another route to performing constant-pH simulations. However, far from being a gratuitous exercise, this approach offers several advantages and features with respect to existing approaches. The implementation is efficient and scalable, and represents one of the few methods that can be plausibly used on very large chemical systems with large numbers of titratable sites. The approach is also general with respect to the model and does not rely on any special treatment of the solvent; this aspect is of paramount importance for membrane simulations, for example. Additional work is ongoing to integrate the method with next-generation force fields such as those that include polarizability, for instance by means of the introduction of Drude oscillators. The method is also agnostic to the details of the equilibrium sampling step and thus permits easy integration with enhanced sampling methods. Additional perturbations could even be included in the nonequilibrium step without significant complication. Lastly, analysis of the method is relatively straightforward, with no fractional states to consider and therefore amenable to reweighting procedures such as WHAM,[43–46] which greatly improves the accuracy and reliability of the observed titration curves.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was performed at the Argonne Leadership Computing Facility, a U.S. Department of Energy Office of Science User Facility, and supported by the U.S. Department of Energy, Office of Science, under Contract No. DE-AC02-06CH11357. This work was also supported by the National Institutes of Health (NIH) via grants U54-GM087519 and P41-RR005969. Additional support was provided by the France and Chicago Collaborating in The Sciences (FACCTS) program (to B.R and C.C.). Computational resources were provided by the University of Chicago Research Computing Center, the Argonne Leadership Computing Facility (ALCF), the National Center for Supercomputing Applications through the Great Lakes Consortium for Petascale Computation (NCSA-GLCPC, to B.R.), and the Grand Equipement National de Calcul Intensif - Centre Informatique National de l'Enseignement Superieur (GENCI-CINES) at Montpelier, France (to C.C.).

Notes and References

1. Karplus M, McCammon JA. Molecular Dynamics Simulations of Biomolecules. Nature Struct Biol. 2002; 9:646–652. [PubMed: 12198485]
2. Karplus M, Kuriyan J. Molecular Dynamics and Protein Function. Proc Natl Acad Sci USA. 2005; 102:6679–6685. [PubMed: 15870208]
3. Dror RO, Dirks RM, Grossman JP, Xu H, Shaw DE. Biomolecular Simulation: A Computational Microscope for Molecular Biology. Annu Rev Biophys. 2012; 41:429–452. [PubMed: 22577825]
4. Nelson, DL., Cox, MM. Lehninger Principles of Biochemistry. 4th. W. H. Freeman; 2005.
5. Webb BA, Chimenti M, Jacobson MP, Barber DL. Dysregulated pH: A Perfect Storm for Cancer Progression. Nat Rev Cancer. 2011; 11:671–677. [PubMed: 21833026]
6. Chen Y, Roux B. Constant–pH Hybrid Nonequilibrium Molecular Dynamics–Monte Carlo Simulation Method. J Chem Theory Comput. 2015; 11:3919–3931. [PubMed: 26300709]
7. Lyubartsev AP, Martsinovksi AA, Shevkunov SV, Vorontsov-Velyaminov PN. New Approach to Monte Carlo Calculation of the Free Energy: Method of Expanded Ensembles. J Chem Phys. 1992; 96:1776.
8. Chodera JD, Shirts MR. Replica Exchange and Expanded Ensemble Simulations as Gibbs Sampling: Simple Improvements for Enhanced Mixing. J Chem Phys. 2011; 135:194110. [PubMed: 22112069]
9. Baptista AM, Teixeira VH, Soares CM. Constant-pH Molecular Dynamics using Stochastic Titration. J Chem Phys. 2002; 117:4184–4200.
10. Mongan J, Case DA, McCammon JA. Constant pH Molecular Dynamics in Generalized Born Implicit Solvent. J Comput Chem. 2004; 25:2038–2048. [PubMed: 15481090]
11. Swails JM, York DM, Roitberg AE. 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. [PubMed: 24803862]
12. Stern HA. Molecular Simulation with Variable Protonation States at Constant pH. J Chem Phys. 2007; 126:164112. [PubMed: 17477594]
13. Nilmeier JP, Crooks GE, Minh DDL, Chodera JD. Nonequilibrium Candidate Monte Carlo is an Efficient Tool for Equilibrium Simulation. Proc Natl Acad Sci USA. 2011; 108:E1009–E1018. [PubMed: 22025687]
14. Chen Y, Roux B. Efficient Hybrid Non-Equilibrium Molecular Dynamics - Monte Carlo Simulations with Symmetric Momentum Reversal. J Chem Phys. 2014; 141:114107. [PubMed: 25240345]
15. Chen Y, Roux B. Generalized Metropolis Acceptance Criterion for Hybrid Non-equilibrium Molecular Dynamics–Monte Carlo Simulations. J Chem Phys. 2015; 142:024101. [PubMed: 25591332]

16. Lee J, Miller BT, Damjanović A, Brooks BR. Constant pH Molecular Dynamics in Explicit Solvent with Enveloping Distribution Sampling and Hamiltonian exchange. J Chem Theory Comput. 2014; 10:2738–2750. [PubMed: 25061443]
17. Lee MS, Salsbury FR Jr, Brooks CL III. Constant-pH Molecular Dynamics using Continuous Titration Coordinates. Proteins. 2004; 56:738–752. [PubMed: 15281127]
18. Donnini S, Tegeler F, Groenhof G, Grubmüller H. Constant pH Molecular Dynamics in Explicit Solvent with λ-Dynamics. J Chem Theory Comput. 2011; 7:1962–1978. [PubMed: 21687785]
19. Huang Y, Chen W, Wallace JA, Shen J. All-Atom Continuous Constant pH Molecular Dynamics With Particle Mesh Ewald and Titratable Water. J Chem Theory Comput. 2016; 12:5411–5421. [PubMed: 27709966]
20. Goh GB, Knight JL, Brooks CL III. pH-Dependent Dynamics of Complex RNA Macromolecules. J Chem Theory Comput. 2013; 9:935–943. [PubMed: 23525495]
21. Dobrev P, Donnini S, Groenhof G, Grubmüller H. Accurate Three States Model for Amino Acids with Two Chemically Coupled Titrating Sites in Explicit Solvent Atomistic Constant pH Simulations and pK_a Calculations. J Chem Theory Comput. 2017; 13:147–160. [PubMed: 27966355]
22. Wallace JA, Shen JK. Continuous Constant pH Molecular Dynamics in Explicit Solvent with pH-based Replica Exchange. J Chem Theory Comput. 2011; 7:2617–2629. [PubMed: 26606635]
23. Khandogin J, Brooks CL III. Constant pH Molecular Dynamics with Proton Tautomerism. Biophys J. 2005; 89:141–157. [PubMed: 15863480]
24. Radak BK, Roux B. Efficiency in Nonequilibrium Molecular Dynamics Monte Carlo Simulations. J Chem Phys. 2016; 145:134109. [PubMed: 27782441]
25. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. Scalable Molecular Dynamics with NAMD. J Comput Chem. 2005; 26:1781–1802. [PubMed: 16222654]
26. Mongan J, Case DA. Biomolecular Simulations at Constant pH. Curr Opin Struct Biol. 2005; 15:157–163. [PubMed: 15837173]
27. Weiss JN. The Hill Equation Revisited: Uses and Misuses. FASEB J. 1997; 11:835–841. [PubMed: 9285481]
28. Liu JS. Peskun's Theorem and a Modified Discrete-State Gibbs Sampler. Biometrika. 1996; 83:681–682.
29. Tanokura M. ^{1}H-NMR Study on the Tautomerism of the Imidazole Ring of Histidine Residues I. Microscopic pK Values and Molar Ratios of Tautomers in Histidine-Containing Peptides. Biochim Biophys Acta. 1983; 742:576–585. [PubMed: 6838890]
30. (a) MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. J Phys Chem B. 1998; 102:3586–3616. [PubMed: 24889800] (b) Klauda JB, Venable RM, Freites JA, O'Connor JW, Tobias DJ, Mondragon-Ramirez C, Vorobyov I, MacKerell JrAD, Pastor RW. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. J Phys Chem B. 2010; 114:7830–7843. [PubMed: 20496934] (c) Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, MacKerell AD Jr. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. J Chem Theory Comput. 2012; 8:3257–3273. [PubMed: 23341755]
31. Essmann U, Perera L, Berkowitz ML, Darden T, Hsing L, Pedersen LG. A Smooth Particle Mesh Ewald Method. J Chem Phys. 1995; 103:8577–8593.
32. Steinbach PJ, Brooks BR. New Spherical-Cutoff Methods for Long-Range Forces in Macromolecular Simulation. J Comp Chem. 1994; 15:667–683.
33. Tuckerman ME, Berne BJ, Martyna GJ. Reversible Multiple Time Scale Molecular Dynamics. J Phys Chem B. 1992; 97:1990–2001.
34. Andersen HC. Rattle: A "Velocity" Version of the Shake Algorithm for Molecular Dynamics Calculations. J Comput Phys. 1983; 52:24–34.

35. Miyamoto S, Kollman PA. SETTLE: An Analytic Version of the SHAKE and RATTLE Algorithms for Rigid Water Models. J Comput Chem. 1992; 13:952–962.

36. Gao J, Kuczera K, Tidor B, Karplus M. Hidden Thermodynamics of Mutant Proteins: A Molecular Dynamics Analysis. Science. 1989; 244:1069–1072. [PubMed: 2727695]

37. Axelsen PH, Li D. Improved Convergence in Dual-Topology Free Energy Calculations through Use of Harmonic Restraints. J Comput Chem. 1998; 19:1278–1283.

38. This is not strictly true for rigid bonds since then the atomic coordinates are correlated. This is ignored in practice and application of SHAKE/RATTLE resolves any discrepancy. Since the momenta are also randomly sampled, the constraints are not satisfied in a deterministic fashion, which should eliminate any bias from this procedure.

39. Sivak DA, Chodera JD, Crooks GE. Time Step Rescaling Recovers Continuous-Time Dynamical Properties for Discrete-Time Langevin Integration of Nonequilibrium Systems. J Phys Chem B. 2015; 118:6466–6474.

40. Brünger A, Brooks CL III, Karplus M. Stochastic Boundary Conditions in Molecular Dynamics Simulations of ST2 Water. Chem Phys Lett. 1984; 105:495–500.

41. Sivak DA, Chodera JD, Crooks GE. Using Nonequilibrium Fluctuation Theorems to Understand and Correct Errors in Equilibrium and Nonequilibrium Simulations of Discrete Langevin Dynamics. Phys Rev X. 2013; 3:011007.

42. Leimkuhler B, Matthews C. Efficient Molecular Dynamics Using Geodesic Integration and Solvent-Solute Splitting. Proc R Soc A. 2016; 472:20160138. [PubMed: 27279779]

43. Tan Z, Gallicchio E, Lapelosa M, Levy RM. Theory of Binless Multi-state Free Energy Estimation with Applications to Protein-ligand Binding. J Chem Phys. 2012; 136:144102. [PubMed: 22502496]

44. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. The Weighted Histogram Analysis Method for Free-energy Calculations on Biomolecules. I. The Method. J Comput Chem. 1992; 13:1011–1021.

45. 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.

46. Shirts MR, Chodera JD. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. J Chem Phys. 2008; 129:124105. [PubMed: 19045004]

47. Radak, B. K.; Suh, D.; Roux, B. *in preparation*

48. (a) Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. J Comput Chem. 2008; 29:1859–1865. [PubMed: 18351591] (b) Jo S, Lim JB, Klauda JB, Im W. CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. Biophys J. 2009; 97:50–58. [PubMed: 19580743]

49. Izrailev, S., Stepaniants, S., Isralewitz, B., Kosztin, D., Lu, H., Molnar, F., Wriggers, W., Schulten, K. Computational Molecular Dynamics: Challenges, Methods, Ideas. In: Deuflhard, P.Hermans, J.Leimkuhler, B.Mark, AE.Skeel, R., Reich, S., editors. Lecture Notes in Computational Science and Engineering. Vol. 4. Springer Verlag; Berlin: 1998. p. 39-65.

50. Lin Y-L, 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. [PubMed: 26586504]

51. Castañeda CA, Fitch CA, Majumdar A, Khangulov V, Schlessman JL, García-Moreno BE. Molecular Determinants of the pK_a Values of Asp and Glu Residues in Staphylococcal Nuclease. Proteins. 2009; 77:570–588. [PubMed: 19533744]

52. Hopkins CW, Le Grand S, Walker RC, Roitberg AE. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. J Chem Theory Comput. 2015; 11:1864–1874. [PubMed: 26574392]

53. (a) Nozaki Y, Tanford C. Intrinsic Dissociation Constants of Aspartyl and Glutamyl Carboxyl Groups. J Biol Chem. 1967; 242:4731–4735. [PubMed: 6061418] (b) Nozaki Y, Tanford C. Examination of Titration behavior. Methods Enzymol. 1967; 11:715–734.

54. Bashford D, Case DA, Dalvit C, Tennant L, Wright PE. Electrostatic Calculations of Side-Chain pK_a Values in Myoglobin and Comparison with NMR Data for Histidines. Biochemistry. 1993; 32:8045–8056. [PubMed: 8347606]

55. Thurlkill RL, Grimsley GR, Scholtz JM, Pace CN. pK Values of the Ionizable Groups of Proteins. Prot Sci. 2006; 15:1214–1218.
56. Grimsley GR, Scholtz JM, Pace CN. A Summary of the Measured pK Values of the Ionizable Groups in Folded Proteins. Prot Sci. 2009; 18:247–251.
57. Teixeira VH, Vila-Viçosa D, Reis PBPS, Machuqueiro M. pK_a Values of Titrable Amino Acids at the Water/Membrane Interface. J Chem Theory Comput. 2016; 12
58. Dashti D, Roitberg A. pH-replica Exchange Molecular Dynamics in Proteins using a Discrete Protonation Method. J Phys Chem B. 2012; 116:8805–8811. [PubMed: 22694266]

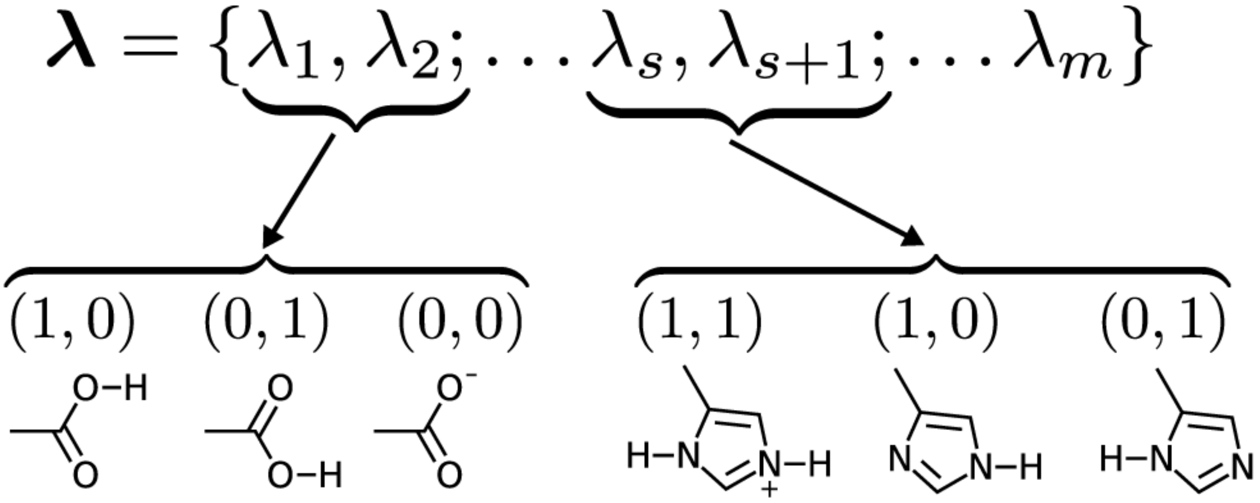


Figure 1.

The protonation state of a titratable system is completely defined by its occupancy vector $\boldsymbol{\lambda}$, where each of its m elements is either a one or a zero depending on whether the given site, s, is or is not occupied, respectively. The protonation state of individual residues is determined by a small subset of the elements of $\boldsymbol{\lambda}$ such that multiple system states may contain the same residue state. The average of a given element of $\boldsymbol{\lambda}$ yields the protonated fraction for that site and corresponds to a *microscopic* $\mathrm{p}K_\mathrm{a}$. Multiple sites may be equivalent such that a *macroscopic* $\mathrm{p}K_\mathrm{a}$ can be determined by grouping two or more sites together (e.g., the neutral states of carboxylate moieties). However, even non-equivalent sites can be grouped into macroscopic transitions, although in these cases the relationship between the two sets of $\mathrm{p}K_\mathrm{a}$ values is not always straightforward (e.g., methyl imidazole).

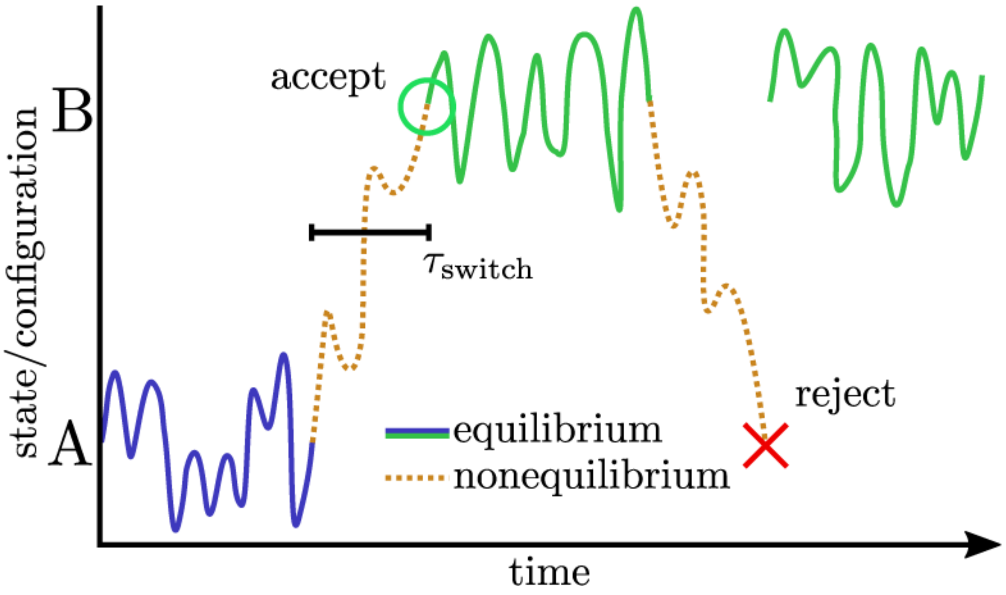


Figure 2.

The constant-pH MD algorithm consists of two part cycles in which standard equilibrium MD (blue and green solid lines) is performed followed by a driven nonequilibrium switch (orange dotted lines), which changes both the configuration and protonation state (arbitrarily labeled A and B). Detailed balance is restored after the nonequilibrium steps by applying a MC procedure in which the new configuration/state is accepted or rejected.

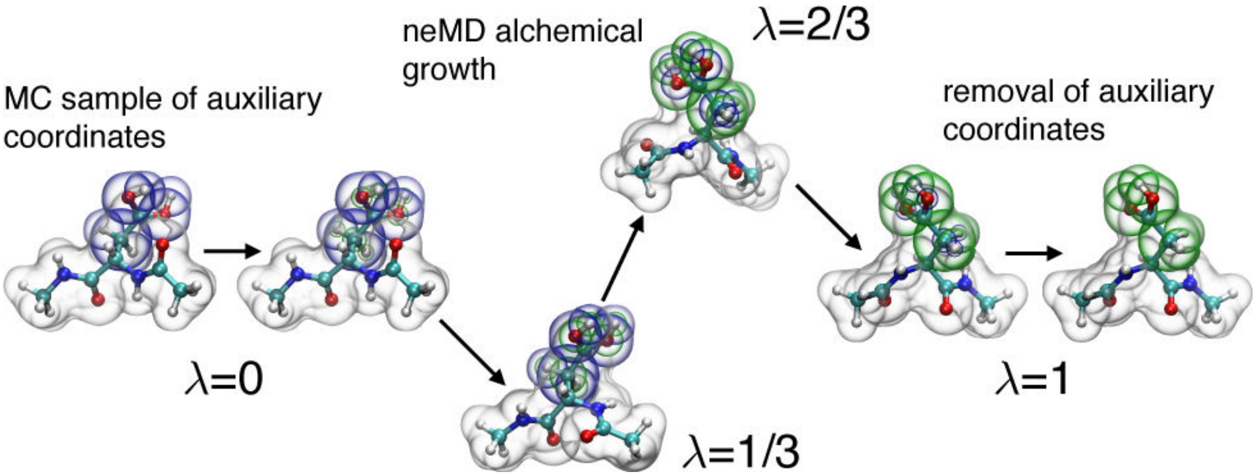


Figure 3.

A switch move contains three main steps: 1) an exact MC sampling of auxilliary sidechain atoms, 2) neMD propagation of the coordinates and coupling constant λ as the original coordinates and state (blue spheres) decouple and the new coordinates and state (green spheres) become coupled, and 3) removal of the non-interacting atoms. If the neMD/MC move is rejected, then the simulation continues from the original coordinates.

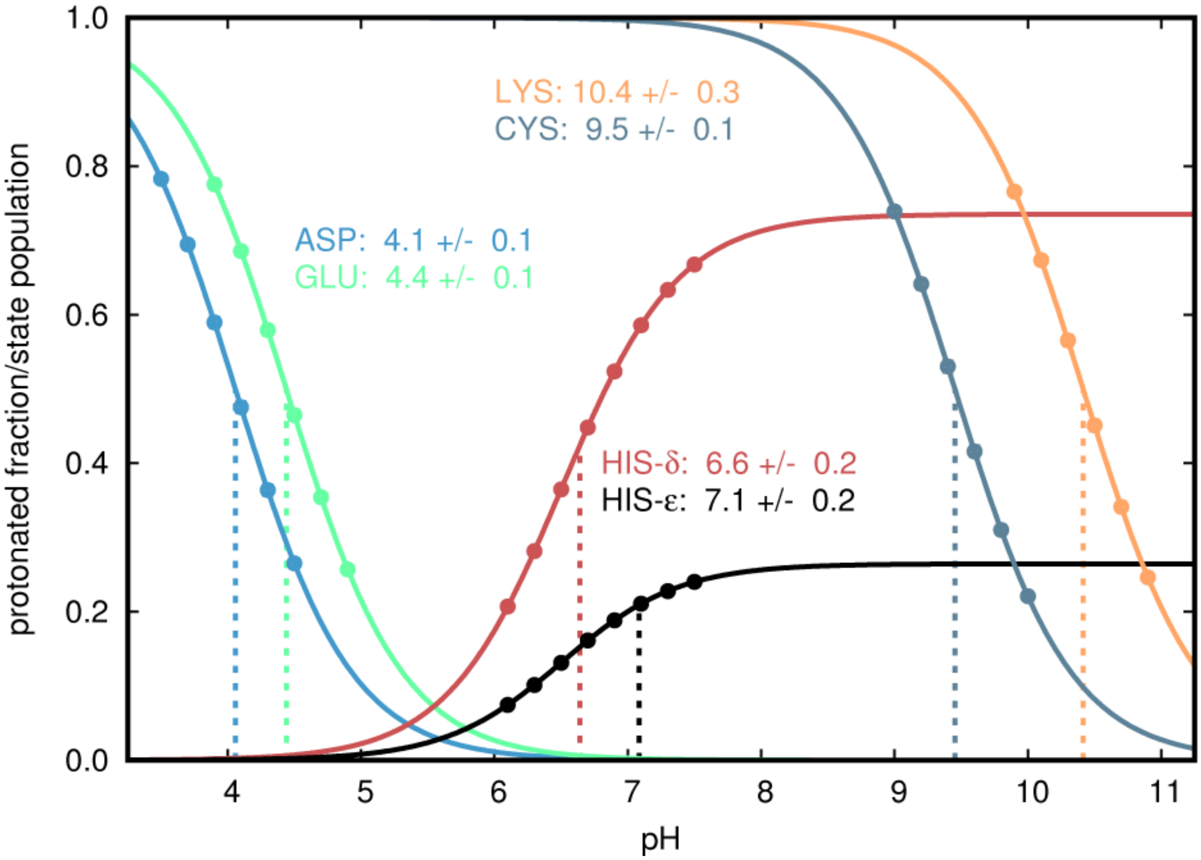


Figure 4.

Titration curves are easily computed for the reference dipeptides after initial parameterization and subsequent constant-pH simulations. Data points represent explicitly sampled pH values while lines represent the analytic curves predicted by UWHAM.

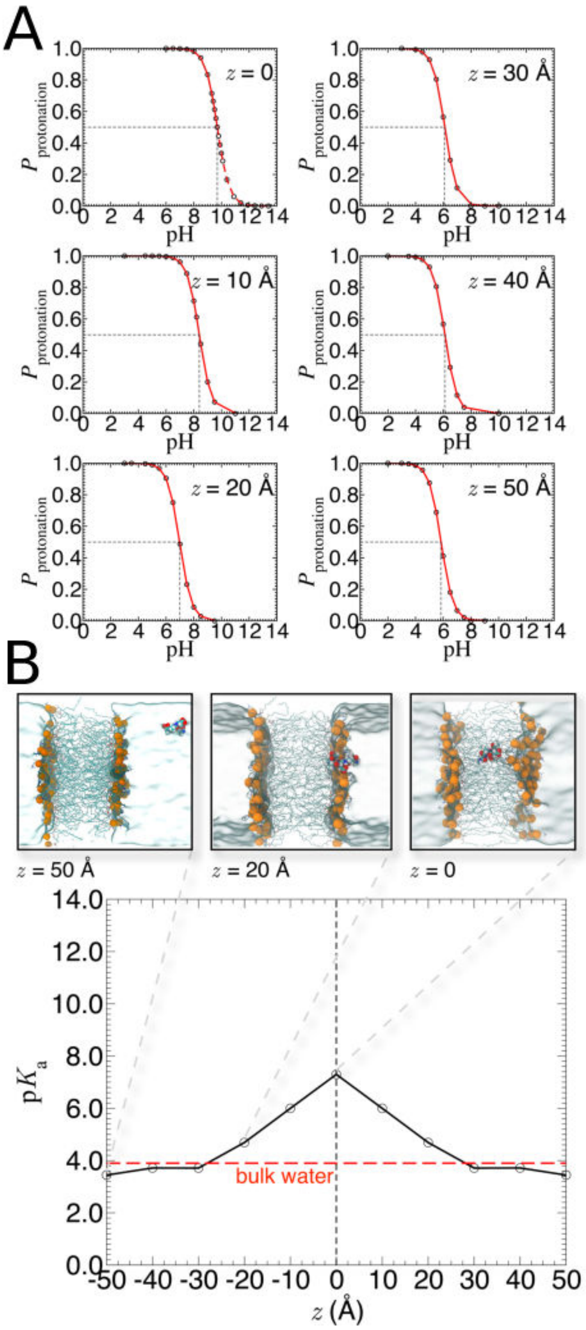


Figure 5.

Translocation of a terminally blocked, titratable pentapeptide, AADAA, across a 1–palmitoyl–2–oleoyl–phosphatidylcholine bilayer was performed by restraining the system at various separations (panel B, top). The insets depict the protonation probability of the central aspartic–acid residue for different positions of pentapeptide along bilayer normal (panel A). Here, z is the Euclidian distance separating the center of mass of the pentapeptide from that of the membrane, projected onto the direction normal to the interface (i.e., $z = 0$ corresponds to the middle of the ~27 Å thick membrane). The dashed red line in panel B corresponds to the pK_a in bulk water.

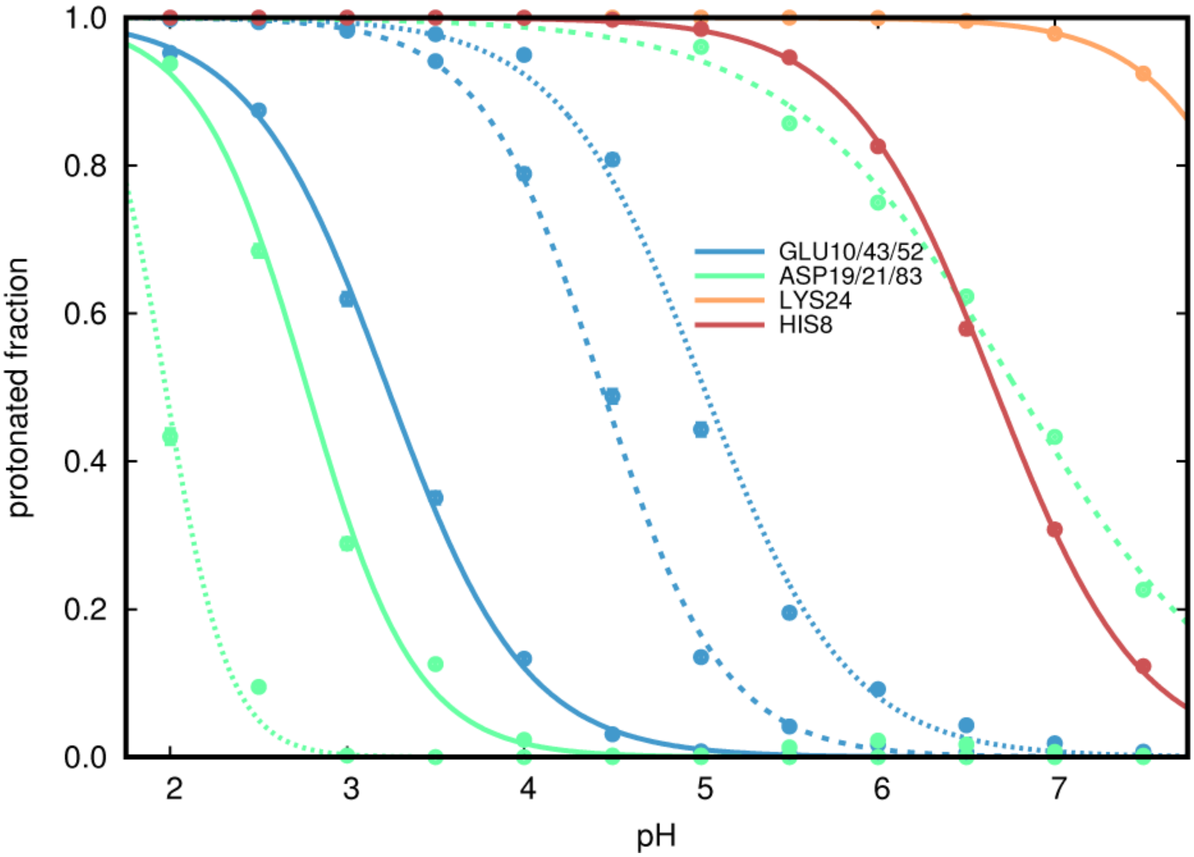


Figure 6.

Representative macroscopic titration curves (8 of 22 total) for SNase indicate a wide range of pK_a values, even amongst similar residues. Residues are colored by type and have different line patterns to denote the same residue in different environments (in ascending order as solid, dashed, and dotted lines).

Table 1

The reference $\mathrm{p}K_\mathrm{a}$ and free energy values needed for a constant-pH simulation are tabulated here for the CHARMM36 force field. Energies are in kcal/mol at 298 K and error bars represent 95% confidence intervals. Most values are defined in Eq. (5). The corrected values, ΔF_corr, represent adjustments made after simulations at constant-pH in order to reproduce the reference value (if needed). The uncertainties for these are essentially the same as for the uncorrected values. The temperature dependence is quantified by fitting ΔF as a function of temperature using a linear expansion of the internal energy, that is, $\Delta U(T) = \Delta U_{298} + \Delta C_{v,298}(T - 298\ K)$.

	$\mathrm{p}K_\mathrm{a}^\mathrm{ref}$	**ΔF**	**ΔU_{298}**	**$\Delta C_{v,298}$**	**ΔF_corr**
ASP	4.0[a]	−50.3 (0.4)	−60.0	−9.7	−50.0
GLU	4.4[a]	−64.6 (0.5)	−78.8	−14.2	−64.4
HIS	6.4[b]				
HIS-δ	6.5[b]	−0.1 (0.4)	2.0	2.1	–
HIS-ε	7.0[b]	−15.9 (0.4)	−23.0	−7.1	–
CYS	9.5[a]	−80.8 (0.3)	−84.6	−3.7	–
LYS	10.4[a]	43.2 (0.4)	54.0	10.7	41.9

[a] From Nozaki and Tanford [53]

[b] From Tanokura [29], adjusted by Bashford et al. [54]

Table 2

Apparent pK_a values for SNase are tabulated from Hill equation fits to the data presented here. Comparison values, where available, are given from both theory and experiment. Error bars have been adjusted to represent 95% confidence intervals. Errors from Huang et al. [19] reported as zero were assumed to be 0.1 units before rescaling.

residue		this work	λ-dynamics[19]	expt.[51]
GLU	10	3.23 (0.60)	3.20 (0.25)	2.82 (0.22)
	43	4.44 (0.07)	4.10 (0.25)	4.32 (0.10)
	52	5.01 (0.26)	4.70 (0.50)	3.93 (0.20)
	57	4.85 (0.33)	4.10 (0.75)	3.49 (0.22)
	67	4.23 (0.80)	4.00 (0.50)	3.76 (0.18)
	73	3.48 (0.92)	3.60 (0.25)	3.31 (0.03)
	75	2.98 (1.31)	2.70 (1.00)	3.26 (0.12)
	101	4.55 (0.45)	4.70 (0.50)	3.81 (0.25)
	122	3.90 (0.64)	4.40 (0.25)	3.89 (0.22)
	129	5.08 (0.61)	5.50 (0.25)	3.75 (0.22)
	135	3.35 (0.48)	2.90 (0.25)	3.76 (0.20)
ASP	19	2.77 (0.76)	3.30 (1.50)	2.21 (0.18)
	21	6.78 (0.99)	6.00 (0.75)	6.54 (0.05)
	40	3.32 (0.52)	2.90 (0.25)	3.87 (0.22)
	77	0.82 (0.50)	<−1.00	<2.20
	83	1.97 (0.72)	<0.00	<2.20
	95	2.74 (0.39)	3.00 (0.25)	2.16 (0.18)
	143	4.41 (0.64)	n/a	3.80 (0.25)
	146	4.01 (0.34)	n/a	3.86 (0.12)
LYS	24	8.43 (0.45)	n/a	n/a
HIS	8	6.66 (0.56)	n/a	n/a
	121	5.36 (0.50)	n/a	n/a