Article

Best Practices in Constant pH MD Simulations: Accuracy and Sampling

Pavel Buslaev,*,# Noora Aho,# Anton Jansen, Paul Bauer, Berk Hess,* and Gerrit Groenhof*

Cite This: *J. Chem. Theory Comput.* 2022, 18, 6134–6147

Read Online

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

ABSTRACT: Various approaches have been proposed to include the effect of pH in molecular dynamics (MD) simulations. Among these, the λ-dynamics approach proposed by Brooks and co-workers [Kong, X.; Brooks III, C. L. *J. Chem. Phys.* **1996**, 105, 2414–2423] can be performed with little computational overhead and hfor each typeence be used to routinely perform MD simulations at microsecond time scales, as shown in the accompanying paper [Aho, N. et al. *J. Chem. Theory Comput.* **2022**, DOI: 10.1021/acs.jctc.2c00516]. At such time scales, however, the accuracy of the molecular mechanics force field and the parametrization becomes critical. Here, we address these issues and provide the community with guidelines on how to set up and perform long time scale constant pH MD simulations. We found that barriers associated with the torsions of side chains in the CHARMM36m force field are too high for reaching convergence in constant

pH MD simulations on microsecond time scales. To avoid the high computational cost of extending the sampling, we propose small modifications to the force field to selectively reduce the torsional barriers. We demonstrate that with such modifications we obtain converged distributions of both protonation and torsional degrees of freedom and hence consistent pK_a estimates, while the sampling of the overall configurational space accessible to proteins is unaffected as compared to normal MD simulations. We also show that the results of constant pH MD depend on the accuracy of the correction potentials. While these potentials are typically obtained by fitting a low-order polynomial to calculated free energy profiles, we find that higher order fits are essential to provide accurate and consistent results. By resolving problems in accuracy and sampling, the work described in this and the accompanying paper paves the way to the widespread application of constant pH MD beyond pK_a prediction.
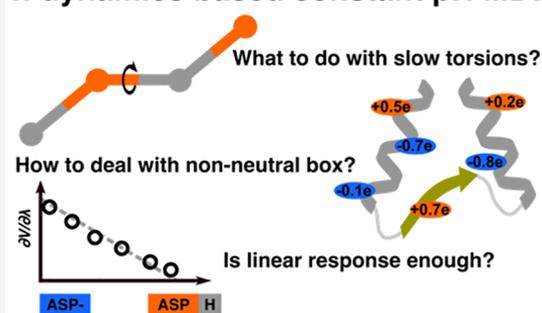
■ INTRODUCTION

Thanks to improvements in algorithms, force fields, and computer hardware, molecular dynamics (MD) simulations have become a versatile tool for investigating the conformational landscape of complex biomolecular systems at the atomic level.[1−5] An important algorithmic improvement has been the explicit inclusion of pH in MD simulations,[6−12] as pH is an important experimental parameter that affects the structure and dynamics of biomolecules. To provide the users of the GROMACS MD package[13] with access to simulations at constant pH, we have implemented the λ-dynamics-based constant pH approach by Brooks and co-workers.[9,14] In contrast to the previous implementation in a fork of GROMACS 3.3,[10] the new implementation, which is described in an accompanying paper,[15] is efficient and constant pH MD simulations can be performed with about 25% for CPU, 30−40% for CPU + GPU computational overhead compared to normal MD simulations irrespective of the number of titratable sites in the system. In the accompanying methodological paper,[15] we also demonstrate that the method can be successfully applied to calculate pK_a values of titratable

sites in a protein. The purpose of this paper is to provide users with guidelines and recommendations on how to set up and perform constant pH MD simulations, including the necessary parametrization steps.

In constant pH MD simulations, titratable groups can dynamically change their protonation state. These changes are driven by interactions between the group and the chemical environment (modeled with a force field) and the aqueous proton concentration (modeled with a pH potential). Because at the force field level a number of contributions to the free energy of (de)protonation are not included explicitly, i.e., quantum mechanical interactions associated with bond breakage and formation as well as the actual proton particle, corrections to the force field are needed in λ-dynamics-based

Table 1. Table of Simulated Systems[a]

system	box size (nm³)	no. of waters	no. of ions	force field
BUF₁	5 × 5 × 5	~4000	11 Na, 11 Cl, 2 Buf	CHARMM36m
BUF₂	3 × 3 × 3	~3000	1 or 2 Buf	CHARMM36m
ADA	5 × 5 × 5	~4000	11 Na, 11 Cl, 1 or 10 Buf	CHARMM36m, CHARMM36m-cph
ADA₃	3 × 3 × 3	~850	2 Na, 2 Cl, 10 Buf	CHARMM36m-cph
ADA₇	7 × 7 × 7	~11100	31 Na, 31 Cl, 10 Buf	CHARMM36m-cph
ADA_low salt	5 × 5 × 5	~4000	4 Na, 4 Cl, 10 Buf	CHARMM36m-cph
ADA_high salt	5 × 5 × 5	~4000	38 Na, 38 Cl, 10 Buf	CHARMM36m-cph
AEA	5 × 5 × 5	~4000	11 Na, 11 Cl, 1 or 10 Buf	CHARMM36m, CHARMM36m-cph
AKA	5 × 5 × 5	~4000	11 Na, 12 Cl, 1 or 10 Buf	CHARMM36m, CHARMM36m-cph
AHA	5 × 5 × 5	~4000	11 Na, 12 Cl, 1 or 10 Buf	CHARMM36m, CHARMM36m-cph
AAA₁	5 × 5 × 5	~4000	11 Na, 12 Cl, 1 or 10 Buf	CHARMM36m, CHARMM36m-cph
AAA₂	5 × 5 × 5	~4000	11 Na, 11 Cl, 1 or 10 Buf	CHARMM36m, CHARMM36m-cph
1CVO₁	6.8 × 7.8 × 7.1	~12400	35 Na, 48 Cl, 20 Buf	CHARMM36m, CHARMM36m-cph
1CVO₂	7.1 × 7.1 × 7.1	~11400	13 Cl, 150 Buf	CHARMM36m
MEMB₁	5.9 × 5.9 × 8.1	~4200	8 Na, 8 Cl, 50 Buf	CHARMM36m
MEMB₂	5.9 × 5.9 × 8.1	~4200	8 Na, 8 Cl, 1 Buf	CHARMM36m
SOL	4.3 × 4.3 × 4.3	~2600	50 Buf	CHARMM36m

[a]We denote the modified CHARMM36m force field as CHARMM36m-cph.

constant pH MD. In GROMACS, these corrections are implemented as analytical functions, $V^{MM}(\lambda_j)$, fitted to the free energy profile associated with the deprotonation of a titratable residue j at the force field level.[15]

The accuracy of such free energy profiles depends not only on how closely the force field model represents the true potential energy surface but also on the convergence of sampling of all other degrees of freedom in the system. Therefore, whereas in normal MD the accuracy of the dynamics depends solely on the quality of the force field, the accuracy of λ-dynamics-based constant pH MD depends additionally on whether all relevant degrees of freedom are sampled sufficiently in the simulations required for parametrizing the correction potentials.

We found that insufficient sampling of the dihedral degrees of freedom in the amino acid side chains can lead to poor convergence in the deprotonation free energy profiles, as also observed by Klimovich and Mobley in simulations without constant pH.[16] We traced the lack of the dihedral sampling to the barriers that separate the minima in the torsion potentials. These barriers are too high to reach a converged sampling of the dihedral free energy landscape on the time scales of typical constant pH MD simulations. Because the interaction between the titratable group and the environment depends critically on the dihedral angles of the side chain, a lack of convergence in these dihedral angles also affects the sampling of the protonation states.

Rather than increasing the time scale of the MD simulations to obtain converged dihedral and protonation state distributions or introducing enhanced sampling techniques,[12,17−21] we propose reducing the barriers for dihedral rotations in a systematic way. We will demonstrate that such optimized dihedral force field parameters improve pK_a estimates of amino acids without compromising the overall conformational sampling of the protein.

With a higher accuracy of the underlying deprotonation free energy profiles, we found that the correct sampling of protonation states also depends critically on the order of the polynomial fit used to obtain an analytical form for the correction potential. We show that the commonly accepted first-order fit,[12] although firmly based on linear response

theory,[22] is not sufficiently accurate and can lead to erroneous protonation dynamics in constant pH MD simulations.

Because the dominant energetic contribution to proton affinity comes from electrostatic interactions,[22] it is of utmost importance to use an accurate description of such interactions. Constant pH MD simulations have been performed with various electrostatic models, including generalized Born,[9] shifted cutoff,[23] and particle mesh Ewald (PME).[10,12] Of these methods, the Ewald summation-based PME method[24,25] is generally considered to provide the most accurate description of the electrostatic interactions in periodic biomolecular systems.[26] Because Ewald summation can only provide accurate results if the simulation box remains neutral,[27] the charge fluctuations associated with the dynamic protonation and deprotonation in constant pH MD simulations need to be compensated to prevent artifacts. Titratable sites can be directly coupled to special particles, modeled as ions or water molecules,[21,28] such that charge is transferred directly between the titratable site and that particle. Alternatively, all sites can be coupled collectively to a sufficiently large number of buffer particles.[29] The latter approach has the advantage that spontaneous fluctuations in the interaction of the buffer particles with their environment affect all titratable sites to the same extent. The disadvantage is that the setup and parametrization of the buffer approach are more involved, as these require selecting the number of buffers and parametrizing their interaction with the rest of the system. To facilitate the use of buffers in constant pH MD, we provide a parametrization strategy aimed at preventing buffer clustering, buffer binding to titratable sites, and buffer permeation into hydrophobic regions. We demonstrate that buffers parametrized with this strategy also avoid finite-size effects associated with the periodicity of small simulation boxes.[30,31]

■ METHODS

Here, we go through all of the important aspects of the simulation setups.

Simulated Systems. We performed standard and constant pH MD simulations of the systems listed in Table 1. The original and modified (described in detail below) CHARMM36m[32,33] force fields were used in all simulations.

The table also presents the box size, the number of CHARMM36 TIP3P water molecules,[34−36] the number of ions, and buffer particles included in each system. The fitting coefficients of the V^{MM} correction potential for the buffer particles were obtained with system BUF_1. To find the optimal charge range and Lennard−Jones parameters for the buffer particles, enhanced sampling simulations with the accelerated weight histogram (AWH) method were performed on system BUF_2. Systems ADA, AEA, AKA, and AHA are alanine tripeptides with capped termini and as the central residue aspartic, glutamic, lysine, and histidine amino acids, respectively. AAA_1 and AAA_2 systems are alanine tripeptides with protonated termini. C- and N-termini were made titratable in AAA_1 and AAA_2 systems, respectively. Two sets of simulations of the cardiotoxin V protein were performed. System $1CVO_1$ was used to calculate the pK_a values of titratable residues, while the larger system $1CVO_2$ was used to compute the radial distribution function of the buffer particles around the protein. The membrane systems $MEMB_1$ and $MEMB_2$ contained 106 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine (POPC) lipids. The starting coordinates and topologies of all systems are provided as Supporting Information.

Simulation Setup. Periodic boundary conditions were applied in all systems. Electrostatic interactions were modeled with the particle mesh Ewald method,[24,25] while van der Waals interactions were modeled with Lennard−Jones potentials which were smoothly switched to zero in the range from 1.0 to 1.2 nm. Simulations were performed at a constant temperature of 300 K, maintained by the v-rescale thermostat,[37] with a time constant of 0.5 ps^{-1} and at a constant pressure of 1 bar, maintained by the Parrinello−Rahman barostat,[38] with a period of 2.0 ps. The leapfrog integrator with an integration step of 2 fs was used. Bond lengths to hydrogens in the solute were constrained with the LINCS algorithm,[39] while the internal degrees of the CHARMM TIP3P water molecules[35,36] were constrained with the SETTLE algorithm.[40] Prior to the constant pH MD simulations, the energy of all systems was minimized using the steepest descent method, followed by a 1 ns equilibration.

Constant pH MD Simulation Setup. In the constant pH MD simulations, the mass of λ-particles was set to 5 atomic units and the temperature was kept constant at 300 K with a separate v-rescale thermostat for the λ degrees of freedom[37] with a time constant of 2.0 ps^{-1}. The single-site representation, defined and described in the accompanying paper,[15] was used for Asp, Glu, Lys, C-ter, and N-ter, whereas the multisite representation, also described in that paper, was used for His.[15,41,42] The multisite representation models each physical state of groups with chemically coupled titratable sites with an independent λ-coordinate and takes chemical coupling into account by applying the linear constraint on these λ-coordinates requiring $\sum_{i \in \text{group}} \lambda_i^{\text{group}} = 1$.[15] The same pH and biasing potentials were used as in Aho et al.[15] In the sampling simulations of single titratable residues, the pH was set equal to the pK_a and the barrier height of the biasing potential was set to zero. The titration of the cardiotoxin V (PDB ID 1CVO[43]) protein was performed by running 10 independent replicas of 100 ns each for 15 equidistantly spaced pH values in the range from 1.0 to 8.0 using both the original and a modified CHARMM36m force field. In the titration simulations, the barrier height of the biasing potential was set

to 7.5 kJ mol^{-1} for groups modeled with a single-site representation and to 5 kJ mol^{-1} for groups modeled with a multisite representation.

Reference Simulations. The constant pH simulations require a correction potential $V^{MM}(\lambda_j)$ for each titratable residue j. These correction potentials are the integrals of polynomial fits to the expectation value of $\langle \partial V / \partial \lambda \rangle_\lambda$ in reference state simulations at fixed λ-values.[15] Thus, after integration, an nth-order polynomial fit to $\langle \partial V / \partial \lambda \rangle_\lambda$ yields an $(n + 1)$th-order polynomial function that represents $V^{MM}(\lambda)$. However, in our implementation, the fit to $\langle \partial V / \partial \lambda \rangle_\lambda$ was used, rather than $V^{MM}(\lambda)$. We thus refer to the fitting order as the order of the polynomial fit to $\langle \partial V / \partial \lambda \rangle_\lambda$.

We performed the reference simulations as follows: The partial charges in the tripeptide systems were linearly interpolated between $\lambda = -0.1$ and $\lambda = 1.1$ with a step of 0.05. For His, all three λ coordinates were changed under the constraint $\lambda_1 + \lambda_2 + \lambda_3 = 1$. For each set of λ-values, called a grid point, we performed an 11 ns MD simulation in which the $\partial V / \partial \lambda_j$ were saved every picosecond and accumulated. The total charge of the system was kept neutral by simultaneously changing the charge of a single buffer particle. The fitting procedure is described in full detail in the accompanying paper.[15]

Dihedral Free Energy Profiles. Because the sampling of protonation states is tightly coupled to the sampling of side chain dihedral degrees of freedom, we computed the free energy profiles associated with the rotation of the dihedrals in the side chain of the central amino acid in the capped tripeptide systems (Table 1) by means of umbrella sampling.[44,45] As the first step, we performed 20 ns MD simulations with a time-dependent potential on the dihedral angle with a force constant of 418.4 kJ mol^{-1} rad^{-2}. The center of this potential was moved from 0° to 360° with a rate of 18° ns^{-1}. From these simulations, frames with dihedral angles closest to 0°, 10°, 20°, etc., were selected as references for the umbrella replicas. The difference between the dihedral angle in the selected frames and the target angle was always below 0.1°. Then, we performed 36 umbrella sampling simulations of 11 ns with a harmonic restraining potential centered at the reference dihedral angle and a force constant of 418.4 kJ mol^{-1} rad^{-2}. We used the WHAM procedure,[46] implemented in GROMACS,[47] to unbias these umbrellas and obtain free energy profiles associated with the full rotation of the dihedral angle.

Dihedral Potential Energy Profiles at the QM and MM Levels. To check the validity of the proposed force field modifications, we computed the potential energy profiles for the N−C$_\alpha$−C$_\beta$−C$_\gamma$ dihedral of aspartic acid with capped residues. The profiles were computed at both quantum mechanical (QM) and molecular mechanical (MM) levels. The QM profiles were computed at the MP2/6-31+G* level of theory using the Firefly QC package,[48] which is partially based on the GAMESS (US)[49] source code. The MM profiles were computed for both the original and the modified CHARMM36m force fields. The potential energy was computed for the N−C$_\alpha$−C$_\beta$−C$_\gamma$ dihedral angle with 10° increments. For each dihedral value, the structures were energy minimized prior to potential energy calculation.

Accelerated Weight Histogram Alchemical Simulations. Buffer particles are used in constant pH MD to maintain the neutrality of the simulated system. Ideally, buffers should not introduce any artifacts due to binding to titratable groups, binding to each other, or penetrating into hydrophobic regions.
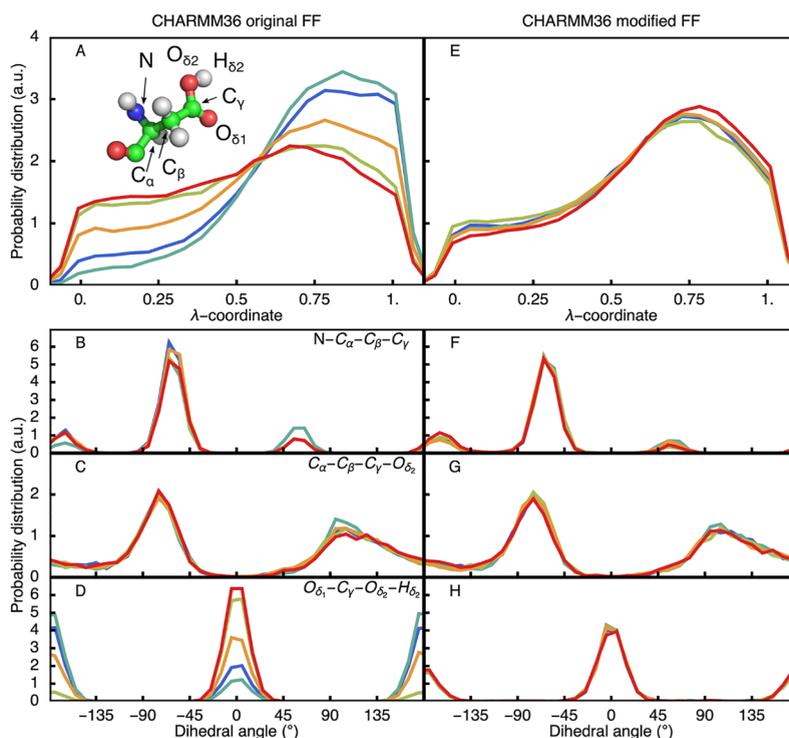
Figure 1. Distributions of the λ-coordinate (A and E) and dihedrals (B−D and F−H) in constant pH MD simulation of Asp with the original (A−D) and modified CHARMM36m (E−H) force field. While the simulations were performed for an ADA tripeptide, only the central aspartic acid is shown for clarity in the inset of A. (A and D) Distributions for third-order fits for $\langle \partial V / \partial \lambda \rangle_\lambda$ obtained with the original force field. Different colors correspond to independent replicas. Distributions for the λ-coordinate (A) as well as distributions for $N−C_\alpha−C_\beta−C_\gamma$ (B) and $O_{\delta_1}-C_\gamma-O_{\delta_2}-H_{\delta_2}$ (D) dihedrals are not identical. Right column (E−H) shows the distributions for the modified CHARMM36m force field with third-order fit for $\langle \partial V / \partial \lambda \rangle_\lambda$. Distributions are identical.

To prevent such behavior, we optimized the charge range and Lennard−Jones parameters of the buffers. To this end, we performed a series of enhanced sampling simulations with the accelerated weight histogram method (AWH).[50,51] In one set of simulations with two buffers in the simulation box (BUF$_2$, Table 1), we quantified the sampling efficiency from the friction metric[50,51] as a function of the absolute charge per buffer particle. In these simulations, the charge of one buffer was changed from 0 to +0.8 while simultaneously the charge of the other buffer was changed from 0 to −0.8 in order to maintain neutrality. The CHARMM36m Lennard−Jones parameters for sodium were used for the buffers in these simulations. In the other set of simulations, we computed the free energy difference between introducing a neutral buffer in water (BUF$_2$) and inside the hydrophobic region of a POPC bilayer system (MEMB$_2$, Table 1) for various values of the Lennard−Jones parameters of the buffer. In these simulations, the Lennard−Jones interactions between the buffer and the rest of the system were increased from noninteracting at $\lambda = 0$ to fully interacting ($\lambda = 1$) in 10 discrete steps. For all systems, we simulated 10 replicas of 10 ns, from which the free energy differences and friction coefficients were obtained as averages over the replicas.

Dihedral Analysis. Distributions of side chain dihedral angles in proteins were derived from publicly shared MD trajectories of proteins with PDB IDs 1U19,[52] 2RH1,[53] 2Y02,[54] and 5UEN[55] obtained from the GPCRMD[56] and SARS-CoV-2 databases (https://covid.molssi.org/).[57] For each trajectory, the distributions of the following dihedral angles were calculated:

(1) $N−C_\alpha−C_\beta−C_\gamma$ in aspartic acid
(2) $N−C_\alpha−C_\beta−C_\gamma$ in glutamic acid
(3) $C_\alpha−C_\beta−C_\gamma−C_\delta$ in glutamic acid
(4) $N−C_\alpha−C_\beta−C_\gamma$ in histidine
(5) $H−O_{\epsilon 2}−C−O_{\epsilon 1}$ in aspartic and glutamic acids.

In this work, we also computed the distributions of these dihedrals from standard MD trajectories of cardiotoxin V (PDB ID 1CVO[43]).

Comparisons of Distributions. To compare two distributions $P_i(x)$ and $P_j(x)$, with the corresponding cumulative distribution functions $F_i(x)$ and $F_j(x)$, we used Kolmogorov−Smirnov statistics (KSS)[58]

$$\text{KSS}(F_i, F_j) = \sup_x [F_i(x) - F_j(x)] \qquad (1)$$

The larger the KSS, the less similar the distributions are. The distributions were considered consistent when $\text{KSS}(F_i, F_j) < 0.03$. The KSS was computed using the script from the PCAlipids package.[59,60]

Titration. To estimate the pK_a values of titratable groups from multiple simulations at various pH values, we computed the average fraction of deprotonated frames (S^{deprot}) over all replicas at each pH value. For a group with a single titratable site, this average was obtained as

$$S^{\text{deprot}}(\text{pH}) = \frac{N^{\text{deprot}}}{N^{\text{prot}} + N^{\text{deprot}}} \qquad (2)$$

where N^{prot} and N^{deprot} are the total number of frames in which the site is protonated and deprotonated, respectively. For titratable sites modeled with the single-site representation, we
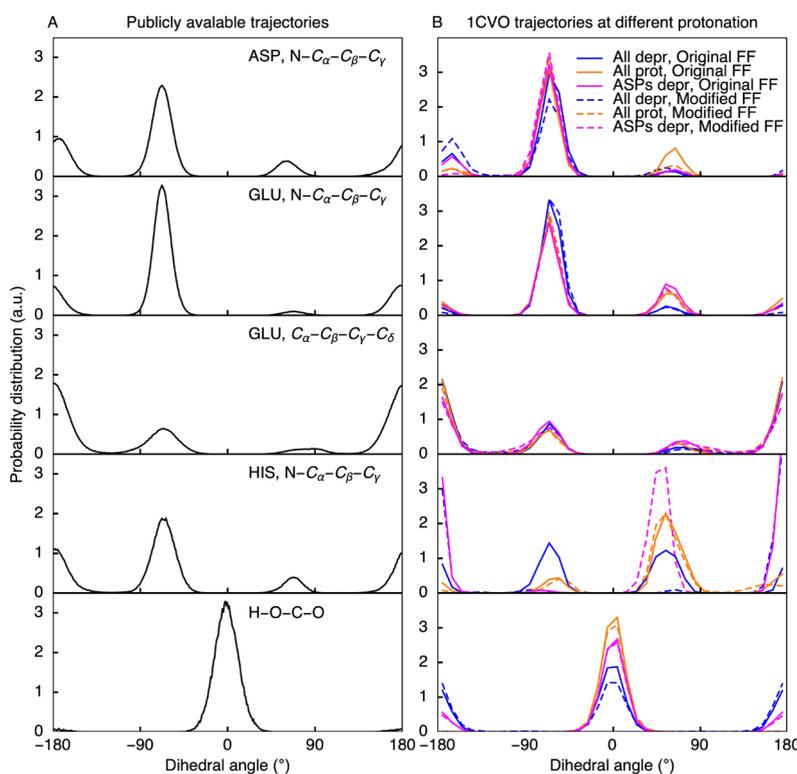
Figure 2. Distributions of dihedral angles for which the torsion potentials were corrected from standard MD simulations. (A) Dihedral distributions from publicly available trajectories[56,57] (total simulation time ≈ 10 μs). Probability density is significant only around local minima. (B) Dihedral distributions obtained from simulations of cardiotoxin V with both the original and the modified barriers for different protonation states of the titratable residues. Local minima are preserved, and no additional configurations are observed.

considered the site protonated if λ is below 0.2 and deprotonated if λ is above 0.8. For sites that are described with the multisite description, we considered a state protonated if the λ associated with the protonated form of the residue is above 0.8 and deprotonated if the λ associated with the deprotonated form of the residue is above 0.8.

To estimate the macroscopic pK_a value of histidine, which contains two titratable sites N_ϵ and N_δ, we calculated for each pH value the average fraction of frames in which the residue is deprotonated at either of the two sites

$$S_{macro}^{deprot}(pH) = 1 - \frac{N^{\lambda_p}}{N^{\lambda_p} + N^{\lambda_\epsilon} + N^{\lambda_\delta}} \quad (3)$$

where N^{λ_p}, N^{λ_ϵ}, and N^{λ_δ} are the number of frames in which $\lambda_p > 0.8$, $\lambda_\epsilon > 0.8$, and $\lambda_\delta > 0.8$.

The averaged fractions at each pH value were fitted to the Henderson−Hasselbalch equation

$$S^{deprot} = \frac{1}{10^{(pK_a - pH)} + 1} \quad (4)$$

which yielded the pK_a values as fitting parameters.

RESULTS AND DISCUSSION

First, we demonstrate that a lack of sampling of the relevant dihedral degrees of freedom in amino acid side chains with the CHARMM36m force field reduces the accuracy of the correction potentials for λ-dynamics. To overcome these convergence problems, we modify the force field by reducing the barriers in the torsion potential and show that this significantly improves the accuracy of the correction potentials

and hence the results of constant pH simulations, including pK_a estimates, without affecting the protein conformational dynamics. After the validation of the modified force field parameters, we show how the buffer particles that maintain the neutrality of the simulation box have to be parametrized to prevent finite-size effects on proton affinities due to periodicity.[30,31]

Sampling. Klimovic and Mobley have shown that calculated hydration free energies of single amino acids depend on the starting conformation.[16] Because a few picoseconds typically suffice to sample bond and angle degrees of freedom in the amino acid as well as the rotational degrees of freedom of the water molecules, we speculate that their observation implies a lack of sampling in the dihedral degrees of freedom in the amino acid side chain. Therefore, we systematically analyzed the convergence of both λ-coordinates and dihedral angles during constant pH simulations of single amino acids in water.

We performed 100 ns constant pH MD simulations at pH = pK_a of systems ADA, AEA, AKA, AHA, AAA$_1$, and AAA$_2$ (Table 1). To enhance the sampling of the λ-coordinate in these systems, we ran the simulations without a barrier in the biasing potential ($V^{bias}(\lambda)$, eq 5 in Aho et al.[15]). The correction potential ($V^{MM}(\lambda)$, eq 5 in Aho et al.[15]) was obtained by fitting a third-order polynomial function to the $\langle \partial V/\partial \lambda \rangle_\lambda$ values of the reference trajectories. We will show later that for accurate and reproducible constant pH MD results, a higher order fit is required. Nevertheless, in spite of its limited accuracy, using the same third-order fit for all system suffices to systematically compare the distributions of the relevant degrees of freedom and assess their convergence.
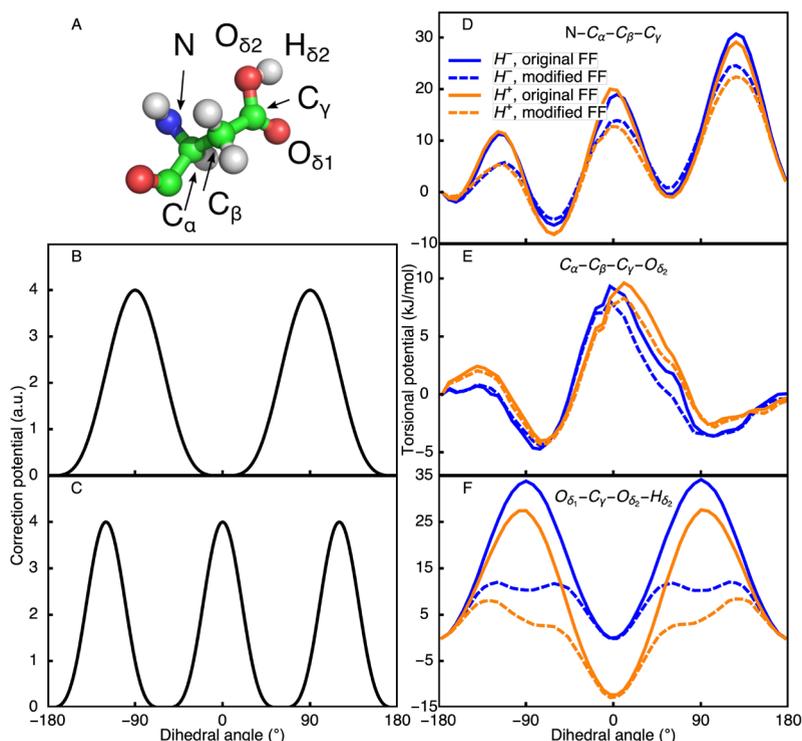
Figure 3. Modification of the torsional barriers in the Asp side chain. (A) Aspartic acid and its atomic nomenclature. (B and C) Corrections added to dihedral torsions of the original force field. Corrections with two (B) and three (C) local minima were used for torsions with two and three local minima, respectively. Heights of the corrections were selected through the iterative process, aimed at achieving consistent λ-distributions without introducing additional local minima in the free energy profiles. (D−F) Original and modified torsional barriers of Asp for both protonated (H^+) and deprotonated (H^-) states.

In Figure 1A, we show the distributions of the λ-coordinate in five constant pH MD replicas of the ADA system with the original CHARMM36m force field parameters. Distributions of the λ-coordinates in the other systems (AEA, AKA AHA, and AAA_1 and AAA_2) are shown in the Supporting Information (SI, Figures S1−S5). The dissimilarity between the λ-distributions in the replicas (maximum KSS between replicas of 0.29, 0.11, 0.04, and 0.095 for ADA, AEA, AHA, and AAA_1, respectively) indicates a lack of convergence. In addition, the distributions of the dihedral angles, shown in Figure 1B−D, are also not identical for all replicas. Because there is no barrier from the biasing potential for the λ-coordinate, we conclude that the lack of convergence in λ is due to insufficient sampling of the dihedral degrees of freedom.

Force Field Modifications. To overcome the lack of sampling, one can increase the simulation time or enhance sampling by means of special algorithms such as replica exchange MD.[61,62] Replica exchange has been applied frequently in the context of constant pH MD with exchanges between replicas at different temperature (T-REMD), pH (pH-REMD), or both.[20,63−66] However, because the protonation state has little effect on the torsion barrier height (Figure 3), changing the pH across the pH ladder would do little to enhance the sampling of the dihedral degrees of freedom. Furthermore, REMD methods are computationally more demanding than performing a single MD simulation and also prevent access to the dynamical properties of the system due to the jumps between replicas. As for some applications the dynamical properties may be highly needed, we consider it important that all relevant protonation states can be sampled correctly in a single constant pH MD trajectory. Because the

sampling of λ-coordinates is tightly coupled to the sampling of the dihedral angles, convergence within a single trajectory can in principle also be reached by lowering the barriers of the torsion potentials.

Previously, modifications to the force field have been introduced for carboxyl groups. To improve the sampling of the syn and anti conformations of the carboxyl proton in Glu, Asp, and the C-terminus, Brooks and co-workers reduced the barrier for this rotation by a factor of 8 and also scaled the carboxyl oxygen radii by 0.95.[9] In contrast, Grubmüller and co-workers modified this torsion potential to prevent sampling the anti-conformation altogether.[67] However, according to our analysis (Figure 1), there is a lack of convergence not only in the carboxyl dihedral angle but also in the other side chain dihedral angles.

Reducing the torsional barriers without affecting the overall sampling of the conformational space is possible only if the regions near such barriers are sparsely sampled. We, therefore, analyzed the distributions of the dihedral angles in the side chains of titratable amino acids in the publicly available trajectories of G-protein-coupled receptors and of SARS-CoV-2 proteins.[56,57] The distributions of these dihedral angles, plotted in Figure 2A, reflect the shape of the underlying torsion potentials with maxima coinciding with local minima of the potential profiles. The low density near barriers suggests that these barriers are rather high and might be reduced without affecting the dihedral distributions.

To achieve convergence of both dihedral angles and λ-coordinates on a 100 ns time scale, we thus altered the maxima of the torsion potentials by adding a small dihedral

angle (ϕ)-dependent correction to the CHARMM36m force field

$$\Delta V_i(\phi) = \epsilon_i\left[\cos(n_i\phi) - (-1)^{n_i}\frac{1}{4}\cos(2n_i\phi)\right] \quad (5)$$

where n_i is the multiplicity of the torsion angle i (i.e., the number of minima) with $n_i = 2$ for conjugated bonds and $n_i = 3$ for aliphatic bonds. The parameter ϵ_i is an empirical coefficient that is optimized such that the barriers are low enough to converge the distribution of ϕ_i without introducing additional minima on the potential energy surface.

For each side chain dihedral angle of the titratable amino acids, the coefficient ϵ_i was optimized in an iterative fashion: After an initial guess, we computed the free energy profiles associated with rotation of the dihedral as well as five unbiased 100 ns trajectories at pH = pK_a with different starting conditions and a biasing potential ($V^{\text{bias}}(\lambda)$, eq 5 in Aho et al.[15]) without barrier. Prior to these constant pH MD simulations, we recomputed the correction potential, $V^{\text{MM}}(\lambda)$, by fitting a third-order polynomial to the $\langle\partial V/\partial\lambda\rangle_\lambda$ values obtained from thermodynamic integration simulations performed with the current value of ϵ_i. Free energy profiles were inspected visually for artificial minima, while distributions of both dihedral angles and λ-coordinates were compared between the five unbiased replica runs based on their similarity. The coefficient ϵ_i was gradually increased until the distributions in the different replicas were sufficiently similar (KSS < 0.03), while at the same time no additional minima appeared in the free energy profiles.

In Figure 3 we show the optimized torsion potentials as well as their effect on the free energy profiles of the dihedral angles in Asp. The corrections and free energy profiles of Glu, His, and the C-terminus are shown in Figures S6−S8 of the SI. With the exception of the C_β−C_γ−O_δ−H dihedral, the corrections introduce no additional minima on the free energy profile of these torsions. Furthermore, as shown on the right panels of Figure 1, the distributions of the dihedrals and λ-coordinates are nearly indistinguishable for all replicas after correction. Because with the corrected potentials the Kolmogorov−Smirnov statistics for Asp, Glu, His, and the C-terminus are 0.028, 0.015, 0.027, and 0.022, we conclude that the corrections improve the convergence of both the λ-coordinates and the dihedral degrees of freedom in constant pH simulations. Note that although the distributions of the λ-coordinate are sufficiently similar, the sampling of these coordinates is not yet uniform (Figure 1E). We will show below that this discrepancy is due to the low order of the polynomial fit used for obtaining the correction potential $V^{\text{MM}}(\lambda)$.

With the exception of the O_{δ_1}−C_γ−O_{δ_2}−H dihedral, the corrections we propose here lead to changes in the torsion barrier of at most 16 kJ mol^{-1} (i.e., for the Glu N−C_α−C_β−C_γ torsion). For many biomolecular force fields, the parameters of the torsion potentials are obtained by fitting suitable periodic functions to energies evaluated at the MP2 level of theory.[68−71] The parameters for each type of torsional potential are simultaneously fitted for multiple amino acids. Therefore, the average root mean squared (RMS) difference between the torsional energy at the CHARMM36m level and that at the MP2 level of theory is on the order of 10 kJ mol^{-1}. The RMS deviation between the modified and the original torsion potentials is at most 8 kJ mol^{-1}, and the RMS deviation

between the ab initio potential at the MP2/6-31+G* level and the N−C_α−C_β−C_γ torsion potential in ASP is reduced from 4 kJ mol^{-1} for the original CHARMM36m force field to 3.5 kJ mol^{-1} for the modified CHARMM36m force field (Figure S9). Therefore, we conclude that with the corrections of the torsion potentials, the modified force field provides an equally good fit to QM potential profiles as the original CHARMM36m force field.[33,71]

We also performed standard MD simulations and simulated five replicas for 100 ns for the two protonation states of the Asp tripeptide in water using both the original and the modified CHARMM36m force field parameters. Without the modifications, the local minima are not consistently sampled in all replicas (Figure S10). In contrast, with the corrections, identical distributions of the dihedral angles are obtained also in standard MD simulations. In addition, the modifications are essential to sample both syn and anti conformations of the carboxyl proton, in line experiment.[72] We note, however, that the correction required for sampling both of these conformations significantly alters the shape of the barrier (Figure 3F). Nevertheless, because of their low mass, proton can tunnel through such barriers, and therefore, we consider the shape and height of the torsional barrier less relevant for this specific dihedral than for the other dihedrals.

Finally, we demonstrate that the modifications do not alter the distributions of the dihedral angles in protein simulations. We performed MD simulations of the 1CVO$_1$ system both with and without the modifications to the torsion potentials of titratable amino acids with either (i) all of these residues protonated, (ii) all deprotonated, or (iii) all Asp residues deprotonated and all other residues protonated. In Figure 2B, we plot the distributions of the dihedral angles for which corrections were introduced. The high similarity between the distributions suggests that the corrections do not lead to the sampling of different dihedral distributions, even if the relative weights of the minima are slightly altered, in particular, for the H−O−C−O dihedral. We conclude, therefore, that the corrections introduced to facilitate sampling of the dihedral and λ-coordinates do not significantly alter the protein conformational landscape and can hence be used to perform both normal and constant pH MD simulations.

Quality of the Correction Potential V^{MM}. *Reference Potential.* To verify that the modified torsion potentials overcome the convergence problems, we performed constant pH simulations at pH = pK_a and without a barrier in the biasing potential. Because with such a setup the potential energy profile for the λ-particle should be flat, we expected a uniform λ-distribution, provided that the dihedral degrees of freedom are sufficiently sampled. However, as shown in Figure 1E, the distributions are identical between replicas but not uniform despite the corrections to the torsion potentials.

Because both the pH-dependent potential $V^{\text{pH}}(\lambda)$ and the biasing potential are flat by construction at pH = pK_a, the deviations must originate from discrepancies between the correction potential $V^{\text{MM}}(\lambda)$ and the underlying free energy profile associated with deprotonation. The correction potential is obtained as a polynomial fit to the $\langle\partial V/\partial\lambda\rangle_\lambda$ values from thermodynamic integration simulations. Because linear response (LR) theory predicts a linear dependence between the hydration free energy and the magnitude of a (point) charge, a first-order fit has often been used to obtain the correction potential for constant pH MD.[9,12] However, even if the change in the charge dominates the free energy of changing the

protonation state, hydrogen-bond rearrangements can contribute as well. Because the effects due to such structural rearrangements are neglected in LR theories, we hypothesized that higher order fits may be necessary for obtaining sufficiently accurate correction potentials.

To test our hypothesis, we investigated the accuracy of the polynomial fit to the correction potential. In Figure 4A, we
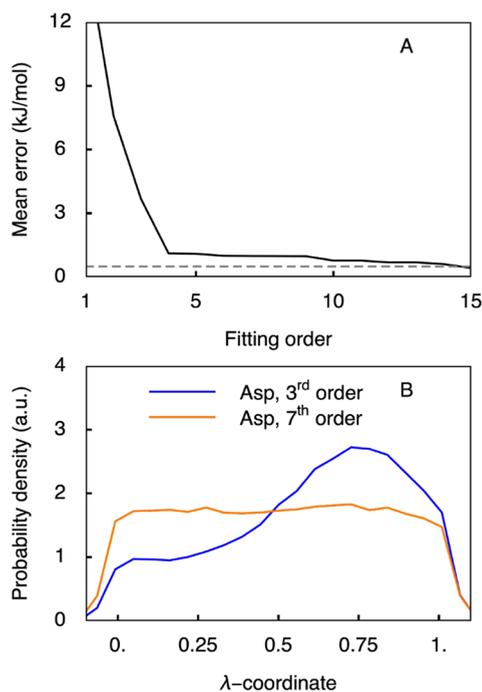
Figure 4. Quality of the $V^{MM}(\lambda)$ correction potential as a function of the order of the polynomial fit to $\langle \partial V/\partial \lambda \rangle_\lambda$ for Asp. (A) Fitting error as a function of fitting order (black line). Gray dashed line shows the average error in the calculated $\langle \partial V^{MM}/\partial \lambda \rangle$. (B) Distributions of λ-coordinates for the third- and seventh-order polynomial fits to $\langle \partial V/\partial \lambda \rangle_\lambda$. Whereas with the lower order fit the distribution is significantly rugged, the distribution becomes nearly flat and uniform on the [0, 1] interval of the λ-coordinate if a seventh-order fit is used.

show the mean error of the correction potential with respect to the computed free energy difference associated with deprotonation as a function of the fitting order. For the LR approximation the fitting errors are higher than 10 kJ/mol. In the worst-case scenario, such errors could lead to deviations in predicted pK_a values of more than one pK_a unit. With an error of 4 kJ mol^{-1}, the third-order fit, used above to address the convergence issues, does not yield a sufficiently accurate representation of the underlying free energy profile. Increasing the order of the polynomial fit reduces this error, and as shown in Figure 4B, at least a seventh-order fit is required to provide a uniform distribution of the λ coordinate for the Asp tripeptide in constant pH MD simulations at pH = pK_a.

Also, for carboxyl groups in the side chains of Glu and in the C-terminus, a polynomial fit to $\langle \partial V/\partial \lambda \rangle_\lambda$ of at least seventh-order is needed to provide a sufficiently accurate correction potential (Figure 4 and Figures S11 and S14 in SI). For the imidazole ring of His with three coupled titratable sites, a seventh-order fit suffices as well (Figure S13), while for the amino bases in the side chain of Lys and the N-terminus, at least an eighth-order fit is required (Figures S12 and S15 in SI). We speculate that the higher order fit is needed for the

latter sites due to the larger change in the charge on the central nitrogen atom from -0.3 e to -0.96 e upon deprotonation. The change in the charge of the carboxylic oxygen from 0.55 e to -0.76 e is smaller as are the changes on the nitrogen atoms of the imidazole ring of His (from -0.36 e to -0.7 e).

Parameterization of Buffer Particles. A change in the protonation state affects the total charge of the simulated system, which can lead to artifacts when Ewald summation is used to treat the electrostatic interactions.[27,31] In our implementation of constant pH MD, we avoid this problem by introducing titratable buffers into the simulation box that compensate for the charge fluctuations of the titratable residues.[29]

In the original implementation of constant pH MD in GROMACS,[10] the buffers were hydronium molecules that compensated for the overall charge fluctuations by changing their charge between 0 and +1 e. To prevent sampling charges beyond this interval, a biasing potential with steep edges at $\lambda = 0$ and 1 was introduced to restrict the range of λ-values. However, this potential still introduces additional forces at the edges of the λ-interval. To avoid the effects of such forces, we use a completely flat biasing potential for the buffers, also outside of the charge interval.

Because changing the charge of a buffer particle in solution induces local rearrangements of the hydrogen-bonding network that in turn could affect the proton affinity of a nearby titratable group, we want to minimize the impact of charging the buffer particles. To determine the charge range in which the buffers do not cause significant hydrogen-bond network rearrangement, we ran AWH simulations with two ions, the charges of which are changed simultaneously in opposite directions (BUF$_2$ system). From the friction metric available in the AWH method,[50] we estimated the local diffusion coefficient, which is related to the efficiency of sampling: The higher the friction, the slower the dynamics and the more sampling is required to reach convergence. We calculated the friction coefficient (Figure 5A) for the coordinate associated with changing the charge on the buffer. For charges higher than 0.5 e, the friction was more than 50% higher than that for zero charges, reflecting longer correlation times and hence slower dynamics. We, therefore, conclude that the optimal range for the buffer charge is between -0.5 e and 0.5 e.

The collective λ-coordinate of the buffer particles is not restricted to a fixed interval by a wall-like potential. To avoid the buffer charge exceeding the optimal range, multiple buffer particles are needed in the simulation box. The optimal number of buffers can be calculated based on the analysis of charge fluctuations performed by Donnini et al.[29]

With a small charge, a buffer particle is apolar. To prevent clustering of such apolar particles in water, permeation into hydrophobic areas, such as membrane interiors, or interactions with the protein, the Lennard–Jones parameters (σ and ϵ) of the buffers were chosen such that the buffers have only repulsive interactions with all other atoms, except water. After experimenting with the parameters for the buffer particles, we settled on a σ of 0.25 nm and an ϵ of 4 kJ mol^{-1}. This choice leads to decreased clustering of buffers, low buffer concentrations in the proximity of titratable sites, and reduced penetration into hydrophobic regions (Figure 5). The resulting free energies of neutral buffer insertion into water and the hydrophobic region of the membrane are -2.09 ± 0.07 and 1.2 ± 0.6 kJ mol^{-1} compared to 8.45 ± 0.05 and 7.8 ± 0.3
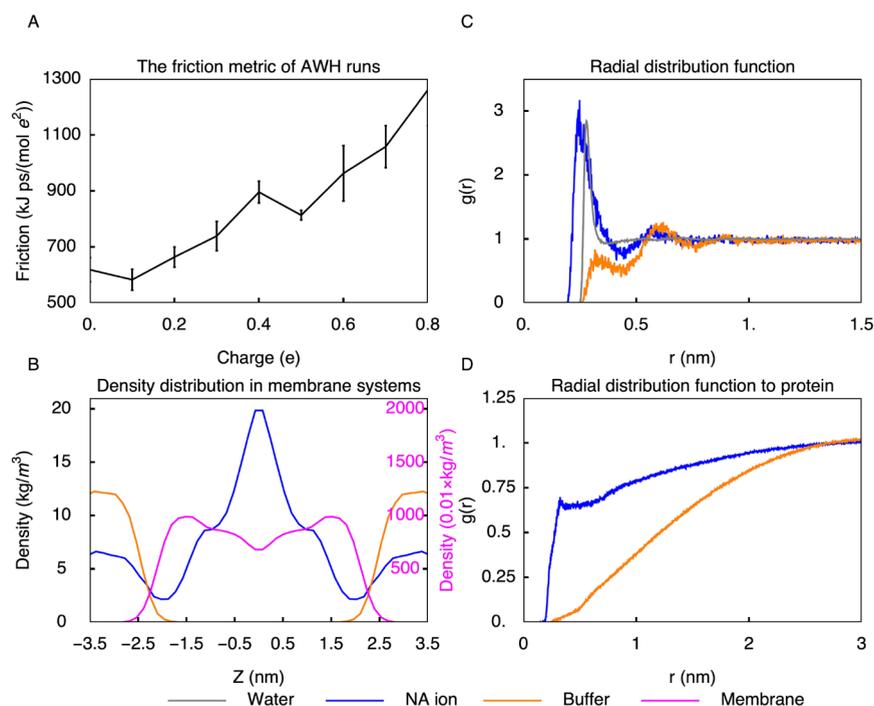
Figure 5. Parameterization of the buffer. (A) AWH friction metric as a function of buffer charge. Since the higher the friction metric is the slower the sampling, the buffers should ideally have low charges. (B) Density distribution of buffers in a membrane system. Optimized buffers do not penetrate into the lipid bilayer, while uncharged sodium ions do. (C) Radial distribution function between buffers. When standard sodium-ion parameters are used as buffers, the tendency to form clusters is high. Optimization of buffer parameters prevents clustering. (D) Radial distribution functions of buffers with respect to the protein. With the original sodium parameters the buffers have a higher tendency to bind to the protein than with the optimized parameters.

kJ mol^{-1} for insertion of an uncharged sodium ion into these regions.

While the primary goal of introducing buffers is to avoid the artifacts associated with a non-neutral periodic simulation box,[27] there can be other artifacts as well.[30,31] In particular, the undersolvation caused by solvent orientational restraints due to periodic boundary conditions[30] could lead to finite-size effects especially for small boxes. To understand if such finite-size effects affect the results of constant pH simulations, we investigated how the distribution of the λ-coordinates depends on the system size. We, therefore, performed constant pH MD simulations for three different box sizes and at various ionic strengths. All simulations were performed at pH = pK_a and without a barrier in the biasing potential. The uniformity of the distributions in these simulations, shown in Figure 6, suggests that for the box sizes tested, the finite-size effects are negligible.

Use Case: Consistent Protein Titrations. To demonstrate that with the modifications of the torsional barriers, a correction potential obtained by fitting at least a seventh-order polynomial to the $\langle \partial V/\partial \lambda \rangle_\lambda$ values of reference simulations, and buffer particles with optimized parameters, it is possible to perform accurate constant pH MD simulations, we calculated the pK_a values of all four cardiotoxin V titratable residues. We performed the pH titration simulations with both the original and the modified CHARMM36m force fields. In Figure 7, we show the titration curves obtained in the simulations and compare them to the experiment. Because there is no exact experimental estimate for the pK_a of ASP59, we only compare the pK_a values obtained for the other residues. The comparison suggests that the force field corrections improve the pK_a estimates, but more importantly, the lower deviation between
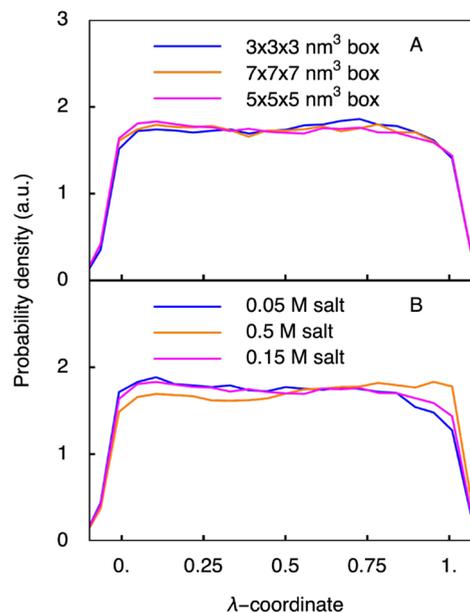
Figure 6. Effect of the box size (A) and ionic strength (B) on the distribution of the λ-coordinate of the ADA systems (ADA, ADA$_3$, ADA$_7$, ADA$_{low\ salt}$, ADA$_{high\ salt}$).

the individual replicas (from 0.12 to 0.07 for ASP42, from 0.16 to 0.05 for ASP59, from 0.1 to 0.03 for GLU17, and from 0.19 to 0.18 for HIS4, Figure 7) suggests that the sampling improves when the modified force field is used.
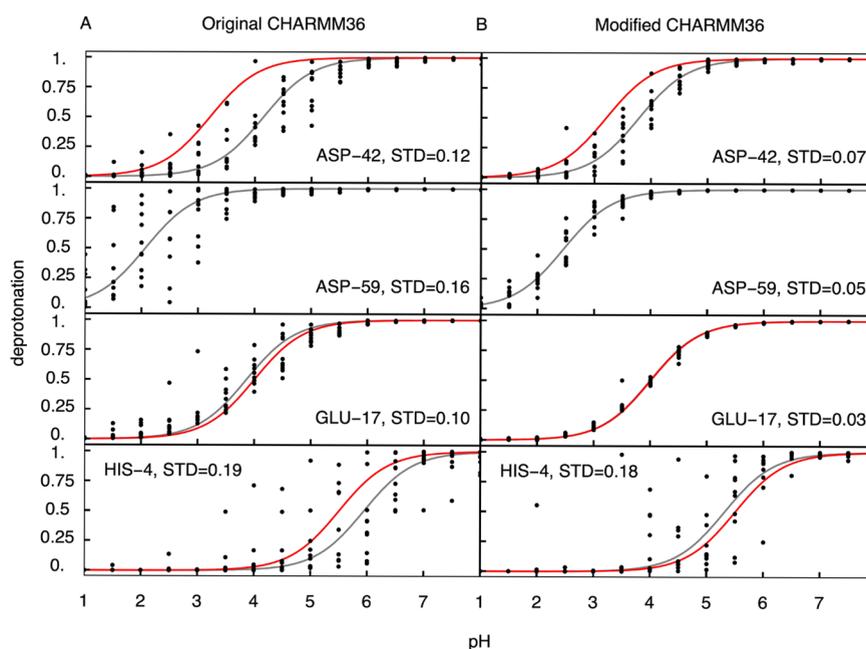
Figure 7. Titration of cardiotoxin V. (A and B) Titration curves of the titratable residues in $1CVO_1$ obtained from constant pH simulations performed with the original and modified CHARMM36 force fields, respectively. Black dots show the deprotonation ratio for the individual replica. Gray lines show the fits to the Henderson−Hasselbalch equation.[73,74] Red lines are Henderson−Hasselbalch curves computed for the experimental pK_a value of the corresponding residue. For ASP59, the exact pK_a is not known. For each curve, the standard deviation between the calculated and the fitted deprotonation ratio, averaged over all pH values and replicas, is shown.

Whereas for Asp and Glu the individual titration replicas are consistent when the modified force field is used, the replicas for HIS4 are not converged. As shown in Figure 8A, HIS4 interacts with TYR12 and PHE10. At pH = 5.5, close to the experimental pK_a of HIS4, we find three dominant conformations of HIS4-TYR12 pair in both normal and constant pH MD trajectories: A close contact (around 5 Å), a medium-range contact (around 6 Å), and a long-range contact (larger than 7 Å, Figure 8B and 8D). The distributions of these distances are not the same in all replicas (Figure 8E), and neither are the distributions of the λ-coordinates associated with the doubly protonated state of HIS4 (Figure 8F). These differences suggest a lack of convergence of the conformational dynamics of the protein in constant pH MD. To test if the protonation state of HIS4 correlates with the distance, we performed standard MD of cardiotoxin V with HIS4 in the three different protonation states (Figure. S16). Because there is no clear correlation between the protonation state and the HIS4-TYR12 distance, we cannot conclude that the protonation states are coupled to the conformation of the pair, at least not directly. Instead, the differences between the replicas suggest a lack of sampling of these conformations. Because the local environment differs between the states, we speculate that this lack of conformational sampling also affects the λ-distributions. We therefore can only conclude that the sampling of these distances would require more than 100 ns to converge. Thus, even if the corrections to the torsion potentials overcome the convergence issues associated with sampling the intrinsic dihedral degrees of freedom in single amino acids, reaching converged sampling of the protonation states in proteins may still require longer time scales if the inherent conformational dynamics is too slow.

Nevertheless, we observe that compared to normal MD, constant pH MD can increase the sampling of the local conformational space of the protein. In Figures 8C and S17 we show that the HIS4 samples configurations more efficiently in constant pH simulations. Specifically, the hydrogen bond between PHE10 and the HIS4 δ-hydrogen is much more stable in normal MD with a fixed protonation of δ-nitrogen (Figure S17), whereas in constant pH MD, the HIS4 also samples configurations in which the H-bond is broken (Figure S17), in particular around pH = pK_a, as evidenced by the distribution of the N−C$_\alpha$−C$_\beta$−C$_\gamma$ dihedral angle in Figure 8C. Thus, by keeping the protonation states flexible, constant pH MD facilitates the sampling of local conformations, which in turn may lead to faster convergence of the global conformational sampling.

■ CONCLUSIONS

It is now possible to run accurate constant pH molecular dynamics simulations on time scales of normal simulations, for example, with the new implementation in the GROMACS package presented in the accompanying paper.[15] Here, we have addressed the accuracy of constant pH simulation at longer time scales. We could demonstrate, on the basis of the CHARMM36m force field, that molecular force fields are not optimal for constant pH simulations because torsion barriers of titratable side chains are too high to reach convergence of the λ-coordinates associated with protonation. To overcome this sampling bottleneck, we proposed a systematic procedure to selectively reduce the barriers of the torsion potentials. In standard MD simulations, these modifications do not introduce noticeable artifacts but facilitate the convergence of side chain conformational sampling. Combined with the optimal fitting of V^{MM}, these force field modifications constitute an essential preparation step for constant pH simulations.
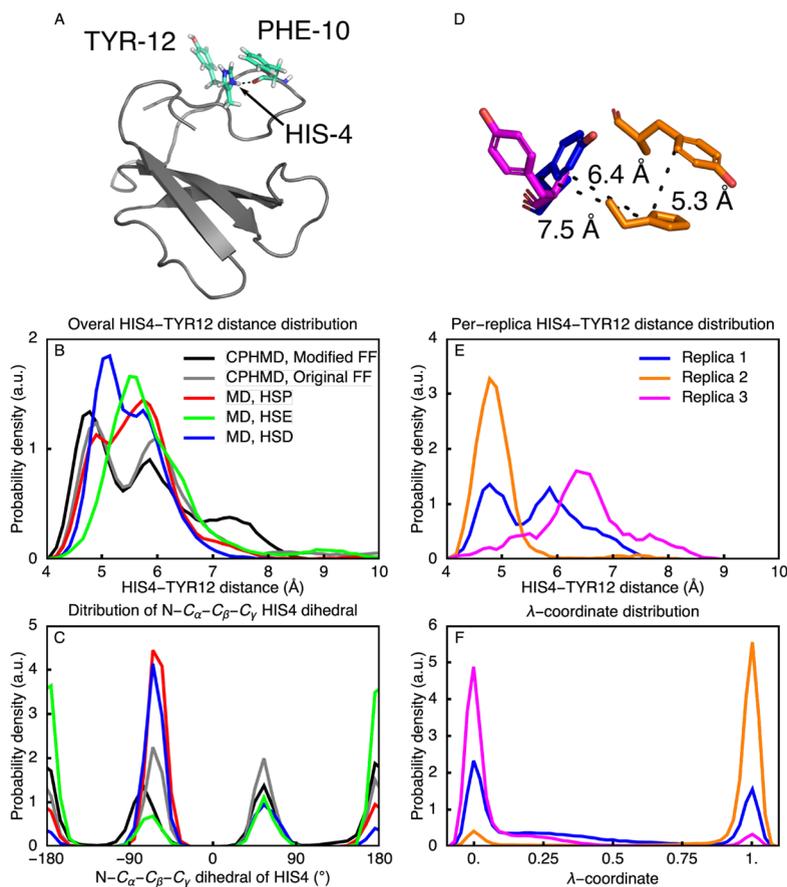
Figure 8. Sampling of cardiotoxin V N-terminal loop. (A) Structure of cardiotoxin V. Protein is shown in cartoon representation with HIS4, PHE10, and TYR12 shown as sticks. (B) Distributions of the distance between the centers of mass of HIS4 and TYR12 for all replicas combined. Combined distributions were calculated for constant pH MD with both the modified and the original force fields as well as for standard MD with HIS4 in all possible protonation states. (C) Distribution of HIS4 $N-C_\alpha-C_\beta-C_\gamma$ torsion. (D) Three observed conformations of TYR12 around HIS4. (E) Distributions of the distance between the centers of mass of HIS4 and TYR12 for individual replicas, computed for constant pH MD with the modified force field. (F) λ distributions of HIS4 protonated state for individual replicas of cardiotoxin V from the constant pH simulations with the modified force field.

The modifications of the CHARMM36m force field and the optimized parameters for the $\partial V^{MM}/\partial \lambda$ of correction potentials for Asp, Glu, His, Lys, and the C- and N-termini are available at https://gitlab.com/gromacs-constantph. Parameters for other force fields and residues will be made available there as well, once these are validated. We kindly ask the community to share with us also any parameters they may derive in their research, so that also these parameters can be made available.

The fork of GROMACS 2021 with constant pH MD implemented, as described here, is available for download free of charge from https://gitlab.com/gromacs-constantph/constantph. In addition to the source code, instructions on how to set up and perform MD simulations are available.

In addition to accurate parameters, it is also essential to keep the simulation system neutral during constant pH MD simulations. To achieve this, we introduced buffer particles with variable charges that dynamically compensate for the charge fluctuations of the titratable residues. To avoid that these buffers cluster, bind to the solute, disrupt hydrogen-bond networks, or penetrate into hydrophobic regions, we proposed a systematic parametrization procedure that can be used for any combination of force field and water model.

We expect the parametrization protocols proposed in this work to facilitate the application of constant pH MD not only

within the user community of GROMACS but also of other MD programs as well. We also want to appeal to force field developers to take constant pH MD into consideration when developing their force fields.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.2c00517.

Input files for MD simulations presented in this work (ZIP)

λ and angle distributions for Glu, His, Lys, and C- and N-termini; torsion corrections for Glu, His, and C-terminus; effects of force field modifications on dihedral dynamics and energy profiles; effects of fitting order on λ-distributions; conformations of HIS4 of cardiotoxin V (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Pavel Buslaev − *Nanoscience Center and Department of Chemistry, University of Jyväskylä, 40014 Jyväskylä,*

Finland; ⊙ orcid.org/0000-0003-2031-4691;
Email: pavel.i.buslaev@jyu.fi

Berk Hess − *Department of Applied Physics and Swedish e-Science Research Center, Science for Life Laboratory, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden;*
Email: hess@kth.se

Gerrit Groenhof − *Nanoscience Center and Department of Chemistry, University of Jyväskylä, 40014 Jyväskylä, Finland;* ⊙ orcid.org/0000-0001-8148-5334;
Email: gerrit.x.groenhof@jyu.fi

Authors

Noora Aho − *Nanoscience Center and Department of Chemistry, University of Jyväskylä, 40014 Jyväskylä, Finland*

Anton Jansen − *Department of Applied Physics, Science for Life Laboratory, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden*

Paul Bauer − *Department of Applied Physics, Science for Life Laboratory, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden;* ⊙ orcid.org/0000-0002-2268-0065

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jctc.2c00517

Author Contributions

[#]P.B. and N.A. are co-first authors and contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99*, 1129−1143.

(2) Groenhof, G.; Modi, V.; Morozov, D. Observe while it happens: catching photoactive proteins in the act with non-adiabatic molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2020**, *61*, 106−112.

(3) Páll, S.; Zhmurov, A.; Bauer, P.; Abraham, M.; Lundborg, M.; Gray, A.; Hess, B.; Lindahl, E. Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS. *J. Chem. Phys.* **2020**, *153*, 134110.

(4) Bottaro, S.; Lindorff-Larsen, K. Biophysical experiments and biomolecular simulations: A perfect match? *Science* **2018**, *361*, 355−360.

(5) Shaw, D. E.; Adams, P. J.; Azaria, A.; Bank, J. A.; Batson, B.; Bell, A.; Bergdorf, M.; Bhatt, J.; Butts, J. A.; Correia, T.; Dirks, R. M.; Dror, R. O.; Eastwood, M. P.; Edwards, B.; Even, A.; Feldmann, P.; Fenn, M.; Fenton, C. H.; Forte, A.; Gagliardo, J.; Gill, G.; Gorlatova, M.; Greskamp, B.; Grossman, J.; Gullingsrud, J.; Harper, A.; Hasenplaugh, W.; Heily, M.; Heshmat, B. C.; Hunt, J.; Ierardi, D. J.; Iserovich, L.; Jackson, B. L.; Johnson, N. P.; Kirk, M. M.; Klepeis, J. L.; Kuskin, J. S.; Mackenzie, K. M.; Mader, R. J.; McGowen, R.; McLaughlin, A.; Moraes, M. A.; Nasr, M. H.; Nociolo, L. J.; O'Donnell, L.; Parker, A.; Peticolas, J. L.; Pocina, G.; Predescu, C.; Quan, T.; Salmon, J. K.; Schwink, C.; Shim, K. S.; Siddique, N.; Spengler, J.; Szalay, T.; Tabladillo, R.; Tartler, R.; Taube, A. G.; Theobald, M.; Towles, B.; Vick, W.; Wang, S. C.; Wazlowski, M.; Weingarten, M. J.; Williams, J.

M.; Yuh, K. A. Anton 3: twenty microseconds of molecular dynamics simulation before lunch. Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. In *SC '21: The International Conference for High Performance Computing, Networking, Storage and Analysis*, St. Louis, MO, USA, Nov 14−19, 2021; de Supinski, B. R., Hall, M. W., Gamblin, T., Eds.; ACM, 2021; pp 1−11.

(6) Baptista, A. M.; Teixeira, V. H.; Soares, C. M. Constant-pH molecular dynamics using stochastic titration. *J. Chem. Phys.* **2002**, *117*, 4184−4200.

(7) Bürgi, R.; Kollman, P. A.; van Gunsteren, W. F. Simulating proteins at constant pH: An approach combining molecular dynamics and Monte Carlo simulation. *Proteins: Struct., Funct., Bioinf.* **2002**, *47*, 469−480.

(8) Mongan, J.; Case, D. A.; McCammon, J. A. Constant pH molecular dynamics in generalized Born implicit solvent. *J. Comput. Chem.* **2004**, *25*, 2038−2048.

(9) Lee, M. S.; Salsbury, F. R., Jr.; Brooks, C. L., III Constant-pH molecular dynamics using continuous titration coordinates. *Proteins: Struct., Funct., Bioinf.* **2004**, *56*, 738−752.

(10) 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.

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

(12) 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.

(13) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19−25.

(14) Kong, X.; Brooks, C. L., III λ-dynamics: A new approach to free energy calculations. *J. Chem. Phys.* **1996**, *105*, 2414−2423.

(15) Aho, N.; Buslaev, P.; Jansen, A.; Bauer, P.; Groenhof, G.; Hess, B. Scalable Constant pH Molecular Dynamics in GROMACS. *J. Chem. Theory Comput.* **2022**, DOI: 10.1021/acs.jctc.2c00516.

(16) Klimovich, P. V.; Mobley, D. L. Predicting hydration free energies using all-atom molecular dynamics simulations and multiple starting conformations. *Journal of Computer-Aided Molecular Design* **2010**, *24*, 307−316.

(17) Vila-Viçosa, D.; Reis, P. B.; Baptista, A. M.; Oostenbrink, C.; Machuqueiro, M. A pH replica exchange scheme in the stochastic titration constant-pH MD method. *J. Chem. Theory Comput.* **2019**, *15*, 3108−3116.

(18) Meng, Y.; Roitberg, A. E. Constant pH Replica Exchange Molecular Dynamics in Biomolecules Using a Discrete Protonation Model. *J. Chem. Theory Comput.* **2010**, *6*, 1401−1412.

(19) 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.

(20) Damjanovic, A.; Miller, B. T.; Okur, A.; Brooks, B. R. Reservoir pH replica exchange. *J. Chem. Phys.* **2018**, *149*, 072321.

(21) Wallace, J. A.; Shen, J. K. Charge-leveling and proper treatment of long-range electrostatics in all-atom molecular dynamics at constant pH. *J. Chem. Phys.* **2012**, *137*, 184105.

(22) Simonson, T.; Carlsson, J.; Case, D. A. Proton binding to proteins: p K a calculations with explicit and implicit solvent models. *J. Am. Chem. Soc.* **2004**, *126*, 4167−4180.

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

(24) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An Nlog (N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(25) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.

(26) Cisneros, G. A.; Karttunen, M.; Ren, P.; Sagui, C. Classical electrostatics for biomolecular simulations. *Chem. Rev.* **2014**, *114*, 779−814.

(27) Hub, J. S.; de Groot, B. L.; Grubmüller, H.; Groenhof, G. Quantifying artifacts in Ewald simulations of inhomogeneous systems with a net charge. *J. Chem. Theory Comput.* **2014**, *10*, 381−390.

(28) Chen, W.; Wallace, J. A.; Yue, Z.; Shen, J. K. Introducing Titratable Water to All-Atom Molecular Dynamics at Constant pH. *Biophys. J.* **2013**, *105*, L15−L17.

(29) Donnini, S.; Ullmann, R. T.; Groenhof, G.; Grubmüller, H. Charge-neutral constant pH molecular dynamics simulations using a parsimonious proton buffer. *J. Chem. Theory Comput.* **2016**, *12*, 1040−1051.

(30) Hünenberger, P. H.; McCammon, J. A. Ewald artifacts in computer simulations of ionic solvation and ion−ion interaction: A continuum electrostatics study. *J. Chem. Phys.* **1999**, *110*, 1856−1872.

(31) Rocklin, G. J.; Mobley, D. L.; Dill, K. A.; Hünenberger, P. H. Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: An accurate correction scheme for electrostatic finite-size effects. *J. Chem. Phys.* **2013**, *139*, 184103.

(32) Huang, J.; MacKerell, A. D., Jr CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of computational chemistry* **2013**, *34*, 2135−2145.

(33) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; De Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71−73.

(34) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulatin liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(35) Durell, S. R.; Brooks, B. R.; Ben-Naim, A. Solvent-induced forces between two hydrophilic groups. *J. Phys. Chem.* **1994**, *98*, 2198−2202.

(36) Neria, E.; Fischer, S.; Karplus, M. Simulation of activation free energies in molecular systems. *J. Chem. Phys.* **1996**, *105*, 1902−1921.

(37) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(38) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.

(39) Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. LINCS: a linear constraint solver for molecular simulations. *Journal of computational chemistry* **1997**, *18*, 1463−1472.

(40) Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of computational chemistry* **1992**, *13*, 952−962.

(41) Knight, J. L.; Brooks, C. L., III λ-Dynamics free energy simulation methods. *Journal of computational chemistry* **2009**, *30*, 1692−1700.

(42) Knight, J. L.; Brooks, C. L., III Multisite λ dynamics for simulated structure−activity relationship studies. *J. Chem. Theory Comput.* **2011**, *7*, 2728−2739.

(43) Singhal, A. K.; Chien, K. Y.; Wu, W. G.; Rule, G. S. Solution structure of cardiotoxin V from Naja naja atra. *Biochemistry* **1993**, *32*, 8036−8044.

(44) Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **1935**, *3*, 300−313.

(45) Torrie, G. M.; Valleau, J. P. Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chem. Phys. Lett.* **1974**, *28*, 578−581.

(46) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of computational chemistry* **1992**, *13*, 1011−1021.

(47) Hub, J. S.; de Groot, B. L.; van der Spoel, D. g_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theory Comput.* **2010**, *6*, 3713−3720.

(48) Granovsky, A. A. *Firefly version 7.1.G*; 2012; http://classic.chem.msu.su/gran/gamess/index.html (accessed 01−07−2022).

(49) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A., Jr General atomic and molecular electronic structure system. *Journal of computational chemistry* **1993**, *14*, 1347−1363.

(50) Lindahl, V.; Lidmar, J.; Hess, B. Accelerated weight histogram method for exploring free energy landscapes. *J. Chem. Phys.* **2014**, *141*, 044110.

(51) Lundborg, M.; Lidmar, J.; Hess, B. The accelerated weight histogram method for alchemical free energy calculations. *J. Chem. Phys.* **2021**, *154*, 204103.

(52) Okada, T.; Sugihara, M.; Bondar, A.-N.; Elstner, M.; Entel, P.; Buss, V. The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. *Journal of molecular biology* **2004**, *342*, 571−583.

(53) Cherezov, V.; Rosenbaum, D. M.; Hanson, M. A.; Rasmussen, S. G.; Thian, F. S.; Kobilka, T. S.; Choi, H.-J.; Kuhn, P.; Weis, W. I.; Kobilka, B. K.; Stevens, R. C. High-resolution crystal structure of an engineered human β2-adrenergic G protein−coupled receptor. *science* **2007**, *318*, 1258−1265.

(54) Warne, T.; Moukhametzianov, R.; Baker, J. G.; Nehmé, R.; Edwards, P. C.; Leslie, A. G.; Schertler, G. F.; Tate, C. G. The structural basis for agonist and partial agonist action on a β1-adrenergic receptor. *Nature* **2011**, *469*, 241−244.

(55) Glukhova, A.; Thal, D. M.; Nguyen, A. T.; Vecchio, E. A.; Jörg, M.; Scammells, P. J.; May, L. T.; Sexton, P. M.; Christopoulos, A. Structure of the adenosine A1 receptor reveals the basis for subtype selectivity. *Cell* **2017**, *168*, 867−877.

(56) Rodriguez-Espigares, I.; Torrens-Fontanals, M.; Tiemann, J. K. S.; Aranda-Garcia, D.; Ramirez-Anguita, J. M.; Stepniewski, T. M.; Worp, N.; Varela-Rial, A.; Morales-Pastor, A.; Medel-Lacruz, B.; Pandy-Szekeres, G.; Mayol, E.; Giorgino, T.; Carlsson, J.; Deupi, X.; Filipek, S.; Filizola, M.; Gomez-Tamayo, J. C.; Gonzalez, A.; Gutierrez-de-Teran, H.; Jimenez-Roses, M.; Jespers, W.; Kapla, J.; Khelashvili, G.; Kolb, P.; Latek, D.; Marti-Solano, M.; Matricon, P.; Matsoukas, M.-T.; Miszta, P.; Olivella, M.; Perez-Benito, L.; Provasi, D.; Rios, S.; Torrecillas, I. R.; Sallander, J.; Sztyler, A.; Vasile, S.; Weinstein, H.; Zachariae, U.; Hildebrand, P. W.; De Fabritiis, G.; Sanz, F.; Gloriam, D. E.; Cordomi, A.; Guixa-Gonzalez, R.; Selent, J.; et al. GPCRmd uncovers the dynamics of the 3D-GPCRome. *Nat. Methods* **2020**, *17*, 777−787.

(57) Pavlova, A.; Zhang, Z.; Acharya, A.; Lynch, D. L.; Pang, Y. T.; Mou, Z.; Parks, J. M.; Chipot, C.; Gumbart, J. C. Machine Learning Reveals the Critical Interactions for SARS-CoV-2 Spike Protein Binding to ACE2. *J. Phys. Chem. Lett.* **2021**, *12* (23), 5494−5502.

(58) Massey, F. J., Jr The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association* **1951**, *46*, 68−78.

(59) Buslaev, P.; Gordeliy, V.; Grudinin, S.; Gushchin, I. Principal component analysis of lipid molecule conformational changes in molecular dynamics simulations. *J. Chem. Theory Comput.* **2016**, *12*, 1019−1028.

(60) Buslaev, P.; Mustafin, K.; Gushchin, I. Principal component analysis highlights the influence of temperature, curvature and cholesterol on conformational dynamics of lipids. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **2020**, *1862*, 183253.

(61) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters* **1999**, *314*, 141−151.

(62) Rosta, E.; Hummer, G. Error and efficiency of replica exchange molecular dynamics simulations. *J. Chem. Phys.* **2009**, *131*, 165102.

(63) 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.

(64) Goh, G. B.; Hulbert, B. S.; Zhou, H.; Brooks, C. L., III Constant pH molecular dynamics of proteins in explicit solvent with proton tautomerism. *Proteins: Struct., Funct., Bioinf.* **2014**, *82*, 1319−1331.

(65) Khandogin, J.; Brooks, C. L. Toward the accurate first-principles prediction of ionization equilibria in proteins. *Biochemistry* **2006**, *45*, 9363−9373.

(66) Itoh, S. G.; Damjanović, A.; Brooks, B. R. pH replica-exchange method based on discrete protonation states. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 3420−3436.

(67) Dobrev, P.; Donnini, S.; Groenhof, G.; Grubmüller, H. Accurate Three States Model for Amino Acids with Two Chemically Coupled Titrating Sites in Explicit Solvent Atomistic Constant pH Simulations and p K a Calculations. *J. Chem. Theory Comput.* **2017**, *13*, 147−160.

(68) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.

(69) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950−1958.

(70) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. *J. Chem. Theory Comput.* **2016**, *12*, 281−296.

(71) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; MacKerell, A. D., Jr Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi 1$ and $\chi 2$ dihedral angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257−3273.

(72) Sofronov, O. O.; Giubertoni, G.; Perez de Alba Ortiz, A.; Ensing, B.; Bakker, H. J. Peptide Side-COOH Groups Have Two Distinct Conformations under Biorelevant Conditions. *journal of physical chemistry letters* **2020**, *11*, 3466−3472.

(73) Chiang, C.-M.; Chien, K.-Y.; Lin, H.-j.; Lin, J.-F.; Yeh, H.-C.; Ho, P.-l.; Wu, W.-g. Conformational change and inactivation of membrane phospholipid-related activity of cardiotoxin V from Taiwan cobra venom at acidic pH. *Biochemistry* **1996**, *35*, 9167−9176.

(74) Chiang, C.-M.; Chang, S.-L.; Lin, H.-j.; Wu, W.-g. The role of acidic amino acid residues in the structural stability of snake cardiotoxins. *Biochemistry* **1996**, *35*, 9177−9186.