Reservoir pH replica exchange

Ana Damjanovic,^{1,2,a)} Benjamin T. Miller,² Asim Okur,² and Bernard R. Brooks²

¹*Department of Biophysics, Johns Hopkins University, 3400 N. Charles Street, Baltimore, Maryland 21218, USA*

²*Laboratory of Computational Biology, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892-5690, USA*

(Received 1 March 2018; accepted 30 May 2018; published online 18 June 2018)

We present the reservoir pH replica exchange (R-pH-REM) method for constant pH simulations. The R-pH-REM method consists of a two-step procedure; the first step involves generation of one or more reservoirs of conformations. Each reservoir is obtained from a standard or enhanced molecular dynamics simulation with a constrained (fixed) protonation state. In the second step, fixed charge constraints are relaxed, as the structures from one or more reservoirs are periodically injected into a constant pH or a pH-replica exchange (pH-REM) simulation. The benefit of this two-step process is that the computationally intensive part of conformational search can be decoupled from constant pH simulations, and various techniques for enhanced conformational sampling can be applied without the need to integrate such techniques into the pH-REM framework. Simulations on blocked Lys, KK, and KAAE peptides were used to demonstrate an agreement between pH-REM and R-pH-REM simulations. While the reservoir simulations are not needed for these small test systems, the real need arises in cases when ionizable molecules can sample two or more conformations separated by a large energy barrier, such that adequate sampling is not achieved on a time scale of standard constant pH simulations. Such problems might be encountered in protein systems that exploit conformational transitions for function. A hypothetical case is studied, a small molecule with a large torsional barrier; while results of pH-REM simulations depend on the starting structure, R-pH-REM calculations on this model system are in excellent agreement with a theoretical model. <https://doi.org/10.1063/1.5027413>

INTRODUCTION

Sampling of conformations and protonation states at different pH values is one of the greatest challenges in the field of biomolecular simulations. One of the approaches to address this issue is through molecular dynamics (MD) simulations at constant pH.¹ Constant pH MD has a theoretical advantage over regular MD in that the protonation state of ionizable groups can change during the course of the trajectory.

Constant pH simulations are also used for calculations of pK_a values of ionizable groups. Accurate calculations of pK_a values are a challenge, especially for internal ionizable groups, which are often found in environments that are very different from water and can thus exhibit large shifts in pK_a values.^{2,3} Any structural changes coupled with protonation/deprotonation need to be accounted for in the pK_a calculations. This is why continuum models that are based on static structures, and in which the dielectric properties of a protein are represented with a single valued constant, are usually unable to self-consistently reproduce pK_a values of internal groups.^{4–6} Improvements to single structure based pK_a calculations were made with methods that account for multiple structures.^{7–9} However, constant pH simulations are theoretically the most accurate because sampling of multiple structures is performed in a pH dependent way.¹

The methods for constant pH simulations fall into two main categories—the discrete protonation state methods^{10–13} and the continuous protonation state methods.^{14–17} In constant pH simulations with discrete protonation states, titratable groups are modeled as either protonated or deprotonated. MD simulations are performed and are periodically interrupted with Monte Carlo (MC) trials to change the protonation states of the titratable groups. If the MC move is accepted, the MD simulation is then continued with the new charges. Several flavors of constant pH simulations with discrete protonation states exist, which use either implicit solvents,^{12,13} a combination of explicit solvents for dynamics and implicit solvents for MC moves,^{10,18} or explicit solvents.^{11,19–23}

Within the framework of continuous protonation state models, titratable groups can have protonation states that can take any value in between fully protonated and deprotonated.^{14–17} In the method of Baptista *et al.*¹⁵ MD simulations are performed with “average” charges on titratable groups (determined with continuum electrostatics) that get periodically updated during simulations. Other methods include the extended Hamiltonian approach,¹⁴ the acidostat method,²⁴ and the λ -dynamics approach.^{16,17,25} The λ -dynamics approach has been used with both implicit water,^{16,25} mixed explicit/implicit water,²⁶ and, more recently, fully explicit water.^{17,27}

Several flavors of constant pH simulations have recently been implemented within the framework of the pH replica exchange method (pH-REM).^{26,28–30} With pH-REM, simultaneous simulations at multiple pH values are performed,

^{a)}Author to whom correspondence should be addressed: ad@jhu.edu. Tel.: (410) 516-5390. FAX: (410) 516-4118.

with periodic attempts to swap pH values between the simulations. This method is in spirit similar to the temperature or Hamiltonian replica exchange. The method has been shown to exhibit smoother titration curves and more efficient sampling.^{28,31}

One of the standing challenges in constant pH simulations is conformational sampling. This is especially true in cases when changes in protonation state are accompanied by changes in conformations. In fact, many proteins exploit this protonation-conformation coupling for function. Prominent examples are ATP synthase³² bacteriorhodopsin,³³ or cytochrome c oxidase.³⁴ In most cases where change in protonation state is coupled with conformational transition, ionizable groups are found in protein interiors, not on surface. Such groups will often experience shifts in pK_a values compared to the pK_a values of same groups in water.

Prominent examples where internal ionizable groups have been characterized in great detail experimentally are variants of staphylococcal nuclease (SNase) with engineered internal ionizable groups. Many variants have been characterized by NMR spectroscopy and crystallography, and some of the variants experience a rather large shift in pK_a values.^{2,3,35–37} Many of such variants show conformational relaxation coupled with changes in pH, and these changes occur on time scales of μ s-ms seconds.^{38–40} Early computational work has suggested that buried states correspond predominantly to the neutral state, while water solvated states correspond to the charged state.^{41–44} SNases variants have become a test case for many computational methods.^{45–50}

Constant pH simulations with hybrid implicit/explicit solvent models were recently performed on several proteins with naturally occurring buried ionizable residues: cytochrome c oxidase,⁵¹ nitrophorin,⁵² sodium-proton antiporter NhaA,⁵³ proton channel,⁵⁴ and inner-membrane transporter AcrB.⁵⁵

We note that in large proteins with buried groups, a true equilibrium pK_a value is often difficult to sample. If the ionizable group samples two or more conformational states, a true equilibrium constant cannot be observed unless a bidirectional transition between the conformations has been sampled multiple times. A similar situation is often seen in free energy calculations when a hysteresis, sometimes significant, is observed based on the starting structure. This would mean, for the constant pH case, that if a pK_a is calculated when only unidirectional conformational transition is observed, the calculated pK_a may not correspond to a true equilibrium value, which could be, in some cases, significantly different. A good test for correctness would be to perform the same calculation from a different starting structure; if the same pK_a is observed, that would be a good indication that the calculated pK_a is converged.

True equilibrium populations between two or more conformations are difficult to achieve in proteins and has been the subject of many developments in enhanced sampling techniques. Recently, constant pH methods have been coupled with various methods for enhanced conformational sampling, such as the temperature replica-exchange method (REM),^{56,57} accelerated MD,⁵⁸ umbrella sampling,^{59,60} and coarse grained methods.⁶¹ These methods have been shown to exhibit better accuracy than unenhanced constant pH molecular dynamics.⁶²

A constant pH method was implemented in Rosetta,^{63,64} and a coarse grained constant pH method was implemented in GROMACS.^{65,66}

We note that the methods above require a constant pH methodology to be coupled with some enhanced sampling methodology. The reservoir pH replica exchange method (R-pH-REM) presented here takes another approach: namely, the problem of conformational sampling gets decoupled from the constant pH methodology. The method consists of two steps: (1) sampling of conformations with fixed protonation states and (2) determining the conformation/protonation equilibrium at various pH values by injecting the reservoir conformations into a pH replica exchange simulation. Perhaps the best way to view this is that the simulations in the first step are done with protonation state constraints, and in the second step those constraints are relaxed, allowing the determination of a proper final protonation state ensemble. This allows for the ability to focus all computational resources into sampling of conformations with a fixed protonation state, which is already a difficult task, even without the constant pH simulations. The reservoir method was inspired by the temperature reservoir T-REM methods,^{67,68} which has been shown to greatly increase the convergence of replica exchange simulations.^{68,69} To demonstrate the accuracy of the R-pH-REM method, we perform simulations on small systems and show that the results are equivalent to the results of regular pH-REM simulations. To demonstrate the usefulness of the method, we perform simulations on a model system with a large energy barrier, and we show that the R-pH-REM results agree very well with a theoretical model.

METHOD

Constant pH algorithm

The constant pH algorithm is based on the combination of MD simulations with periodic MC samplings of protonation states, as proposed by Mongan *et al.*¹³ and recently implemented in the Chemistry at Harvard Molecular Mechanics (CHARMM) molecular simulation program.^{28,70,71} MD simulations are used to sample the conformational space, while at regular intervals MC trials to change protonation states of ionizable groups are performed.⁷² We note that the deprotonated state is modeled with the proton present, but with the charge distribution of the deprotonated state. To determine the transition probability in MC simulations, the transition free energy for protonating a residue (ΔF) for a given MD conformation is needed,^{10,11,13}

$$\Delta F = kT(pH - pK_{a,w}) \ln 10 + \Delta F_{ele} + \Delta F_{ele,w}. \quad (1)$$

Here, k is the Boltzmann constant, T is the temperature, and $pK_{a,w}$ is the experimentally determined pK_a value of a model compound in aqueous solution. ΔF_{ele} and $\Delta F_{ele,w}$ are the electrostatic components of the free energy differences between the protonated and deprotonated states of the titratable residue in the protein and in the model compound (a blocked amino acid) in aqueous solution, respectively. The above equation is based on the division of the total transition free energy into electrostatic and non-electrostatic components, as discussed by Mongan *et al.*¹³ In the current model, the non-electrostatic

components including the free energy of the bond between the proton and the titratable residue and the proton solvation free energy are assumed to cancel out between the titratable residue in the protein and in aqueous solution. The electrostatic component of the free energy $\Delta F_{\text{ele},w}$ is approximated as the difference in the electrostatic energy calculated with the charges of the current state and the charges of the proposed state. Van der Waals radii are approximated to stay unchanged during titration. An implicit solvent model is used to describe the aqueous environment.

The Metropolis scheme is used to calculate the probability of transition from the deprotonated state to the protonated state,

$$w(d \rightarrow p) = \begin{cases} 1, & \Delta F \leq 0, \\ e^{-\beta \Delta F}, & \Delta F > 0. \end{cases} \quad (2)$$

Here, $\beta = 1/kT$.

The pH replica exchange algorithm

The pH replica exchange method for the discrete protonation state model has been introduced by Itoh *et al.*²⁸ Replicas of the target system are simulated at different pH values with periodic attempts to exchange the pH value between adjacent replicas. The exchange of pH values between the replicas is equivalent to the exchange of conformations and protonation states between the replicas. A schematic representation of the pH replica exchange method is presented in Fig. 1.

The system consists of M non-interacting replicas, where the replica i ($i = 1, \dots, M$) has a coordinate vector q_i , momentum vector p_i , and a pH value of pH_i ($i = 1, \dots, M$), and is denoted as $X_i^l = (q_i, p_i, N_i^p, \text{pH}_i)$. The system has N^t titratable residues, and N_i^p titratable residues are protonated in the replica i . For an exchange between the replica i at pH_i and the replica j at pH_m , the detailed balance condition is defined by

$$P(X_i^l)P(X_j^m)w(X_i^l, X_j^m \rightarrow X_i^m, X_j^l) = P(X_i^m)P(X_j^l)w(X_i^m, X_j^l \rightarrow X_i^l, X_j^m). \quad (3)$$

Here P is the equilibrium probability and w is the transition probability. In the grand canonical ensemble of protons, the equilibrium probability P with a kinetic energy K , potential energy V at the temperature T , and pH of pH_i is given by^{28,73}

$$P(X_i^{\text{pH}_i}) = \frac{1}{\Xi} e^{-\beta(K(p_i, N_i^p) + V(q_i, N_i^p) + kT \text{pH}_i \ln 10 N_i^p)}, \quad (4)$$

where Ξ is the grand canonical partition function. We note that all replicas are kept at the same constant temperature T , e.g., by a Nose-Hoover or a Langevin thermostat. From the above equations, the ratio of transition probabilities is

$$\frac{w(X_i^l, X_j^m \rightarrow X_i^m, X_j^l)}{w(X_i^m, X_j^l \rightarrow X_i^l, X_j^m)} = e^{-\Delta}, \quad (5)$$

where

$$\Delta = \ln 10 (\text{pH}_m - \text{pH}_i)(N_i^p - N_j^p). \quad (6)$$

The exchange probability is thus obtained as²⁸

$$w(X_i^l, X_j^m \rightarrow X_i^m, X_j^l) = \begin{cases} 1, & \Delta \leq 0, \\ e^{-\Delta}, & \Delta > 0. \end{cases} \quad (7)$$

During a pH replica exchange procedure, a change in protonation states is not attempted when a replica-exchange trial is being performed, which keeps the replica-exchange Markov chain uncoupled from the constant pH Markov chain.

Reservoir pH replica exchange algorithm

Within the context of the reservoir pH replica exchange algorithm (R-pH-REM), one or more reservoirs of conformations are generated. Each reservoir has fixed protonation states and is assumed to populate a Boltzmann ensemble for that particular protonation state. A reservoir can be obtained from long molecular dynamics simulations with fixed charges or from some enhanced sampling methodology such as replica exchange or the potential of mean force (PMF). The purpose of the reservoirs is to pre-sample all accessible

FIG. 1. Scheme of pH-REM and R-pH-REM simulations. In pH-REM simulations, conformations and protonation states are exchanged between replicas run at different pH values. With R-pH-REM simulations, in addition to the regular exchanges between the replicas, conformations and protonation states from the reservoirs are inserted into the simulations of replicas. The scheme shows two reservoirs, R1 and R2, where structures from R1 are injected into the replica at pH1 and structures from R2 are injected into the replica at pH4.

areas of conformational space at a given protonation state value.

After the reservoirs have been generated, the reservoir constant pH procedure is initiated: The constant pH simulations are interrupted after a certain pre-determined number of steps, and an attempt to replace the current system conformation with a random reservoir structure is performed. The attempt is always accepted if the protonation state of all ionizable residues matches exactly with the protonation state of all ionizable residues in the reservoir. If protonation states of all residues do not match, the attempt is rejected. During this process, new conformations are not inserted into the reservoir itself. When an attempt to insert a reservoir structure is made for the replica coupled to the reservoir, a pH swap with another replica is not attempted.

In this paper, we have investigated using more than one reservoir, and each reservoir was coupled to a simulation at a different pH value. For computational efficiency, the reservoir should be coupled to a pH value at which the reservoir protonation state is expected to be significantly populated because otherwise many insertion attempts will be rejected. While we have not considered inserting structures from multiple reservoirs into a simulation at the same pH value, this should, in principle, be possible. If multiple reservoirs are coupled to the same pH, just like the structure is randomly chosen, one of the multiple reservoirs should be chosen at random for an attempt. However, coupling more than one or two reservoirs with a simulation at a single pH may not be efficient. In principle, there could be a reservoir for each protonation state, although, again, that may not be practical or efficient for two reasons: one is that multiple reservoirs need to be generated, and another one is that the attempts to accept the reservoir structure are accepted only if there is a match between the protonation state in the simulation and the protonation state of the reservoir.

This procedure is supposed to preserve the Boltzmann ensemble of each protonation state in the simulation since from a practical perspective the reservoirs are considered to be already sufficiently converged to a Boltzmann distribution. After the exchange attempt, simulations are continued as usual, until a new exchange attempt. We note that insertion attempts should not be weighted with a Boltzmann factor, as is usually the case in temperature replica exchange simulations, as the reservoir itself is at the same temperature as the temperature of the constant pH simulation. In fact, Boltzmann weighting the attempts would introduce an error.

The quality of the results depends on the quality of the reservoir. Here we use the Boltzmann type reservoir,⁶⁸ rather than the non-Boltzmann reservoir⁷⁴ and avoid some of the sampling problems and bias associated with the non-Boltzmann reservoir.⁶⁹ In practice, the reservoirs are generated from a finite time MD simulation or a simulation involving enhanced sampling. For large proteins, the reservoir ensemble may not be fully converged. However, for systems that exhibit multiple conformational states separated by large energy barriers, as long as all relevant conformations of the protonated and deprotonated states are captured in correct proportions, the current method provides a significant improvement over the pH-REM method.

Assumptions

The method is formally exact, but in practice it is not. There are two approximations that could limit its accuracy and utility when applying this method to larger molecules: (1) the fixed protonation state reservoir(s) represent a canonical ensemble containing all major conformational basins in correct populations and (2) the reservoirs contain at least some structures for all major conformations for every protonation state in the final ensemble. These approximations are mitigated by employing larger reservoirs that better represent a proper ensemble and by choosing reservoir sets that will include all major conformations.

In general, there is no need to generate reservoirs for all possible combinations of protonation states. In a protein, most ionizable residues will be at the surface and will not exhibit large conformational transitions as a function of pH. Thus, a single reservoir with only one significantly populated protonation state of these residues will be suitable, as the conformational space at different pH values can easily be sampled on the time scale of a constant pH simulation.

However, if titration of some residue is coupled to a conformational transition that occurs on a longer time scale, some exploratory simulations might be needed. For example, if a biomolecule, say protein, exists in two major conformations that are coupled to a titration of a particular residue, exploratory simulation will be needed to figure out which protonation state of that residue is a better choice for the reservoir. Similarly, brief exploratory constant pH simulations might be needed to figure out which pH value corresponds best to the reservoir protonation state and to choose that pH value to be coupled to the reservoir.

As far as we are aware, such a two-step constant pH procedure presented here has not been reported. Our methodology allows a variety of enhanced sampling techniques to be combined with constant pH simulations without the need for a complex amalgamation of disparate methods.

Simulation setup

The CHARMM22 protein force field⁷⁵ was used in all simulations, together with Generalized Born (GB) implicit solvent models. Initially, the GBMV⁷⁶ variety was employed. The solvent dielectric, ϵ_{solv} , was set to 80, the solute dielectric, ϵ_{solute} , was set to 1, and electrostatic interactions were screened with the effective salt concentration of 0.1 M. Nonbonded electrostatic and van der Waals cutoffs for protein simulations were applied with a switching function starting at 16 Å and ending at 18 Å. For the nonpolar surface area term, $U_{\text{nonpolar}} = \sigma * SA$, σ was set to 0.02 kcal/(mol Å²), and SA, the solvent-accessible surface area, was calculated with a probe radius of 1.4 Å according to the algorithm specified in the work of Lee *et al.*⁷⁶

The calculated reference energies ($\Delta F_{\text{ele,w}}$) depend on the solvent model used. Previously, we have reported calculated $\Delta F_{\text{ele,w}}$ values with the GBMV method, in the absence of salt.²⁸ To be consistent with the CHARMM22 protein force field,⁷⁵ we have recalculated the $\Delta F_{\text{ele,w}}$ value for Lys to be −62.7 kcal/mol in the absence of salt. Here we report the

calculated $\Delta F_{\text{ele,w}}$ values in the presence of 0.1 M salt: 36.4 kcal/mol for Asp, 40.6 kcal/mol for Glu, -63.7 kcal/mol for Lys and -19.1 , and -4.0 kcal/mol for His- δ and His- ϵ , respectively. $\Delta F_{\text{ele,w}}$ values were calculated as follows: multiple 3 ns long pH-REM calculations for each model compound were performed with slightly different values of $\Delta F_{\text{ele,w}}$. A $\Delta F_{\text{ele,w}}$ value yielding the pK_a that was closest to the model compound pK_a was then chosen. The model compound pK_a values are the same as reported by Itoh *et al.*²⁸ We note that all simulations involving Lys residues (Lys model compound, KK, and KAAE peptides) were performed in the absence of salt.

All simulations were performed with the CHARMM program.^{70,71} During the constant pH MD simulations, the SHAKE algorithm was used to constrain hydrogen to heavy atom bond lengths. The time step was taken to be 1.0 fs. The temperature of 300 K was maintained with the Langevin thermostat, with the friction coefficient $\gamma = 0.5 \text{ ps}^{-1}$. Attempts to exchange pH values were performed every 100 steps, while constant pH Monte Carlo moves were attempted every 50 steps.

RESULTS AND DISCUSSION

Model compound: Lys residue

Accuracy of the R-pH-REM method was benchmarked against the pH-REM method. As the simplest test case, we have calculated pK_a values from a 10 ns long simulation of blocked Lys residues. Simulations were performed in the pH range 8–13, with an increment of 1 pH unit. With the plain pH-REM method, the calculated pK_a value was 10.40.

Reservoirs were generated from 50 ns long MD simulations with Lys protonated (charged) and Lys deprotonated (uncharged). Two reservoir simulations were performed: one with the protonated reservoir, coupled to a simulation at pH = 8 (R_{p8}), and one with the deprotonated reservoir, coupled to a simulation at pH = 13 (R_{d13}). Calculated pK_a values indicate a good agreement between the pH-REM and R-pH-REM methods: 10.40 (pH-REM), 10.40 (R_{p8}), and 10.41 (R_{d13}). The calculated Hill coefficient in all three cases was 0.99.

KK peptide

Next, we perform simulations with a Lys-Lys dipeptide. This is a small system, so convergence can be achieved, but the presence of another ionizable group in the vicinity will result in the shift of the pK_a values. Three different reservoirs were generated through three 100 ns long MD simulations with three different protonation states. In the first reservoir, the two Lys molecules were deprotonated (R_{dd}), in the second reservoir, they were protonated (R_{pp}), and in the third reservoir, Lys-1 was deprotonated and was Lys-2 protonated (R_{dp}).

Regular pH-REM simulations were performed, with replicas spaced between pH values 7 to 12 in increments of 1 pH unit. Total simulation time was 50 ns. The pK_a values of the two Lys molecules were determined to be 10.51 and 10.37.

TABLE I. pK_a values of Lys-1 and Lys-2 in the KK peptide, calculated with different reservoirs, coupled to replicas at different pH values.

	No R	R_{dd11}	R_{dd12}	R_{dp9}	R_{pp7}	R_{pp8}	R_{pp9}
pK_{a1}	10.51	10.52	10.51	10.52	10.50	10.50	10.51
pK_{a2}	10.37	10.37	10.37	10.38	10.39	10.39	10.39

For R-pH-REM simulations, we have tested different reservoirs, coupled to replicas at different pH values. Table I shows the pK_a values of Lys-1 and Lys-2 calculated from different reservoir simulations: R_{dd11} (reservoir R_{dd} coupled to the replica at pH 11), R_{dd12} (reservoir R_{dd} coupled to the replica at pH 12), R_{dp9} (reservoir R_{dp} coupled to the replica at pH 9), R_{pp7} (reservoir R_{pp} coupled to the replica at pH 7), R_{pp8} (reservoir R_{pp} coupled to the replica at pH 8), and R_{pp9} (reservoir R_{pp} coupled to the replica at pH 9). Note that we have coupled reservoirs to replicas at pH values where we were expecting a significant population of that particular protonation state (for example, R_{dd} was coupled to higher pH replicas).

Table I indicates that the pK_a values calculated without the reservoir and with all reservoirs are in excellent agreement (within 0.02 pK_a units). The calculated Hill coefficient in all simulations was either 0.82 or 0.83.

Figure 2 shows protonated fractions of all simulations. The green dots show superpositions of protonated fractions of 3 different simulations of the Lys model (without the reservoir, with R_d , and with R_p). The green line is a fit to each protonated fraction. All 3 points are on top of each other, and all three best fit lines are virtually indistinguishable and appear as a single line. The red dots show protonated fractions of Lys-1 in the KK peptide from 7 different simulations, the simulation without the reservoir, and a simulation with 6 reservoirs discussed above. The blue dots show protonated fractions of Lys-2 in the KK peptide from 7 different simulations, the simulation without the reservoir, and a simulation with 6 reservoirs discussed above. Both red and blue lines are superpositions of 7 best fit lines and appear as a single line, suggesting that it is virtually impossible to distinguish best fit lines

FIG. 2. Titration curve of the Lys model compound (green), Lys-1 in the KK peptide (red), and Lys-2 in the KK peptide (blue). The line and points for the Lys model compound are superpositions of 3 different simulations (1 without reservoir and 2 reservoir simulations). The lines and points for the KK peptide are superpositions of 7 different simulations (1 without reservoir and 6 reservoir simulations).

FIG. 3. Convergence of protonated fractions at pH = 11. Protonated fractions were calculated as averages over 50 ps. Simulations without reservoir are shown with thick lines; red for Lys-1 and blue for Lys-2. Simulations with reservoirs are shown with thin black lines.

obtained from simulations with different reservoirs and without reservoirs.

Figure 3 shows good convergence of the simulations. These plots show that reservoir calculations fluctuate less and converge faster than the simulations without the reservoir.

KAAE peptide

The structure of the Lys-Ala-Ala-Glu peptide was generated with CHARMMing.⁷⁷ In all simulations, Glu was modeled as charged, while the charge on Lys was allowed to vary. This is a reasonable assumption since the pK_a of Glu is approximately 4.4; thus in the range of pH where Lys titrates, Glu is expected to be charged. To investigate the issue of structural hysteresis, two different starting conformations were generated from MD equilibration runs performed at different charge states of Lys: an “ion-paired” conformation with a salt bridge between Lys and Glu and a “separated” conformation in which Lys and Glu were spatially separated. The two starting conformations, “charged” and “neutral,” are shown in Fig. 4. Because the system is small, we do expect that the initial bias due to the starting structure will disappear quickly.

Temperature replica exchange simulations with fixed charges were used to generate two structure reservoirs: R_p with protonated (charged) Lys and R_d with deprotonated (neutral) Lys. The R_p reservoir was coupled to the replica at

pH = 8, while R_d was coupled to a replica at pH = 13. The following temperatures were used in the T-REM protocol: 300, 343, 393, 450, 515, and 590 K, and only the lowest temperature replica was used to generate the reservoir. Only structures from the final 10 ns of a 30 ns long run were used in the reservoir.

Regular pH-REM simulations were performed, with replicas spaced between pH values 8 to 13 in increments of 1 pH units. Total simulation time was 100 ns. For R-pH-REM simulations, the same pH range was used. The following R-pH-Rex simulations were performed: simulations with the R_d reservoir coupled to a replica at pH = 13 and the R_p reservoir coupled to a replica at pH = 8 and with two reservoirs, R_d and R_p . pK_a values were calculated from a fit to the protonated fraction determined over the entire 100 ns run. The calculated pK_a values started from the “ion-paired” and from the “separated” starting conformation are shown in Table II. The Table shows the results from simulations with no reservoirs (No R) and with reservoirs.

The Table reveals that all pK_a values are shifted up from the pK_a value of a Lys model compound (10.4), as expected from an interaction with a negatively charged Glu residue. While the pK_a values from simulations without the reservoirs and with the R_p reservoir are in good agreement with each other, pK_a values from simulations with the R_d reservoir are slightly lower. The reason for the slight difference is that the reservoir simulations themselves, obtained as a 10 ns average from a 30 ns long temperature replica exchange simulation, are not converged and may thus slightly bias the results. Figure 5 shows the histogram of the distance between Lys:NZ and Glu:CD atoms in simulations of two reservoirs (top) and in simulations without the reservoir (middle and bottom). The

FIG. 4. Two starting conformations for constant pH simulations of the KAAE peptide. Figure produced with VMD.⁷⁸

TABLE II. pK_a values calculated from the “ion-paired” and the “separated” starting structures. Results are shown for simulations with no reservoirs, with the R_d reservoir, with the R_p reservoir, and with two reservoirs.

	No R	R_{d13}	R_{p8}	R_{d13}, R_{p8}
Ion-paired	10.80	10.64	10.81	10.78
Separated	10.76	10.67	10.80	10.79

FIG. 5. Histogram of distances (binned in steps of 2 Å) between Lys-1:NZ and Glu-4:CD. Top left and right figures show the results for R_p and R_d , respectively. The middle and bottom figures represent simulations with no reservoir at pH 8 (left) and 13 (right), started from the ion-paired structure (middle) and separated structure (bottom).

figures on the left should match if converged, as well as figures on the right. Figure 5 indicates that distribution of conformations in the R_d reservoir (top right) slightly deviated from the distributions at pH = 13 in non-reservoir simulations (middle and bottom right)—namely, the reservoir slightly underestimates the states that are in the range 5–9 Å. A closer distance between Lys and Glu would result in a larger shift in the pK_a value from the model compound, and thus the reservoir slightly underestimates the pK_a shift.

To study the speed of convergence, averaged protonated fractions at pH = 11 are plotted in Fig. 6. Simulations started from the “ion-paired” conformation were compared with simulations started from the “separated” conformation. Simulations without any reservoirs exhibit initial hysteresis and within 3 ns

come close to each other. However, after that time, they continue to differ from each other. This is consistent with the lack of overall convergence of the non-reservoir simulations. We note that this result is not surprising, as a four-residue peptide is expected to converge on a longer than 100 ns time scale. In many biological examples of proteins that exhibit conformational changes on a long time scale, simulations that are not fully converged will be a necessary evil, and thus we believe that this sort of analysis with a non-fully converged system and a non-fully converged reservoir is fully relevant.

Constant pH simulations with the R_p reservoir (red) converge within about 3 ns, while simulations with the R_d reservoir (blue) converge somewhat slower, within about 10 ns. We

FIG. 6. Protonated fractions at pH = 11. Protonated fractions are averages over 50 ps. Simulations without reservoir are shown with thick black lines. Simulations with reservoirs are shown with thin lines: blue are coupled to R_{d13} , red to R_{p8} , and green to both reservoirs.

believe that this is because the R_d reservoir does not contain any ion paired conformations which dominates the protonated state conformational space (Fig. 5). This is to be expected as the Lys is uncharged and cannot form an ion pair. On the contrary, the R_p reservoir contains both ion-paired and non-ion-paired conformations and thus presents a better representation of the conformational space of both protonation states. The use of two reservoirs, though, should guarantee that all relevant conformational space is sampled. Furthermore, the simulations with two reservoirs converge the fastest, within 1 ns, and fluctuate the least. Thus, while a single reservoir that covers all conformational space might be sufficient, we recommend the use of multiple reservoirs.

Small molecule with a large conformational barrier

The reservoir method becomes most useful in systems that exhibit two or more conformational states separated by a large energy barrier, when very long simulations or enhanced sampling methods are needed. The speed of the constant pH simulations at this point still does not match the speed of classical, fixed charge MD or the power of enhanced sampling MD. To demonstrate the power of the reservoir method, we have constructed a small molecule (shown in Fig. 7) that contains ionizable groups connected by a double bond. The molecule can, in principle, sample both cis and trans conformations; however, rotation about such a double bond is not easily achieved even with fixed charge MD simulations.

This molecule was constructed using the graphical atom builder of the CHARMMing Web server,⁷⁷ which utilizes parameters produced by the ParamChem server (<https://www.paramchem.org>) to generate force-field parameters for use with the CHARMM Generalized Force Field (CGenFF).^{79,80} The force field parameters are shown in the [supplementary material](#). The charge penalties on the protonated and deprotonated variants were 90.052 and 34.106, while their respective parameter penalties were 214.5 and 110. Although these penalties are large, representing very poor overlap with other molecules in ParamChem's database, the purpose of this example system is merely to illustrate how the R-pH-REM method can overcome barriers to sampling. This demonstration is not dependent on the accuracy of the specific parameters used. To study titration of the nitrogen moiety, we have used ethylamine as a model compound. Its pK_a value is known to be 10.7. Thermodynamic integration (TI) was used to find the free energy difference between the protonated and deprotonated states, which was calculated to be -50.9 kcal/mol.

FIG. 7. The cis (left) and trans (right) conformations of the molecule used in this study.

pH-REM simulations of the titration of the amino moiety of this molecule was performed in the pH range 4–13. For simplicity, we looked at the titration of the amino moiety only and fixed the charges on the carboxylic group to a charged state. Attempts to exchange pH values were performed every 1000 steps. Averages over the 50 ns long simulation runs started from the cis and trans conformations yield two different pK_a values, 8.05 (cis) and 6.88 (trans).

To construct a reservoir for this system, we first performed the potential of mean force (PMF) for the rotation about the double bond using the adaptive umbrella sampling method implemented by CHARMM's ADUMB command. The central dihedral angle involving the carboxylic carbon atom, the double bond atoms, and the nitrogen atom of the amino moiety was used as the degree of freedom for umbrella sampling, with a total of 36 bins. Based on this calculation, the population of the cis and trans states was determined. These calculations revealed that when the amino moiety is charged, the cis state is only 7% populated, while the trans state is 93% populated. However, when the amino moiety is neutral, both states are equally populated.

To generate the low and high pH reservoirs, we performed four sets of 250 ns long molecular dynamics simulations. The charges of the amino moiety were those of the protonated state in one reservoir and of the deprotonated state in another. We note that in the actual simulation, the proton is not removed; only the charges are set to the charges of the protonated and deprotonated forms of amino-acid. Because the energy barrier for the cis-trans transition is high (~ 20.5 kcal/mol for both the protonated and deprotonated states), simulations started in either state will stay in that state. Thus, we performed simulations started from both cis and trans conformations, in each protonation state. To generate the correct reservoir with deprotonated states, we mixed conformations from simulations of the deprotonated state started from the cis and trans states in the equal proportion. Similarly, to build the correct reservoir with protonated states, we mixed the conformations from simulations with the charged state in the proportion of 7% and 93%, as determined from the PMF calculation. Each reservoir contained a total of 20 000 structures. We will refer to those two reservoirs as R_p and R_d for reservoir protonated and reservoir deprotonated, respectively.

R-pH-REM simulations were then performed starting from both the cis and the trans conformation. Attempts to exchange pH were performed every 1000 steps. We performed a total of 5 different reservoir simulations for both cis and trans starting conformations: simulations with a protonated (charged) reservoir only, coupled to $pH = 4$ (R_{p4}),

TABLE III. pK_a values calculated from cis and trans starting structures. The reservoir type (Rd vs Rp) and the pH of the replica that was coupled to a reservoir are indicated. Simulations without a reservoir are in the column labeled No R. Results from the theoretical four-state model are also shown.

Start	No R	R_{p4}	R_{d13}	R_{p4}, R_{d13}	R_{p6}, R_{d9}	$R_{p7}, R_{d8.5}$	Model
cis	8.05	7.81	7.76	7.77	7.77	7.79	7.77
trans	6.88	7.79	7.75	7.77	7.77	7.79	7.77

FIG. 8. Protonated fractions at pH = 8. Thick black lines show simulations without reservoirs. Red shows Rd-13 simulations, blue Rp-8 simulations, and green are all simulations with two reservoirs.

simulations with deprotonated (neutral) reservoir only (R_{d13}), coupled to pH = 13, simulations with both reservoirs, coupled to pH = 4 and pH = 13 (R_{p4}, R_{d13}), simulations with both reservoirs, coupled to pH = 6 and pH = 9, respectively (R_{p6}, R_{d9}), and simulations with both reservoirs, coupled to pH = 7 and pH = 8.5, respectively ($R_{p7}, R_{d8.5}$). Average protonation states from over 50 ns long R-pH-REM runs are shown in Table III.

Convergence of the simulations is shown in Fig. 8. The figure shows that a reasonable accuracy of the simulations can be achieved within about 10 ns and that simulations with all reservoirs agree well with each other.

Four state model

Another way to check whether the pK_a value calculated with the reservoir method is correct is to approximate the system with a four state model. Within the framework of the four state model, the molecule can exist in two conformations, trans (1) and cis (2), and two protonation states, deprotonated (d) and protonated (p). The four states will thus be labeled as D_1 , P_1 , D_2 , and P_2 . The grand canonical partition function for this system is

$$Z = 1 + e^{-\ln 10(pH - pK_{a1})} + e^{-\beta \Delta \epsilon_d} + e^{-\beta \Delta \epsilon_d} e^{-\ln 10(pH - pK_{a2})}. \quad (8)$$

Here we have assumed that the reference state for this calculation is D_1 and that its energy is chosen to be 0. pK_{a1} and pK_{a2} are the microscopic pK_a values of the two conformations. β is $1/kT$ and $\Delta \epsilon_d$ is the free energy difference between conformations 2 and 1 when the group is deprotonated. The probabilities of populating each of the four states can be calculated as

$$\begin{aligned} p(D_1) &= \frac{1}{Z}, \\ p(D_2) &= \frac{e^{-\beta \Delta \epsilon_d}}{Z}, \\ p(P_1) &= \frac{e^{-\ln 10(pH - pK_{a1})}}{Z}, \\ p(P_2) &= e^{-\beta \Delta \epsilon_d} \frac{e^{-\ln 10(pH - pK_{a2})}}{Z}. \end{aligned} \quad (9)$$

The total probability of observing a protonated state is equal to $p(P_1) + p(P_2)$ and depends on pK_{a1} , pK_{a2} , and $\Delta \epsilon_d$. The apparent pK_a value of this four state system can be determined from the condition

$$p(P_1) + p(P_2) = p(D_1) + p(D_2) \quad (10)$$

and results in

$$pK_a = \log(10^{pK_{a1}} p(D_1) + 10^{pK_{a2}} p(D_2)). \quad (11)$$

The above expression is equivalent to the expressions presented in recent publications.^{81–83} The $\Delta \epsilon_p$, i.e., the free energy difference between states 2 and 1 when the group is protonated, can be determined from the thermodynamic cycle.

From the above expression, based on the microscopic pK_a values of 6.88 and 8.05 of the molecule studied above, and equal population of cis and trans states when the amino moiety is deprotonated, the apparent pK_a value is calculated to be 7.77. This is in excellent agreement with the pK_a values determined through the reservoir calculations.

DISCUSSION

Sampling of relevant conformations in constant pH simulations is a problem in biomolecular systems where ionizable residues can exist in two or more conformations separated by large energy barriers. This is the case for many proteins that exploit conformational change for function. Our method separates the task of sampling of protonation states from the task of sampling of conformations. The method consists of two steps; in the first step, one, two, or more reservoirs of conformations, with fixed protonation states, are generated. These reservoirs can be generated with standard molecular dynamics simulations, including simulations on massively parallel supercomputers, such as Anton. Furthermore, reservoirs can be generated with enhanced sampling methods such as temperature replica exchange, PMF calculations, metadynamics, etc., without the need of integrating these methods into the constant pH methodology. In the second step, the reservoir conformations are injected into constant pH simulations through a Monte

Carlo procedure—namely, after a given number of steps, an attempt to accept a reservoir structure is made. If the protonation state of the reservoir matches that of the current state in the constant pH simulation, the MC move is accepted. Note that this procedure yields a proper ensemble if the reservoir represents a Boltzmann or near Boltzmann ensemble. The reservoir should be coupled to a pH value at which this protonation state is expected to have a significant population, to ensure a large number of acceptances. Multiple reservoirs can be coupled to multiple pH values.

The method was applied to study blocked amino acid Lys, a Lys-Lys (KK) peptide, a four-residue peptide (KAAE), and a small molecule with cis and trans conformations separated by a large energy barrier between them. For a single amino acid, and the KK peptide, the protonation profiles calculated with and without the reservoir are indistinguishable.

In the case of the KAAE peptide, we see some small differences in results obtained with two reservoirs. This is because the reservoir simulations themselves are not fully converged, and a reservoir that is not fully Boltzmann may bias the simulations. We note that the issue can be alleviated by introducing another reservoir into the simulation, which effectively results in better sampling of conformations at different protonation states. Reservoir simulations converge much faster than simulations without reservoirs (which do not fully converge on the time scale of the simulations). The use of multiple reservoirs is advantageous not only in improving the accuracy of the simulations but also improves the convergence of simulations because relevant conformations get sampled faster.

The small molecule with a large conformational barrier is the test case where regular pH-REM simulations cannot converge on a reasonable time scale, as the barrier of 20.5 kcal/mol would take on the order of seconds to cross in standard MD simulations. We show that we can use the reservoir method to obtain the pK_a values that are otherwise unavailable from regular constant pH simulations. To generate the equilibrium populations of the cis and trans states for the protonated and deprotonated reservoir, we utilized the potential of mean force calculations to determine the populations of the cis and trans conformers. We then generated the reservoir that reproduces these populations. The calculations with such reservoirs, by utilizing both single and double reservoirs, were in excellent agreement with each other and with a theoretical model.

We note that in real world examples such as in cases of large proteins it is likely that the reservoir will not be fully converged to the Boltzmann distribution. This will introduce reservoir bias. However, the bias introduced due to such a reservoir will be much smaller than the error obtained by not being able to sample the relevant conformational states at all. In principle, any molecular dynamics simulation that is not fully converged, constant pH or not, will bias the calculation of free energies.

The applications of the R-pH-REM methodology go beyond the simple pK_a determinations. Constant pH methodologies have the power of obtaining protonation states at a given pH value and can offer valuable insights into mechanisms of function in some proteins. For example, in enzymes,

such calculations can offer insights into the mechanism of catalysis.⁸⁴ In energy and charge transport proteins, the method can be used to understand the mechanisms of electron, proton, or ion transport. It can also be used when studying binding affinities since a change in the protonation state often accompanies binding.

Currently, the R-pH-REM method is implemented for implicit solvent constant pH methods based on a combination of MD and MC. However, the method should work with most constant pH methods and should ideally be performed with explicit solvent molecules. For constant pH methods based on the Monte Carlo procedure to exchange protonation states, such as those that combine explicit water simulations for conformational sampling and implicit water potentials for the MC procedure,^{10,18} as well as the recent methodology based on non-equilibrium simulations,^{21,22} exactly the same procedure should be used as in this paper; namely, the reservoir structures should be accepted when the protonation states match. We are not sure if this procedure can easily be applied on constant pH simulations based on continuous protonation states, as the reservoirs correspond to discrete protonation states and those are never sampled in continuous constant pH simulations. For this to work, an approximation must be made to accept structures that are close to a discrete protonation state. For the constant pH simulations based on the enveloping distribution sampling, the simulations corresponding to the base or non-elevated simulation temperatures (usually 300 K) sample the correct Boltzmann populations, and only those should be connected to the reservoirs. In the near future, we are planning to present a version of the R-pH-REM method, based on the enveloping distribution sampling in explicit solvents^{20,30} and on larger biomolecular systems.

CONCLUSIONS

The R-pH-REM method is useful in constant pH simulation of molecules or biomolecules that exhibit two or more conformational states separated by large energy barriers, as is the case for many proteins where function is coupled with conformational change. The idea behind the method is to perform simulations in two steps. In the first step, simulations with fixed protonation states are performed to generate a reservoir that reflects a proper canonical distribution at that protonation state. In the second step, the reservoir structures are injected in a constant pH simulation, to determine the proper protonation state ensemble in a formally exact manner. The method thus allows the use of varied sampling techniques to generate reservoirs without the need to integrate such sampling techniques into the constant pH methodology.

SUPPLEMENTARY MATERIAL

See [supplementary material](#) for the topology file for the small molecule.

ACKNOWLEDGMENTS

This research was supported in part by the Intramural research program of the NHLBI, NIH. We thank Alex Sodt, Juyong Lee, and Gerhard Koenig for useful discussions. We

also acknowledge the use of the LoBoS cluster and the help of the LoBoS staff.

- ¹J. Mongan and D. A. Case, "Biomolecular simulations at constant pH," *Curr. Opin. Struct. Biol.* **15**, 157–163 (2005).
- ²D. G. Isom, C. A. Castañeda, B. R. Cannon, P. D. Velu, and B. García-Moreno E, "Charges in the hydrophobic interior of proteins," *Proc. Natl. Acad. Sci. U. S. A.* **107**, 16096–16100 (2010).
- ³D. G. Isom, C. A. Castañeda, B. R. Cannon, and B. García-Moreno E, "Large shifts in pK_a values of lysine residues buried inside a protein," *Proc. Natl. Acad. Sci. U. S. A.* **108**, 5260–5265 (2011).
- ⁴C. A. Fitch, D. A. Karp, K. K. Lee, W. E. Stites, E. E. Lattman, and B. García-Moreno E, "Experimental pK_a values of buried residues: Analysis with continuum methods and role of water penetration," *Biophys. J.* **82**, 3289–3304 (2002).
- ⁵B. E. García-Moreno, J. J. Dwyer, A. G. Gittis, E. E. Lattman, D. S. Spencer, and W. E. Stites, "Experimental measurement of the effective dielectric in the hydrophobic core of a protein," *Biophys. Chem.* **64**, 211–224 (1997).
- ⁶C. N. Schutz and A. Warshel, "What are the dielectric constants of proteins and how to validate electrostatic models?," *Proteins: Struct., Funct., Bioinf.* **44**, 400–417 (2001).
- ⁷J. Antosiewicz, J. A. McCammon, and M. K. Gilson, "The determinants of pK_as in proteins," *Biochemistry* **35**, 7819–7833 (1996).
- ⁸H. W. T. van Vlijmen, M. Schaefer, and M. Karplus, "Improving the accuracy of protein pK_a calculations: Conformational averaging versus the average structure," *Proteins: Struct., Funct., Bioinf.* **33**, 145–158 (1998).
- ⁹R. E. Georgescu, E. G. Alexov, and M. R. Gunner, "Combining conformational flexibility and continuum electrostatics for calculating pK_as in proteins," *Biophys. J.* **83**, 1731–1748 (2002).
- ¹⁰A. M. Baptista, V. H. Teixeira, and C. M. Soares, "Constant-pH molecular dynamics using stochastic titration," *J. Chem. Phys.* **117**, 4184–4200 (2002).
- ¹¹R. Bürgi, P. A. Kollman, and W. F. van Gunsteren, "Simulating proteins at constant pH: An approach combining molecular dynamics and Monte Carlo simulation," *Proteins: Struct., Funct., Bioinf.* **47**, 469–480 (2002).
- ¹²M. Dlugosz and J. M. Antosiewicz, "Constant-pH molecular dynamics simulations: A test case of succinic acid," *Chem. Phys.* **302**, 161–170 (2004).
- ¹³J. Mongan, D. A. Case, and J. A. McCammon, "Constant pH molecular dynamics in generalized Born implicit solvent," *J. Comput. Chem.* **25**, 2038–2048 (2004).
- ¹⁴J. E. Mertz and B. M. Pettitt, "Molecular dynamics at a constant pH," *Int. J. High Perform. Comput. Appl.* **8**, 47–53 (1994).
- ¹⁵A. M. Baptista, P. J. Martel, and S. B. Petersen, "Simulation of protein conformational freedom as a function of pH: Constant-pH molecular dynamics using implicit titration," *Proteins: Struct., Funct., Bioinf.* **27**, 523–544 (1997).
- ¹⁶M. S. Lee, F. R. Salsbury, and C. L. Brooks, "Constant-pH molecular dynamics using continuous titration coordinates," *Proteins: Struct., Funct., Bioinf.* **56**, 738–752 (2004).
- ¹⁷S. Donnini, F. Tegeler, G. Groenhof, and H. Grubmüller, "Constant pH molecular dynamics in explicit solvent with λ -dynamics," *J. Chem. Theory Comput.* **7**, 1962–1978 (2011).
- ¹⁸J. M. Swails, D. M. York, and A. E. Roitberg, "Constant pH replica exchange molecular dynamics in explicit solvent using discrete protonation states: Implementation, testing, and validation," *J. Chem. Theory Comput.* **10**, 1341–1352 (2014).
- ¹⁹H. A. Stern, "Molecular simulation with variable protonation states at constant pH," *J. Chem. Phys.* **126**, 164112 (2007).
- ²⁰J. Lee, B. T. Miller, A. Damjanovic, and B. R. Brooks, "Constant pH molecular dynamics in explicit solvent with enveloping distribution sampling and Hamiltonian exchange," *J. Chem. Theory Comput.* **10**, 2738–2750 (2014).
- ²¹Y. Chen and B. Roux, "Constant-pH hybrid nonequilibrium molecular dynamics-Monte Carlo simulation method," *J. Chem. Theory Comput.* **11**, 3919–3931 (2015).
- ²²B. K. Radak, C. Chipot, D. Suh, S. Jo, W. Jiang, J. C. Phillips, K. Schulten, and B. Roux, "Constant-pH molecular dynamics simulations for large biomolecular systems," *J. Chem. Theory Comput.* **13**, 5933–5944 (2017).
- ²³X. Wu and B. R. Brooks, "A virtual mixture approach to the study of multi-state equilibrium: Application to constant pH simulation in explicit water," *PLoS Comput. Biol.* **11**, e1004480 (2015).
- ²⁴U. Borjesson and P. H. Hunenberger, "Explicit-solvent molecular dynamics simulation at constant pH: Methodology and application to small amines," *J. Chem. Phys.* **114**, 9706–9719 (2001).
- ²⁵J. Khandogin and C. L. Brooks, "Constant pH molecular dynamics with proton tautomerism," *Biophys. J.* **89**, 141–157 (2005).
- ²⁶J. A. Wallace and J. K. Shen, "Continuous constant pH molecular dynamics in explicit solvent with pH-based replica exchange," *J. Chem. Theory Comput.* **7**, 2617–2629 (2011).
- ²⁷G. B. Goh, B. S. Hulbert, H. Zhou, and C. L. Brooks, "Constant pH molecular dynamics of proteins in explicit solvent with proton tautomerism," *Proteins: Struct., Funct., Bioinf.* **82**, 1319 (2014).
- ²⁸S. G. Itoh, A. Damjanovic, and B. R. Brooks, "PH replica exchange method based on discrete protonation states," *Proteins: Struct., Funct., Bioinf.* **79**, 3420–3436 (2011).
- ²⁹D. S. Dashti, Y. Meng, and A. E. Roitberg, "pH-replica exchange molecular dynamics in proteins using a discrete protonation method," *J. Phys. Chem. B* **116**, 8805–8811 (2012).
- ³⁰J. Lee, B. T. Miller, A. Damjanovic, and B. R. Brooks, "Enhancing constant-pH simulation in explicit solvent with a two-dimensional replica exchange method," *J. Chem. Theory Comput.* **11**, 2560–2574 (2015).
- ³¹J. M. Swails and A. E. Roitberg, "Enhancing conformation and protonation state sampling of hen egg white lysozyme using pH replica exchange molecular dynamics," *J. Chem. Theory Comput.* **8**, 4393–4404 (2012).
- ³²V. K. Rastogi and M. E. Girvin, "Structural changes linked to proton translocation by subunit c of the ATP synthase," *Nature* **402**, 263–268 (1999).
- ³³J. K. Lanyi, "Bacteriorhodopsin," *Annu. Rev. Physiol.* **66**, 665–688 (2004).
- ³⁴S. Yoshikawa, K. Shinzawa-Itoh, R. Nakashima, R. Yaono, E. Yamashita, N. Inoue, M. Yao, M. J. Fei, C. P. Libeu, T. Mizushima, H. Yamaguchi, T. Tomizaki, and T. Tsukihara, "Redox-coupled crystal structural changes in bovine heart cytochrome c oxidase," *Science* **280**, 1723–1729 (1998).
- ³⁵D. G. Isom, B. R. Cannon, C. A. Castañeda, and A. Robinson, "High tolerance for ionizable residues in the hydrophobic interior of proteins," *Proc. Natl. Acad. Sci. U. S. A.* **105**, 17784–17788 (2008).
- ³⁶C. A. Castañeda, C. A. Fitch, A. Majumdar, V. Khangulov, J. L. Schlessman, and B. E. García-Moreno, "Molecular determinants of the pK_a values of Asp and Glu residues in staphylococcal nuclease," *Proteins: Struct., Funct., Bioinf.* **77**, 570–588 (2009).
- ³⁷C. M. Kougantakis, E. M. Grasso, A. C. Robinson, J. A. Caro, J. L. Schlessman, A. Majumdar, and B. García-Moreno E, "Anomalous properties of Lys residues buried in the hydrophobic interior of a protein revealed with 15N-detect NMR spectroscopy," *J. Phys. Chem. Lett.* **9**, 383–387 (2018).
- ³⁸M. S. Chimenti, C. A. Castañeda, A. Majumdar, and B. García-Moreno E, "Structural origins of high apparent dielectric constants experienced by ionizable groups in the hydrophobic core of a protein," *J. Mol. Biol.* **405**, 361–377 (2011).
- ³⁹M. Chimenti, V. Khangulov, A. Robinson, A. Heroux, A. Majumdar, J. Schlessman, and B. García-Moreno, "Structural reorganization triggered by charging of Lys residues in the hydrophobic interior of a protein," *Structure* **20**, 1071–1085 (2012).
- ⁴⁰D. E. Richman, A. Majumdar, and B. García-Moreno E, "Conformational reorganization coupled to the ionization of internal Lys residues in proteins," *Biochemistry* **54**, 5888–5897 (2015).
- ⁴¹A. Damjanovic, B. García-Moreno, E. E. Lattman, and A. E. García, "Molecular dynamics study of water penetration in staphylococcal nuclease," *Proteins: Struct., Funct., Bioinf.* **60**, 433–449 (2005).
- ⁴²A. Damjanović, J. L. Schlessman, C. A. Fitch, A. E. García, and B. García-Moreno E, "Role of flexibility and polarity as determinants of the hydration of internal cavities and pockets in proteins," *Biophys. J.* **93**, 2791–2804 (2007).
- ⁴³A. Damjanovic, X. Wu, B. García-Moreno E, and B. R. Brooks, "Backbone relaxation coupled to the ionization of internal groups in proteins: A self-guided Langevin dynamics study," *Biophys. J.* **95**, 4091–4101 (2008).
- ⁴⁴A. Damjanovic, B. R. Brooks, and B. García-Moreno E, "Conformational relaxation and water penetration coupled to ionization of internal groups in proteins," *J. Phys. Chem. A* **115**, 4042–4053 (2011).
- ⁴⁵E. J. Arthur, J. D. Yesselman, and C. L. Brooks, "Predicting extreme pK_a shifts in staphylococcal nuclease mutants with constant pH molecular dynamics," *Proteins: Struct., Funct., Bioinf.* **79**, 3276–3286 (2011).
- ⁴⁶J. A. Wallace, Y. Wang, C. Shi, K. J. Pastoor, B. Nguyen, K. Xia, and J. K. Shen, "Toward accurate prediction of pK_a values for internal protein residues: The importance of conformational relaxation and desolvation energy," *Proteins: Struct., Funct., Bioinf.* **79**, 3364–3373 (2011).

- ⁴⁷S. L. Williams, P. G. Blachly, and J. A. McCammon, "Measuring the successes and deficiencies of constant pH molecular dynamics: A blind prediction study," *Proteins: Struct., Funct., Bioinf.* **79**, 3381–3388 (2011).
- ⁴⁸G. B. Goh, E. N. Laricheva, and C. L. Brooks, "Uncovering pH-dependent transient states of proteins with buried ionizable residues," *J. Am. Chem. Soc.* **136**, 8496–8499 (2014).
- ⁴⁹Y. Zheng and Q. Cui, "Microscopic mechanisms that govern the titration response and pK_a values of buried residues in staphylococcal nuclease mutants," *Proteins: Struct., Funct., Bioinf.* **85**, 268–281 (2017).
- ⁵⁰X. Wu, J. Lee, and B. R. Brooks, "Origin of pK_a shifts of internal lysine residues in SNase studied via equal-molar VMMS simulations in explicit water," *J. Phys. Chem. B* **121**, 3318–3330 (2017).
- ⁵¹A. S. Oliveira, S. R. R. Campos, A. M. Baptista, and C. M. Soares, "Coupling between protonation and conformation in cytochrome c oxidase: Insights from constant-pH MD simulations," *Biochim. Biophys. Acta, Bioenerg.* **1857**, 759–771 (2016).
- ⁵²N. V. Di Russo, M. A. Marti, and A. E. Roitberg, "Underlying thermodynamics of pH-dependent allostery," *J. Phys. Chem. B* **118**, 12818–12826 (2014).
- ⁵³Y. Huang, W. Chen, D. L. Dotson, O. Beckstein, and J. Shen, "Mechanism of pH-dependent activation of the sodium-proton antiporter NhaA," *Nat. Commun.* **7**, 12940 (2016).
- ⁵⁴W. Chen, Y. Huang, and J. Shen, "Conformational activation of a transmembrane proton channel from constant pH molecular dynamics," *J. Phys. Chem. Lett.* **7**, 3961–3966 (2016).
- ⁵⁵Z. Yue, W. Chen, H. I. Zgurskaya, and J. Shen, "Constant pH molecular dynamics reveals how proton release drives the conformational transition of a transmembrane efflux pump," *J. Chem. Theory Comput.* **13**, 6405–6414 (2017).
- ⁵⁶J. Khandogin and C. L. Brooks, "Toward the accurate first-principles prediction of ionization equilibria in proteins," *Biochemistry* **45**, 9363–9373 (2006).
- ⁵⁷Y. Meng and A. E. Roitberg, "Constant pH replica exchange molecular dynamics in biomolecules using a discrete protonation model," *J. Chem. Theory Comput.* **6**, 1401–1412 (2010).
- ⁵⁸S. L. Williams, C. de Oliveira, F. Augusto, and J. A. McCammon, "Coupling constant pH molecular dynamics with accelerated molecular dynamics," *J. Chem. Theory Comput.* **6**, 560–568 (2010).
- ⁵⁹E. R. May, K. Arora, and C. L. Brooks, "pH-induced stability switching of the bacteriophage HK97 maturation pathway," *J. Am. Chem. Soc.* **136**, 3097–3107 (2013).
- ⁶⁰E. N. Laricheva, G. B. Goh, A. Dickson, and C. L. Brooks, "pH-dependent transient conformational states control optical properties in cyan fluorescent protein," *J. Am. Chem. Soc.* **137**, 2892–2900 (2015).
- ⁶¹L. S. Ahlstrom, S. M. Law, A. Dickson, and C. L. Brooks, "Multiscale modeling of a conditionally disordered pH-sensing chaperone," *J. Mol. Biol.* **427**, 1670–1680 (2015).
- ⁶²J. A. Wallace and J. K. Shen, in *Predicting pK_a Values with Continuous Constant pH Molecular Dynamics*, Methods in Enzymology, edited by M. L. Johnson, G. K. Ackers, and Jo M. Holt (Academic Press, 2009), Chap. 19, Vol. 466, pp. 455–475.
- ⁶³Y. Song, "Exploring conformational changes coupled to ionization states using a hybrid Rosetta-MCCE protocol," *Proteins: Struct., Funct., Bioinf.* **79**, 3356–3363 (2011).
- ⁶⁴K. Kilambi and J. Gray, "Rapid calculation of protein pK_a values using Rosetta," *Biophys. J.* **103**, 587–595 (2012).
- ⁶⁵M. Enciso, C. Schutte, and L. Delle Site, "A pH-dependent coarse-grained model for peptides," *Soft Matter* **9**, 6118–6127 (2013).
- ⁶⁶W. F. D. Bennett, A. W. Chen, S. Donnini, G. Groenhof, and D. P. Tieleman, "Constant pH simulations with the coarse-grained MARTINI model: Application to oleic acid aggregates," *Can. J. Chem.* **91**, 839–846 (2013).
- ⁶⁷H. Li, G. Li, B. A. Berg, and W. Yang, "Finite reservoir replica exchange to enhance canonical sampling in rugged energy surfaces," *J. Chem. Phys.* **125**, 144902 (2006).
- ⁶⁸A. Okur, D. R. Roe, G. Cui, V. Hornak, and C. Simmerling, "Improving convergence of replica-exchange simulations through coupling to a high-temperature structure reservoir," *J. Chem. Theory Comput.* **3**, 557–568 (2007).
- ⁶⁹A. Okur, B. T. Miller, K. Joo, J. Lee, and B. R. Brooks, "Generating reservoir conformations for replica exchange through the use of the conformational space annealing method," *J. Chem. Theory Comput.* **9**, 1115–1124 (2013).
- ⁷⁰B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations," *J. Comput. Chem.* **4**, 187–217 (1983).
- ⁷¹B. R. Brooks *et al.*, "CHARMM: The biomolecular simulation program," *J. Comput. Chem.* **30**, 1545–1614 (2009).
- ⁷²N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *J. Chem. Phys.* **21**, 1087 (1953).
- ⁷³T. Yamaguchi, T. Kiuchi, T. Matsuoaka, and S. Koda, "Multi-pH Monte Carlo simulation of coil-globule transition of weak polyelectrolyte," *Bull. Chem. Soc. Jpn.* **78**, 2098–2104 (2005).
- ⁷⁴A. E. Roitberg, A. Okur, and C. Simmerling, "Coupling of replica exchange simulations to a non-Boltzmann structure reservoir," *J. Phys. Chem. B* **111**, 2415–2418 (2007).
- ⁷⁵A. D. MacKerell *et al.*, "All-atom empirical potential for molecular modeling and dynamics studies of proteins," *J. Phys. Chem. B* **102**, 3586–3616 (1998).
- ⁷⁶M. S. Lee, M. Feig, F. R. Salsbury, and C. L. Brooks, "New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations," *J. Comput. Chem.* **24**, 1348–1356 (2003).
- ⁷⁷B. T. Miller, R. P. Singh, J. B. Klauda, M. Hodošček, B. R. Brooks, and H. L. Woodcock III, "CHARMMing: A new, flexible Web portal for CHARMM," *J. Chem. Inf. Model.* **48**, 1920–1929 (2008).
- ⁷⁸W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual molecular dynamics," *J. Mol. Graph.* **14**, 33–38 (1996).
- ⁷⁹K. Vanommeslaeghe and A. D. MacKerell, "Automation of the CHARMM general force field (CGenFF) I: Bond perception and atom typing," *J. Chem. Inf. Model.* **52**, 3144–3154 (2012).
- ⁸⁰K. Vanommeslaeghe, E. P. Raman, and A. D. MacKerell, "Automation of the CHARMM general force field (CGenFF) II: Assignment of bonded parameters and partial atomic charges," *J. Chem. Inf. Model.* **52**, 3155–3168 (2012).
- ⁸¹V. J. Hilser, B. Garcia-Moreno, E. T. G. Oas, G. Kapp, and S. T. Whitten, "A statistical thermodynamic model of the protein ensemble," *Chem. Rev.* **106**, 1545–1558 (2006).
- ⁸²N. V. Di Russo, D. A. Estrin, M. A. Marti, and A. E. Roitberg, "pH-dependent conformational changes in proteins and their effect on experimental pK_as: The case of nitrophorin 4," *PLoS Comput. Biol.* **8**, e1002761 (2012).
- ⁸³C. Shi, J. Wallace, and J. Shen, "Thermodynamic coupling of protonation and conformational equilibria in proteins: Theory and simulation," *Biophys. J.* **102**, 1590–1597 (2012).
- ⁸⁴T. Dissanayake, J. M. Swails, M. E. Harris, A. E. Roitberg, and D. M. York, "Interpretation of pH-activity profiles for acid–base catalysis from molecular simulations," *Biochemistry* **54**, 1307–1313 (2015).