GPU-Accelerated Implementation of Continuous Constant pH Molecular Dynamics in Amber: $pK_a$ Predictions with Single-pH Simulations

Robert C. Harris and Jana Shen*

Department of Pharmaceutical Sciences, University of Maryland School of Pharmacy, Baltimore, MD 21201

E-mail: jana.shen@rx.umaryland.edu
Abstract

We present a GPU implementation of the continuous constant pH molecular dynamics (CpHMD) based on the most recent generalized Born implicit-solvent model in the pmemd engine of the Amber molecular dynamics package. A series of 2-ns single-pH simulations were performed for over 120 titratable residues in 10 benchmark proteins. The calculated pKₐ’s showed a root-mean-square deviation of 0.80 and correlation coefficient of 0.83 with respect to experiment. 90% of the pKₐ’s were converged with estimated errors below 0.1 pH units. Surprisingly, this level of accuracy is slightly better than our previous replica-exchange simulations with an exchange attempt frequency of 2 ps⁻¹ (Huang, Harris and Shen, J Chem Info Model, 2018). Interestingly, for the linked titration sites in two enzymes, the single-pH titration reproduced the experimental macroscopic pKₐ values to within 1 pH unit, whereas the aforementioned replica-exchange simulations significantly underestimated the pKₐ splitting, due to insufficient conformational relaxation in response to pH changes. A 2-ns single-pH simulation of a 400-residue protein takes about one hour on a single NVIDIA GeForce RTX 2080 graphics card, which is over 1000 times faster than a CpHMD run on a single CPU core of a high-performance computing cluster node. These data suggest that the GPU-accelerated continuous CpHMD has the accuracy and speed for routine deployment in a variety of applications, from assisting MD simulations with protonation state assignment to offering pH-dependent corrections of binding free energies and identifying reactive hot spots for covalent drug design.
INTRODUCTION

Molecular dynamics (MD) simulations are widely used in structure-based drug design and to understand biophysical processes on a molecular level,\textsuperscript{1,2} but one weakness of the most common simulations is that they require the protonation states of the titratable sites to be fixed. Typically, the protonation states of these residues are set to match those of model compounds representing single amino acids in water at pH 7, but these states may not be correct, as the p\textsubscript{K}\textsubscript{a} of a titratable site in the protein environment may differ significantly from the model p\textsubscript{K}\textsubscript{a} in solution. Sometimes, structure-based p\textsubscript{K}\textsubscript{a} prediction tools, such as the empirical method Propka\textsuperscript{3} and the Poisson-Boltzmann solver based methods APBS-PDB2PQR,\textsuperscript{4} H++,\textsuperscript{5} and DelPhiPKa,\textsuperscript{6} are used to obtain estimates of the protonation states prior to running simulations; however, these tools do not directly account for the dynamic flexibility of the protein, which may lead to incorrect assignment of protonation states. Most importantly, even if the initial assignment is correct, protonation states may change in the course of conformational dynamics, as demonstrated in pH-dependent protein folding,\textsuperscript{7,8} protein-ligand binding,\textsuperscript{9–11} enzyme catalysis,\textsuperscript{12} and ion/substrate transport across the membrane.\textsuperscript{13,14}

One solution to the above problems is to use constant pH molecular dynamics (CpHMD) methods to determine protonation states on the fly during the MD simulation.\textsuperscript{15–20} CpHMD methods treat the pH as a macroscopic thermodynamic variable of the simulation, similar to how temperature and pressure are commonly handled in MD simulations. By linking the protonation states of the titratable residues to the solute conformation, the effects of conformational flexibility on the p\textsubscript{K}\textsubscript{a} estimates are explicitly taken into account. Several CpHMD methods have been developed in the past, and they generally fall into two categories: The Monte-Carlo based methods,\textsuperscript{15,16,18,20} which maintain a constant pH by periodically attempting Metropolis Monte Carlo moves to protonate or deprotonate titratable sites and continuous CpHMD,\textsuperscript{17,21–26} which is rooted in the $\lambda$-dynamics method for free energy calculations\textsuperscript{27} and treats the protonation states of the titratable sites as continuous dynamic variables of the system. Continuous CpHMD lets systems escape local energy minima by allowing transient access to partially protonated states which are unphysical and must be discarded from the final p\textsubscript{K}\textsubscript{a} calculations. In contrast, Monte-Carlo based CpHMD allows only fully protonated or deprotonated states, but as a result they tend to converge more slowly. Additionally, the proper way to extend these methods to fully explicit solvent is unclear.\textsuperscript{19,20} A more extensive discussion of the various CpHMD methods can be found in the reviews.\textsuperscript{28,29}

Both Monte-Carlo and continuous CpHMD simulations have historically required some importance-sampling technique, such as temperature-\textsuperscript{30,31} or pH-based replica exchange,\textsuperscript{22,32} to obtain converged p\textsubscript{K}\textsubscript{a} estimates. However, at the present time, replica exchange is not necessarily attractive for computation.
on Graphics Processing Units (GPUs), as in some molecular dynamics packages, including Amber, each replica requires a separate graphics card, and to ensure frequent exchange between neighboring replicas, a large number of parallel replicas are sometimes needed. Such simulations would require a relatively expensive GPU cluster and could prevent the method from being widely used.

Here we present the first implementation of the generalized Born (GB) based continuous CpHMD method in the GPU-accelerated *pmemd* program of the Amber molecular dynamics (MD) package. The performance of the code was benchmarked in pKₐ predictions. Considering that common users may not have access to a GPU node equipped with multiple GPU cards, single pH titration simulations were used. We restricted the simulation length to 2 ns to facilitate comparison with our previous work, where the pKₐ's were calculated using pH replica-exchange (pH-REX) titration simulations of 2 ns per replica on the CPUs. We note that even though the time scale for conformational rearrangements in implicit solvent is much shorter than in explicit solvent, 2 ns may be insufficient for conformational relaxation to specific pH conditions, particularly for buried residues and those undergoing strong electrostatic or hydrogen-bond interactions. However, the objective of this work is to demonstrate the tool not for constant pH simulations but for pKₐ predictions given a user-specified simulation time, e.g., 2 ns. Surprisingly, despite the short simulation length, the calculated pKₐ values of ten benchmark proteins agree well with our previous pH-REX results and experimental data. Interestingly, although the microscopic pKₐ's for the linked residues in two enzymes were not converged, the single-pH simulations were able to better reproduce the experimentally observed macroscopic pKₐ splitting as compared to the previous pH-REX simulations with a high exchange attempt frequency (2 ps⁻¹). On a single NVIDIA 2080 graphics card, a 2-ns single-pH simulation of a 400-residue protein takes about one hour. Thus, we envision the GPU-accelerated continuous CpHMD tool will be routinely applied to predict pKₐ's on commodity hardware.

**METHODS AND PROTOCOLS**

The Compute Unified Device Architecture (CUDA) code of the GB molecular dynamics method in the *pmemd.cuda* program of Amber (version 2018) was modified. The energies, forces, and other numbers computed with the code matched those given by our CPU implementation, subject to the precision differences between the GPU and CPU code. As in the CPU implementation, the code is currently limited to using the GB-Neck2 model for both conformational and titration dynamics. We performed single-pH simulations on 11 proteins: the 36-residue villin headpiece subdomain (HP36, pdbid 1VII), 45-residue binding domain of 2-oxoglutarate dehydrogenase multi-enzyme complex (BBL, pdbid 1W4H), 56-residue
N-terminal domain of ribosomal protein L9 (NTL9, pdbid 1CQU), 56-residue turkey ovomucoid third domain (OMTKY, pdbid 1OMU), 105-residue reduced form of human thioredoxin (pdbid 1ERT), 129-residue hen egg-white lysozyme (HEWL, pdbid 2LZT), 143-residue hyperstable ∆+PHS variant of staphylococcal nuclease (SNase, pdbid 3BDC), 124-residue ribonuclease A (RNase A, pdbid 7RSA), 155-residue E. coli ribonuclease H (RNase H1, pdbid 2RN2), 185-residue oxidized form of Bacillus circulans xylanase (xylanase, pdbid 1BCX), and 389-residue unbound β-secretase 1 catalytic domain (BACE1, pdbid 1SGZ). These proteins have been previously used to validate our CPU implementation of the same continuous CpHMD method.35

The initial structures for the simulations were taken from our previous paper.35 We started from the PDB coordinates by adding acetylated N terminus and amidated C terminus caps, building any disulfide bonds, and adding hydrogens with the CHARMM program (version c42a1).37 The protonation states of the titratable residues were set so that Asp/Glu were deprotonated and His/Lys/Arg/Cys/Tyr were protonated. The structures then underwent 50 steps of steepest descent minimization in GBSW implicit solvent38 with a harmonic force constant of 50 kcal/mol/Å² applied to each heavy atom. Next, dummy atoms were added to the Asp/Glu residues, and the structure was minimized for 10 steps of steepest descent and 10 steps of Newton-Raphson minimization. These final structures were then converted to the structure files with the Leap utility in Amber.34

The GPU single-pH simulations used the Amber ff14SB force field39 and the GB-Neck2 implicit-solvent model.36 All bonds containing hydrogens were constrained with the SHAKE algorithm,40 the salt concentration was set to 0.15 M, and a 2 fs timestep was used. For the 2 proteins without His residues (HP36 and NTL9), 14 single-pH simulations were run with pH 1-7.5 in 0.5 unit increments, and for the other proteins 18 simulations were run with pH 1-9.5 also in 0.5 unit increments. Each simulation was run for 2 ns. To calculate pKᵦ’s, the probability of deprotonation (unprotonated fraction) was fit to the generalized Henderson-Hasselbalch (HH) equation, as in our previous paper.35 For simplicity, the word generalized will be omitted in later discussions. The statistical errors in the pKᵦ’s were estimated from the covariances of the fit parameters.

The CpHMD settings and options were the same as in our previous replica-exchange simulations, except that the latter also employed a pH replica-exchange protocol, in which exchanges between adjacent pH replicas were attempted every 250 MD steps. In the current work, we also performed additional replica-exchange simulations for SNase and RNase H with exchanges attempted every 500 and 1000 MD steps. Larger pH ranges were used in the replica-exchange simulations. For the 2 proteins without His, 16 replicas were used with pH 0-7.5 in increments of 0.5 units, and for the others 20 replicas with pH ranging from 0-9.5 in increments of 0.5 units were used.35
RESULTS AND DISCUSSION

Titration simulations of model peptides. First, we verified that the GPU-based single-pH titration simulations with the model parameters from our previous work\(^{35}\) can reproduce the reference experimental pK\(_{a}\)'s of model alanine pentapeptides, CH\(_3\)CO-Ala-Ala-X-Ala-Ala-NH\(_2\), where X represents Asp, Glu, or His. These reference pK\(_{a}\)'s are also referred to elsewhere as model or solution pK\(_{a}\)'s. For each pentapeptide, 5 independent 2-ns titration simulations were performed at each of the 8 evenly spaced pH conditions with an interval of 0.5 units, in the pH ranges 2–5.5 for Asp, 2.5–6 for Glu, and 5–8.5 for His. Fig. 1 displays the unprotonated fraction versus pH (titration data) for each residue and the least-square fit to the HH equation. Fitting of all 40 data points for each peptide returned a pK\(_{a}\) of 3.71 for Asp, 4.17 for Glu, and 6.33 for His. These pK\(_{a}\)'s are within 0.1 units of the previous CPU-based replica-exchange pK\(_{a}\)'s and within 0.2 units of the experimental values (Table 1). This level of agreement is acceptable and therefore we decided that we could use the previously derived model parameters for the titration simulations of proteins.

![Figure 1](image-url)

**Figure 1:** Simulated titration data for the model alanine pentapeptides in solution. Unprotonated fraction as a function of pH for Asp (a), Glu (b), and His (c). Green curves are the best fits to the HH equation. The calculated pK\(_{a}\)'s based on all five sets of titration data and the fitting errors are shown.

Error estimates for the model pK\(_{a}\)'s. From the least-square fit of five titration datasets (40 data points) to a single HH equation, an error for the pK\(_{a}\), i.e., fitting parameter, can be obtained. This fitting error is below 0.03 for model Asp, Glu, and His, demonstrating a high precision of the pK\(_{a}\) (Fig. 1). We were also interested in the error (reproducibility) of the calculated pK\(_{a}\) given only one titration dataset (one data point for each pH). To address this question, a bootstrap method was used, where the unprotonated fractions at different pH were combined to generate every possible combination out of 5\(^8\) virtual sets of titration data, where 5 is the number of simulations per pH and 8 is the number of pH conditions. Thus, 5\(^8\) pK\(_{a}\)'s were obtained, from which an average and standard deviation were calculated. As expected, the bootstrap procedure gave identical pK\(_{a}\)'s but larger errors compared to those obtained from the best fits of
Table 1: Calculated $pK_a$’s of the model alanine pentapeptides from the GPU single-pH simulations compared to the CPU replica-exchange results and experiment.

<table>
<thead>
<tr>
<th></th>
<th>GPU Fit$^a$</th>
<th>GPU Bootstrap$^b$</th>
<th>CPU$^c$</th>
<th>Expt$^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp</td>
<td>3.7 ± 0.03</td>
<td>3.7 ± 0.10</td>
<td>3.8 ± 0.05</td>
<td>3.67 ± 0.04</td>
</tr>
<tr>
<td>Glu</td>
<td>4.2 ± 0.02</td>
<td>4.2 ± 0.06</td>
<td>4.1 ± 0.05</td>
<td>4.25 ± 0.05</td>
</tr>
<tr>
<td>His</td>
<td>6.3 ± 0.03</td>
<td>6.3 ± 0.10</td>
<td>6.4 ± 0.03</td>
<td>6.54 ± 0.04</td>
</tr>
<tr>
<td>HID$^d$</td>
<td>7.0 ± 0.03</td>
<td>7.0 ± 0.08</td>
<td>7.0 ± 0.05</td>
<td>n/d</td>
</tr>
<tr>
<td>HIE$^e$</td>
<td>6.4 ± 0.03</td>
<td>6.4 ± 0.11</td>
<td>6.5 ± 0.02</td>
<td>n/d</td>
</tr>
</tbody>
</table>

$^a$pK$_a$’s and errors from the fitting of all 40 data points.
$^b$pK$_a$’s and errors from a bootstrap procedure (see main text).
$^c$pK$_a$’s from our previous replica-exchange simulations with errors calculated as the standard deviations of the pK$_a$’s from five sets of replica-exchange simulations.
$^d$NMR-determined pK$_a$’s by Thurlkill et al. These pK$_a$’s were used as the reference (model) values in our simulations.
$^e$The pK$_a$ of HID is associated with HIP $\rightleftharpoons$ HID. The pK$_a$ of HIE is associated with HIP $\rightleftharpoons$ HIE.

all five datasets (Table 1). The bootstrap errors are only somewhat larger than the standard deviations from the CPU replica-exchanges simulations, which is surprising, as previous work based on the continuous CpHMD in CHARMM, which uses the GBSW model and CHARMM22 protein force field as well as the discrete CpHMD in Amber, which employs the GB-OBC model (igb=2) and Amber99SB force field indicated that the single-pH model titration data are very noisy, with errors much larger than those from the replica-exchange simulations. We will continue the discussion of pK$_a$ convergence and error in the context of protein titration simulations.

Timing of the GPU-accelerated simulations. To demonstrate the feasibility of GPU CpHMD titration for routine use on a single desktop equipped with a graphics card, we examined the wall clock time of 2-ns simulations of the benchmark proteins on a Nvidia GeForce RTX 2080 graphics card. The GPU wall clock time increases with the system size, from 5 minutes for the 36-residue HP39 to 17 minutes for the 185-residue xylanase, and about one hour for the 389-residue BACE1 (Fig. 2a). Compared to a single core on the AMD Opteron 6276 processor cluster node, the wall-clock speedup on the GPU increases with system size and reaches about 2400 times. Thus, CpHMD simulations no longer require a high-performance computing cluster and are significantly more cost effective on commodity GPUs.

We compared the speed of the GPU CpHMD titration with the GPU-accelerated GBSW-based continuous CpHMD implemented by Arthur and Brooks III within the CHARMM-OpenMM interface. The latter implementation was reported to output a trajectory of about 81 ns/day for SNase on a Nvidia GeForce
GTX 780 Ti graphics card.\textsuperscript{45} Using our implementation, SNase ran at 250 ns/day on an RTX 2080 card. Considering the difference in single-precision floating point operations per second (FLOPS) between 780 Ti and 2080 cards (5.3 vs. 10.1) and the Amber benchmark data for GB simulations on various graphics cards (http://ambermd.org/GPUPerformance.php), we estimated the speed of our GPU implementation to be slightly faster than the implementation in the CHARMM-OpenMM interface.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{Timing of the GPU-based titration simulations. (a) The time (in hours) required to run a 2-ns single-pH simulation on a single NVIDIA GeForce 2080 graphics card for the 11 proteins discussed in the main text. (b) GPU speedup relative to one core on the AMD Opteron 6276 processor node with 64 cores. The simulations ran on 2–8 cores and the time was extrapolated to one core.}
\end{figure}

Protein titration simulations: overall comparison to the replica-exchange simulations and experiment. We performed single-pH titration simulations of 10 benchmark proteins, comprising 4 small proteins and 6 enzymes, on a single GPU card. Each protein was run for 2 ns at single pH values between 1 and 7.5/9.5, with a pH spacing of 0.5 units. We compared the calculated pH\textsubscript{a}s to our previous replica-exchange simulations on the CPUs,\textsuperscript{45} which used the same amount of sampling (2 ns per replica, with replicas separately by 0.5 pH units). Assuringly, the pH\textsubscript{a}s are very similar, with a root-mean-square deviation (RMSD) of 0.54, linear regression slope (\textit{m}) of 0.97 and Pearson’s correlation coefficient (R) of 0.93 (Fig. 3a). In terms of the pH\textsubscript{a} shifts with respect to the model pH\textsubscript{a}s, a comparison between the single-pH simulations on the GPU and replica-exchange simulations on the CPUs gives \textit{m} of 0.77 and R of 0.84. Although part of the differences between the GPU and CPU calculated pH\textsubscript{a}s may be attributed to the numerical precision differences between the methods, i.e., the GPU uses mixed precision, whereas the CPU uses double precision, most of the differences are due to the use (or not) of replica exchange. Superficially, one might think that the replica exchange results would be more reliable, as enhanced sampling should improve convergence, but this is not always true (see later discussion).

A comparison of the calculated pH\textsubscript{a}s from the single-pH simulations and experiment yields a RMSD of 0.80, \textit{m} of 1.0 and R of 0.83 (Fig. 3c). A comparison of the corresponding pH\textsubscript{a} shifts to experiment yields \textit{m} of 0.75 and R of 0.68 (Fig. 3d). These statistics indicate that the single-pH simulations give pH\textsubscript{a}s
in slightly better agreement with experiment, as compared to the previous replica-exchange simulations,\textsuperscript{35} which had a RMSD of 0.87, R of 0.81 for pK\textsubscript{a}'s, and R of 0.61 for pK\textsubscript{a} shifts (see SI). This finding is surprising, as previous work based on both continuous and discrete constant pH MD indicated that the pK\textsubscript{a} accuracies from single-pH simulations are much lower than those from temperature or pH replica-exchange simulations.\textsuperscript{30–32}

Now we examine the calculated pK\textsubscript{a}'s of Asp, Glu, and His individually by calculating the comparison statistics (Table 2) and by plotting the histograms of the pK\textsubscript{a} differences between the GPU single-pH and CPU replica-exchange simulations or experiment (Fig. 4). Assuringly, the histograms of the differences between the single-pH and replica-exchange pK\textsubscript{a}'s for Asp, Glu, and His are peaked around 0 and span a range of -1 to 1 unit, with Asp showing a few data counts between 1-2. Both the comparison statistics and histograms of the differences between the single-pH and experimental pK\textsubscript{a}'s (or pK\textsubscript{a} errors) indicate differences among the three types of residues. The distribution of the pK\textsubscript{a} errors for His is broad and essentially flat, in contrast with those for Asp and Glu, which are narrower and peaked around 0. The finding of larger errors for His is consistent with our previous replica-exchange data\textsuperscript{35} and suggests suboptimal GB parameters. In our previous work,\textsuperscript{35} we found that the original GB-Neck2 parameters gave overly strong salt bridges involving His, which prompted us to reduce the imidazole hydrogen Born radius from 1.3 to 1.17 Å, following the modification made to the Born radius of the guanidinium hydrogens of arginine which led to a weakening of the salt-bridge strength in the GB-Neck2 parameterization work.\textsuperscript{36} Our modification of the His Born radius brought nearly all calculated His pK\textsubscript{a}'s down and closer to experiment, but the remaining discrepancies between simulation and experiment suggest a need for further improvement.

Although Asp and Glu differ in only one methylene group, the histograms of the pK\textsubscript{a} errors show noticeable differences. Most of the pK\textsubscript{a} errors for Glu are within ±0.5 units and do not extend beyond ±1; however, there are several larger pK\textsubscript{a} errors (above 2 units in magnitude) for Asp and the histogram appears to be somewhat skewed to the left (more negative errors). The sidechain of Asp is shorter than that of Glu by one methylene group and as a result, the carboxylate oxygen can form a hydrogen bond with its own or a neighbor backbone amide group. These sidechain-to-backbone hydrogen bonds were observed in our simulations with both Amber and CHARMM force fields. The hydrogen bond formation or breakage perturbs the protonation state of Asp and contributes to the larger fluctuation in the model peptide titration (Table 1) and perhaps also the protein titration simulations (Fig. 4a). In the parameterization of the GB-Neck2 model,\textsuperscript{36} the strengths of the Glu···Lys and Glu···Arg salt bridges were examined; however, the strengths of the Asp···Lys and Asp···Arg salt bridges were not. As the structural differences in Asp and Glu lead to changes in the interactions formed by the sidechains, perhaps a slight adjustment is needed for the Asp parameters, which are currently the same as for Glu.
Figure 3: Comparison of the $pK_a$'s and $pK_a$ shifts from the GPU single-pH simulations to the CPU replica-exchange simulations or experiment. The GPU $pK_a$ estimates compared to (a) CPU and (b) experiment and the corresponding plots ((c) and (d)) for the $pK_a$ shifts. The model $pK_a$'s from Thurlkill et al.\textsuperscript{41} were used for calculating the shifts. Data for Asp, Glu, and His residues are colored magenta, dark red, and blue, respectively. The diagonal ($y = x$) and linear regression lines are colored black and green, respectively. RMSD, regression slope ($m$) and Pearson’s correlation coefficient (R) are shown. The regression was performed with a zero intercept (R only changed in the second decimal point). In c and d, $pK_a$'s without precisely measured experimental values were excluded. In d, the data for Asp26 is hidden for clarity (the experimental $pK_a$ shift > 4).
Figure 4: (a-c) Comparisons of the Asp, Glu, and His pKₐ's from the GPU and CPU simulations (ΔpKₐ(CPU) = GPU pKₐ – CPU pKₐ). (d-f) Comparisons of the GPU results and experiment (ΔpKₐ(Expt) = GPU pKₐ – Expt pKₐ). (g-i) Histograms of the error estimates (δpKₐ) from the least-squares fits. The linked residues, Asp19/Asp21 in SNase and Asp10/Asp79 in RNase H, are excluded. (j-l) Histograms of the error estimates (δpKₐ) obtained from the difference between the 2-ns and 1.5-ns pKₐ estimates. The linked residues, Asp19/Asp21 of SNase and Asp10/Asp79 of RNase H, are excluded. For clarity, data point for Asp26 of Thioredoxin (δpKₐ = -0.34) is hidden.
Table 2: Comparisons of Asp, Glu, and His $pK_a$'s from the GPU single-pH simulations and CPU replica-exchange simulations and experiment$^a$

<table>
<thead>
<tr>
<th>Res</th>
<th>N</th>
<th>maxd</th>
<th>mud</th>
<th>rmsd</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>48</td>
<td>2.0</td>
<td>0.49</td>
<td>0.66</td>
<td>0.84</td>
</tr>
<tr>
<td>Glu</td>
<td>54</td>
<td>1.3</td>
<td>0.28</td>
<td>0.42</td>
<td>0.84</td>
</tr>
<tr>
<td>His</td>
<td>18</td>
<td>1.1</td>
<td>0.35</td>
<td>0.44</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>42</td>
<td>2.7</td>
<td>0.69</td>
<td>0.92</td>
<td>0.78</td>
</tr>
<tr>
<td>Glu</td>
<td>54</td>
<td>2.1</td>
<td>0.45</td>
<td>0.61</td>
<td>0.64</td>
</tr>
<tr>
<td>His</td>
<td>18</td>
<td>1.7</td>
<td>0.90</td>
<td>1.04</td>
<td>-0.08</td>
</tr>
</tbody>
</table>

$^a$The maximum unsigned deviation (maxd), mean unsigned deviation (mud), root-mean-square deviation (rmsd), and Pearson's correlation coefficient (R) are listed. Only residues with precisely measured experimental values are included in the single pH vs. experiment comparisons. BACE1 residues are excluded.

Convergence and mixed fractions in the single-pH simulations of proteins. To assess the convergence of protonation-state sampling and $pK_a$'s, we examined the running estimates of the unprotonated fractions at different pH and the running estimates of the $pK_a$ values (Fig. S6). The unprotonated fractions of the majority of the 120 residues with experimental values and the corresponding $pK_a$'s appear to stabilize within 2 ns. The convergence behavior seems comparable to the previous replica-exchange simulations, consistent with the model peptide titration data discussed earlier. We note that $pK_a$, which is calculated by fitting unprotonated fractions from all pH conditions, converges more quickly than unprotonated fractions. For example, the $pK_a$ of Asp162 in BBL is converged, although the unprotonated fraction at one of the pH conditions appears to slightly increase between 1 ns and 2 ns (Fig. S6).

To quantify the convergence of the calculated $pK_a$'s, we plotted the histograms of the convergence error estimates, obtained as the $pK_a$ differences between the 2-ns and 1.5-ns $pK_a$ simulations for Asp, Glu, and His (Fig. 4g–i). The errors for Glu and His $pK_a$'s fall in the range of ±0.1. Except Asp19/Asp21 in SNase, Asp10/Asp70 in RNase H, and Asp26 in thioredoxin, the errors for Asp $pK_a$'s fall in the range of ±0.2. The slower convergence of the Asp $pK_a$'s compared to Glu $pK_a$'s is consistent with the larger deviations between the calculated and experimental $pK_a$'s of Asp. Curiously, the titration data of Asp19/Asp21 in SNase and Asp10/Asp70 in RNase H do not follow a single sigmoidal trend. We will come back to the discussion of these residues.

One potential drawback of continuous CpHMD is the presence of mixed states, defined here as $\lambda$ values between 0.2 and 0.8. Although mixed states are necessary and their presence can accelerate sampling,
His150 is deeply buried, with an experimental pK_a sampling. The worst case is His150 of xylanase, which had a mixed fraction of 35 to 55% at pH below 6. His114 of RNase H, and His12/His48 of RNaseA, and none of them displayed converged protonation-state in the continuous CpHMD simulations in CHARMM. It remains to be seen if they decrease in much longer trajectories.

Table 3: Protein pK_a’s calculated from the GPU single-pH simulations compared to the CPU replica-exchange simulations and experiment

<table>
<thead>
<tr>
<th>Residue</th>
<th>Exp</th>
<th>CPU</th>
<th>GPU</th>
<th>Residue</th>
<th>Exp</th>
<th>CPU</th>
<th>GPU</th>
<th>Residue</th>
<th>Exp</th>
<th>CPU</th>
<th>GPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp129</td>
<td>3.9</td>
<td>2.8</td>
<td>3.0(0.09)</td>
<td>Asp44</td>
<td>3.1</td>
<td>2.1</td>
<td>2.4(0.06)</td>
<td>Asp8</td>
<td>2.7</td>
<td>2.5</td>
<td>2.9(0.08)</td>
</tr>
<tr>
<td>Glu141</td>
<td>4.5</td>
<td>4.1</td>
<td>3.9(0.04)</td>
<td>Glu45</td>
<td>4.0</td>
<td>3.7</td>
<td>3.8(0.02)</td>
<td>Glu11</td>
<td>4.1</td>
<td>3.9</td>
<td>4.1(0.02)</td>
</tr>
<tr>
<td>His142</td>
<td>6.5</td>
<td>6.9</td>
<td>7.2(0.01)</td>
<td>Asp46</td>
<td>3.5</td>
<td>3.6</td>
<td>3.2(0.03)</td>
<td>Glu20</td>
<td>3.2</td>
<td>3.7</td>
<td>3.8(0.05)</td>
</tr>
<tr>
<td>Asp145</td>
<td>3.7</td>
<td>2.6</td>
<td>2.5(0.06)</td>
<td>Glu72</td>
<td>4.4</td>
<td>4.1</td>
<td>4.1(0.05)</td>
<td>Asp28</td>
<td>2.3</td>
<td>3.6</td>
<td>4.3(0.05)</td>
</tr>
<tr>
<td>Glu161</td>
<td>3.7</td>
<td>3.3</td>
<td>3.6(0.07)</td>
<td>Glu44</td>
<td>4.8</td>
<td>4.6</td>
<td>4.5(0.02)</td>
<td>Asp48</td>
<td>1.4</td>
<td>1.8</td>
<td>1.8(0.02)</td>
</tr>
<tr>
<td>Asp162</td>
<td>3.2</td>
<td>3.2</td>
<td>2.6(0.06)</td>
<td>RNase A</td>
<td>7.5</td>
<td>6.6</td>
<td>6.2(0.07)</td>
<td>Glu53</td>
<td>7.5</td>
<td>6.6</td>
<td>6.2(0.07)</td>
</tr>
<tr>
<td>Glu164</td>
<td>4.5</td>
<td>4.0</td>
<td>4.1(0.10)</td>
<td>Glu2</td>
<td>2.8</td>
<td>3.2</td>
<td>3.2(0.09)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His166</td>
<td>5.4</td>
<td>6.0</td>
<td>5.8(0.04)</td>
<td>Glu9</td>
<td>4.0</td>
<td>3.4</td>
<td>3.8(0.03)</td>
<td>Asp10</td>
<td>6.1</td>
<td>4.8</td>
<td>7.1(0.18)</td>
</tr>
<tr>
<td>NT9</td>
<td></td>
<td></td>
<td></td>
<td>Asp14</td>
<td>2.0</td>
<td>2.4</td>
<td>3.8(0.08)</td>
<td>Asp10*</td>
<td>6.1</td>
<td>4.8</td>
<td>7.1(0.18)</td>
</tr>
<tr>
<td>Glu6</td>
<td>4.8</td>
<td>3.9</td>
<td>4.8(0.11)</td>
<td>His119</td>
<td>6.1</td>
<td>6.1</td>
<td>5.8(0.03)</td>
<td>Asp94</td>
<td>3.2</td>
<td>3.2</td>
<td>3.0(0.07)</td>
</tr>
<tr>
<td>Glu13</td>
<td>4.4</td>
<td>4.4</td>
<td>4.5(0.03)</td>
<td>Asp121</td>
<td>3.1</td>
<td>3.5</td>
<td>2.4(0.02)</td>
<td>Asp102</td>
<td>&lt;2.0</td>
<td>3.4</td>
<td>3.5(0.15)</td>
</tr>
<tr>
<td>Asp16</td>
<td>4.0</td>
<td>4.0</td>
<td>4.1(0.06)</td>
<td>Asp108</td>
<td>3.2</td>
<td>3.1</td>
<td>3.0(0.04)</td>
<td>Glu67</td>
<td>3.8</td>
<td>4.5</td>
<td>4.2(0.01)</td>
</tr>
<tr>
<td>Asp20</td>
<td>3.8</td>
<td>2.9</td>
<td>3.0(0.01)</td>
<td>Xylanase</td>
<td>His114</td>
<td>&lt;5.0</td>
<td>7.0</td>
<td>6.8(0.05)</td>
<td>Glu119</td>
<td>3.3</td>
<td>3.9</td>
</tr>
<tr>
<td>Asp26</td>
<td>9.9</td>
<td>6.2</td>
<td>7.0(0.27)</td>
<td>Asp5</td>
<td>3.0</td>
<td>3.3</td>
<td>4.2(0.06)</td>
<td>Glu119</td>
<td>4.1</td>
<td>3.9</td>
<td>3.8(0.04)</td>
</tr>
<tr>
<td>His43</td>
<td>6.1</td>
<td>5.9</td>
<td>3.0(0.03)</td>
<td>Asp12</td>
<td>2.5</td>
<td>2.5</td>
<td>3.1(0.06)</td>
<td>His124</td>
<td>7.1</td>
<td>6.2</td>
<td>5.5(0.06)</td>
</tr>
<tr>
<td>Glu47</td>
<td>4.1</td>
<td>4.3</td>
<td>4.0(0.04)</td>
<td>Glu79</td>
<td>4.6</td>
<td>5.1</td>
<td>5.6(0.10)</td>
<td>His127</td>
<td>7.9</td>
<td>6.6</td>
<td>6.4(0.02)</td>
</tr>
<tr>
<td>Glu56</td>
<td>3.3</td>
<td>4.5</td>
<td>3.9(0.21)</td>
<td>Asp84</td>
<td>&lt;2.0</td>
<td>3.4</td>
<td>3.9(0.08)</td>
<td>Glu129</td>
<td>3.6</td>
<td>4.6</td>
<td>3.3(0.05)</td>
</tr>
<tr>
<td>Asp58</td>
<td>5.3</td>
<td>3.8</td>
<td>4.6(0.25)</td>
<td>Asp102</td>
<td>&lt;2.0</td>
<td>3.3</td>
<td>4.3(0.08)</td>
<td>Glu131</td>
<td>4.3</td>
<td>4.3</td>
<td>4.5(0.05)</td>
</tr>
<tr>
<td>Asp60</td>
<td>2.8</td>
<td>4.6</td>
<td>3.0(0.11)</td>
<td>Asp107</td>
<td>2.7</td>
<td>3.1</td>
<td>3.6(0.14)</td>
<td>Asp134</td>
<td>4.1</td>
<td>4.3</td>
<td>4.3(0.12)</td>
</tr>
<tr>
<td>Asp61</td>
<td>4.2</td>
<td>4.6</td>
<td>4.1(0.05)</td>
<td>Asp120</td>
<td>3.2</td>
<td>4.0</td>
<td>3.6(0.06)</td>
<td>Glu135</td>
<td>4.3</td>
<td>4.2</td>
<td>4.2(0.04)</td>
</tr>
<tr>
<td>Asp64</td>
<td>3.2</td>
<td>3.1</td>
<td>2.8(0.03)</td>
<td>Asp132</td>
<td>3.6</td>
<td>3.4</td>
<td>3.4(0.07)</td>
<td>Glu147</td>
<td>4.2</td>
<td>3.9</td>
<td>3.7(0.09)</td>
</tr>
<tr>
<td>Glu68</td>
<td>4.9</td>
<td>4.3</td>
<td>4.2(0.05)</td>
<td>His150</td>
<td>2.3</td>
<td>4.8</td>
<td>3.9(0.08)</td>
<td>Asp148</td>
<td>&lt;2.0</td>
<td>2.4</td>
<td>2.8(0.05)</td>
</tr>
<tr>
<td>Glu70</td>
<td>4.6</td>
<td>5.0</td>
<td>5.1(0.02)</td>
<td>His157</td>
<td>6.5</td>
<td>7.3</td>
<td>7.5(0.03)</td>
<td>Glu154</td>
<td>4.4</td>
<td>3.8</td>
<td>3.6(0.05)</td>
</tr>
<tr>
<td>Glu88</td>
<td>3.7</td>
<td>3.8</td>
<td>3.3(0.04)</td>
<td>Glu173</td>
<td>6.7</td>
<td>7.0</td>
<td>6.9(0.13)</td>
<td>all</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu95</td>
<td>4.1</td>
<td>3.5</td>
<td>3.9(0.02)</td>
<td>Glu98</td>
<td>3.9</td>
<td>3.9</td>
<td>3.6(0.07)</td>
<td>BACE1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu103</td>
<td>4.4</td>
<td>4.7</td>
<td>4.6(0.01)</td>
<td>Glu103</td>
<td>5.2</td>
<td>4.2</td>
<td>1</td>
<td>Asp32*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu103</td>
<td>4.4</td>
<td>4.7</td>
<td>4.6(0.01)</td>
<td>Glu103</td>
<td>5.2</td>
<td>4.2</td>
<td>1</td>
<td>Asp228*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*aExperimental pK_a’s were taken from our previous work. The CPU data were taken from our previous replica-exchange simulations with an exchange attempt frequency (EAF) of 2 ps^-1. The errors in the pK_a’s estimated from the least-square fits are given in parentheses. The pK_a’s of residues with linked titration were obtained from the best fits to the coupled titration model (Equation 1). They are unphysical and a high occupancy of them may lower the accuracy of a CpHMD simulation. We were puzzled that the previous replica-exchange simulations showed somewhat higher mixed fractions than the continuous CpHMD simulations in CHARMM. In contrast, the mixed fractions in the current single-pH simulations are very low, below 10% at all pH for a majority of residues (Fig. S6), consistent with the CpHMD simulations in CHARMM with a lower replica exchange frequency (see later discussion). A few residues which showed a mixed fraction above 30% are: Asp26 of thioredoxin, Glu79/Asp84/His150 of xylanase, His114 of RNase H, and His12/His48 of RNaseA, and none of them displayed converged protonation-state sampling. The worst case is His150 of xylanase, which had a mixed fraction of 35 to 55% at pH below 6. His150 is deeply buried, with an experimental pK_a below 2.3, compared to the estimate of 3.9 from our 2-ns simulations. 2 ns is likely too short to complete the conformational rearrangement accompanying the protonation of His150. While the present data show that the mixed fractions are correlated with insufficient sampling, it remains to be seen if they decrease in much longer trajectories.
Error estimates for the protein pKₐ's. Unlike replica-exchange titration, single-pH simulations are independent of each other and therefore the pKₐ error obtained in the least-square fit of one set of titration data may be a good representation of the statistical error. To test this hypothesis, we performed an additional three sets of single-pH simulations at 9 evenly spaced pH conditions between 5 and 9 for RNase A to explore the statistical errors of the His pKₐ's. Table 4 lists the calculated pKₐ's and least-square fitting errors based on the original set of titration data and the pKₐ's and errors from the bootstrap procedure based on a total of 3⁹ virtual sets of titration data. The calculated pKₐ's deviate by 0.1 units at most and the fitting errors using one dataset are similar and highly correlated with the bootstrap errors. Although this test is not a rigorous proof, it suggests that the fitting error derived from a single set of titration simulations can be an approximate representation of the statistical error. The ease of obtaining error estimates is an advantage for single-pH over replica-exchange simulations, for which deriving error estimates is less straightforward.

We list the pKₐ fitting errors for all titratable residues in Table 3. We list the pKₐ fitting errors for all titratable residues in Table 3. The corresponding histograms are given in Figure 4j–l. Except for a handful of residues with fitting errors around 0.2–0.3, the errors are smaller than 0.1, demonstrating that the statistical noise in the calculated pKₐ's from the single-pH simulations is generally low.

### Table 4: Error analysis for the pKₐ's of histidines in RNase A.

<table>
<thead>
<tr>
<th>Residue</th>
<th>Fit⁺</th>
<th>Bootstrap⁺</th>
</tr>
</thead>
<tbody>
<tr>
<td>H12</td>
<td>7.5±0.11</td>
<td>7.5±0.08</td>
</tr>
<tr>
<td>H48</td>
<td>7.6±0.14</td>
<td>7.7±0.13</td>
</tr>
<tr>
<td>H105</td>
<td>6.7±0.04</td>
<td>6.8±0.06</td>
</tr>
<tr>
<td>H119</td>
<td>5.6±0.07</td>
<td>5.7±0.06</td>
</tr>
</tbody>
</table>

⁺pKₐ's and fitting errors based on the first set of simulations. ⁺pKₐ's and errors estimated from the bootstrap method using three sets of simulations.

Single-pH simulations captured linked titration and stepwise pKₐ's for SNase and RNase H. A close examination of the residues that do not display a single sigmoidal trend, Asp19/Asp21 of SNase and Asp10/Asp70 of RNase H, revealed linked (coupled) titration, i.e., protonation of one residue is linked to the protonation of the other. As a result, the titration plot of one residue shows evidence of pH-dependent titration of the other residue and the transition from the fully protonated to the fully deprotonated persists over a very wide pH range (>4 pH units for Asp19/Asp21 in RNase and Asp10/Asp70 in RNase H, see Fig. 5 and Fig. 6). For linked titration, the residue-specific pKₐ's cannot be experimentally determined. Rather, the signals, e.g., NMR chemical shifts, are fit to a coupled two-proton model, which results in two macroscopic pKₐ's representing the stepwise titration events.⁴⁶ To obtain macroscopic pKₐ's from simulations, the average number of protons bound to the two residues (Nₗₚₒₜ) at different pH values can be
fit to an analogous statistical mechanics model,\textsuperscript{24,47,48}

\[ N_{\text{prot}} = \frac{10^{pK_2 - pH} + 2 \times 10^{pK_1 + pK_2 - 2pH}}{1 + 10^{pK_2 - pH} + 10^{pK_1 + pK_2 - 2pH}}, \]  

(1)

where the denominator is the partition function, and the fitting parameters, \( pK_1 \) and \( pK_2 \), are the macroscopic \( pK_a \)'s. The resulting \( pK_a \)'s from the single-pH simulations for the linked residues in SNase are 1.5 and 6.1 (Fig. 5i) and for those in RNase H are 1.7 and 7.1 (Fig. 6i).

To identify which residue should be assigned the higher \( pK_a \), we can observe which residue becomes protonated first as pH decreases. For SNase, since the occupancy of Asp19\(^{-} \)/Asp21\(^{+} \) is higher than Asp19\(^{+} \)/Asp21\(^{-} \) at pH above 4.5, i.e., Asp21 protonates first (Fig. 5m), Asp21 has the higher \( pK_a \). Thus, we assigned Asp19 and Asp21 the \( pK_a \)'s of 1.5 and 6.1, in excellent agreement with the respective experimental assignment of 2.2 and 6.5. Similarly for RNase H, since the occupancy of Asp10\(^{+} \)/Asp70\(^{-} \) is higher than Asp10\(^{-} \)/Asp70\(^{+} \) at pH above 6 (Fig. 5j), we assigned Asp10 and Asp70 the \( pK_a \)'s of 7.1 and 1.7, in good agreement with the respective experimental assignment of 6.1 and 2.6.

**A high exchange attempt frequency leads to insufficient conformational relaxation.**

Interestingly, the residue-specific titration data of Asp19/Asp21 in SNase (Fig. 5b and f) and Asp10/Asp70 in RNase H (Fig. 5b and e) from our previous replica-exchange simulations did not offer noticeable evidence of coupling. A comparison to the single-pH titration data reveals that the occupancies of the singly protonated states are much lower while those of the mixed states are much higher, up to 30% for SNase and up to 45% for RNase H in the pH range where two singly protonated states seem to compete in the single-pH simulations. The decreased amount of coupling seen in the replica-exchange simulations is consistent with the underestimation of the higher \( pK_a \) and the splitting between the low and high \( pK_a \) observed in experiment. Since the fraction of mixed states was much higher in the replica-exchange compared to single-pH simulations, we hypothesized that the underestimation of \( pK_a \) splitting could reflect the replicas having insufficient time to relax to the specific pH conditions in the replica-exchange simulations. In these simulations,\textsuperscript{35} the exchange attempt frequency (EAF) was every 250 MD steps (or 2 ps\(^{-1} \)), rather than every 500 (1 ps\(^{-1} \)) or 1000 MD steps (0.5 ps\(^{-1} \)), as used in our previous CHARMM-based CpHMD. Our decision to increase EAF was inspired by the reported observation that high EAF improved \( pK_a \) estimates in the discrete-CpHMD simulations in Amber.\textsuperscript{32} We will come back to this discussion later.

To test if the lack of conformational relaxation is responsible for the decreased amount of coupling in our previous replica-exchange simulations, we performed additional replica-exchange simulations with EAFs of 1 ps\(^{-1} \) and 0.5 ps\(^{-1} \) for SNase and EAF of 1 ps\(^{-1} \) for RNase H. As seen from Fig. 5 and 6, at the lower EAFs,
the occupancies of the singly protonated states increase to similar levels as in the single-pH simulations, while the mixed fractions decrease to below 10% as in the single-pH simulations. Importantly, the higher macroscopic pK_a's increase by about 2 units, to similar values as from the single-pH simulations of SNase and RNase H. The increased amount of coupling is also reflected in the noisier residue-specific titration data, which no longer conform to the one-proton HH equation. Additionally, the solvent-accessible surface areas (SASA) and number of hydrogen bonds of the two residues predicted by simulations with the lower EAFs of 1 and 0.5 ps^{-1} agree more closely with the single-pH results than the simulations with the high EAF of 2 ps^{-1} (Fig. S1 and Fig. S2), which offers a direct piece of evidence for the insufficient conformational relaxation for the coupled residue pairs with the EAF of 2 ps^{-1}. Interestingly, for both proteins, the nucleophile (residue with the lower pK_a) has a larger SASA and number of hydrogen bonds compared to the proton donor (higher pK_a), in agreement with the findings from CpHMD simulations using the CHARMM force field and the GBSW model.49

Having established that a high EAF leads to insufficient conformational relaxation for linked residues, we wondered if the pK_a's of other independent residues are affected as well. A comparison of the pK_a's from the SNase simulations with different EAFs and single pH values (no exchange) showed that, although the differences are much smaller than for the couple residues, the pK_a's from the two slower exchange (EAF of 1 and 0.5 ps^{-1}) simulations agree with each other and with the single-pH simulations much better as compared to the fast exchange (EAF of 2 ps^{-1}) simulations (Table S1). A similar trend is observed for the RNase H simulations (Table S1). These data are consistent with the findings in regards to the linked titration and support the notion that an EAF of 2 ps^{-1} may not offer sufficient conformational relaxation.

**Linked titration in a much larger protein requires much longer simulations.** To further test the ability of single-pH simulations to quantitatively capture the experimental apparent pK_a's of coupled sites in enzymes, we performed 2-ns titration simulations of BACE1, a 389-residue protein, more than three times as large as HEWL, a frequently used benchmark protein for pK_a calculations,16,32 and more than twice as large as xylanase, the largest protein among the 10 benchmark proteins studied here. Our previous replica-exchange simulations with an EAF of 2 ps^{-1}49 and the hybrid-solvent continuous CpHMD in CHARMM12 correctly predicted the pK_a order for the catalytic aspartyl dyad but underestimated the higher pK_a and the amount of coupling. The current single-pH simulations also reproduced the experimental pK_a order; however, Asp228, the residue with the lower pK_a did not protonate even at pH 1 (the lowest simulation pH), due to the persistent hydrogen bond with Asp32 (data not shown). Considering the size of the protein and the strength of the hydrogen bond,49 significantly longer simulations (than 2 ns) or enhanced sampling may be required. We compared all the single-pH calculated pK_a's and the corresponding replica-exchange
results (Table S1). Although most pKₐ's are very similar (within 0.2-0.3 units), Asp83 shows the largest difference, with the single-pH pKₐ of 2.2, which is 3.1 units lower than the replica-exchange pKₐ. The lower pKₐ in the single-pH simulations was due to a salt-bridge-like interaction with Arg96 that was not observed in the replica-exchange simulation.
Figure 5: Titration data of the linked residues D19 and D21 in SNase from single-pH and replica-exchange simulations. From top to bottom rows, single-pH and replica-exchange simulations with different exchange attempt frequencies are shown. (a-h) Unprotonated fraction as functions of pH and best fit to the residue-specific HH equation for D19 (a-d) and D21 (e-h). Each data point is a pie chart showing the probability of the four protonation states: D19\(^{H}/\)D21\(^{H}\) (blue), D19\(^{-}/\)D21\(^{-}\) (red), D19\(^{H}/\)D21\(^{-}\) (magenta), and D19\(^{-}/\)D21\(^{H}\) (orange). Cyan indicates that one of the residues is in a mixed state. (i-l) Number of protons as a function of pH and best fit to the linked-titration model. (m-p) Probabilities of the singly-protonated and mixed states as functions of pH. The color scheme is the same as in (a-h). The macroscopic experimental \(pK_a\)'s are 2.2 and 6.5 for D19 and D21, respectively.
Figure 6: Simulated titration of the linked residues D10 and D70 in RNase H from single-pH and replica-exchange simulations. From top to bottom rows, single-pH and replica-exchange simulations with different exchange attempt frequencies are shown. (a-h) Unprotonated fraction as a function of pH and best fit to the residue-specific HH equation for D10 (a-d) and D70 (e-h). Each data point is a pie chart showing the probability of the four protonation states: D10$^\text{H}$/D70$^\text{H}$ (blue), D10$^-$/$D70^-$ (red), D10$^\text{H}$/D70$^-$ (magenta), and D10(-)/D70$^\text{H}$ (orange). Cyan indicates that one of the residues is in a mixed state. (i-l) Number of protons as a function of pH and best fit to the linked-titration model. (m-p) Probabilities of the singly-protonated and mixed states as functions of pH. Color scheme is the same as in (a-h). The experimental macroscopic pK$_a$'s are 6.1 and 2.6 for D10 and D70, respectively.
CONCLUDING REMARKS

We presented a GPU implementation of the generalized Born continuous CpHMD in the Amber pmemd.cuda program. A 2-ns single-pH simulation of a 400-residue protein runs in about 1 hour on a single NVIDIA GeForce 2080 graphics card, which represents a three-orders-of-magnitude speedup compared to the wall-clock time on a single CPU core of a high-performance computing cluster node (AMD Opteron 6276). The calculated $pK_a$’s of 10 proteins from 2-ns single-pH simulations were well converged and in close agreement (RMSD of 0.54) with our previous replica-exchange simulations on CPUs. Compared to the experimental $pK_a$’s, the single-pH simulations gave a RMSD of 0.80 units, which is slightly lower than the RMSD of 0.87 from our previous replica-exchange simulations. Surprisingly, the errors in the protein $pK_a$’s from the single-pH simulations were small, mostly below 0.1 and never above 0.2 units. This finding is in stark contrast to the reported data from the continuous CpHMD in CHARMM and the discrete CpHMD in Amber, which found that errors from single-pH simulations were much larger than those from replica-exchange simulations. Perhaps the most surprising finding is that the single-pH simulations were able to correctly predict the $pK_a$ order of the linked residues in SNase and RNase H and quantitatively reproduced the macroscopic $pK_a$’s. In contrast, our previous replica-exchange simulations, which used an exchange attempt frequency (EAF) of 2 ps$^{-1}$ underestimated the higher $pK_a$ by over 2 units and gave a significant amount of mixed states which were not present in the single-pH simulations. Additional replica-exchange simulations with less frequent exchange attempts gave similar $pK_a$’s and conformational behavior as the single-pH simulations, suggesting that an EAF of 2 ps$^{-1}$ does not allow sufficient conformational relaxation at specific pH conditions.

The above finding does not contradict a previous study using the GB-OBC based discrete CpHMD, which found that an EAF of 50 ps$^{-1}$ significantly improved the $pK_a$ of a salt-bridged residue (Asp66 in HEWL), which did not titrate with an EAF of 0.5 ps$^{-1}$. While our single-pH simulations did not encounter this issue (Table 3), likely because salt bridges are significantly weakened in the newer GB-Neck2 model, we suggest that conformational relaxation time for buried residues, such as the aforementioned linked pairs, is much longer than solvent-exposed ones. As a result, too frequent exchanges may not allow internal residues to adjust themselves and the local environment to a change in pH. This may explain why the $pK_a$’s of buried residues are more sensitive to the EAF, while $pK_a$’s of solvent-exposed residues are not. For example, a deeply buried Asp26 in thioredoxin has a calculated $pK_a$ of 6.2 from replica-exchange and 7.2 from single-pH simulations. The latter is in better agreement with the experimental $pK_a$ of 9.9.

It is indisputable that enhanced sampling in either temperature or pH space can accelerate $pK_a$ convergence. The $pK_a$ results from the 2-ns single-pH simulations of BACE1 illustrated this point. Although
the pK$_a$ order of the catalytic dyad, Asp32/Asp228, was correctly predicted, Asp228 did not protonate at pH 1 (the lowest simulation pH), due to a persistent hydrogen bond, which requires enhanced sampling or significantly longer sampling time to break. On the other hand, the extent of coupling and pK$_a$ splitting were significantly underestimated in the replica-exchange simulations with an EAF of 2 ps$^{-1}$. Thus, our future work will explore the usage of either longer single-pH simulations or an optimum EAF to make the most accurate and efficient predictions of pK$_a$'s for tough cases. We will also benchmark the method using a much larger dataset. Another area of interest is the improvement of His pK$_a$'s, which have the largest errors in the current dataset. Although there remains room for improvement, our present data demonstrate that single-pH simulations can be routinely performed on a desktop computer equipped with a single GPU card for a variety of applications, from assisting MD stimulations with protonation state assignment to offering pH-dependent corrections of binding free energies$^{10}$ and nucleophilic hot spots for covalent drug design.$^{50}$

SOFTWARE AVAILABILITY

The software will be distributed as a part of the Amber molecular dynamics package (Amber20).

Supporting Information Available

A supplemental table, additional analysis for linked titration, and residue-specific titration plots as well as convergence analyses for all residues of the 11 proteins are included.

ACKNOWLEDGEMENTS

The authors acknowledge National Institutes of Health (R01GM098818 and R01GM118772) and National Science Foundation (CBET143595) for funding.

References


Graphical TOC Entry