JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

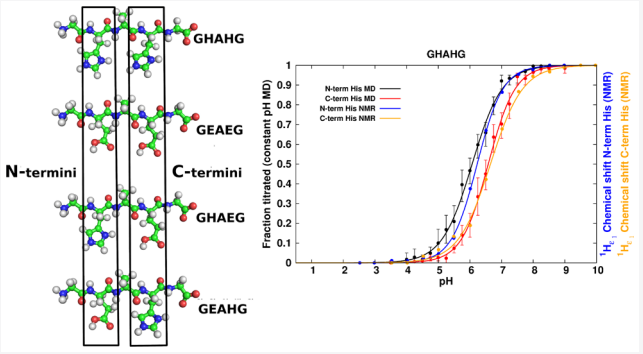
Probing the Accuracy of Explicit Solvent Constant pH Molecular Dynamics Simulations for Peptides

Plamen Dobrev, Sahithya Phani Babu Vemulapalli, Nilamoni Nath, Christian Griesinger, and Helmut Grubmüller*

Cite This: *J. Chem. Theory Comput.* 2020, 16, 2561−2569 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Protonation states of titratable amino acids play a key role in many biomolecular processes. Knowledge of protonatable residue charges at a given pH is essential for a correct understanding of protein catalysis, inter- and intramolecular interactions, substrate binding, and protein dynamics for instance. However, acquiring experimental values for individual amino acid protonation states of complex systems is not straightforward; therefore, several *in silico* approaches have been developed to tackle this issue. In this work, we assess the accuracy of our previously developed constant pH MD approach by comparing our theoretically obtained pK_a values for titratable residues with experimental values from an equivalent NMR study. We selected a set of four pentapeptides, of adequately small size to ensure comprehensive sampling, but concurrently, due to their charge composition, posing a challenge for protonation state calculation. The comparison of the pK_a values shows good agreement of the experimental and the theoretical approach with a largest difference of 0.25 pK_a units. Further, the corresponding titration curves are in fair agreement, although the shift of the Hill coefficient from a value of 1 was not always reproduced in simulations. The phase space overlap in Cartesian space between trajectories generated in constant pH and standard MD simulations is fair and suggests that our constant pH MD approach reasonably well preserves the dynamics of the system, allowing dynamic protonation MD simulations without introducing structural artifacts.



1. INTRODUCTION

The protonation state of titratable amino acids is often a determining factor in a variety of biomolecular processes. Knowing the actual charge of an amino acid residue at a given pH is particularly essential for studying electrostatically driven reactions. These charged residues define the spatial distribution of the electrostatic potential inside a molecule, which is crucial for enzyme catalysis, protein dynamics, or inter- and intramolecular interactions, for example.

However, due to an often large number of protonatable sites in such systems, titration curves of individual residues are not straightforwardly accessible in experiments. For that reason, several *in silico* approaches have been developed. Most of them rely on calculating the shift of the deprotonation free energy of an amino acid model compound in solution (usually referred to as "reference state") from the deprotonation free energy of the amino acid residue in a protein. Depending on the electrostatic scheme used, these methods fall in two different categories: continuum or all-atom. The first category comprises Poisson−Boltzmann (PB) or Generalized Born (GB) models, where the deprotonation free energy is calculated using fixed atomic structures in a dielectric environment. The continuum electrostatics (GB) scheme is further used in conjunction with molecular dynamics (MD) methods that utilize dynamic protonation states (termed "constant pH MD").[1−4] In these models, the protonation state of an amino acid changes continuously[1,2,5] by propagating the motion of a virtual particle (usually referred to as "λ-particle") that interpolates between two different protonation Hamiltonians, typically in implicit solvent MD simulations. The forces acting on both the particles in Cartesian as well as λ-space are calculated using continuum electrostatics.

The second category contains methods that rely on all-atom electrostatics in MD simulations. One group of them also uses λ-particle propagation, where all electrostatic interactions are treated explicitly, either via Particle Mesh Ewald (PME)[6−9] or a shifted Coulomb potential, which yields a value of zero at a given cutoff distance.[10−12] Another group of methods in that category uses the λ as a parameter in non-equilibrium explicit

Received: December 12, 2019
Published: March 19, 2020

https://dx.doi.org/10.1021/acs.jctc.9b01232
J. Chem. Theory Comput. 2020, 16, 2561−2569

solvent MD, for externally changing the protonation state of one group at a time followed by a Monte Carlo (MC) step, which rejects or accepts the protonation state switch.[13−15] All-atom simulations allow the explicit treatment of hydrogen bonds between solvent and solute, which is essential for the correct deprotonation free energy calculation, hence for the precise estimation of amino acid protonation states.

As an alternative, in a number of methods, both above-mentioned approaches have been combined and all-atom MD protocols developed in which the electrostatics in Cartesian space is calculated using PME and in λ-space using PB[4,16] or GB.[17] In this formalism, the protonation states of the amino acids are changed either discretely via MC sampling where the λ-particle adopts values of only 0 or 1[4,16,18] or propagates continuously.[17]

Previously, we developed a method for constant pH MD simulations in explicit solvent with PME electrostatics in both λ- and Cartesian space.[6−8] In the present study, we assess the accuracy of our approach and further optimize our simulation protocol and relevant parameters. The accuracy evaluation is achieved by comparing calculated pK_a values to ones determined by Heteronuclear Single Quantum Correlation (HSQC) NMR experiments for a set of test peptides selected to pose a challenge to the calculations. The test set comprises four pentapeptide systems, each containing two titratable residues and charged N- and C-termini. Our choice was motivated by three main criteria: First, the pK_a values in both the reference state of the amino acids and the peptides should be precisely measurable in HSQC NMR titration experiments, to provide a reliable comparison for the theoretically calculated values. Second, the size of the systems should be small enough such that their conformations are exhaustively sampled in MD simulations, and therefore, lack of convergence is not an issue. Third, the systems should involve coupling of titratable sites, which implies close proximity of the residues, as well as charged N- and C-termini ensuring strong intermolecular charge interactions that poses a challenging test case for the constant pH MD ability to correctly model electrostatic interactions. Based on these criteria, accurate results for the test systems would suggest that the used constant pH method would also accurately perform for larger biomolecular systems such as proteins.

Evaluating the capability of our constant pH MD implementation for predicting pK_a values solely based on structural data is the goal of this study. Hence we independently performed experimental and theoretical titrations and consequently compared the resulting pK_a value data sets after autonomous evaluation. Following this strategy, we verified the ability of our constant pH MD implementation to be used as a standard algorithm in MD simulations in near future.

2. METHODS

2.1. Pentapeptides. Four different pentapeptides, each containing two titratable residues, were used as test systems (Figure 1). The sequences were composed of the Gly-X-Ala-X-Gly motif, where X represents either a glutamate (Glu) or a histidine (His) residue. The rationale for selecting sequences that contain both two Glu and two His as well as combinations of one His and one Glu was to test if our method reproduces the protonation behavior of the amino acids depending on whether they interact with residues whose reference pK_a values are different or identical to their own. The positions of both residues with respect to the charged N- and C-termini were interchanged so that each of the two was placed next to a positive as well as a negative charge.

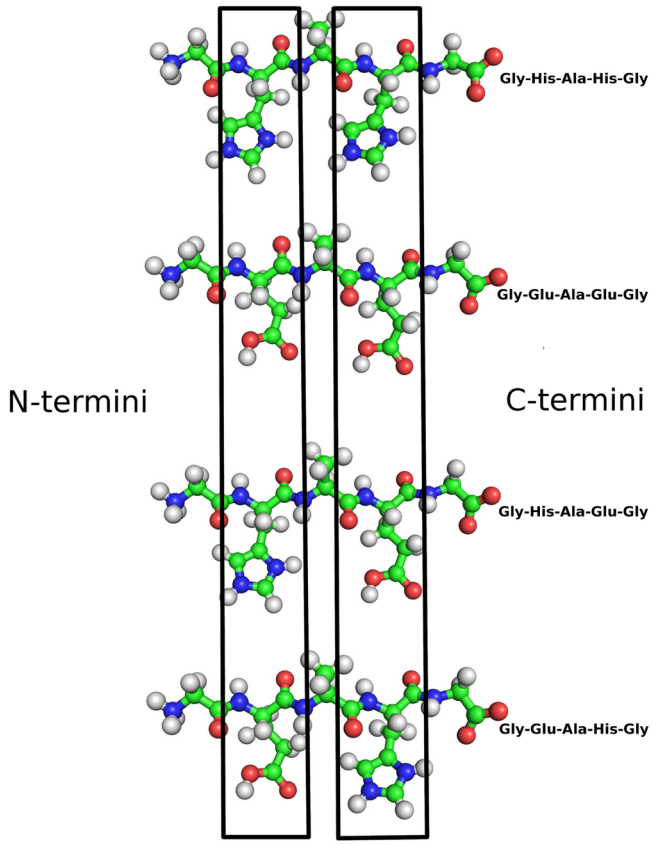


Figure 1. Pentapeptide test set. The peptides are composed of a Gly-X-Ala-X-Gly motif where the two titratable residues, denoted as X, correspond to either a Glu or a His residue and are highlighted with black rectangles.

2.2. Constant pH MD. Our previously developed three state model in the framework of our constant pH MD module[6−8] for the GROMACS 3.3.3[19−21] MD software package was used for the *in silico* titrations. A separate λ-particle describes the protonation state of each titratable site, as the propagation of the particles interpolates the respective sites between different protonation forms. The force acting on the λ-particle is $-\mathrm{d}H/\mathrm{d}\lambda$, where H is the all-atom Hamiltonian of the system, given by

$$H = (1-\lambda)H_A + \lambda H_B + f(\lambda)RT\ln 10(\mathrm{p}K_{a,ref} - \mathrm{pH}) + V_{MM}(\lambda) + \frac{m}{2}\dot{\lambda}^2 \quad (1)$$

and H_A and H_B are the Hamiltonians of the protonated and deprotonated states, respectively. $V_{MM}(\lambda)$ is a correction term introduced to account for the lack of an electron rearrangement term in force field simulations;[6,8] m and $\dot{\lambda}$ are the mass and the velocity of the λ-particle, respectively. The third term of the equation is a function of λ that ensures the correct ratio between protonated and deprotonated fractions of the titrated amino acid in its reference state (corresponding to the N- and C-methyl capped residues in solution). p$K_{a,ref}$ is the experimental pK_a value of the residue reference state, and pH is the pH at which the simulation is carried out. Here we chose the function $f(\lambda)$ based on a previously developed

https://dx.doi.org/10.1021/acs.jctc.9b01232
J. Chem. Theory Comput. 2020, 16, 2561−2569

double well biasing potential,[7] which was introduced to focus the distribution of λ mostly to the physically meaningful states $\lambda = 0$ (protonated) and $\lambda = 1$ (deprotonated). In the present work, $f(\lambda)$ is further enhanced to serve as a pH correction potential, which provides the free energy difference between the protonation forms of the titrated compound reference state by setting different depths for the two minima of the potential

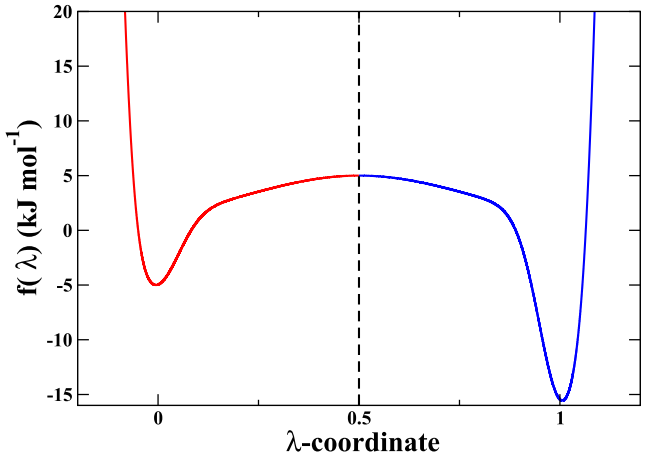


Figure 2. Biasing potential $f(\lambda)$ in eq 1. The barrier between the two minima decreases the occupancy of the nonphysical intermediate values of the λ-particle. In addition, the different minima depths are set such that the correct fraction between protonated (red) and deprotonated (blue) forms is achieved, according to the experimental pK_a value of the titrated compound and the pH at which the simulation is carried out.

(Figure 2). The potential (for derivation see Donnini et al.[7]) has the form

$$f(\lambda) = -k_1 \exp\left[-\frac{(\lambda - \mu_2 - b)^2}{2a^2}\right] - k_0 \exp\left[-\frac{(\lambda - \mu_1 + b)^2}{2a^2}\right] + d \exp\left[-\frac{(\lambda - 0.5)^2}{2s^2}\right] + 0.5w(1 - \mathrm{erf}[r(\lambda + m)]) + (1 + \mathrm{erf}[r(\lambda - 1 - m)]) \quad (2)$$

where the first two terms are two negative Gaussian functions corresponding to the two minima that generate λ-distributions whose average values are positioned at μ_1 and μ_2. The depths of the minima, k_0 and k_1, were chosen such that the free energy difference between the two intervals of the λ-coordinate ($\lambda < 0.5$ and $\lambda > 0.5$) corresponds to the free energy difference between the protonated and the deprotonated form of the amino acid (according to the reference p$K_{a,ref}$ value of the residue and the pH at which the simulation is carried out).

The width of the two minima also affects the population of the two potonation forms. To take this effect into account, the partition functions Z_0 and Z_1 of the two λ-intervals were calculated, and the free energy difference was corrected by $-RT \ln(Z_1/Z_0)$.

The third term of $f(\lambda)$ is a Gaussian function that generates a barrier with a height d (calculated from the baseline of the two minima) and the fourth term creates steep potential walls with height $w = 100$ kJ·mol^{-1} at the end of the λ-interval keeping the λ-particle inside. The barrier width s is set to $0.3(\mu_2 - \mu_1)$, and the a and b parameters are adjusted iteratively as described in Donnini et al.[7] The r and m parameters determine the steepness of the walls and depend on w.[7] The barrier height is used to control the transition rate between protonation states, whereas the positions of μ_1 and μ_2 are used to correct for noninteger average charges in the two protonation intervals (see Section 2.3). All other parameters described above are calculated automatically by our constant pH MD module. For details see Donnini et al.[7] As a result, the only parameters the user needs to specify are the initial minima positions, the initial barrier height, and the pH at which the simulation is carried out.

Because the two residues Glu and His have two chemically coupled titratable sites, the two oxygen atoms of the carboxyl group and the two nitrogen atoms of the imidazole ring, respectively, these cannot be described with a single λ-coordinate (eq 1). Therefore, to describe these two chemically coupled titratable sites, our previously developed three state model[8] was applied that uses two λ-coordinates: Along the first, the amino acid changes its protonation state, and along the second, its tautomeric form, thus moving among four protonation Hamiltonians. Because the two deprotonated Hamiltonians for Glu and the two protonated Hamiltonians for His are identical, only three protonation states are to be considered, as defined by the Hamiltonian

$$H = (1 - \lambda_1)H_A + \lambda_1[(1 - \lambda_2)H_B + \lambda_2 H_C] \quad (3)$$

where H_B and H_C are the Hamiltonians of the two deprotonated states, and for Glu

$$H = (1 - \lambda_1)[(1 - \lambda_2)H_A + \lambda_2 H_C] + \lambda_1 H_B \quad (4)$$

where H_A and H_C are the Hamiltonians of the two protonated states. The pH correction term is given as

$$f(\lambda_1)RT \ln 10(\mathrm{p}K_{a,ref} - \mathrm{pH}) + f(\lambda_2)RT \ln 10(\mathrm{p}K_{a_2} - \mathrm{p}K_{a_1}) \quad (5)$$

where p$K_{a,ref}$ is the experimental pK_a value of the titrated compound in its reference state, pH denotes the pH at which the simulation is carried out, and pK_{a_1} and pK_{a_2} correspond to the microscopic pK_a values of the two chemically coupled titratable sites.

As an approximation (see Section 3), the high energy "anti-" conformation of the hydrogen atom was excluded in our constant pH simulations. This exclusion was achieved by altering the dihedral parameters of the group as described in our previous work.[8]

2.3. Preservation of an Integer Charge of the Protonated and Deprotonated Forms. Due to the charge interactions among the amino acids in the peptides, the average of the λ-particle distribution often deviates from the physical values of 0 and 1 in the protonated and deprotonated λ-intervals, respectively. In the present work, we used the minima position of the biasing potential $f(\lambda)$ to achieve average integer charges in the two protonation intervals, $\lambda < 0.5$ and $\lambda > 0.5$. To this end, before each production titration run, a calibration run was performed, during which optimized locations of the minima were obtained ensuring average λ-values significantly closer to integer values.

The locations of potential minima were adjusted as follows: initially, the two minima of the biasing potential were positioned such that the average λ (μ_1 and μ_2), calculated from the distributions generated by the two minima, equaled $\lambda = 0$ and $\lambda = 1$ (Figure 3). The barrier height was set to 10 kJ·

https://dx.doi.org/10.1021/acs.jctc.9b01232
J. Chem. Theory Comput. 2020, 16, 2561−2569

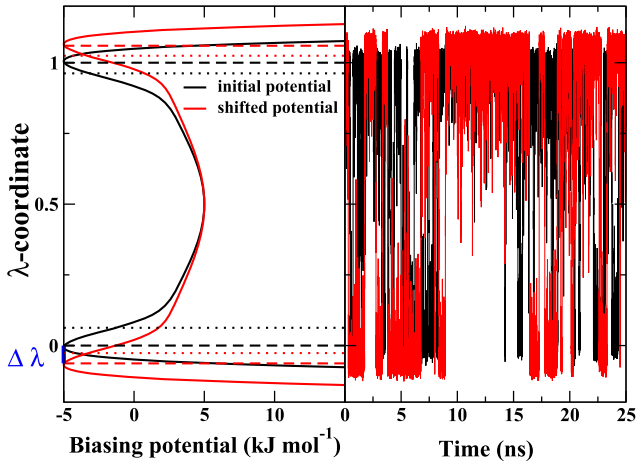


Figure 3. Preservation of the integer charge of two protonation forms. The left panel shows the initial biasing potential, with minima positioned, such that $\mu_1 = 0$ and $\mu_2 = 1$ (black line), and the one with new minima calculated via λ-distribution reweighting (red line). Dashed lines depict the position of the potential minima, whereas dotted lines show the average λ-value for the protonated and deprotonated λ-intervals. The shift ($\Delta\lambda$) of the potential minimum along the λ-coordinate (see eq 6) is highlighted in blue. The right panel shows the λ-trajectories with the initial potential (black line) and the one with new potential minima (red line).

mol^{-1}[18] relative to the depth of the higher minimum. After the calibration run was carried out, the λ-trajectories, for which the physical values of $\lambda = 0$ and $\lambda = 1$ were outside the error estimation of the average λ in the respective λ-intervals, were reweighted (eq 6) with a set of biasing potentials (Figure 3). The minima positions of each potential were shifted by 0.001 λ-units with respect to the previous. The shifting of the potential was performed until, after reweighting the averages $\langle\lambda\rangle$ in the $\lambda < 0.5$ and $\lambda > 0.5$ regions, equaled 0 and 1 (within the precision allowed by the numerical calculation of the double well potential[7]), respectively. The average $\langle\lambda\rangle$ for every shifted potential was calculated as

$$\langle\lambda\rangle = \frac{\sum_{n=1}^{n=N} \lambda_n e^{-\beta(V_{\Delta\lambda}-V)}}{\sum_{n=1}^{n=N} e^{-\beta(V_{\Delta\lambda}-V)}} \quad (6)$$

where V is the initial potential with μ_1 and μ_2 positioned at $\lambda = 0$ and $\lambda = 1$, respectively. $V_{\Delta\lambda}$ is the potential at an iteration step in which the minima are shifted by $\Delta\lambda$, respectively. N is the number of λ-points, along the λ-trajectory, for the respective λ-interval. In cases where the shift of the minima position was not adequate to achieve integer λ-values during reweighting (mostly in pH regions close to the pK_a of the respective residue), the barrier was increased. The reweighting procedure was done separately for each of the two minima and for both λ-coordinates of the three state model. Only the λ-trajectories from the second titration were used for the pK_a calculation.

2.4. Simulation Details. All simulations were carried out with the Amber99sb[22] force field (ported[23,24] to GROMACS 3.3.3) and the SPCE[25] water model. Virtual sites[26] and a 4 fs time step were used in all simulations. All bonds of the peptides and the coupled water molecules were constrained with the LINCS[27] constraint algorithm, and the bonds of all other water molecules were constrained with the SETTLE[28] algorithm. The simulations were carried out in a cubic simulation box with a 150 mM salt (NaCl) concentration (same as in NMR experiment). The temperature was set to 300 K, via the Nosè-Hoover thermostat[29,30] and the pressure to 1 bar via the Parrinelo−Rahman barostat.[31,32] PME electrostatics[33,34] was used with a cutoff of 1 nm for the direct Coulomb interactions and a spacing of 0.12 nm for the Fourier grid. A cutoff of 1 nm was used for the Lennard-Jones potential.

The titrating reaction coordinate of the two titratable residues was coupled to neutralizing water molecules, ensuring an uncharged simulation box.[8] To minimize the electrostatic interactions between the peptide and the coupled water molecules, we made sure that the time averaged distances between any of the peptide atoms and any of the two coupled waters or their periodic images as well as the distance between the two water molecules themselves was not shorter than 3.5 Debye−Hückel lengths, resulting in a 6 nm simulation box size. During the simulations, the coupled water molecules were fixed in space by restraining their oxygen atoms using a harmonic potential with a force constant of 1000 kJ·mol^{-1}·nm^{-1}. The same force constant was used to restrain the C_α atom of the Ala residue of each peptide. The mass of the λ-particle was set to 20 atomic mass units and its temperature to 300 K via the Andersen[35] thermostat with a 6 ps^{-1} coupling constant.

All peptides were titrated separately in a series of constant pH MD simulations at pH values between pH 3.5 and pH 9 for GHAHG, pH 2 and pH 6.5 for GEAEG, and pH 2 and pH 9 for GEAHG and GHAEG, in steps of 0.25 pH units, which was increased to 0.5 pH units for the two outermost values. Reference macroscopic pK_a values of 4.08 for Glu and 6.54 for His were used, which were determined in an NMR experiment by titrating His and Glu amino acid residues with methyl-capped N- and C-termini and following the same experimental protocol used for titrating the four peptides. Capped amino acid residues were also used as a reference state in the constant pH MD simulation protocol (for details see Dobrev et al.[8]). The difference between the two microscopic pK_a values for His was taken from Tanokura.[36] The microscopic pK_a values for Glu were calculated from the macroscopic ones as described in Dobrev et al.[8]

2.5. NMR Determination of the Pentapeptide pK_a Values. The four pentapeptides GHAHG, GEAHG, GHAEG, and GEAEG were purchased from PeptideSynthetics Research Ltd. (Hampshire, U.K.) with >98% purity and were used without further purification. All NMR titrations were carried out on lyophilized peptide powders dissolved in D_2O to ∼10 mM in the presence of 150 mM NaCl. Acetone was added as a pH independent internal ^1H (δ_{1H} = 2.22 ppm) and ^{13}C (δ_{13C} = 30.9 ppm) chemical shift reference standard. The pH of the peptide solution was adjusted with aliquots of HCl and NaOH. After measuring the pH, the peptide solution was transferred back to an NMR tube and $^{13}C-^1H$ HSQC (Heteronuclear Single Quantum Correlation, Bruker pulse program: hsqcetgpsi2) acquired.[38−40] The pH of the peptide solution was measured before and after an NMR experiment using Mettler Toledo SevenEasy S20 pH meter, and the average value of the two measurements was used as the pH of the sample to eliminate the error due to fluctuations in the pH. A pH of 0.4 units was subtracted from the experimentally determined pH of the samples to account for the deuterium isotope effect.[37] Subsequently after completing the measurement, the sample was transferred from an NMR tube to a 1.5 mL reaction tube,

https://dx.doi.org/10.1021/acs.jctc.9b01232
J. Chem. Theory Comput. 2020, 16, 2561−2569

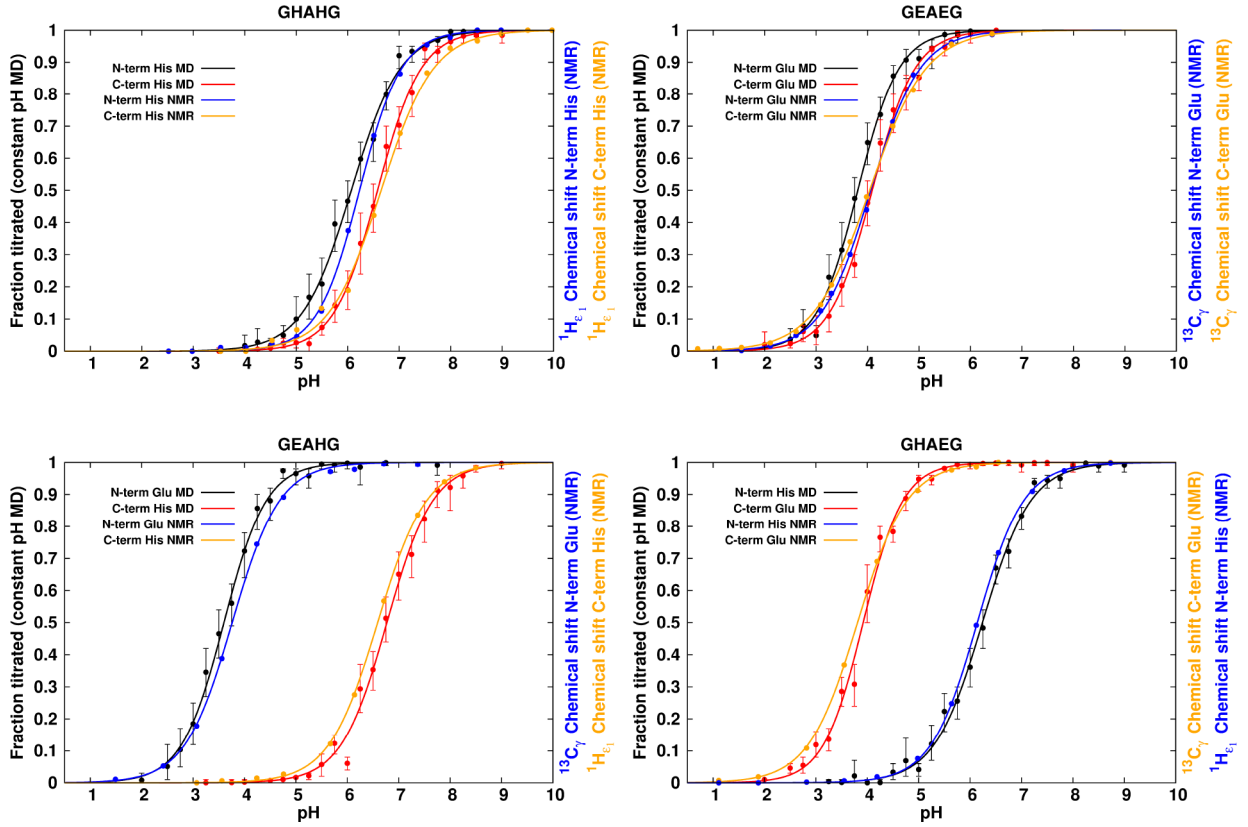


Figure 4. Experimental vs theoretical titration results. The titration curves of all four peptides obtained *in silico* (black and red) and in NMR experiments (blue and orange) are depicted. The continuous line in each plot was obtained via plotting the Hill curve for the theoretical or via using the adapted Henderson−Hasselbalch equation (eq 7) for the experimental titrations, using the pK_a values and Hill coefficients calculated from constant pH MD or NMR experiments, respectively. The *Y*-axes of the experimental curves are scaled to the 0 to 1 interval to ensure clearer comparison with the *in silico* data.

Table 1. pK_a Values, Hill Coefficients, Shifts from the Residues Reference pK_a Values (ΔpK_a), and Shift of the Theoretical from the Experimental pK_a Values ($\Delta\Delta$pK_a) of the Four Pentapeptides Estimated via NMR (Using the Chemical Shifts of C_γ and H_{ϵ_1} Atoms for Glu and His, Respectively) and Constant pH MD[a]

		Amber99sb			NMR			
		pK_a	ΔpK_a	Hill coefficient	pK_a	ΔpK_a	Hill coefficient	$\Delta\Delta$pK_a
GHAHG	N-term His	6.07 ± 0.05	−0.47	0.91 +0.07 −0.06	6.23 ± 0.01	−0.32	1.06 ± 0.03	−0.16
	C-term His	6.58 ± 0.05	+0.04	1.02 +0.09 −0.07	6.66 ± 0.03	+0.11	0.85 ± 0.04	−0.08
GEAEG	N-term Glu	3.81 ± 0.05	−0.27	1.08 +0.1 −0.09	4.06 ± 0.01	−0.01	0.90 ± 0.02	−0.25
	C-term Glu	4.09 +0.05 −0.04	+0.01	1.04 +0.09 −0.08	4.04 ± 0.01	−0.03	0.79 ± 0.02	+0.05
GEAHG	N-term Glu	3.60 ± 0.05	−0.48	1.05 +0.1 −0.08	3.76 ± 0.02	−0.31	0.94 ± 0.03	−0.16
	C-term His	6.77 ± 0.05	+0.23	0.97 +0.08 −0.07	6.59 ± 0.02	+0.04	0.91 ± 0.03	+0.18
GHAEG	N-term His	6.24 ± 0.04	−0.30	0.92 +0.06 −0.05	6.14 ± 0.01	−0.41	0.96 ± 0.01	+0.10
	C-term Glu	3.92 ± 0.04	−0.16	1.12 +0.08 −0.07	3.81 ± 0.01	−0.26	0.90 ± 0.02	+0.11

[a]The reference pK_a values used in constant pH MD simulations were 4.081 and 6.54 for Glu and His, respectively, which were calculated as average from all atoms in the residue. These values are slightly different from the 4.073 and 6.55 for the C_γ and H_{ϵ_1} of the Glu and His, respectively, considered to be closer to the real values. Although the differences are too small to have any effect on the conclusions, in the table the ΔpK_a values are calculated using the former set for constant pH MD and the latter set for experiment.

and the pH was determined using a Mettler Toledo SevenEasy S20 pH-meter. All pH measurements were carried out at a temperature of 300 K. The same reaction tube and NMR tube were used throughout the titration for each peptide. A three point calibration was conducted using standard buffers with pH = 4.0, 7.0, and 10.0 before measuring pH of the peptide solution.

NMR experiments were carried out on Bruker Avance NEO 800 MHz, Bruker Avance 700 MHz (Oxford), and Bruker Avance NEO 600 MHz spectrometers equipped with a 3 mm cryoprobe, 5 mm cryoprobe prodigy, and 5 mm cryoprobe prodigy, respectively, at 300 K. ^{1}H and ^{13}C chemical shifts of all peptides were unambiguously assigned using 1D ^{1}H and 2D (Double-Quantum Filtered Correlation Spectroscopy (DQFCOSY), Total Correlation Spectroscopy (TOCSY), Nuclear Overhauser Effect Spectroscopy (NOESY), and HSQC) NMR spectral analyses. The variation of ^{1}H and ^{13}C chemical shifts as a function of pH was monitored by

https://dx.doi.org/10.1021/acs.jctc.9b01232
J. Chem. Theory Comput. 2020, 16, 2561−2569

measuring 1D ^{1}H and $^{13}C-^{1}H$ HSQC NMR experiments, respectively. HSQC spectra were acquired in phase sensitive mode using echo/antiecho-TPPI gradient selection. One bond carbon−proton coupling ($^{1}J_{CH}$) was optimized to 145 Hz, and a trim pulse of 2 ms was used. NMR data was processed using Topspin 3.5 pl5 (Bruker, Germany).

The modified Henderson−Hasselbalch equation[41,42] was used to account for the observed change in the $^{1}H_{\epsilon_1}$ and $^{13}C_{\gamma}$ chemical shifts of histidine and glutamate, respectively, as a function of pH

$$\delta = \delta_{low} - \frac{\delta_{low} - \delta_{high}}{1 + 10^{n(\mathrm{p}K_a - \mathrm{p}H)}} \quad (7)$$

where the pK_a denotes the ionization constant, δ_{low} is the chemical shift at low pH, δ_{high} is the chemical shift at high pH, and n is the Hill coefficient. The parameters were obtained from the nonlinear least-squares fits (performed using OriginPro 8, v8.0724, Northampton, MA, U.S.A.) of experimental data to the modified Henderson−Hasselbalch equation.

3. RESULTS AND DISCUSSION

The main goal of the current work was to test how accurately our all-atom force field constant pH MD simulations predict the titration behavior of biological systems. As test systems, we chose four peptides that, due to their small size, allowed comprehensive sampling. At the same time, due to their charge composition, they posed a challenging case for pK_a calculation. The *in silico* calculated pK_a values and titration curves were compared to those independently determined in NMR experiments for the C_{γ} (Glu) and H_{ϵ_1} (His) atoms (see Figure 4). For the pK_a error estimation, multiple (50000) Hill equation fits were performed within the error estimates of the λ-value for each pH point (Table 1) as described in Dobrev et al.[8]

The close agreement between experimental and theoretical pK_a values, with a largest difference of 0.25 pK_a units, shows that we are indeed able to calculate the magnitude of the electrostatic interactions with high precision, what makes our implementation an applicable approach in MD simulations. Further, the correct prediction of amino acid pK_a shifts of residues within the peptides, compared to their pK_a values in solution, suggests that the here presented constant pH MD method reliably calculates the correct sign of the electrostatic interaction between amino acid residues, the only exception being the C-terminal Glu of the GEAEG peptide. The reference pK_a value of Glu in that case, however, is practically identical and falls well within the error estimation of the *in silico* value for Glu in the peptide. Therefore, no solid claim can be made about the pK_a shift of the latter.

The electrostatic interactions among the titratable residues and charged termini define the pK_a shift of the former. In the peptides with two different titratable residues (GEAHG and GHAEG), the pK_a values of the Glu residues are shifted down (ΔpK_a in Table 1) to smaller values compared to the reference value indicating a strong interaction with the nearby positively charged His. The fact that the direction of the shift does not depend on the actual position of the residues with respect to the charged N- or C-terminus further suggests that the Glu residue in these peptides most strongly interacts with the positively charged His residue rather than the terminal charges. The larger pK_a shift in the case of GEAHG indicates an additional effect due to the proximity of the positive N-terminus exerted on the Glu residue.

In contrast to the Glu, the pK_a values of the His residues in both peptides shift in different directions with respect to the reference suggesting a stronger interaction with the charged termini than with the negatively charged Glu. This observation is further supported by the fact the pK_a value of the His residue shifts up in proximity to the C-terminus and down in proximity to the N-terminus.

In peptides with two identical titratable residues (GHAHG and GEAEG), the pK_a shifts of the two His residues in the GHAHG peptide are in opposite directions, indicating stronger interaction of the residues with the charged termini than with each other. The experimental pK_a values of both Glu residues in the GEAEG peptide shift down with respect to their reference. However, due to the small differences between the pK_a values of the N-terminal Glu in the peptide and the Glu reference pK_a value, their error bars practically overlap. Therefore, it cannot be concluded which interaction is stronger, the one among the adjacent residues or the charged termini. Nonetheless, the pK_a values calculated via constant pH MD are close to the experimental ones. The *in silico* values show a more significant down shift for the Glu closer to the N-terminus, most probably due to their electrostatic interaction.

In contrast to the pK_a values, the Hill coefficients calculated from our constant pH MD simulation titration curves differ significantly from those determined by NMR. We considered two possible causes for this discrepancy. First, it might be caused by insufficient sampling, most likely due to direct and long-lived contacts between Na/Cl ions and protonatable groups in the simulations. We do not consider this explanation likely, because neither such long-lived contacts nor any other sign of insufficient sampling were observed. Second, because the Hill coefficients reflect the degree of titrating cooperativity between the two protonatable amino acids, the discrepancy might be due to how different atoms of one titrating residue "sense" (due to the different chemical environment) the protonation of the neighboring titrating residue in the NMR experiments.

To test how the choice of probe atoms affects the experimental Hill coefficients, we calculated additional titration curves from the chemical shifts observed for the H_{δ_2} (His) and C_{β} (Glu) atoms. Indeed, whereas the pK_a values obtained for those atoms are practically identical to the ones obtained for H_{ϵ_1} and C_{γ} shown in Table 1 (with the largest difference between the two sets of 0.06 pK_a units and retaining the sign of ΔpK_a in all groups), the values of the Hill coefficients show marked differences of up to 0.15 in both cases, similar in size to the actually observed deviations in Table 1. Further, in some cases, e.g., like with the N-terminal His of the GHAHG peptide and the N-terminal Glu of the GEAHG peptide, the direction of cooperativity changes, switching from values below 1 to above and *vice versa*. Also, the NMR measured Hill coefficient of the isolated Glu model compound, where no cooperativity exists, is 0.92 on average rather than 1 as expected. Due to this uncertainty in the interpretation of the NMR experiment, we prefer not to draw conclusions on the accuracy of our model based on a Hill coefficient comparison.

Because here we mainly focused at comparing pK_a values between experiment and constant pH MD and because only these are robust for different atoms in the NMR experiments,

https://dx.doi.org/10.1021/acs.jctc.9b01232
J. Chem. Theory Comput. 2020, 16, 2561−2569

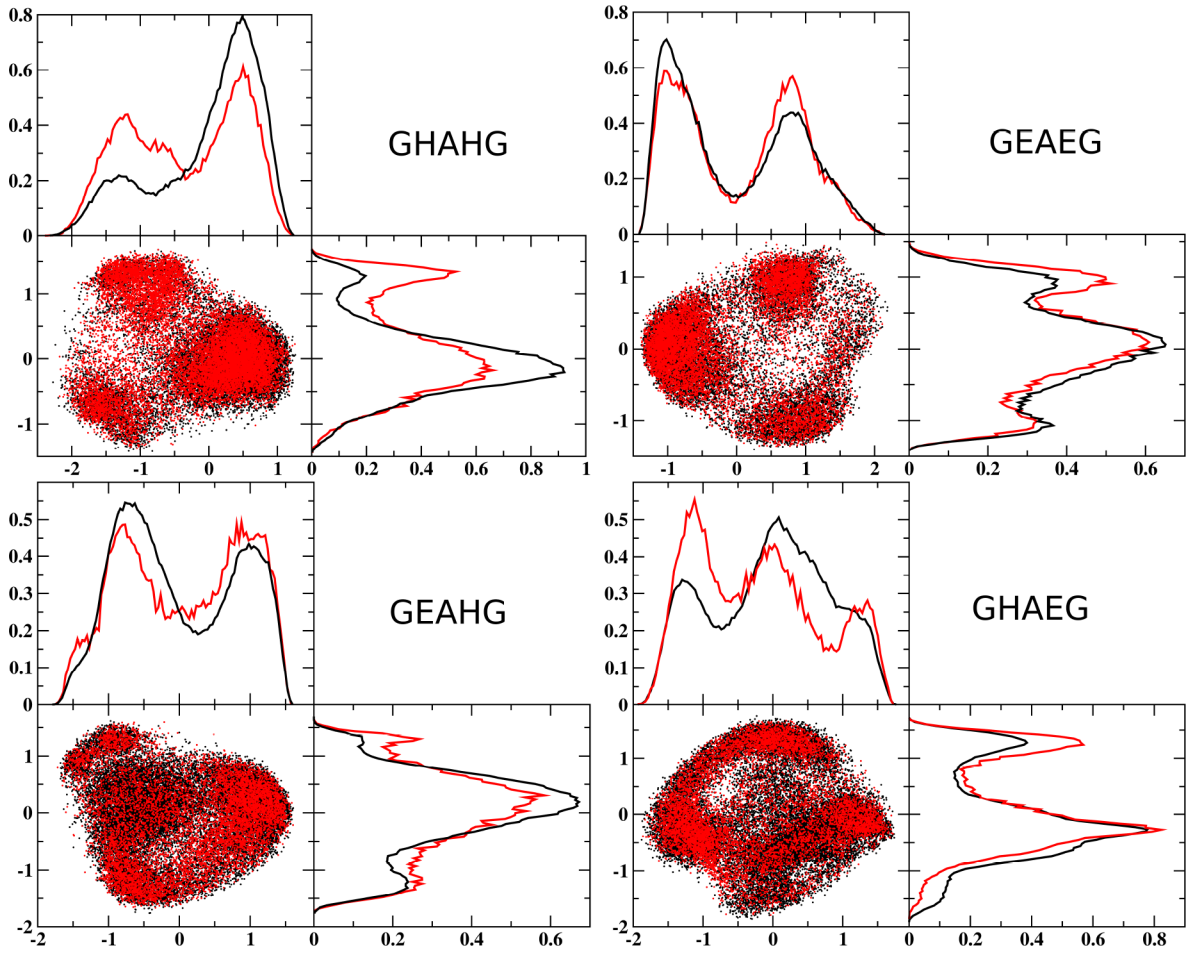


Figure 5. Subspace analysis of atomic trajectories obtained in regular and constant pH MD. In each graph is depicted the subspace of a different pentapeptide. Projection of the atomic trajectories obtained in constant pH MD (black) and regular MD (red) on the first and second eigenvectors of a covariance matrix generated from the regular MD trajectories are shown in the lower left panel in each graph. The probability distribution of the projections along the first and the second eigenvectors are depicted on the upper left and lower right panel of the graphs, respectively.

resolving this experimental uncertainty of the Hill coefficient is outside the scope of this paper.

In the current constant pH MD implementation, we approximate the description of the carboxyl group with three protonation states, thus excluding the high energy "anti-" orientation of the carboxyl hydrogen atoms. However, in the case of the peptide systems studied in this work, due to the low reorientation barrier of the carboxyl group (being exposed in solution) and the high energy of the "anti-" conformation, it is very unlikely that a structure would appear which can be realized only via adopting the high energy hydrogen conformation rather than via reorientation. Therefore, we do not believe that the approximation biases the calculated pK_a values.

Our constant pH MD approach uses a continuous protonation reaction coordinate. As a result, the titrated compounds spend a certain fraction of time in states with noninteger charges. To estimate to what extent these partial charges bias the Cartesian ensemble during titration, we compared the phase space overlap of the peptide systems sampled in regular and constant pH MD. For that purpose, we used regular MD trajectories concatenated to include all possible protonation forms of the two residues in each peptide and the analogous constant pH MD trajectories concatenated for all pH values. Figure 5 shows the peptide heavy atom projections of the two simulation sets on the first and the second eigenvectors of the covariance matrix generated from the regular MD trajectories. The regular MD probabilities of the different protonation forms along the vectors were weighted such that they matched the probabilities of the respective forms observed in constant pH MD. The projection distributions overlap reasonably well for all peptides, showing no artificial states arising in Cartesian space due to noninteger charges. The different heights of the maxima and minima in the two cases do not affect the protonation behavior of the peptides, because no correlation was found between the agreement of experimental and theoretical protonation estimation and the phase space overlap in regular and in constant pH MD.

Another consequence of using the continuous reaction coordinate is that the average charges of the intervals $\lambda < 0.5$ and $\lambda > 0.5$ do not have integer values of 0 and 1, respectively. To address this issue, we performed two sets of titrations: in the first training set, the distributions of the λ-particle in both intervals were estimated, and in the second, productive set the λ-distributions from the first set were used to obtain new biasing potential minima to ensure charges significantly closer to integer values in the two protonation intervals. The root-mean-square deviations (RMSD) of the average λ from 0 and 1 in both protonation intervals improve from 0.045 and 0.047 in

https://dx.doi.org/10.1021/acs.jctc.9b01232
J. Chem. Theory Comput. 2020, 16, 2561−2569

the initial titration set to 0.026 and 0.028 in the productive set, respectively. It must be noted that, due to the large pH range spanned in the titrations and the fact that the largest deviation from integer values occurs only in simulations in which the pH is close to the pK_a of the residue, the noncorrected values do not significantly differ from 0 and 1. Even without taking this fact into account, the RMSD values suggest that the integer charge preservation procedure introduced in the current work is effective. Additionally it is computationally inexpensive, at only very little computational cost.

The effect of the above-mentioned correction on the calculated pK_a values is small: The largest difference is 0.1 pK_a units (improvement in the corrected simulations) in the case of C-terminal Glu of the GEAEG peptide. In all other cases the differences are within statistical error. Nonetheless, we use the pK_a values only as an assessment of how well our constant pH MD approach reproduces the electrostatics of the system. The main goal of our method is to generate accurate Cartesian ensembles, while at the same time it allows dynamic protonation of titratable residues. So even if the pK_a values are accurate without correction, we would still like to reproduce the structural ensemble generated with charges as close to the physical state as possible.

4. CONCLUSIONS

The accurate description of protonation states in atomistic simulations of biological macromolecules is crucial for a quantitative understanding of many of their key properties, particularly electrostatically driven processes. Here we have assessed how accurately our recently developed constant pH MD simulation method reproduces experimentally determined protonation states. To this aim, we used constant pH MD simulation to predict pK_a values of titratable residues in four challenging test peptides with electrostatically coupled titratable groups and subsequently compared these with the respective pK_a values measured by NMR in a doubly blinded manner.

Indeed, the predicted pK_a values deviated from the measured ones by less than 0.25 units, underscoring the high accuracy of the constant pH simulations. In all cases, the direction of the obtained pK_a shifts from the reference values agrees within statistical error in both experiment and theoretical calculation. This agreement suggests that our constant pH MD approach can realistically reproduce the positive or negative contributions from different residues to the deprotonation free energy of titratable amino acids also in larger biomolecular systems.

Less good agreement between NMR and our simulations was seen for the Hill coefficients, which, however, can be explained by limited accuracy of the NMR values. In particular, the chemical shift of the "probe atom" used to measure the protonation state of the titratable groups seems to be also affected by other electrostatic interactions to nearby groups, as evidenced by the observation that the measured Hill coefficient changes when different "probe atoms" are used by an amount equal to or even larger than the observed discrepancy between experiment and simulation. In all cases, the measured pK_a values remained robust. We therefore do not think that these discrepancies point to simulation artifacts.

One possible additional concern is the use of a continuous protonation coordinate, which can give rise to unphysical, partially protonated states. Analysis of the charge distribution on the titratable sites in our constant pH MD simulations of the test peptides shows, however, that—as intended by construction—our implementation preserves integer charges of the protonation forms to a large extent, and that the fraction of unphysical partially charged states is small enough to leave the resulting Boltzmann ensemble of peptide structures largely unaffected. In particular, our test simulations have shown that the Boltzmann ensemble generated from the constant pH simulations can be described by a properly weighted superposition of conventional force field simulations with fixed protonation states.

In summary, the good agreement between experiment and simulation suggests that our continuous-coordinate constant pH MD simulation approach provides a more accurate description of biomolecular electrostatics and energetics than fixed charge simulations, also and in particular for large biomolecular systems. Further, it allows for precise control of the pH which otherwise would not be possible.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.9b01232.

Peptide pK_a values and Hill coefficients calculated from the chemical shifts of the C_β and H_{δ_2} atoms for Glu and His, respectively (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Helmut Grubmüller − *Max-Planck-Institut fur Biophysikalische Chemie, Theoretical and computational biophysics, Gottingen 37077, Germany;* orcid.org/0000-0002-3270-3144; Email: hgrubmu@gwdg.de

Authors

Plamen Dobrev − *Max-Planck-Institut fur Biophysikalische Chemie, Theoretical and computational biophysics, Gottingen 37077, Germany;* orcid.org/0000-0002-4235-4945

Sahithya Phani Babu Vemulapalli − *Max Planck Institute for Biophysical Chemistry, NMR-based Structural Biology, Gottingen 37077, Germany*

Nilamoni Nath − *Max Planck Institute for Biophysical Chemistry, NMR-based Structural Biology, Gottingen 37077, Germany; Gauhati University, Department of Chemistry, Guwahati 781014, Assam, India*

Christian Griesinger − *Max Planck Institute for Biophysical Chemistry, NMR-based Structural Biology, Gottingen 37077, Germany;* orcid.org/0000-0002-1266-4344

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.9b01232

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The work was supported by the Max Planck Society. Help by Serena Donnini with the implementation of the double well biasing potential in the constant pH MD module is gratefully acknowledged.

REFERENCES

(1) Lee, M. S.; Salsbury, F. R., Jr.; Brooks, C. L., 3rd. Constant-pH Molecular Dynamics Using Continuous Titration Coordinates. *Proteins: Struct., Funct., Genet.* **2004**, *56*, 738−752.

(2) Khandogin, J.; Brooks, C. L., 3rd. Constant pH Molecular Dynamics with Proton Tautomerism. *Biophys. J.* **2005**, *89* (1), 141−57.

(3) Simonson, T.; Carlsson, J.; Case, D. Proton binding to proteins: pK_a Calculations with Explicit and Implicit Solvent Models. *J. Am. Chem. Soc.* **2004**, *126*, 4167−4180.

(4) Mongan, J.; Case, D.; McCammon, J. Constant pH Molecular Dynamics in Generalized Born Implicit Solvent. *J. Comput. Chem.* **2004**, *25*, 2038−2048.

(5) Mertz, J. E.; Pettitt, B. M. Molecular Dynamics at a constant pH. *Int. J. Supercomput. Appl. High Perform. Eng.* **1994**, *8*, 47−53.

(6) 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.

(7) Donnini, S.; Ullmann, T.; Groenhof, G.; Grubmüller, H. Charge-Neutral Constant pH Molecular Dynamics Simulations Using a Parsimonious Proton buffer. *J. Chem. Theory Comput.* **2016**, *12* (3), 1040−1051.

(8) Dobrev, P.; Donnini, S.; Groenhof, G.; Grubmüller, H. Accurate Three State 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.

(9) Huang, Y.; Chen, W.; Wallace, J. A.; Shen, J. All-Atom Continuous Constant pH Molecular Dynamics with Particle Mesh Ewald and Titratable Water. *J. Chem. Theory Comput.* **2016**, *12*, 5411−5421.

(10) Goh, G. B.; Knight, J. L.; Brooks, C. L., 3rd. Constant pH Molecular Dynamics Simulations of Nucleic Acids in Explicit Solvent. *J. Chem. Theory Comput.* **2012**, *8* (1), 36−46.

(11) Goh, G. B.; Hulbert, B. S.; Zhou, H.; Brooks, C. L., 3rd. Constant pH Molecular Dynamics of Proteins in Explicit Solvent with Proton Tautomerism. *Proteins: Struct., Funct., Genet.* **2014**, *82*, 1319−1331.

(12) Lee, J.; Miller, B. T.; Damjanovic, A.; Brooks, B. R. Constant pH Molecular Dynamics in Explicit Solvent with Enveloping Distribution Sampling and Hamiltonian Exchange. *J. Chem. Theory Comput.* **2014**, *10*, 2738−2750.

(13) Radak, B. K.; Chipot, C.; Suh, D.; Jo, S.; Jiang, W.; Phillips, J. C.; Schulten, K.; Roux, B. Constant-pH Molecular Dynamics Simulations for Large Biomolecular Systems. *J. Chem. Theory Comput.* **2017**, *13*, 5933−5944.

(14) Chen, Y.; Roux, B. Constant-pH Hybrid Nonequilibrium Molecular Dynamics-Monte Carlo Simulation Method. *J. Chem. Theory Comput.* **2015**, *11*, 3919−3931.

(15) Stern, H. Molecular Simulation with Variable Protonation States at Constant pH. *J. Chem. Phys.* **2007**, *126*, 164112.

(16) Baptista, A. M.; Teixeira, V. H.; Soares, C. M. Constant-pH Molecular Dynamics Using Stochastic Titration. *J. Chem. Phys.* **2002**, *117*, 4184−4200.

(17) Wallace, J. A.; Shen, J. K. Continuous Constant pH Molecular Dynamics in Explicit Solvent with pH-Based Replica Exchange. *J. Chem. Theory Comput.* **2011**, *7*, 2617−2629.

(18) Swails, J. M.; York, D. M.; Roitberg, A. E. Constant pH Replica Exchange Molecular Dynamics in Explicit Solvent Using Discrete Protonation States: Implementation, Testing, and Validation. *J. Chem. Theory Comput.* **2014**, *10*, 1341−1352.

(19) Lindahl, E.; Hess, B.; van der Spoel, D. Gromacs 3.0: A Package for Molecular Simulation and Trajectory Analysis. *J. Mol. Model.* **2001**, *7*, 306−317.

(20) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91*, 43−56.

(21) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.; Berendsen, H. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.

(22) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 712−725.

(23) Sorin, E.; Pande, V. Exploring the Helix-Coil Transition via All-Atom Equilibrium Ensemble Simulations. *Biophys. J.* **2005**, *88*, 2472−2493.

(24) DePaul, A.; Thompson, E.; Patel, S.; Haldeman, K.; Sorin, E. Equilibrium Conformational Dynamics in an RNA Tetraloop from Massively Parallel Molecular Dynamics. *Nucleic Acids Res.* **2010**, *38*, 4856−4867.

(25) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269−6271.

(26) Feenstra, K.; Hess, B.; Berendsen, H. Improving Efficiency of Large Time-Scale Molecular Dynamics Simulations of Hydrogen-Rich System. *J. Comput. Chem.* **1999**, *20*, 786−798.

(27) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1172.

(28) Miyamoto, S.; Kollman, P. SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952−962.

(29) Nosè, S. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, *52*, 255−268.

(30) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A: At., Mol., Opt. Phys.* **1985**, *31*, 1695−1697.

(31) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.

(32) Nosè, S.; Klein, M. L. Constant Pressure Molecular Dynamics for Molecular Systems. *Mol. Phys.* **1983**, *50*, 1055−1076.

(33) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(34) Essmann, U.; Perera, L.; Berkowitz, M.; Darden, T.; Lee, H.; Pedersen, L. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8592.

(35) Andersen, H. C. Molecular Dynamics Simulations at Constant Pressure and/or Temperature. *J. Chem. Phys.* **1980**, *72*, 2384−2393.

(36) Tanokura, M. ^{1}H-NMR Study on the Tautomerism of the Imidazole Ring of Histidine. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1983**, *742* (3), 576−585.

(37) Glasoe, P.; Long, F. Use of Glass Electrodes to Measure Acidities in Deuterium Oxide. *J. Phys. Chem.* **1960**, *64*, 188−190.

(38) Palmer, A., III; Cavanagh, J.; Wright, P.; Rance, M. Sensitivity Improvement in Proton-Detected Two-Dimensional Heteronuclear Correlation NMR Spectroscopy. *J. Magn. Reson.* **1991**, *93*, 151−170.

(39) Kay, L.; Keifer, P.; Saarinen, T. Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy with Improved Sensitivity. *J. Am. Chem. Soc.* **1992**, *114*, 10663−10665.

(40) Schleucher, J.; Schwendinger, M.; Sattler, M.; Schmidt, P.; Schedletzky, O.; Glaser, S.; Sørensen, O.; Griesinger, C. A General Enhancement Scheme in Heteronuclear Multidimensional NMR Employing Pulsed Field Gradients. *J. Biomol. NMR* **1994**, *4*, 301−306.

(41) Markley, J. Observation of Histidine Residues in Proteins by Nuclear Magnetic Resonance Spectroscopy. *Acc. Chem. Res.* **1975**, *8*, 70−80.

(42) Dames, S.; Kammerer, R.; Moskau, D.; Engel, J.; Alexandrescu, A. Contributions of the Ionization States of Acidic Residues to the Stability of the Coiled Coil Domain of Matrilin-1. *FEBS Lett.* **1999**, *446*, 75−80.

https://dx.doi.org/10.1021/acs.jctc.9b01232
J. Chem. Theory Comput. 2020, 16, 2561−2569