Accurate Three States Model for Amino Acids with Two Chemically Coupled Titrating Sites in Explicit Solvent Atomistic Constant pH Simulations and pK_a Calculations

Plamen Dobrev,[†] Serena Donnini,[‡] Gerrit Groenhof,[§] and Helmut Grubmüller[*,†]

[†]Theoretical & Computational Biophysics Department, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

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

S *Supporting Information*

ABSTRACT: Correct protonation of titratable groups in biomolecules is crucial for their accurate description by molecular dynamics simulations. In the context of constant pH simulations, an additional protonation degree of freedom is introduced for each titratable site, allowing the protonation state to change dynamically with changing structure or electrostatics. Here, we extend previous approaches for an accurate description of chemically coupled titrating sites. A second reaction coordinate is used to switch between two tautomeric states of an amino acid with chemically coupled titratable sites, such as aspartate (Asp), glutamate (Glu), and histidine (His). To this aim, we test a scheme involving three
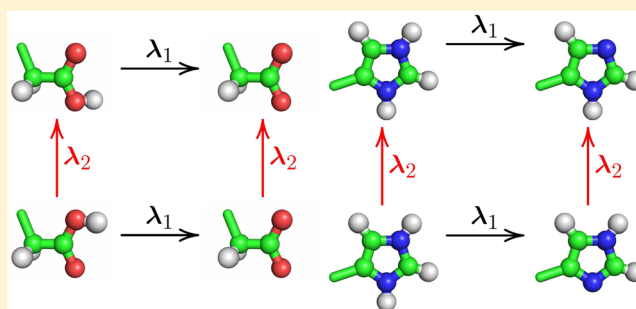
protonation states. To facilitate charge neutrality as required for periodic boundary conditions and Particle Mesh Ewald (PME) electrostatics, titration of each respective amino acid is coupled to a "water" molecule that is charged in the opposite direction. Additionally, a force field modification for Amber99sb is introduced and tested for the description of carboxyl group protonation. Our three states model is tested by titration simulations of Asp, Glu, and His, yielding a good agreement, reproducing the correct geometry of the groups in their different protonation forms. We further show that the ion concentration change due to the neutralizing "water" molecules does not significantly affect the protonation free energies of the titratable groups, suggesting that the three states model provides a good description of biomolecular dynamics at constant pH.

1. INTRODUCTION

In proteins and other biological macromolecules, the charges of the protonatable amino acid residues contribute crucially for determining the spatial distribution of the electrostatic potential. Therefore, changes in the titration states of the residues strongly couple to protein conformational transitions and dynamics. Both the electrostatics and the dynamics are crucial determinants for enzymatic catalysis, protein–protein interactions, substrate binding, and other biologically relevant processes. Moreover, correct protonation states of the titratable residues are essential for a proper description of the entropic and enthalpic properties of the system. If the electrostatic environment of such a residue changes due to a conformational change of the protein, a different protonation state may become more favorable, thus causing the amino acid to change charge. However, in conventional force field based molecular dynamics (MD) simulations, the charges of titratable amino acids are assigned at the beginning of the simulation and are subsequently kept constant. This model of fixed charges does not account for the charge change of a protonatable group caused by the change of its electrostatic environment.

One way of addressing this issue in MD simulations is to include the reaction coordinate of the protonation of each titratable group as an additional degree of freedom of the system, commonly referred to as λ, which describes the change between a protonated and a deprotonated state in terms of a virtual particle.[1−3,6,8,9] A simple two states model cannot describe systems with two and more chemically coupled titratable sites. For histidine (His) or amino acid residues containing carboxyl groups, such as glutamate (Glu) or aspartate (Asp), two or more λ-coordinates are required to describe the correct titration behavior of the group.

This problem has previously been approached by adding a second reaction coordinate. Donnini et al.[6] introduced a second (de)protonation coordinate for each of the two chemically coupled sites of a histidine, and the coupling is explicitly accounted for in the Hamiltonian of the system. When applied to carboxyl groups, however, the drawback of this approach is that it includes the nonphysical double protonated form. An

alternative approach has been proposed earlier for implicit solvent[3] and later for explicit solvent cutoff electrostatics simulations[8,9] and Generalized Born based explicit and hybrid solvent simulations,[4,5] where one of the λ-coordinates describes the amino acid protonation, whereas the other determines which of the two chemically coupled sites is protonated. In practice, this approach reduces the protonation space from four to three states.

Here, we extend this scheme to explicit solvent simulations with Particle Mesh Ewald (PME) electrostatics,[6] while at the same time, we benefit from the advantage that no *a posteriori* corrections are required to obtain the pK_a of the titrated sites. Similar to Khandogin and Brooks,[3] the microscopic pK_a values of the chemically coupled sites are required, but in our approach different pH correction is used.

Similarly to previous approaches,[12−14] we preserve the total charge of the system by coupling the titrating degree of freedom, λ, to a distant "water" molecule, the charge of which is changing in the opposite direction. Maintaining a neutral charge during constant pH MD is particularly important when PME electrostaticsis is used.[10,11,43] The approach of coupling a protonatable site to a counterion or a "water" molecule is used in free energy MD[12] as well as in constant pH MD[13,14] simulations. An additional advantage of using a three states model over four states[6] is that one of the λ-coordinates switches between the two tautomers, and therefore no net charge needs to be compensated by a counter charge. Thus, the number of coupled "water" molecules or counterions is reduced which is particularly advantageous when the system contains multiple titratable groups.

For ease of notations, we here describe this approach mainly for amino acids with carboxyl groups. Because the double deprotonated form of His has a very high pK_a value, which is not observed under physiological conditions, the three states model can also be used for His in a straightforward manner. We note that the two chemically coupled titrating sites of Asp and Glu have identical pK_a values, whereas in His the two microscopic pK_a values differ by approximately 0.4 pK_a units.

To assess the accuracy of the three states model in explicit solvent simulations, we titrated Glu, Asp, and His residues and calculated the microscopic pK_a values of each titrating site as well as the macroscopic pK_a value of the entire group. In addition, we further extended the λ-dynamics approach in explicit solvent simulations with the aim of improving its accuracy, by (a) introducing a new nonlinear pH correction term in the Hamiltonian for constant pH MD and (b) improving the calculation of the force field correction term. Moreover, we suggest an improved estimate for the pK_a error obtained from constant pH MD by using the pK_a probability distribution. Further, we adapted the bonded parameters of the carboxyl group such that they can be used with the three states model.

2. METHODS

Here, we describe all methodological advances and design decisions that in total allow a thermodynamically consistent description of compounds with two chemically coupled titrating sites using a three states model. All methods described here were implemented within the constant pH MD module[6,7] for the MD simulation software package GROMACS-3.3.3.[17,21,22]

2.1. Constant pH MD Simulations with the λ-Dynamics Approach. For constant pH MD simulations, we

followed the previously developed protocol,[6] where the Hamiltonian of the system is defined as

$$H = (1 - \lambda)H_A + \lambda H_B + f(\lambda)RT \ln 10(pK_{a,ref}^{exp} - pH) + V_{MM}^{corr}(\lambda) + U(\lambda) + \frac{m}{2}\dot{\lambda}^2 \tag{1}$$

The λ-particle moves between 0 and 1, thereby interpolating the Hamiltonian of the system between protonated (A) and deprotonated (B) states. $pK_{a,ref}^{exp}$ is the experimental pK_a value of the reference state of the titrated compound (typically the free amino acid in water). The pH correction term is a function of λ ($f(\lambda)$) which describes the contribution to the deprotonation free energy of the titratable site in solution from (i) bond breaking and formation and (ii) pH. V_{MM}^{corr} describes the force field contribution to the deprotonation free energy of the reference state of the titratable site, and it is subtracted from the Hamiltonian, allowing a realistic description of the deprotonation equilibria. The purpose of the biasing potential $U(\lambda)$ is to increase the populations of λ near 0 and 1 relative to intermediate nonphysical values; m and $\dot{\lambda}$ are the mass and the velocity of the λ-particle, respectively.

The above formalism can be applied only to titrating systems where the titratable groups only interact via Coulomb interactions but are otherwise chemically uncoupled. Therefore, groups with more than two protonation forms, such as carboxyl groups or His residues, cannot be described within this limited framework. Furthermore, even if the coupled sites are symmetric, as the two oxygen atoms of a carboxyl group, considering only one of them for protonation with a two states model would result in an artificially high barrier between protonated and deprotonated states due to additional sampling required in Cartesian space. For these reasons, the ability that both sites protonate is of particular importance when the movement of the titrating compound is sterically restricted.

2.2. Three States Model. Figure 1 A shows the two titratable atoms of a carboxyl group. The λ_1 coordinate titrates the group, whereas the λ_2 coordinate switches between the tautomers.

Upon deprotonation, two initial reactants result in the same end product (Figure 1 B), and the Hamiltonian in eq 1 reads

$$H = (1 - \lambda_1)[(1 - \lambda_2)H_A + \lambda_2 H_C] + \lambda_1 H_B \tag{2}$$

where H_A and H_C are the Hamiltonians of the two protonated states, and H_B is the Hamiltonian of the deprotonated state. Here, the pH correction is given by

$$V_{pH,corr} = f(\lambda_1)RT \ln 10(pK_{a,ref}^{exp} - pH) + f(\lambda_2)RT \ln 10(pK_{a_2} - pK_{a_1}) \tag{3}$$

where $pK_{a,ref}$ is the experimental pK_a of the reference state of the titrated compound. This pK_a value is the macroscopic pK_a of the compound and is calculated as

$$K_a = \frac{1}{\frac{1}{K_{a_1}} + \frac{1}{K_{a_2}}} \tag{4}$$

and therefore

$$pK_a = \log(10^{pK_{a_1}} + 10^{pK_{a_2}}) \tag{5}$$

irrespective of which oxygen carries the proton. In the above equations, pK_{a_1} and pK_{a_2} are the microscopic pK_a values of the two individual oxygen atoms, and K_{a_1} and K_{a_2} are the microscopic equilibrium constants for the two oxygens,
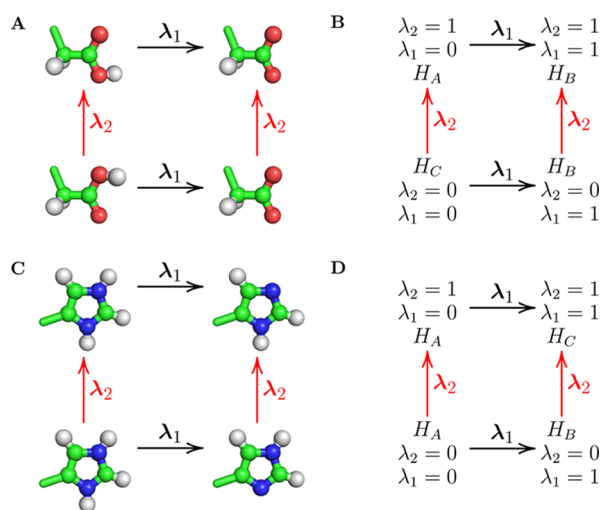
Figure 1. Three states model describing the protonation states of amino acid residues with carboxyl groups (upper row) and of His (lower row). Panels A and C show different protonation states depicted by the different positions of the hydrogen atom at the two oxygen or nitrogen atoms. B and D show the corresponding Hamiltonians of the individual states. The titrating reaction coordinate (λ_1) is depicted with black arrows and the switching one (λ_2) with red arrows. For residues with carboxyl groups (panel B), H_A and H_C are the Hamiltonians of the two protonated states, and H_B is the Hamiltonian of the deprotonated state. In case of His (panel D), H_A is the Hamiltonian of the protonated state, and H_B and H_C are the Hamiltonians of the two deprotonated states.

respectively. For carboxyl groups, the two titrating sites are symmetric, their microscopic pK_a values are identical, and $pK_a = pK_{a_{1,2}} + \log 2$.

For His, there are two nitrogen atoms in the imidazole ring which can be protonated as depicted in Figure 1 C. Here, in contrast to residues with carboxyl groups, the two deprotonated sites are different, hence the Hamiltonian of the system is given as

$$H = (1 - \lambda_1)H_A + \lambda_1[(1 - \lambda_2)H_B + \lambda_2 H_C] \tag{6}$$

where H_A is the Hamiltonian of the protonated state, and H_B and H_C are the Hamiltonians of the two deprotonated states. The pH correction potential is the same as defined in eq 3 for the carboxyl group. Here, the relation between the macroscopic and microscopic equilibrium constants is defined as $K_a = K_{a_1} + K_{a_2}$, and for the macroscopic pK_a one obtains

$$pK_a = -\log(10^{-pK_{a_1}} + 10^{-pK_{a_2}}) \tag{7}$$

2.3. Force Field Correction Term. The force field correction term V_{MM}^{corr} in eq 1 was obtained from the free energy profile of deprotonating the titratable site in solution. The free energy profiles were obtained by performing constant pH MD simulations at fixed values of the λ_1 and λ_2 coordinates. In these simulations, λ was increased from 0 to 1 in 10 steps of 0.1. Additionally, the free energy values for $\lambda_1, \lambda_2 = -0.05$ and $\lambda_1, \lambda_2 = 1.05$ were also included to cover the full range spanned by the λ-particles (See Section 2.5). The derivatives of the potential energy ($\partial H/\partial \lambda_1$, $\partial H/\partial \lambda_2$) were collected at discrete (λ_1/λ_2) positions and integrated to yield a two-dimensional free energy profile. To improve the numerical accuracy of the free energy profile, it was calculated both, by first integrating over λ_1 and then λ_2 as well as *vice versa*. Both results were then used to

construct a single plane for the two λ-particles. We note that, as shown by Hünenberger,[15] the $\partial H/\partial \lambda$ slopes are nonlinear. Specifically, we found that in explicit solvent PME simulations the plane ($\tilde{G}$) was approximated most accurately by a fifth-order polynomial function

$$G(\lambda_1, \lambda_2) \approx \tilde{G} = \sum_{r=0}^{r=5} \sum_{s=0}^{s=5} a_{rs}\lambda_1^r\lambda_2^s \tag{8}$$

The parameters were obtained by minimizing the squared difference D between the free energy profiles obtained from integrating via λ_1 and λ_2

$$
\begin{aligned}
D = &\sum_{i=1}^{n-1} \sum_{j=1}^{n} [\tilde{G}(\lambda_{1(i+1)}, \lambda_{2(j)}) - \tilde{G}(\lambda_{1(i)}, \lambda_{2(j)}) - \Delta G_{j,i\to i+1}]^2 \\
&+ \sum_{i=1}^{n} \sum_{j=1}^{n-1} [\tilde{G}(\lambda_{1(i)}, \lambda_{2(j+1)}) - \tilde{G}(\lambda_{1(i)}, \lambda_{2(j)}) - \Delta G_{i,j\to j+1}]^2
\end{aligned}
\tag{9}
$$

with respect to each coefficient. G denotes the free energy calculated from the simulations, and the i and j indices denote the positions of λ_1 and λ_2 along the titrating and switching coordinate, respectively. Thus, in addition to improved statistics, a single free energy surface for both λ-particles is defined, ensuring the energy conservation in the titrated system.

2.4. Neutralizing the System Net Charge upon Titration. To keep the system neural upon charging a titrating residue, similar to the work of Shen,[13,14] we coupled the titration coordinate to a "water" molecule that can switch between a positive hydronium ion and a neutral state (Figure 2), thereby compensating the net charge of the titratable site.
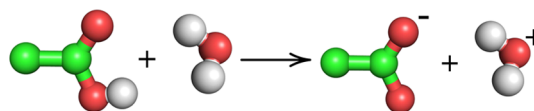
Figure 2. Coupling scheme of an amino acid residue with a carboxyl group to neutralizing a "water" molecule. Shown is the charge transfer from the carboxyl group to the coupled "water" molecule. Only the carboxyl group and the adjacent C atom are shown.

The protonation of the coupled "water" molecule contributes to the deprotonation free energy of the residue it is coupled to. This contribution therefore was included within the V_{MM} correction potential in eq 1. The underlying assumption of this approach is that this contribution is the same in the reference simulations (V_{MM}) as in the protein. To ensure that the environment of the coupled "water" was the same in the two simulations, this "water" molecule was kept far from the protein by restraining its position.

To test if the limited diffusion of the coupled "water" molecule and the change of the solution ionic concentration due to the appearance of a charged "water" molecule affect the deprotonation free energy, test simulations were performed as described in Section 2.10.2.

2.5. Biasing Potential. To ensure that the system stays most of the time in physical states, i.e., $\lambda \approx 0$ (protonated) and $\lambda \approx 1$ (deprotonated), we included a double-well potential as described in ref 7. The form of the biasing potential allows fluctuations of the λ-particle outside the $0 > \lambda < 1$ interval. The aim of the extended interval was to achieve a λ-distribution, the

respective averages of which are at $\lambda \approx 0$ and $\lambda \approx 1$, respectively. The height of this barrier of this potential also serves to control the transition rate between the two states.

2.6. pH Correction Potential. The pH correction potential describes the free energy difference between the protonated and deprotonated form of the reference state of the titratable site. It is a function of λ ($f(\lambda)$ in eq 1) which adds $RT \ln 10(pK_a - pH)$ kJ·mol^{-1} to the deprotonated state Hamiltonian.

In this work, we implemented a new pH correction function ($f(\lambda)$), which is a sigmoidal spline (Figure 3), whose values are
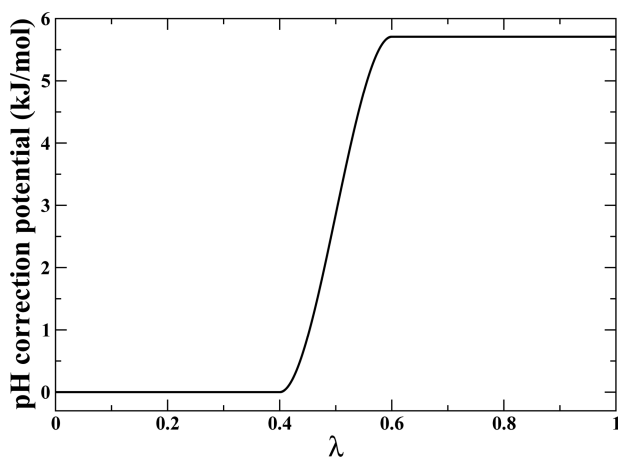

Figure 3. Form of the pH correction function $f(\lambda)$ for a difference of $pK_a - pH = 1$.

constant toward the end of the λ-interval. The sigmoidal form was chosen for two reasons. First, the constant parts of the correction potential in the vicinity of physical protonation states, $\lambda = 1$ and $\lambda = 0$, avoid artificial perturbation of these states. Second, the smooth transition between $\lambda = 0.4$ and $\lambda = 0.6$ avoids numerical instability.

2.7. pK$_a$ Calculation and Error Estimation. Here, we also developed an improved scheme for pK$_a$ calculation and error assessment from constant pH simulations based on our previous approach for protonation and deprotonation rate constant estimation.[6] The statistical accuracy of the calculated pK$_a$ values is now assessed from significance intervals obtained from a distribution of pK$_a$ estimates. This distribution was obtained via a bootstrapping process as follows. Multiple titration simulations were performed at different pH values (see Section 2.10), and the probability distribution $p(\lambda)$ for each pH value was estimated from the probability of having a certain combination of deprotonation (k_0) and protonation (k_1) rate constants ($p(k_0, k_1)$)

$$p(\lambda) = \int\int \delta\left(\lambda - \frac{k_0}{k_0 + k_1}\right) p(k_0, k_1) dk_0 dk_1 \tag{10}$$

The accuracy of the rate constant estimation is defined by the number of deprotonation and protonation events, whereas the times spent in the respective protonation states define the ratio of the constants. The probability of a certain k_0, k_1 combination, given an observed number n_0 and n_1 of deprotonation and protonation events, respectively, is[6]

$$p(k_0, k_1) = \frac{T_0^{n_0+1} T_1^{n_1+1}}{n_0! n_1!} k_0^{n_0} k_1^{n_1} \exp[-T_0 k_0 - T_1 k_1] \tag{11}$$

where T_0 and T_1 are the times spent in the protonated or deprotonated state, respectively. The λ-distribution for every pH was calculated from eqs 10 and 11 and was used to generate a set of 20000 λ-values for each pH. Values that fall outside the 95% confidence interval of the λ-distributions were excluded from the sets. A random λ-value was picked from each of the sets, a Henderson−Hasselbalch fit was performed, and a pK$_a$ value calculated. This process was repeated 50000 times yielding a pK$_a$ probability distribution. The error estimate of the pK$_a$ values was finally derived from the 95% confidence interval of the obtained distributions. The pK$_a$ itself was calculated as pK$_a = \sum p_i pK_{ai}$, where p_i is the probability of a certain pK$_{ai}$.

To avoid including nonphysical intermediate values of λ in the statistics described above, only snapshots where either $\lambda < 0.1$ or $\lambda > 0.9$ were considered for pK$_a$ calculation. The fraction of used snapshots from the different titration simulations, in this work, ranged from ca. 72% to ca. 75% of the overall snapshot count for the respective simulation.

The pK$_a$ values served as an estimate of how accurately the force field correction term (V_{MM}) describes the two-dimensional free energy landscape of the three states model. To evaluate the accuracy of the V_{MM} calculation procedure and if the three states model correctly describes the protonation behavior of the two chemically coupled titrating sites, separate pK$_a$ calculations were carried out for the entire carboxyl group and for each of the two oxygen atoms, respectively. For the microscopic pK$_a$ values of the two oxygen atoms, a state was considered as protonated only when the proton was present on the respective oxygen whose pK$_a$ was calculated. The macroscopic pK$_a$ calculation of the entire group was based on the overall time spent in the protonated and deprotonated state for the entire group, irrespective of which oxygen atom was protonated. To assess if the three states model can reproduce the correct partitioning among the protonation forms (for Asp and Glu), we tested if the macroscopic and the microscopic pK$_a$ values agree with the ones obtained via eqs 5 and 7. The same approach was applied to the His residue.

2.8. Force Field. *2.8.1. "Chimeric" Carboxyl Group.* Ideally one would have to use two different sets of bonded and Lennard-Jones (LJ) parameters, respectively, for protonated and deprotonated states of the carboxyl group. However, in our constant pH MD implementation for chemically coupled titratable sites,[6] all bonded and nonbonded parameters remain unchanged and correspond to the protonated state, except for the partial charges which change appropriately upon protonation. We have shown that the contribution from the bonded interactions for a two states titratable site is small.[6]

For carboxylic acids, conventional force fields describe a protonation state where only one of the oxygens is protonated. Moreover, in the standard Amber99sb force field[24] used here, the two oxygen atoms are not equivalent due to the fact that one carries a proton. Because we intend to consider both oxygens for titration, we added a "chimeric" residue to the Amber99sb force field. This "chimeric" residue has two equivalent oxygen atoms with bonded and LJ parameters of a protonated oxygen. As described above, these parameters were not changed during the course of the constant pH MD simulations. One hydrogen atom was bound to each of the oxygen atoms. The switching λ-coordinate ensures that only one hydrogen atom can carry a charge at a time. Figure 4 illustrates the bond angles of the "chimeric" group and of a standard protonated carboxyl group. Note that the sum of the angles is now decreased by 10° compared to the original force
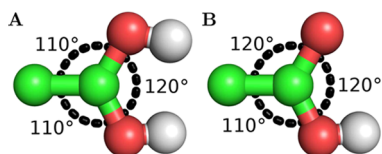
Figure 4. Carboxyl group angle comparison. The C−O angles of the **A** "chimeric" and **B** single protonated (original force field) carboxyl groups differ by 10°.

field. To test how this change affects the planarity of the group, we calculated the distribution of the dihedral angle, which is defined by the carbon atom of the side chain, the carboxyl carbon atom, and the two oxygen atoms in the "chimeric" and the standard carboxyl group (see Subsection 3.3).

Another issue that affects constant pH MD simulations of carboxyl groups is related to the *syn* and *anti* conformations of the titratable proton (Figure 5). The experimental free energy of the *anti* conformation, measured in solid-state NMR,[3,18] is ca. 11.72 kJ·mol^{-1} higher than that of the *syn* conformation. In quantum calculations,[20] the *syn* conformation is shown to be more favorable by ≈4.60−7.53 kJ·mol^{-1}. Therefore, the *anti* conformer is rarely observed. However, during constant pH MD, upon deprotonation, the hydrogen atom remains bound to the carboxyl group but has no charge. Figure 5 B shows that the torsion potential favors the *anti* conformation by about 15 kJ·mol^{-1}. In the deprotonated state, therefore, the *anti* conformation is expected to become more favorable and is indeed frequently observed in constant pH MD simulations.

The potential depicted in Figure 5 A was calculated by rotating the hydrogen atom in an aspartic carboxyl group in vacuum in 10° steps. The potential energy with contributions from the entire force field was obtained for all rotamers at the 0-th step of a one step MD simulation. In contrast, the

potentials shown in Figure 5 B were analytically calculated from the force field parameters. The dihedral potentials in the Amber99sb port[25,26] used in this work (see Section 2.8.2) are defined as Ryckaert-Bellemans functions.[19] Two dihedral potentials define the orientation of the hydrogen atom in the carboxyl group as shown in Figure 5 B. Due to the different force field definition of the 0-th position of the two angles, the $V_{C-C-O-H}$ potential, depicted in black, was shifted by 180° in the figure to enable direct comparison of both potentials and their minima

$$V_{O-C-O-H} = 27.196 - 7.9496 \cos \psi - 19.2464 \cos^2 \psi \tag{12}$$

$$V_{C-C-O-H} = 19.2464 - 19.2464 \cos^2 \psi + 180° \tag{13}$$

where ψ is the dihedral angle of the bond defined by the denoted atoms.

The occurrence of the *anti* conformation can introduce artificial bias in the simulation. This potential artifact originates from the fact that the deprotonation free energy of the *anti* conformer can be rather different from that of the *syn* one. Typically, the deprotonation of the *anti* conformer is not considered when computing the V_{MM} correction term in the Hamiltonian of constant pH MD (eq 1). The difference between the effective potential of the system and the V_{MM} term (eq 1) therefore does not describe the *anti* conformation, which occurs during constant pH MD, thereby introducing a bias. In practice, to avoid occurrence of the *anti* conformation, we increased the energy of the *anti* orientation by altering the $V_{C-C-O-H}$ torsion potential

$$V_{C-C-O-H} = -19.2464 \cos \psi \tag{14}$$
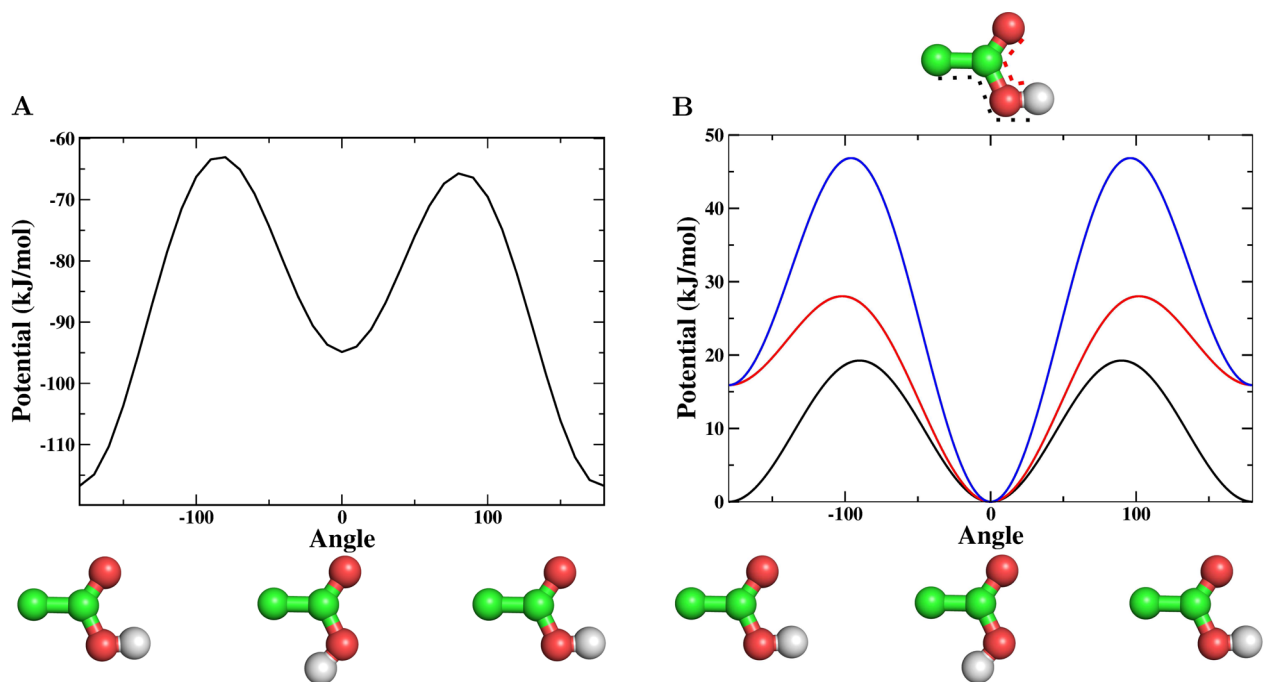
Figure 5. Amber99sb force field rotation potentials of carboxyl H atom. Potential energy profiles in vacuum along the C−C−O−H dihedral angle of Asp. **A**. All bonded and nonbonded contributions were taken into account. **B**. Potential energy contributions from only the C−C−O−H and O−C−O−H (black and red dotted lines in the structure representation above the graph, respectively) dihedral potentials. The sum of the two dihedral potentials is depicted as the blue line. In both panels, the corresponding rotamers are depicted below the graph in ball and stick representations.

151
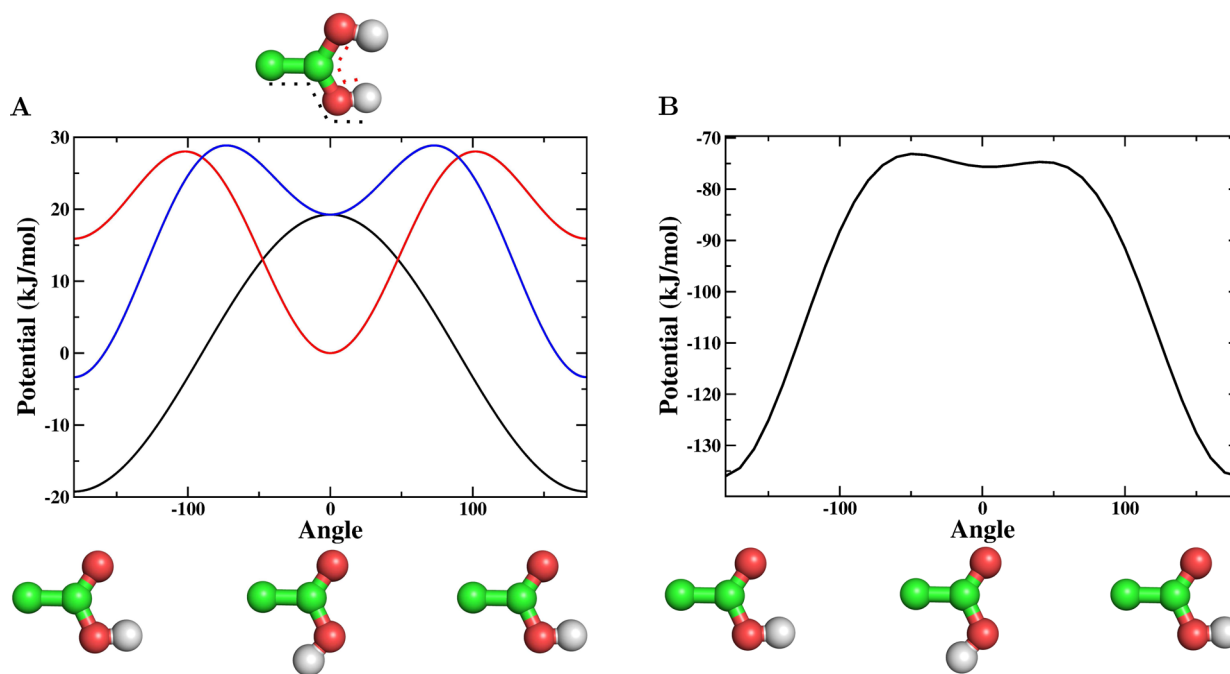
Figure 6. Altered rotation potentials of carboxyl H. **A**. The altered dihedral potential of the C−C−O−H angle (eq 14) is depicted as a black line, the O−C−O−H angle (eq 12) is depicted as a red line, and their sum is depicted as a blue line. **B**. The potential of Asp in vacuum with altered dihedral potential as a function of the C−C−O−H dihedral angle. The corresponding rotamers are plotted below the graphs, and the black and red dotted lines of the chemical structure in the above left graph depict the dihedral angles whose colors correspond to the colors of the potentials.

Figure 6 A and B shows the resulting dihedral potential and the full force field potential including the coulomb interactions, respectively.

The dihedral potentials and the charge interactions described so far for Amber99sb would result in a similar behavior of the hydrogen atom in other force fields such as GROMOS 53A6.[34] We note that in our standard MD simulations in solution the hydrogen atom of the protonated standard Amber99sb Asp assumes the *anti* conformation during the simulation. After the transition, it never (or very rarely) samples the *syn* orientation again (contrary to the experimental data or other force fields such as GROMOS 53A6 or OPLS[33]). The procedure for limiting the *anti* orientation of the deprotonated hydrogen during constant pH MD, which we introduced in this work, eliminates this abnormally high occupancy of the high energy conformer for the protonated form of the group as well. A similar approach can be applied to those cases as well.

In cases of strong interactions between the carboxyl hydrogen and its surrounding, the high energy *anti* orientation can occur.[35] One approach for including the *anti* conformer in the current constant pH MD implementation of chemically coupled sites is to perform the reference free energy simulations with the hydrogen atom oriented in the *anti* position. Favoring this orientation can be achieved by altering the dihedral potentials in eqs 12 and 13.

2.8.2. Water Model, Hydronium, and Titratable Amino Acids. The Amber99sb[24] force field used in the simulations was ported[25,26] to GROMACS 3.3.3 together with the SPCE[27] water model. Hydrogen virtual sites[28] with a 4 fs time step were used.

Transforming a neutralizing "water" molecule into a hydronium ion requires the appearance of a third hydrogen atom on it. To avoid the technical problems arising from this process, we used the geometry of a regular SPCE water model instead of a full hydronium ion. Atomic charges of this positively charged "water" molecule were calculated with Gaussian03[42] using a Hartree−Fock level of theory with a 6-31G* basis set. Even though the positivity charged "water" is not a physical model for a hydronium ion, it fulfills the requirement of compensating for the charge of the titratable site, it is coupled to, and it has the advantage of being a very simple model. In principle, any titratable site can be used to maintain neutrality of the system. Further, to avoid strong electrostatic interactions, the coupled "water" was kept far from its counterpart; thus, the geometry of the "hydronium" ion is not important because it interacts only with bulk solution molecules.

Following the work of McCammon,[40] the backbone charges were kept unchanged irrespective of the protonation state of the side chain. However, in our model, the backbone atoms had the same charge as the backbone atoms of the deprotonated form for Asp and Glu and the protonated form for His. Changing only the side chain charges during titration results in excess charge due to the fact that the charges do not sum up to 0 in the protonated form of Asp and Glu and the deprotonated form of His. The additional charge required for summation to a whole charge was distributed equally among the side chain atoms (see the Supporting Information for all atomic charges). The backbone charges of the protonated and the deprotonated form of the residues were very similar. Therefore, after the excess charge was divided over all side chain atoms, the charge change per atom was very small. For all residues in this work, the largest excess charge per atom was observed for Asp where each side chain atom received +0.021 additional charge (see Table S1). This approach was chosen instead of a quantum mechanical calculation of the charges, because we wanted to stay as close as possible to the original Amber99sb charges due to the fact that the force field had been optimized to work with them. The interactions between the two hydrogen atoms in the "chimeric" carboxyl group were excluded. The bonded and the

LJ parameters of neither the carboxyl nor the imidazole groups were changed during constant pH MD. The parameters used for the His were the ones of its double protonated form.

2.9. Model Compounds and Reference Free Energy Simulations. The force field correction term V_{MM} in the Hamiltonian in eq 1 is the deprotonation free energy of the titratable site in solution, as reference state of the titratable site. The reference or model compound was the amino acid free in solution with -C and -N termini capped with methyl groups. The caps were linked to the termini via peptide bonds thus modeling an amino acid residue in the protein backbone.

All reference free energies were calculated via constant pH MD simulations[6] during which the $\partial H/\partial \lambda_1$ and $\partial H/\partial \lambda_2$ were collected at different values of λ_1 and λ_2. The free energy profiles along the two λ-coordinates were calculated in 2D in 13 discrete steps by integrating the $dH/d\lambda$ (see Section 2.3). Both the titrating residue and the coupled "water" molecule were included in the reference simulations so that the contribution of the "water" charge change to the deprotonation free energy of the model compound was appropriately included.

2.10. Simulation Details and Ion Test Simulation Setups. *2.10.1. Simulation Details.* All free energy and titration simulations were performed using the software package GROMACS 3.3.3[17,21,22] together with the constant pH MD module.[6] The test simulations of the "chimeric" carboxyl group were performed with GROMACS 4.5.7.[23] All simulations were carried out at 300 K using the Nosè-Hoover thermostat[36,37] and at 1 bar pressure using the Parrinelo-Rahman barostat.[38,39] Bond lengths of water molecules, with the exception of the protonatable "water", were constrained with SETTLE,[30] whereas all other bonds were constrained using the LINCS constraint algorithm.[29] PME electrostatics[31,32] was used, with a cutoff of 1 nm for the direct Coulomb interactions and 0.12 nm spacing for the Fourier grid. The λ dependence of the Ewald energy was calculated using eq 24 in our previous work.[6] At every step we calculated the potentials of the protonated and deprotonated Hamiltonians of each site, while the charges of all other sites were obtained by linear interpolation between the charges of their protonated and deprotonated forms along the corresponding λ-coordinates. In this work, chemically coupled titratable sites Asp, Glu, and His were considered which were described using three states. In this case, at every step, we calculated the potentials of the three Hamiltonians in eqs 2 and 6 and obtain the forces acting on each of the two λ-particles using eqs 20 and 21 from our previous work.[6] For each potential an Ewald grid[31,32] was evaluated. A cutoff of 1 nm was used for the LJ potential. The height of the biasing potential was set to 10 kJ·mol^{-1}. The mass of the λ-particle was set to 20 u, and its temperature was kept at 300 K by coupling it to a heat bath via the Andersen thermostat[6] with a coupling constant of 6 ps^{-1}.

All titration simulations were performed on the model compounds of the amino acids used for the reference free energy calculations. The reference and constant pH MD simulations were carried out with an identical setup: the model compound and the coupled "water" were kept at a distance, during both reference and constant pH MD simulations, of ≈ 6 D lengths (ca. 4.7 nm) for the respective salt concentration (0.15 mol·L^{-1}) to minimize their interaction. The simulations were carried out in a cubic simulation box with 6 nm side length. To both the C-α atom of the amino acid and the oxygen atom of the coupled "water", a harmonic potential with a force constant of 1000 kJ·mol^{-1}·nm^{-2} was applied, preventing

conformational sampling bias by allowing free rotation of all other bonds.

All reference free energy simulations were performed using constant pH MD simulations with fixed λ-values. The simulation lengths were 30 to 40 ns per λ-point.

The titrations of the amino acids were performed in 0.25 pH unit steps in intervals symmetrical to their pK_a values. In the case of Asp, the titration interval started from pH = 1.5 to pH = 6.5, in the case of Glu, the titration interval started from pH = 1.5 to pH = 6.75, and in the case of His, the titration interval started from pH = 3.75 to pH = 8.75. The space between the last two pH points at both ends of the pH interval was increased to 0.5 pH units. To reduce statistical error, ten simulations for Asp, six for Glu, and eight for His were performed independently for each pH value totalling ca. 300 ns for Asp and ca. 200 ns for Glu and His.

The reference macroscopic pK_a values for Asp and Glu, 4.00 and 4.40, respectively, were taken from ref 16 implying microscopic values of $pK_a - \log 2$. The macroscopic reference pK_a value of 6.38 for His was taken from ref 41 with microscopic pK_a values of 6.53 and 6.92 for N_δ and N_e, respectively.

Each simulation system contained 20 Na$^+$ and 20 Cl$^-$ ions (19 Na$^+$ for the His simulations system), corresponding to a physiological salt concentration of 0.15 mol·L^{-1}.

2.10.2. Effect of the Ions on the Deprotonation Free Energy. The appearance of neutralizing, positively charged "water" molecules during constant pH MD simulations changes the number of ions in the simulation box. To test to what extent this shift of the ion concentration affects the free energy of deprotonation, we performed free energy calculations of Asp in water with increasing ion concentrations from 0 to 1 mol·L^{-1} in 0.1 steps. The step size was decreased near physiological concentration (0.15 mol·L^{-1}), and values of 0.11, 0.12, 0.13, 0.14, and 0.15 mol·L^{-1} were also included. Furthermore, to estimate the convergence of the deprotonation free energy, all calculations were carried out using simulation lengths of 1, 5, 20, and 40 ns.

We also tested how the lowered diffusion of the neutralizing "water" molecules due to the restraint would affect their interaction with the titrating sites. To estimate that effect, we calculated the deprotonation free energy of Asp as a function of ion concentration (from 0.2 to 1 mol·L^{-1}) in a setup with a concentration of freely diffusing ions of 0.15 mol·L^{-1}. The additional ions of the respective concentrations were kept at fixed positions as far as possible away from each other and from the deprotonating residue, thus modeling fixed charges appearing during constant pH MD. The movement of the fixed ions was restrained via a harmonic potential with a 1000 kJ·mol^{-1}·nm^{-2} spring constant. In all these test simulations described above, the deprotonating Asp was not coupled to a "water" molecule to ensure a constant ion concentration throughout the simulation.

The effect of the distance between the neutralizing "water" molecule and the titrating residue on the deprotonation free energy was tested in a separate set of simulations where the residue was coupled to a "water" molecule that was kept at distances of 1, 1.5, 2, and 4 nm using an umbrella potential with a 1000 kJ·mol^{-1}·nm^{-2} spring constant.

The free energies in all ion test simulations were calculated using discrete thermodynamic integration (TI) where the value of the reaction (λ) coordinate was changed from 0 to 1 with step size of 0.1. The simulation box size was identical to one of
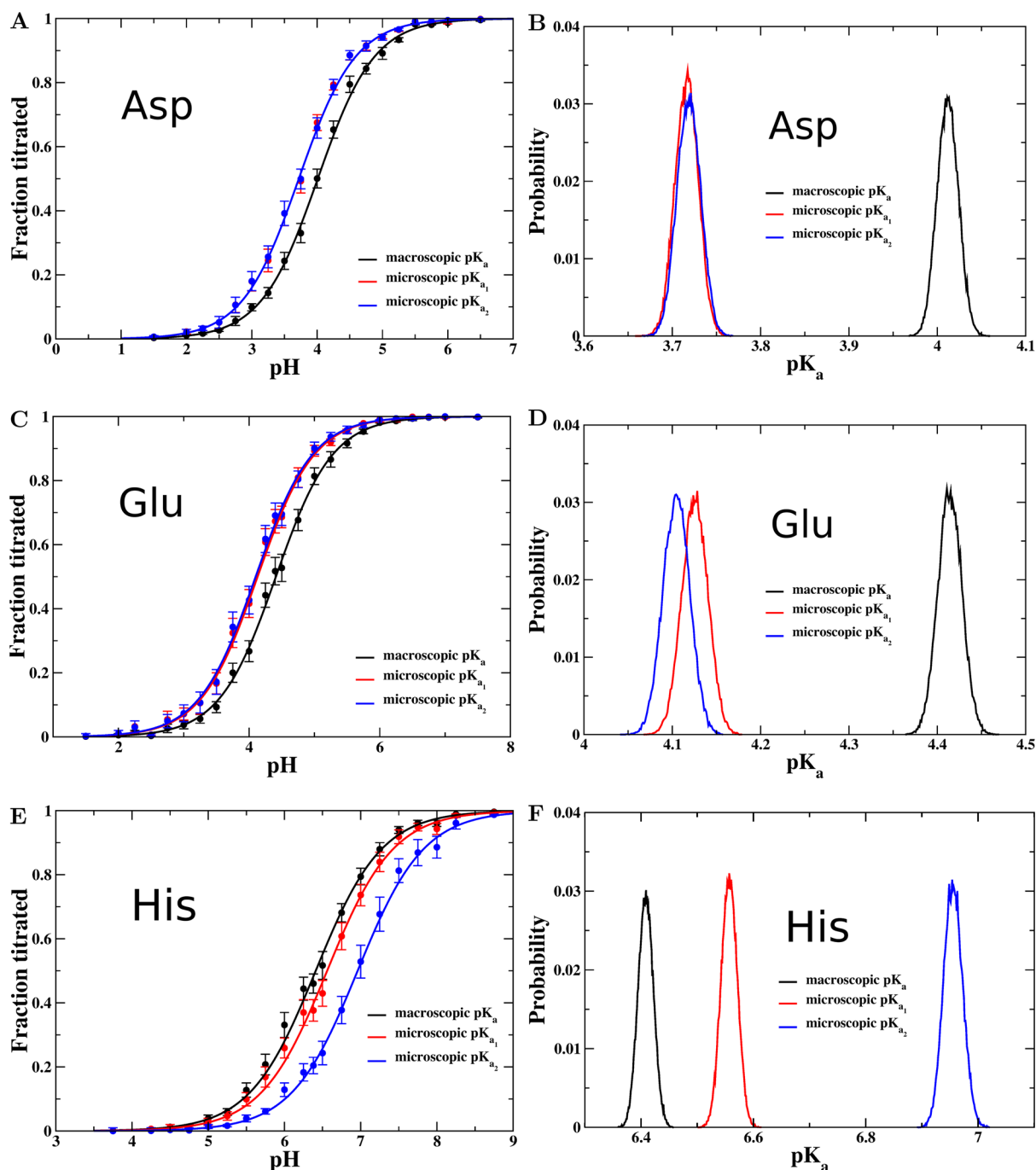
Figure 7. Titration simulations results. The macroscopic (black) and microscopic (red and blue) titration curves of Asp, Glu, and His are depicted in panels A, C, and E, respectively. The error bars of the titrated fractions were obtained from the λ-distribution for the given pH (see Methods section 2.7). The calculated macroscopic pK_a values were 4.01, 4.41, and 6.41 for Asp, Glu, and His, respectively. The two microscopic pK_a values were for Asp 3.72 and 3.72, for Glu 4.13 and 4.10, and for His 6.56 for N_δ and 6.96 for N_ϵ. Panels B, D, and F show the microscopic (red and blue) and macroscopic (black) pK_a probability distributions of Asp, Glu, and His, respectively, obtained from multiple Henderson-Hasselbalch fits as described in the Methods section 2.7.

the titration simulations. The $d\mathrm{H}/d\lambda$ values, obtained during the first 50 ps of every simulation, were not used in the free energy calculations.

3. RESULTS AND DISCUSSION

3.1. pK_a Calculations. To test if the three states model accurately describes the microscopic titration behavior of Asp, Glu, and His, specifically the Hendersen-Hasselbalch curves of both the two chemically coupled titratable sites as well as of the

entire group, we performed pH titration simulations on the model compounds of Asp, Glu, and His and calculated their microscopic and macroscopic pK_a values.

Further, as the reference and the simulated systems are identical, it is expected that the calculated pK_a values match exactly the reference ones. The shift from the reference pK_a values serves as an estimate of how accurately our force field correction procedure describes the two-dimensional free energy surface of chemically coupled titratable sites.

Table 1. Reference and Constant pH MD Calculated pK_a Values of Asp, Glu, and His

	Asp		Glu		His	
	constant pH MD	reference	constant pH MD	reference	constant pH MD	reference
macro pK_a	4.01 ± 0.02	4.00	4.41 ± 0.03	4.40	6.41 ± 0.03	6.38
micro pK_{a_1}	$3.72 + 0.03 - 0.02$	3.70	4.13 ± 0.03	4.10	6.56 ± 0.03	6.53
micro pK_{a_2}	$3.72 + 0.03 - 0.02$	3.70	4.10 ± 0.03	4.10	$6.96 + 0.04 - 0.03$	6.92
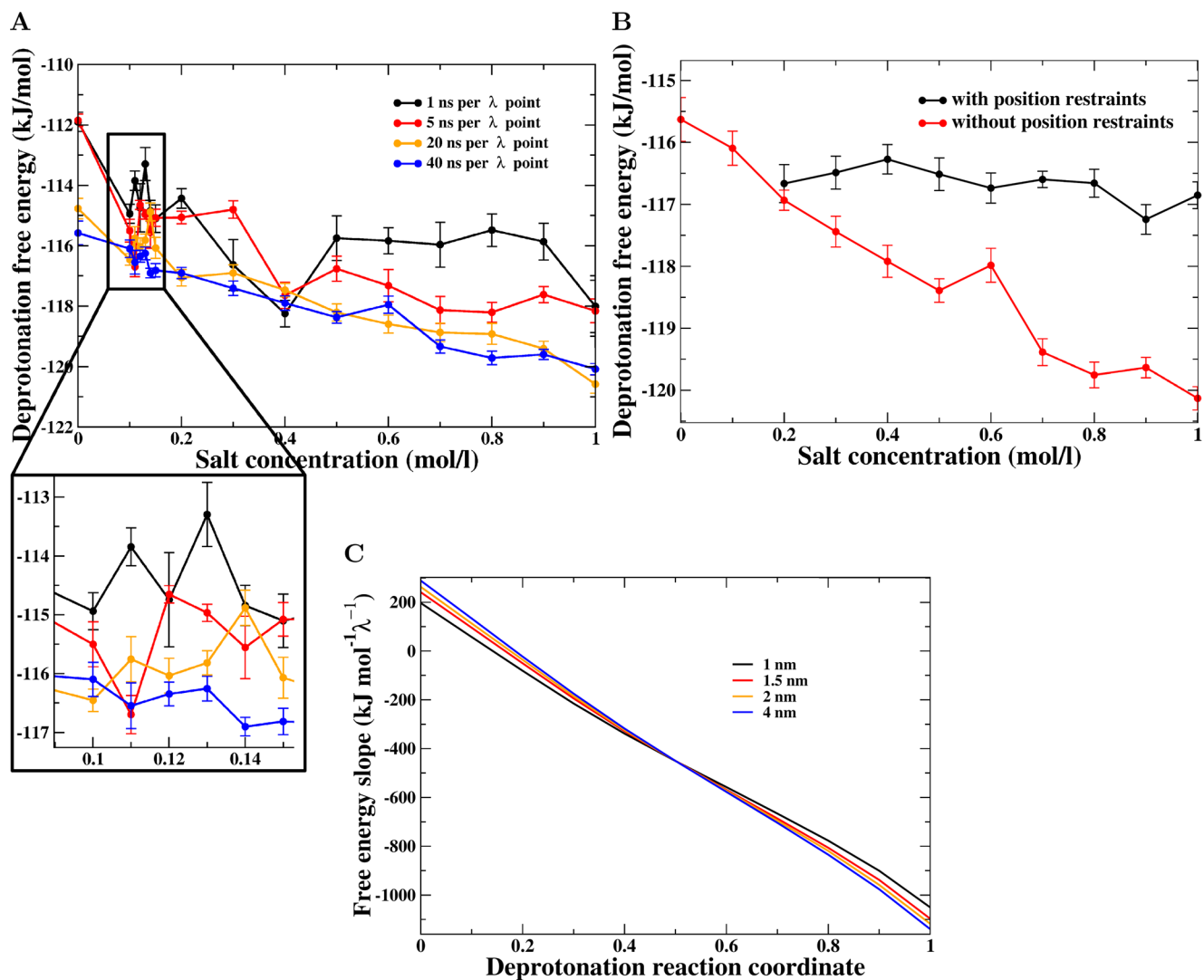
Figure 8. Deprotonation free energies for different ion concentrations, mobilities, and distances between titrating residue and neutralizing "water" molecule. **A.** The free energy of deprotonation as a function of ion concentration with different simulation lengths of the TI (depicted with different colors). The inset shows the deprotonation free energies in salt concentrations near the physiological value (0.1, 0.11, 0.12, 0.13, 0.14, 0.15 mol·L^{-1}). The free energy values in that region were calculated with smaller step size. **B.** The effect of restraining the ion movement on the deprotonation free energy shown in black color for restrained and in red color for nonrestrained ions. **C.** The dependence of the deprotonation free energy slope as a function of λ on the distance between the titrating site and the coupled "water" molecule.

The effect of the coupling between the neutralizing "water" molecule and the titrated residue on the transition rate between protonated and deprotonated states was also tested by comparing the number of transitions between the two states in coupled and noncoupled systems of one tautomer Asp.

Figure 1 illustrates the relevant protonation states considered in the simulations, the two neutral tautomers and the negatively charged form of Asp and Glu and the positively charged form and the two neutral tautomers of His. Constant pH simulations were performed at different pH values for each amino acid in explicit solvent at physiological salt concentrations as described

in the Methods section. Figure 7 shows the micro- and macroscopic pK_a distributions (right panels) and titration curves (left panels) of Asp, Glu, and His, calculated from constant pH MD, from which the pK_a values shown in Table 1 were obtained (see Methods section 2.7). The Hendersen-Hasselbalch curves in Figure 7 were plotted using the pK_a values in Table 1 and overlaid with the average λ (calculated from the λ-distribution which was obtained via eq 10) and its error estimation for each pH.

The calculated macroscopic pK_a values 4.01, 4.41, and 6.41 for Asp, Glu, and His are in good agreement with the reference

values of 4.00, 4.40, and 6.38, respectively. The small errors of the average λ-values and the narrow pK_a probability distribution in Figure 7 B, D, and F and Figure 7 A, C, and E, respectively, as well as the small errors of the pK_a estimations, which are within 0.04 pK_a units (Table 1), suggest sufficient statistical convergence of the titration simulations. Therefore, it can be concluded that the introduced procedure for calculating the force field correction term V_{MM} (eq 1) reproduces well the two-dimensional free energy landscape defined by the three protonation forms. The relative difference between the calculated macroscopic and microscopic pK_a values for Asp and Glu is 0.3. This difference agrees with the log 2 factor expected from the protonation equilibrium between the two tautomeric forms of a carboxyl group. The same agreement is seen for His, which, taken together, suggests that our three states model accurately describes the respective tautomeric forms.

As can also be seen in Figure 7 A, C, and E, the average λ-values calculated from the individual constant pH simulations reproduce the Henderson−Hasselbalch behavior very well, mostly within the estimated statistical error. To quantify the agreement, the Hill equation was fitted to the average λ for both the microscopic and macroscopic titration curves of the three amino acids. All obtained Hill coefficients fell between 0.94 and 1.02, providing further support that the titration behavior of amino acids in solution, even for extreme pH values, is very well described by our three states model.

It can also be seen that the error of the average λ-values in Figure 7 A, C, and E decreases toward the end of the 0−1 intervals. This effect is not due to enhanced sampling but rather due to the fact that the position of the titrating λ-particle distribution, for a certain pH, depends nonlinearly on the respective times spent in the protonated and deprotonated states and transitions (eq 11), such that large changes affect the average λ-value to a lesser extent, and hence the statistical error is reduced.

Next, we asked to which extent the coupling of the titratable groups to position-restrained "water" molecules (described in Section 2.4) affects the protonation transition rates and, hence, the achieved statistical accuracy. In 200 ns long simulations of Asp with only one tautomeric form, 605 transitions between the two protonation states without coupling were observed, whereas only 470 transitions were observed in simulations with coupling under otherwise identical conditions. We note that for an uncoupled system a fourth-order polynomial fit represents the deprotonation free energy profile more accurately than a fifth-order fit. Because a correct free energy profile is of crucial importance for counting the correct number of transitions, a fourth-order fit was used when the Asp residue was not coupled to a water molecule in these calculations. The force field correction terms V_{MM}, specifically calculated for each of the two simulation setups, ensure there is no additional barrier in the coupled system. Therefore, such a reduced transition rate is likely due to the lower attempt frequency of the system to protonate or deprotonate when two molecules change their charge instead of one.

3.2. Effect of the Coupled "Water" Molecules.
A possible concern of the above-described coupling to distant "water" molecules is that the charged "water" molecules change the effective ion concentration. To this aim, we carried out deprotonation free energy simulations of one tautomer Asp, at different salt (NaCl) concentrations, different ion mobility, and various distances between the coupled "water" and the solute.

In addition, we performed deprotonation simulations of different lengths to estimate the time required for sufficient ion and amino acid conformational sampling.

Figure 8 A shows the obtained free energy change of deprotonation as a function of salt concentration. The different colors in the figure correspond to different simulation lengths (1, 5, 20, and 40 ns). From the inset, it can be seen that an increase of the ion concentration from 0.1 to 0.15 $mol \cdot L^{-1}$ (which for the used simulation box size requires the addition of 7 positive (Na^+) and 7 negative (Cl^-) ions) affects the deprotonation free energy by less than 1.0 $kJ \cdot mol^{-1}$. Such a weak dependence of the deprotonation free energy on the ion concentration suggests, providing the box size is sufficient to maintain the concentration shift in the magnitude described above, that the protonation behavior of the titrated residues will not be severely affected by the neutralizing "water" molecules.

We note that the trajectories with smaller lengths exhibit considerable scatter of the deprotonation free energy values, which is significantly larger than the individual error bars (specifically between 1 and 5 ns). This discrepancy suggests that the errors are not converged due to either (1) insufficient sampling of the interaction between the ions and Asp or (2) to insufficient conformational sampling of the residue within the individual short simulations. To estimate the effect of both factors on deprotonation free energy convergence, we performed deprotonation free energy calculations of Asp with position restraints on all heavy atoms of the amino acid. The position restraints ensured that the only factor determining the free energy convergence was the sampling of the amino acid−ion interactions. Simulations with 1, 5, 20, and 40 ns length were carried out at two salt (NaCl) concentrations: 0.15 $mol \cdot L^{-1}$ and 0.3 $mol \cdot L^{-1}$. The differences in the calculated free energies between the relatively short 1 and 5 ns long simulation and the longer 20 and 40 ns simulations at both salt concentrations are less than 1 $kJ \cdot mol^{-1}$. These differences are much smaller than the ones observed in Figure 8 A. These calculations suggest that the contribution from intramolecular interactions, of the deprotonating amino acid, to the deprotonation free energy, converges much more slowly than the contribution from the amino acid-ion interaction. This observation is also supported by the discrepancies in the free energy values, for different simulation lengths, in Figure 8 A, even when no ions were present in the solution. By performing principle component analysis and deprotonation free energy calculations, with restraints only on the backbone atoms, it was found that the slowest degrees of freedom, hence limiting factor in the model compound sampling, are the ones of the backbone. The convergence of the deprotonation free energy values for different simulations lengths (see Figure 8 A) suggests that simulation times of 40 ns per λ-point are required for sufficient sampling of the amino acid conformations. Of course, the optimal length of the reference simulations may differ among different water models and force fields.

The neutralizing "water" molecules, which change into "hydronium" ions during constant pH MD, are position-restricted to avoid artificial interaction with the solvated biomolecule. We therefore asked, next, if this restricted diffusion affects the deprotonation free energies. To study the effect of the restrained ions on the deprotonation free energies, we have therefore simulated a 0.15 $mol \cdot L^{-1}$ NaCl with additional restrained ions at various concentrations. The fixed ions were placed in such a way so that each ion was as far away as possible from the titrating residue or from the other fixed

ions. To this end the ions were placed on their positions one after the other. For each ion we scanned the axes of the simulation box with a 0.4 nm step and chose the location, which maximizes the minimal distance between the ion and the amino acid atoms or already placed ions for the given box size. The distances to the periodic images of the fixed ions and the amino acid were also included in the calculations so that the interactions among the fixed ions and their periodic images were minimized as well. Figure 8 B compares the deprotonation free energy as a function of salt (NaCl) concentrations of two setups, one with freely diffusing Na^+ and Cl^- ions and the other with restrained ions. As can be seen, the deprotonation free energy changes markedly only for freely diffusing ions. In contrast, no significant concentration dependence is observed in simulations with restrained ions.

We suggest that this effect is due to the reduced interaction between the ions and the deprotonating residue when the ions are fixed at their positions. Indeed, a simple calculation of the average number of van der Waals contacts, per snapshot, between the two partially negatively charged oxygen atoms of the Asp residue and the positively charged nonrestrained Na^+ ions (increasing from 0.054 at 0.2 mol·L^{-1} ion concentration to 0.077 at 1 mol·L^{-1} in the restraint setup versus an increase from 0.058 to 0.291 for the same concentrations in the freely diffusing setup, counting the contacts in all λ steps of the deprotonation TI simulations) supports this notion. We conclude that the effect of charging the neutralizing "water" molecules on the deprotonation free energy of the titrated residues, during constant pH MD simulations, is negligible.

In all ion concentration test simulations, described above, the Asp was not coupled to a "water" molecule ensuring that no additional ion appeared in the solution and the ion concentration remained constant. Because of the identical net charge created in all simulations, the offset in the calculated deprotonation free energy is identical, making the results comparable.

Besides ion concentration, the effect of the distance between the coupled partners on the deprotonation free energy was also tested. This effect is essential because if multiple titratable residues are to be simulated (e.g., peptides or proteins), several neutralizing "water" molecules would be required, placed at different distances from the residues they are coupled to. If the deprotonation free energy depends on the distance between the coupled partners, that would make the force field correction term (V_{MM}^{corr}) in eq 1 specific for each titratable site in the protein.

The dependence of the deprotonation free energy on the distance between the titrated residue and the coupled "water" molecule is shown in Table 2 and Figure 8 C. The free energy values in Table 2 are practically identical, suggesting that the distance between the two molecules does not affect the free energy difference between the two protonation forms, at least beyond 1 nm and at a salt concentration of 0.15 mol·L^{-1}.

Table 2. Effect of the Distance between the Coupled Molecules on the Deprotonation Free Energy

| distance (nm)	free energy (kJ·mol^{-1})
1	−435.60 ± 0.13
1.5	−435.43 ± 0.11
2.0	−435.39 ± 0.13
4.0	−434.94 ± 0.08

However, the free energy slopes in Figure 8 C change with the distance between the coupled molecules. The different TI slopes show, that despite the coupled "water" molecules being placed at different distances from their coupled residues (as shown by Table 2), these distances should not change in the course of the simulation. Otherwise the slope of their deprotonation TI curves would change with time, yielding a different free energy of deprotonation depending on the particular trajectory. The different slopes also suggest that the energetic barrier between the two states varies with distance. By adjusting the biasing potential height for the particular coupled pair, one can compensate for these differences.

3.3. Chimeric Carboxyl Group. Special consideration of the carboxyl group in Asp and Glu, as described in the Methods section 2.8.1, completes our three states model. Because the carboxyl group has two symmetric titratable positions, a "chimeric" carboxyl group is defined, for which the two bonds between the carbon and the two oxygen atoms are identical and an additional hydrogen atom is introduced. To assess the accuracy of this approximation, we compared the distributions of the C−C−O bond and the C−C−O−O extraplanar and the C−C−O−H dihedral angles (Figure 9) with those obtained with the standard Amber99sb force field.[24] As can be seen in Figure 9 A, the bond angle distributions of the deprotonated "chimeric" and standard Amber99sb Asp agree well. As expected, the bond angle distribution of the deprotonated "chimeric" carboxyl group is between those of the two oxygen atoms of the protonated standard force field Asp. The positions of the protonated and deprotonated "chimeric" angle distributions with respect to the ones of a standard Asp suggest that the chosen angle parameters indeed provide a reasonable approximation of the "chimeric" C−C−O angle to the C−C−O angles of all forms in the standard force field.

For the extraplanar C−C−O−O angle (Figure 9 B), our approximation does not affect its average value, albeit a slight narrowing of the angle distribution is seen. Therefore, the overall planarity is preserved.

Finally, the C−C−O−H dihedral angle might be affected by our modification, which aimed at removing the high energy *anti* conformer of the carboxyl group as described in the Methods section 2.8.1. However, as seen in Figure 9 C, no significant difference is observed in the respective distributions.

To assess the effect of the above slight geometry changes on the ability of the carboxyl group to form hydrogen bonds, we compared the number of hydrogen bonds formed between the oxygen atoms (or the hydroxyl group in the case of protonated forms) and the surrounding water molecules in the "chimeric" and standard Amber99sb force field carboxyl groups. Indeed, very similar numbers for the deprotonated forms (an average of 2.87 versus 2.85 hydrogen bonds for the "chimeric" and standard force field group, respectively) are seen, as is expected from the very similar bond and extraplanar angle distributions. In contrast, a larger deviation is seen for the protonated form (1.15 versus 1.22 hydrogen bonds to the carbonyl oxygen and 1.18 vs 1.39 to the hydroxyl group on average). We assume this deviation is due to the "chimeric" nature of the respective bond angle distribution (black line in Figure 9 A), which serves to approximate the two different (green and blue lines) bond distributions of the standard Amber99sb force field. Therefore, in cases where the geometry of the hydrogen bonding is crucial, e.g. enzyme active sites, it is necessary that the ability of the "chimeric" group to form the correct hydrogen bonds is additionally tested.
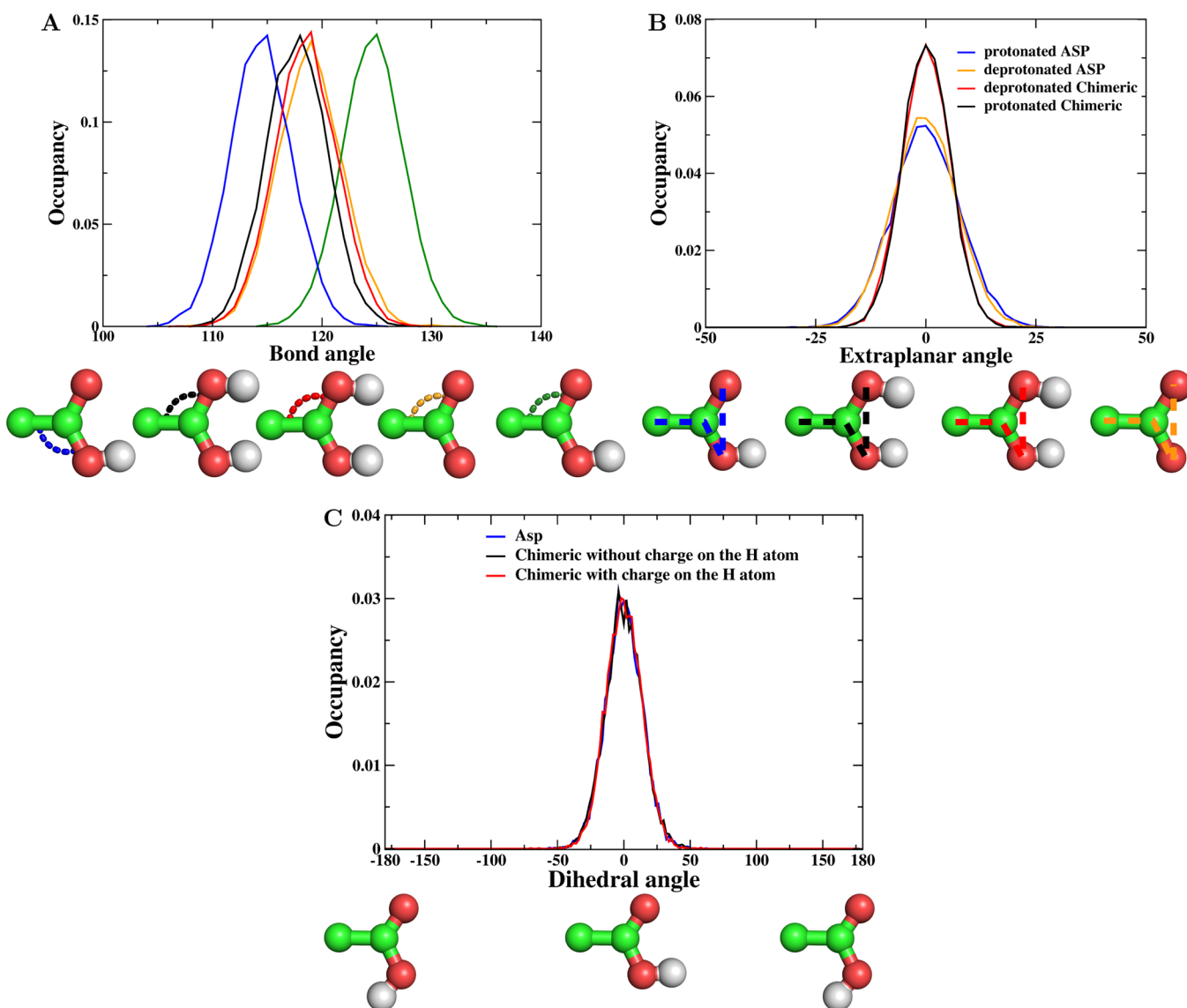
Figure 9. Carboxyl group angle calculations. **A**. The C−C−O bond angle distribution of the "chimeric", protonated, and deprotonated carboxyl groups. The color coding of the graph corresponds to the angles of the different forms of the carboxyl group depicted as dotted lines in the structures below. The red and black angle notations correspond to the deprotonated and protonated "chimeric" carboxyl groups, respectively. The blue, green, and orange angle notations depict the corresponding angles for the standard force field carboxyl group. **B**. Distribution of the extraplanar angle in the deprotonated and protonated "chimeric" (red and black lines, respectively) and protonated and deprotonated (blue and orange lines, respectively) forms of the standard force field carboxyl group. The dashed lines in the molecule structures below depict the angles whose distribution has been plotted with corresponding colors. The average angles in the case of the standard protonated and deprotonated Asp are 0.086° ± 0.090 and −0.192° ± 0.095, respectively, and 0.119° ± 0.065 and 0.070° ± 0.191 in the case of the Asp with "chimeric" protonated and deprotonated carboxyl groups, respectively. **C**. C−C−O−H dihedral angle distribution. The original force field is shown in blue. The red and the black lines show the hydrogen angle distribution of the protonated "chimeric" carboxyl group for the hydrogen atom which carries a charge and the one which does not, respectively. The rotamers shown below the graph correspond to the angle value on the axis.

4. CONCLUSION

We have developed a three states model to describe the deprotonation of titratable groups with proton tautomers in explicit solvent constant pH MD simulations. An accurate description of protonation equilibria of tautomers is particularly important for those carboxyl groups in proteins for which the rotation of the side chain is restricted. At the same time, the model excludes the nonphysical double protonated form, which would require an artificial pK_a value, hence incorrect, partitioning among the protonated states. Further, in our three states model, only one reaction coordinate needs to be coupled to a charge neutralizing "water" molecule, that is

advantageous for biomolecules with many titratable sites, for which the placement of the respective neutralizing "water" molecules in the limited space of the simulation box becomes critical.

To assess the ability of the three states model to describe the titration behavior of both carboxyl containing residues, such as Asp and Glu, as well as His, we performed extensive titration simulations of the amino acids and calculated the corresponding macroscopic and microscopic pK_a values. Both the obtained shifts from the reference pK_a values as well as the well preserved difference between the microscopic and macroscopic pK_a values suggest that our three states model is capable of

yielding correct partitioning between the two tautomeric forms of titratable residues with two chemically coupled titrating sites. Further, the Hill coefficients for all microscopic and macroscopic titration curves are close to 1, which shows that the three states model provides not only an accurate description of pK_a shifts but also of the complete Henderson−Hasselbalch behavior of the titratable sites.

To improve the accuracy of the force field correction term and to ensure energy conservation, we introduced a procedure to minimize the difference between the free energy profiles calculated along the two λ-coordinates. The agreement between the calculated macroscopic pK_a values and the reference ones suggests that the procedure can reproduce well the free energy landscape of the three states model.

Charge preservation of the system is crucial for MD simulations with periodic boundary conditions and PME electrostatics.[43] During constant pH MD simulations, however, titratable amino acids change their protonation state, thus changing the total charge of the simulation system. Similar to previous work, where the titrated compound was coupled to a "water" molecule the charge of which changes in the opposite direction to that of the titrating compound,[13,14] here, the titrating λ-coordinate of the three states model was coupled and additionally the coupled "water" molecule was restraint. Because this process effectively changes the ion concentration of the solution, we have assessed its effect on the pK_a of the titratable residues and found it small enough to conclude that this scheme does not severely affect protonation equilibria. Further, for neutralizing "water" molecules with distances larger than 1 nm to the titratable groups, no distance dependency of the protonation free energies was seen, such that calculation of only one force field correction profile for each type of titratable amino acid (Asp, Glu, and His) suffices.

One further effect of the coupling to "water" molecules is a reduced transition rate between the respective protonation states. In our test simulations, however, the resulting increase of statistical error for the calculated pK_a values was only small.

Because a pK_a unit corresponds to only 5.8 kJ·mol^{-1}, accurate free energies are required. For our test simulation, using the Amber99sb force field and the SPCE water model, 30 to 40 ns long simulations for each λ-value provided sufficient sampling.

For the carboxyl group, the Amber99sb force field needed to be modified to allow protonation of both oxygen atoms. The bond and dihedral angle distributions of angles crucial for determining the geometry of the carboxyl group show that the behavior of the protonated form is affected to a larger extent than the deprotonated form, yielding a different number of hydrogen bonds with the surrounding water molecules. Therefore, additional testing is required when the geometry of the hydrogen bonding is crucial for a given process, e.g. if the amino acid plays a role in enzyme catalysis.

Taken together, our extensive tests suggest that our three states model, combined with the charge preservation scheme, should provide a sufficiently accurate description of titratable groups with two chemically coupled protonation sites and thus enable one to carry out efficient constant pH MD simulations of large biomolecules in explicit solvent and with PME electrostatics.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.6b00807.

Partial atomic charges of Asp, Glu, and His in their protonated and deprotonated states (PDF)

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: hgrubmu@gwdg.de
Notes
The authors declare no competing financial interest.

■ REFERENCES

(1) Mertz, J. E.; Pettitt, B. M. *Int. J. Supercomput. Appl. High Perform. Eng.* **1994**, *8*, 47−53.
(2) Lee, M. S.; Salsbury, F. R., Jr.; Brooks, C. L., 3rd *Proteins: Struct., Funct., Genet.* **2004**, *56*, 738−752.
(3) Khandogin, J.; Brooks, C. L., 3rd *Biophys. J.* **2005**, *89* (1), 141−57.
(4) Wallace, J. A.; Shen, J. K. *J. Chem. Theory Comput.* **2011**, *7*, 2617−2629.
(5) Morrow, B.; Wang, Y.; Wallace, J.; Koenig, P.; Shen, J. *J. Phys. Chem. B* **2011**, *115*, 14980−14990.
(6) Donnini, S.; Tegeler, F.; Groenhof, G.; Grubmüller, H. *J. Chem. Theory Comput.* **2011**, *7*, 1962−1978.
(7) Donnini, S.; Ullmann, T.; Groenhof, G.; Grubmüller, H. *J. Chem. Theory Comput.* **2016**, *12* (3), 1040−1051.
(8) Goh, G. B.; Knight, J. L.; Brooks, C. L., 3rd *J. Chem. Theory Comput.* **2012**, *8* (1), 36−46.
(9) Goh, G. B.; Hulbert, B. S.; Zhou, H.; Brooks, C. L., 3rd *Proteins: Struct., Funct., Genet.* **2014**, *82* (7), 1319−1331.
(10) Herce, H. D.; Garcia, E. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 20805−20810.
(11) Yesylevskyy, S.; Marrink, S.; Mark, A. E. *Biophys. J.* **2009**, *97* (1), 40−49.
(12) Dixit, S. B.; Chipot, C. *J. Phys. Chem. A* **2001**, *105*, 9795−9799.
(13) Wallace, J. A.; Shen, J. K. *J. Chem. Phys.* **2012**, *137*, 184105.
(14) Chen, W.; Wallace, J.; Yue, Z.; Shen, J. *Biophys. J.* **2013**, *105* (4), L15−L17.
(15) Börjesson, U.; Hünenberger, P. *J. Chem. Phys.* **2001**, *114*, 9706−9719.
(16) Creighton, T. *Proteins: Structure and molecular properties*; W. H. Freeman and Company: New York, 1993; Chapter 1, p 6.
(17) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306−317.
(18) Gu, Z.; Ridenour, C. F.; Bronnimann, C. E.; Iwashita, T.; McDermott, A. *J. Am. Chem. Soc.* **1996**, *118*, 822−829.
(19) Ryckaert, J. P.; Bellemans, A. *Faraday Discuss. Chem. Soc.* **1978**, *66*, 95−106.
(20) Chen, J.; Noodleman, L.; Case, D.; Bashford, D. *J. Phys. Chem.* **1994**, *98*, 11059−11068.
(21) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43−56.
(22) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.; Berendsen, H. *J. Comput. Chem.* **2005**, *26*, 1701−1718.
(23) Pronk, S.; Pàll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics* **2013**, *29* (7), 845−854.
(24) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 712−725.

(25) Sorin, E.; Pande, V. *Biophys. J.* **2005**, *88*, 2472−2493.

(26) DePaul, A.; Thompson, E.; Patel, S.; Haldeman, K.; Sorin, E. *Nucleic Acids Res.* **2010**, *38*, 4856−4867.

(27) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269−6271.

(28) Feenstra, K.; Hess, B.; Berendsen, H. *J. Comput. Chem.* **1999**, *20*, 786−798.

(29) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463−1472.

(30) Miyamoto, S.; Kollman, P. *J. Comput. Chem.* **1992**, *13*, 952−962.

(31) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(32) Essmann, U.; Perera, L.; Berkowitz, M.; Darden, T.; Lee, H.; Pedersen, L. *J. Chem. Phys.* **1995**, *103*, 8577−8592.

(33) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118* (45), 11225−11236.

(34) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25* (13), 1656−1676.

(35) Borgstahl, G.; Williams, D.; Getzoff, E. *Biochemistry* **1995**, *34* (19), 6278−6287.

(36) Nosè, S. *Mol. Phys.* **1984**, *52*, 255−268.

(37) Hoover, W. G. *Phys. Rev. A: At., Mol., Opt. Phys.* **1985**, *31*, 1695−1697.

(38) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182−7190.

(39) Nosè, S.; Klein, M. L. *Mol. Phys.* **1983**, *50*, 1055−1076.

(40) Mongan, J.; Case, D.; McCammon, J. *J. Comput. Chem.* **2004**, *25*, 2038−2048.

(41) Tanokura, M. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1983**, *742* (3), 576−585.

(42) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision C.02*; Gaussian, Inc.: Wallingford, CT, 2004.

(43) Hub, J. S.; de Groot, B. L.; Grubmüller, H.; Groenhof, G. *J. Chem. Theory Comput.* **2014**, *10*, 381−390.