Titr-DMD – A Rapid, Coarse-Grained Quasi-All-Atom Constant *pH* Molecular Dynamics Framework

David J. Reilley,[†] Jian Wang,[‡] Nikolay V. Dokholyan,[‡,§] Anastassia N. Alexandrova[*,†,∥]

[†] Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095-1569, United States

[‡] Department of Pharmacology, Department of Biochemistry & Molecular Biology, Penn State University College of Medicine, Hershey, Pennsylvania 17033, United States

[§] Departments of Chemistry and Biomedical Engineering, Pennsylvania State University, University Park, Pennsylvania 16802, United States

[∥] California NanoSystems Institute, Los Angeles, California 90095-1569, United States

KEYWORDS: CpHMD, pK_a, DMD, Dynamics

ABSTRACT: The *pH*-dependence of enzyme fold stability and catalytic activity is a fundamentally dynamic, structural property which is difficult to study. The challenges and expense of investigating dynamic, atomic scale behavior experimentally means that computational methods, particularly constant *pH* molecular dynamics (CpHMD), are well situated tools for this. However, these methods often struggle with affordable sampling of sufficiently long timescales while also obtaining accurate pK_a prediction and verifying the structures they generate. We introduce Titr-DMD, an affordable CpHMD method that combines the quasi-all-atom coarse-grained discrete molecular dynamics (DMD) method for conformational sampling with Propka for pK_a prediction, to circumvent these issues. The combination enables rapid sampling on limited computational resources, while simulations are still performed at atomic scale. We benchmark the method on a set of proteins with experimentally attested pK_a and on the *pH* triggered conformational change in a staphylococcal nuclease mutant, a rare experimental study of such behavior. Our results show Titr-DMD to be an effective and inexpensive method to study *pH*-coupled protein dynamics.

Introduction

Solution *pH* is a chemical property with an immense effect on protein behaviors that are difficult to study at the atomic scale. Peak protein fold stability and catalytic activity are both dependent on an often narrow range of *pH*. Understanding the sequential and structural underpinning of these preferences contributes to the design and application of enzymes, particularly extremophile enzymes – which would allow for their use in harsher reaction conditions in industrial catalysis,[1–4] and answers a wide range of questions of medical interest as precise *pH* regulation is critical for cellular homeostatis.[5–7] However, this understanding demands atomistic information of fundamentally dynamic phenomena. *pH*-dependent dynamics is challenging to study experimentally, requiring a combination of techniques such as NMR monitored *pH*-titration, circular dichroism (CD) spectroscopy, and X-ray crystallography none of which alone provide the complete picture. Experimental complexity leaves computational investigation[8] as a critical tool to fill in the gaps.

Successful computational methods that assess *pH*-dependent protein behavior must accurately couple amino acid protonation state change with conformational dynamics. Typically, continuum electrostatic methods describe the protonation states of amino acids, assessing the free energy of protonation and deprotonation events or pK_a. Various solutions to the Poisson-Boltzmann equation can provide this,[9] especially the generalized Born model.[10,11] Simpler electrostatic methods are used as well. Tools such as UHBD,[12] H++,[13] and Propka[14,15] predict the pK_a of amino acid residues. Other tools,[16,17] including FPTS[18] rely on Monte Carlo simulations to sample protonation states as well as solvent and/or ion configurations in some cases. All of these methods are useful to study many *pH*-dependent protein properties, including charge regulation during complexation and prediction of some titration curves.[19] However, these methods operate with largely static structures for the protein with little or no backbone motion, and so can not fully capture *pH* dependent dynamic behavior on their own. Molecular dynamics (MD) can provide the missing conformational sampling. Such combinations are known as constant *pH* molecular dynamics (CpHMD); these methods generally use electrostatic methods to model the protonation state changes of amino acids over the course of a molecular dynamics simulation.

The appropriate sampling of *pH*-coupled dynamics is difficult to achieve for all CpHMD methods and challenging to verify. The choice of solvation model is central to sampling and broadly breaks CpHMD methods into two categories: those using explicit solvation and those using implicit solvation. Explicit solvent based methods can provide greater accuracy through atomistic solvent treatment,[20,21] but sufficient

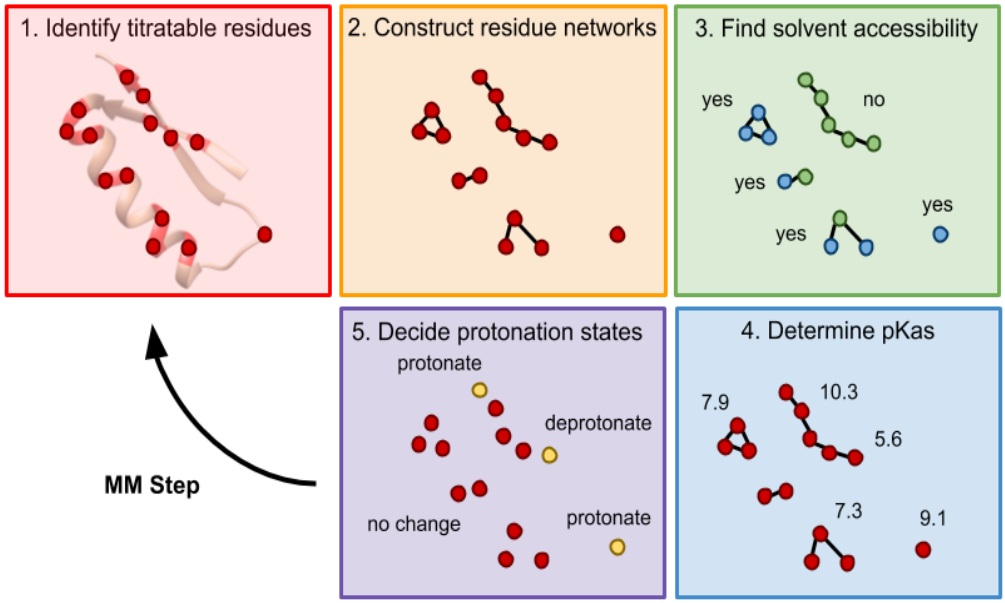


Figure 1: Schematic of the Titr-feature algorithm. The algorithm runs between short DMD (or any molecular mechanics) simulations to assign discrete protonation states.

sampling is difficult to achieve, as both conformational and protonation states need to be sampled. Furthermore, protonation sampling is affected by poor overlap between solvent configurations such that protonation state changes are often immediately rejected. To counter this, many groups have applied λ-dynamics, based off pioneering work by Brooks et al.[22] (in turn based on earlier work with other thermodynamic properties in mind),[23,24] which treats the protonation state of individual amino acid sites as continuous degrees of freedom rather than discrete ones sampled distinctly.[25–27] Other efforts focus on enhancing/accelerating conformational sampling through GPU processing[28] or replica exchange.[29–31] Implicit solvent-based methods offer increased sampling without acceleration techniques by treating the surrounding solution as a simple dielectric medium.[32–35] Implicit solvent therefore avoids the issue of solvent configuration sampling altogether.

Another approach to improve sampling is the use of coarse-grained (CG) molecular dynamics methods. CG can be implemented with either explicit or implicit solvent. Most often, CG models reduce the number of particles needed in a simulation by condensing atoms into supra-atomic beads. CpHMD methods based on these kinds of CG force fields, including Martini,[36] HiRE-RNA,[37] and OPEP6,[38] have been recently developed. Use of supra-atomic beads is, of course, more approximate that all-atom methods so this class of CpHMD methods is most attractive for particularly large systems, such as multi-protein complexes, or long timescales, such as those of protein refolding. Another form of CG is to simplify the MD force field potentials, reducing the number of calculations needed each timestep. This sort of CG is done in methods such as discrete molecular dynamics (DMD)[39–41] with square-well potentials used in place of continuous ones. Such CG potentials allow for quasi-all-atom simulations (with only some non-polar hydrogen excluded from full atomistic treatment), unlike other CG models. To our knowledge, no CpHMD methods based on the CG potential paradigm of DMD yet exist.

Ultimately, regardless of protonation scheme, solvation, and use of CG, verification of the generated ensemble of conformational and protonation states is of great important in all CpHMD methods. This is not trivial due to paucity of complementary experimental results. More plentiful indirect evidence, such as reconstruction of titration curves or estimation of experimental pK_a values, is not sufficient on its own for verification. Available results used in the past include limited helicity and secondary structural information from CD and NMR spectroscopy,[38,42–44] and occasionally X-ray crystal structures that demonstrate *pH*-dependent differences.[45] Such verification is of critical importance when first introducing a method.

We present here Titr-DMD as an undemanding method for the investigation of *pH* dependent protein behavior. Our method dynamically updates the protonation states of a DMD simulation[39–41] using pK_a predicted for instantaneous structures along its trajectory as probabilities. In the current implementation Titr-DMD uses pK_a values generated through the semi-empirical electrostatics method Propka, but is not restricted to that specific tool. It follows a generally similar approach to the early CpHMD promulgated by Baptista et al.[20] However, DMD's CG square-well potentials and implicit solvation provide rapid conformational sampling at atomic resolution on limited resources, while periodic protonation state reassessment based on Propka confers extensive protonation state sampling. Our program is highly modular for easy modification as better approaches for instantaneous pK_a prediction develop. We benchmark Titr-DMD on both its ability to calculate ensemble pK_a compared to experiment and on its ability to recapitulate the *pH*-dependent conformational change found experimentally in staphylococcal nuclease (SNase), a rare, well-described system.[46,47] Titr-DMD proves to be an effective, affordable method to study *pH*-dependent protein dynamic behavior at atomic scale.

Theory and Methods

Titr-DMD method. This method combines rapid DMD[39] conformational sampling with a custom algorithm to resolve protonation based on Propka3.1[14] pK_a predictions. Simulations are performed iteratively, alternating between a short DMD simulation and a titration (Titr)-feature that discretely assigns protonation states. The algorithm for the Titr-feature itself comprises five steps: (1) titratable residues are

identified, (2) contact networks are constructed from the identified residues, (3) the solvent accessibility of each network is determined, (4) the probability of protonation state change is determined for each network or residue, (5) protonation state changes are determined by a Monte Carlo step (Figure 1).

The intervals between protonation state reassessment are run just long enough so that protonation and deprotonation are equilibrated over the DMD simulation timescale. As isolated proton transfer events, including many individual reaction steps in proteins,[48] generally occur on the femtosecond to picosecond timescale,[49] 200 DMD steps (which is ~10 ps) suffices – a comfortable separation of 1-3 orders of magnitude. The size of the timestep allows for both the consistent and meaningful application of theory, discussed more thoroughly throughout the rest of the description of the Titr-DMD method algorithm, and extensive sampling of a system's potential protonation states. A higher reassessment frequency is therefore unnecessary and computationally expensive; while additional time spent on the Titr-feature itself is minimal, a higher frequency requires more, shorter DMD simulations and thus more time overhead during the program initialization.

Selection of titratable residues is based on their solution pK_a values. The amino acids aspartate, glutamate, histidine, cysteine, tyrosine, lysine, and arginine are the only ones to have side-chain solution pK_a values in the physiological range of *pH* 1-13 and so are the only ones considered. While significant shifts in pK_a often occur when an amino acid is part of a protein, all other residues have side-chain pK_a that fall far enough out of the physiological range to be largely irrelevant in the vast majority of systems. For the same reason, only the first protonation/deprotonation event is considered for the included amino acids; states such as doubly deprotonated lysine or doubly protonated glutamine are inaccessible. The C-terminal carboxylate and N-terminal amine could be titrated as well, but are not currently implemented due to missing DMD potentials for their less preferred states.

Contact networks are constructed on the basis of the proximity of titratable residues. First, interacting pairs of residues are identified based on their (de)protonatable heteroatoms that are within a certain cutoff distance, r_p, of each other. The protonation contact distance r_p is selected as 3.5 Å to be consistent with the DMD definition of a long hydrogen bond. Each thus defined network represents a series of residues close enough that in the timeframe of the DMD phase of the Titr-DMD simulation the proton exchange is equilibrated between them and lies firmly under thermodynamic control.

Solvent accessibility of each residue contact network is determined in a manner consistent with Propka, which defines a specific residue as buried or exposed based on its contact number, *w(N)*. *w(N)* is determined by the number of heavy atoms, *N*, within 15 Å of the residue's charge center according to

$$w(N)=\begin{cases} 0 & \text{if } N \le N_{min} \\ \dfrac{N-N_{min}}{N_{max}-N_{min}} & \text{if } N_{min} < N < N_{max} \\ 1 & \text{if } N \ge N_{max} \end{cases} \quad (1)$$

The residue is thus 0% buried if $N \le 280$ (N_{min}) and 100% buried if $N \ge 560$ (N_{max}).[14] In the Titr-feature, a network is considered solvent accessible if any residue in it is below a certain cutoff. As proton exchange is equilibrated within a network, so long as one residue is solvent accessible the rest of the network can freely exchange protons with solvent. The best value of the solvent access cutoff is a parameter in the model and is often system dependent. We find that the most appropriate value for the solvent access cutoff could range from 45% to 75% and matters most in systems with important, frequently buried residues. Alternative approaches to the solvent access cutoff are also possible. We discuss this fully in the future development of Titr-DMD section and within our test system simulations.

The probability of a protonation state change is assessed for each titratable residue based on instantaneous pK_a and the residue network information. In the current implementation, Propka3.1 is used for pK_a prediction, based on the latest structure from the preceding DMD trajectory. The protonation state change probability is then assessed for each residue. It is calculated differently depending on whether the residue is in a solvent accessible or inaccessible network. For a solvent accessible network the probability is based on the *pH* of the solution with which the residue can freely exchange protons (solvent is treated implicitly in DMD). This probability is based off the Henderson-Hasselbash equation

$$pH = pK_a + \log\left(\frac{[D]}{[P]}\right) \quad (2)$$

where [*D*] is the concentration of the deprotonated state and [*P*] is the concentration of the protonated state. Therefore, the probability of adopting the protonated state, P_P, can be defined as

$$P_P = \frac{[P]}{[P]+[D]} = \frac{10^{pK_a - pH}}{1+10^{pK_a - pH}} \quad (3)$$

In the solvent inaccessible case, only the titratable protons already present in the network can be exchanged. Buried residues not part of a network are therefore unable to change protonation state, unless, over the course of a simulation, they become solvent accessible or merge with another network. The probabilities of protonation state changes for the residues in a contact network are thus coupled; protonation state changes must be determined for the whole network at once, rather than residue by residue. Solvent inaccessible networks therefore require full enumeration of all proton configurations across the network. The preference of a proton to localize on any individual residue is determined by its pK_a, but with comparison to the competing residues in the network rather than the solution *pH*. To calculate the probability of a configuration, let *R* be the set of all residues in a network and *n* be the number of titratable protons in that network. Let $T|n(R)$ be the set of all possible proton configurations *S*, *Q*, ... such that $T|n(R) = \{S \in T(R) : |S| = n\}$. Then for every $S \in T|n(R)$ the probability of adopting that proton configuration is

$$P_C(R,S) = \frac{\prod_{s \in S} 10^{pK_a(s)}}{\sum_{Q \in T|_n(R)} \prod_{q \in Q} 10^{pK_{a,T}(q)}} \quad (4)$$

The weighting term for each proton configuration is the product of 10 raised to the pK_a of each residue that holds a proton in that state ($s \in S$, $q \in Q$, ...). Equation 4 is used to calculate the probability of each possible configuration.

Finally, protonation state changes are decided discretely by a single Monte-Carlo step based on the probabilities generated for each network. As with the probabilities, the decision differs slightly between solvent exposed and buried networks. For solvent accessible networks, a decimal between 0 and 1 is randomly generated for each residue and compared to its decimal probability. If it is above that probability the residue is unpro-

tonated, and if below it is protonated. The solvent accessible approach holds regardless of what the previous protonation state was. For solvent inaccessible networks, the decimal probabilities of all potential protonation configurations are put in a sequential order. A probability range for each configuration, *S*, is then defined as from P_L up to $P_L + P_C(R,S)$, where P_L is the sum of all configuration probabilities already considered and $P_C(R,S)$ is that of the current configuration. A decimal between 0 and 1 is then randomly generated, and the configuration is decided based on which range the random number falls within. Probabilities are generated and protonation states are decided just once during the Titr-feature step before moving on to another DMD simulation step. Any changes from the previous structure are then made, with hydrogen removed when necessary and DMD placing any new hydrogen on the appropriate heteroatoms. The structure is then ready for the next DMD simulation.

A correction is needed to maintain consistency across DMD energies in a Titr-DMD trajectory. As the Titr-feature may add and remove hydrogen by exchange with implicit solvent, the chemical composition of the system can change. As protonation state changes are done through an external program, the energy associated with them are not directly taken into account in the DMD Hamiltonian which only sees the loss and gain of hydrogen. Consequentially, the correction does not affect how protonation state changes are made, but is simply for analysis of the energy trajectory. One approach for an energy correction would be to use a value for the solvation energy of a proton, but that can not be obtained directly from experiment and can only be determined by extrapolation.[50] Values that can be obtained for the solvation energy (-264.3[51] and -265.9 kcal/mol[50,52]) are large compared to the DMD energy changes associated with structural fluctuations (ca. 100 kcal/mol). Unmodified use of the proton solvation free energy would result in unphysical behavior – Titr-DMD would always deprotonate any residue. Appropriate scaling of the solvation energy is one solution. For the current implementation of Titr-DMD an energy correction for each iteration is obtained instead based on the Propka pK_a of all residues with protonation states that deviate from the original structure. For each protonation state, take the following acid dissociation reaction

$$PRTN \rightleftharpoons PRTN^- + H^+ \quad (5)$$

where *PRTN* is the original protein and $PRTN^-$ is the new state. The free energy of reaction 5 can be written as

$$\Delta G_{deprot} = G(PRTN^-) + G(H^+) - G(PRTN) \quad (6)$$

Additionally, the K_a of the reaction is defined as

$$K_a = 10^{-\Delta G_{deprot}/RT} \quad (7)$$

Hence

$$G(PRTN^-) + G(H^+) = G(PRTN) - RTln(10^{-pK_a}) \quad (8)$$

where *G*(*PRTN*) is the uncorrected DMD energy and the left-hand side of the reaction is a corrected energy for a comparable system with the same chemical composition. For the protonation reaction, casting *PRTN* as $PRTN^+$ and $PRTN^-$ as *PRTN* in the original reaction gives the equation

$$G(PRTN^+) - G(H^+) = G(PRTN) + RTln(10^{-pK_a}) \quad (9)$$

The energy associated with each protonation state change from the PRTN structure can therefore be written as $\pm RTln(10^{-pKa})$, positive for protonation and negative for deprotonation. The corrections are on the scale of 2-20 kcal/mol, consistent with DMD energy fluctuations. Correction terms are calculated for each iteration and summed with its DMD energy for the corrected energy.

Current limitations of Titr-DMD. The scope of Titr-DMD leaves it with a few limitations, which are worth describing here. Its reliance on Propka and DMD implicit solvent means that it does not take interactions with ions in solution into account. Other methods rely on Debye-Huckel theory to do this.[18] As covered in the Results and Discussion section, benchmarking suggests that Propka does not always provide accurate pK_a's for certain, specific residues, namely cysteine and aspartate residues with very acidic pK_a's (around 1.0). However, Propka does quite well with glutamate and asparate residues with pK_as' near to or higher than the solution value (above 4.5). The protonation of cysteine in disulfide bridges – and thus breaking of disulfide bridges – is not allowed in the current implementation. Titr-DMD does not assess protonation state changes to the C-terminal carboxylate or N-terminal amine. Titr-DMD has so far not been used to study catalytic protonation and deprotonation events, so it is unclear how well it can describe highly coupled, hydrogen bonding residues often involved in these processes.[53] However, Titr-DMD may obtain reasonably accurate pK_a's for catalytically coupled residues as both DMD and Propka contain hydrogen bonding terms. Investigations of such behavior are beyond the scope of this initial publication and left for future studies.

Future development of Titr-DMD. The modularity of Titr-DMD allows for easy adaptation and refinement. Changes to the method do not require reparameterization of the forcefield. Future developments of Propka or any other tool to calculate the instantaneous pK_a of a protein conformer can be exchanged to generate the probabilities of protonation state change and improve the quantitative accuracy of the feature. The Titr-feature could even be paired with another molecular mechanics method besides DMD, so long as it is in implicit solvent for consistency with the probabilities of protonation state change. Alternatives to the somewhat system dependent solvent access cutoff are also of interest. One is to use the Propka buried percentage as a scalar probability of solvent accessibility rather than assign a sharp cutoff, while another is based on the solvent-accessible surface (SAS) determined by reduced surface. The SAS method defines the contour of the protein that can be accessed by solvent by rolling a sphere with the van der Waals radius of the solvent (the 'probe') across the protein, avoiding the van der Waals radii of the other atoms.[54] The solvent accessibility of any residue can be determined by measuring the distance of its titratable group to the nearest vertex of the water SAS. If the vertex is within the van der Waals radius of the titratable group, it is solvent accessible. We are currently investigating a SAS approach for future developments.

Benchmark systems. Most of the systems considered for pK_a prediction have been studied extensively both experimentally and with other computational methods (Figure 2). Hen egg-white lysozyme (HEWL) was used as it is a prototypical system for CpHMD benchmarking. The input structure of the protein was taken from the Protein Data Bank (PDB ID 1LZN).[55] All solvent molecules were removed for the simulation – water, nitrogen trioxide, and the sodium ion. As HEWL only reports experimental pK_a for GLU, ASP, LYS, TYR, and a single HIS residue, both human thioredoxin (HTRX) and human muscle creatine kinase (HMCK) were simulated as well. HTRX brought in another HIS residue to the dataset alongside

many more GLU and ASP. Two CYS residues in HTRX were not considered, as they are involved in a disulfide bridge. HMCK only added one CYS residue to the dataset, but was included as it is one of the largest proteins with an experimentally identified amino acid pK_a at 381 residues (compared to 105 residues for HTRX and 129 for HEWL). A Staphylococcal nuclease mutant, V66K (SNase V66K), was included as it contains a buried LYS66 residue that is deprotonated at neutral *pH*. The initial structure used for HTRX was PDB ID 1ERT,[56] with all water molecules removed and the rotamers labeled 'A' used when more than one was recorded. As it is unclear whether the 320-331 loop of HMCK is unstructured or an alpha helix, two structures were used. The unstructured case was based on the A chain of PDB ID 1U6R,[57] mutated back to the WT sequence with the substrate ADP, inhibitor (diaminomethyl-methyl-amino)-acetic acid, all water, nitrogen trioxide, and magnesium ions removed. The alpha loop structure was the same except the 320-331 loop was replaced with the 321-332 loop of the A-chain from PDB ID 3B6R.[58] The structure used for SNase V66K was PDB ID 2SNM[69] with thymidine-3',5'-diphosphate, water molecules, and the calcium ion removed. All experimental reference pK_a were drawn from the PKAD database.[59] The pK_a values used ultimately come from Bartik et al.[60] and Webb et al.[61] for HEWL, from Forman-Kay et al.[62] and Qin et al.[63] for HTRX, Wang et al.[64] for HMCK, and Fitch et al. for SNase V66K.[65] The pK_a predictions from our simulations measure error and deviations to the average of these datasets for each residue with more than one reported value. Titr-DMD pK_a predictions were also compared to existing methods. For HTRX, values were obtained from Harris et al.,[28] an explicit solvent replica exchange CpHMD method. Comparisons for HEWL were made based on a truncated set of residues that was also assessed by the explicit solvent Vila-Viçosa et al.[30] and Goh et al.[25] replica exchange CpHMD methods, the implicit solvent implementation of the Wallace et al.[29] replica exchange CpHMD method, a CG CpHMD method using supra-atomic beads called OPEP6,[38] and the Monte-Carlo method FPTS.[34] The dataset includes mostly asparatate and glutamate residues as well as one histidine residue. Comparisons for the buried LYS66 of SNase V66K were made with the Wallace et al. implicit solvent-based CpHMD[29] and FPTS.[34] There was no other result to compare to for HMCK, but assessment of cysteine residues is unusual. Finally, for all the benchmark systems results were also compared to the NULL model. The NULL model does not involve any simulation but is used to calculate error with solution pK_a values assigned to each amino acid. In our case, we assign amino acids the reference solution pK_a values used by Propka. Beating the NULL model is important for any pK_a prediction tool as failure to do so means that the tool does not even qualitatively capture the pK_a shifting effect of the protein environment on residues.

The system used to assess *pH*-conformational coupling was SNase V66K, a well characterized system (Figure 2D). Experimental information about protein conformational dynamics, including in the context of *pH* change, is difficult to obtain. As discussed in the introduction, the study of SNase mutants is a rare example with available experimental data on dynamics. A combination of NMR, CD, and titration suggests that the protonation of LYS66 is concurrent with and may be coupled to the unraveling of the first loop of the alpha helix on which it is located.[46,66–68] The V66K mutant was selected as it demonstrates an extreme pK_a shift of 10.5 down to 6.4 – which alongside the conformational coupling is a real challenge for any CpHMD method.

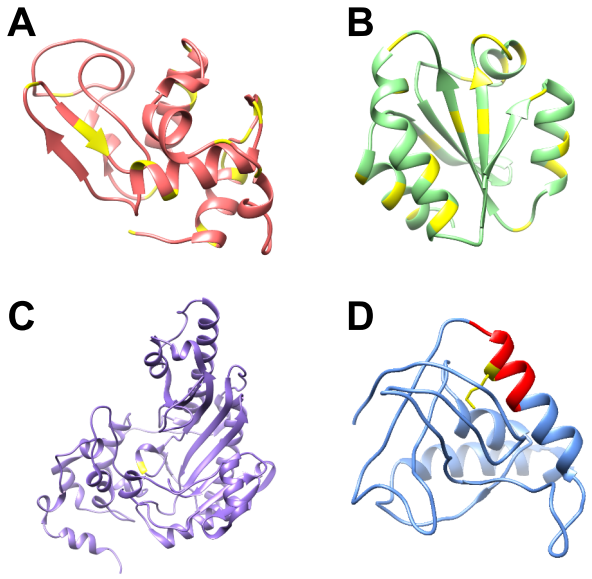


Figure 2: Ribbon diagrams of protein test systems for Titr-DMD benchmarking: HEWL (A), HTRX (B), HMCK (C), and SNase mutant V66K (D). The residues whose pK_a's are considered and compared to experiment are highlighted in yellow. In the case of the SNase mutant, this is the buried LYS66 residue. The alpha helical loop to which it belongs and that unravels is shown in red.

Titr-DMD settings. Benchmarking simulations differ slightly between those done to estimate pK_a values and those that assess *pH*-conformational coupling. The *pH*-conformational coupling simulations were longer and hotter to achieve the necessary sampling. DMD simulations without the Titr-feature were also run for the *pH*-conformation coupling system as a control – to make sure conformational changes are *pH* dependent. The pK_a estimating simulations were run for 2,000,000 DMD timesteps (roughly 100 ns, defined empirically) at 50 K (note that temperature in DMD is defined specifically, and does not directly correspond to the physical temperature).[39] A high heat exchange of 10.0 was used for thermal stability because Titr-DMD consists of many short DMD simulations – a more typical, low value has a destabilizing effect. As discussed with the description of the method, a standard protonation contact distance of 3.5 Å was used, as well as the standard protonation state reassessment frequency of 200 steps. The solvent access cutoff was 75%, which is discussed in more detail in the Supporting Information. The *pH*-conformation coupling simulations were run for a longer 4,000,000 DMD timesteps (roughly 200 ns) with solvent access cutoff values of 65% and 45% ultimately selected and a temperature of 150K for increased mobility. The other settings were the same as for the pK_a prediction. The DMD control simulations without the Titr-feature were performed for the same time and temperature as the *pH*-conformational coupling simulations.

A total of 45 Titr-DMD and 4 DMD simulations were performed for benchmarking. Simulations were done for HEWL at *pH* 3, 5, 7, and 9, for HTRX at *pH* 3, 5, and 7, and for HMCK both with the unstructured and alpha helical 320-331 loop at *pH* 9. The *pH* values were selected to straddle the pK_a of residues with experimentally reported values. Simulations were run for SNase at *pH* 4.6, 5.7, and 7. These values are

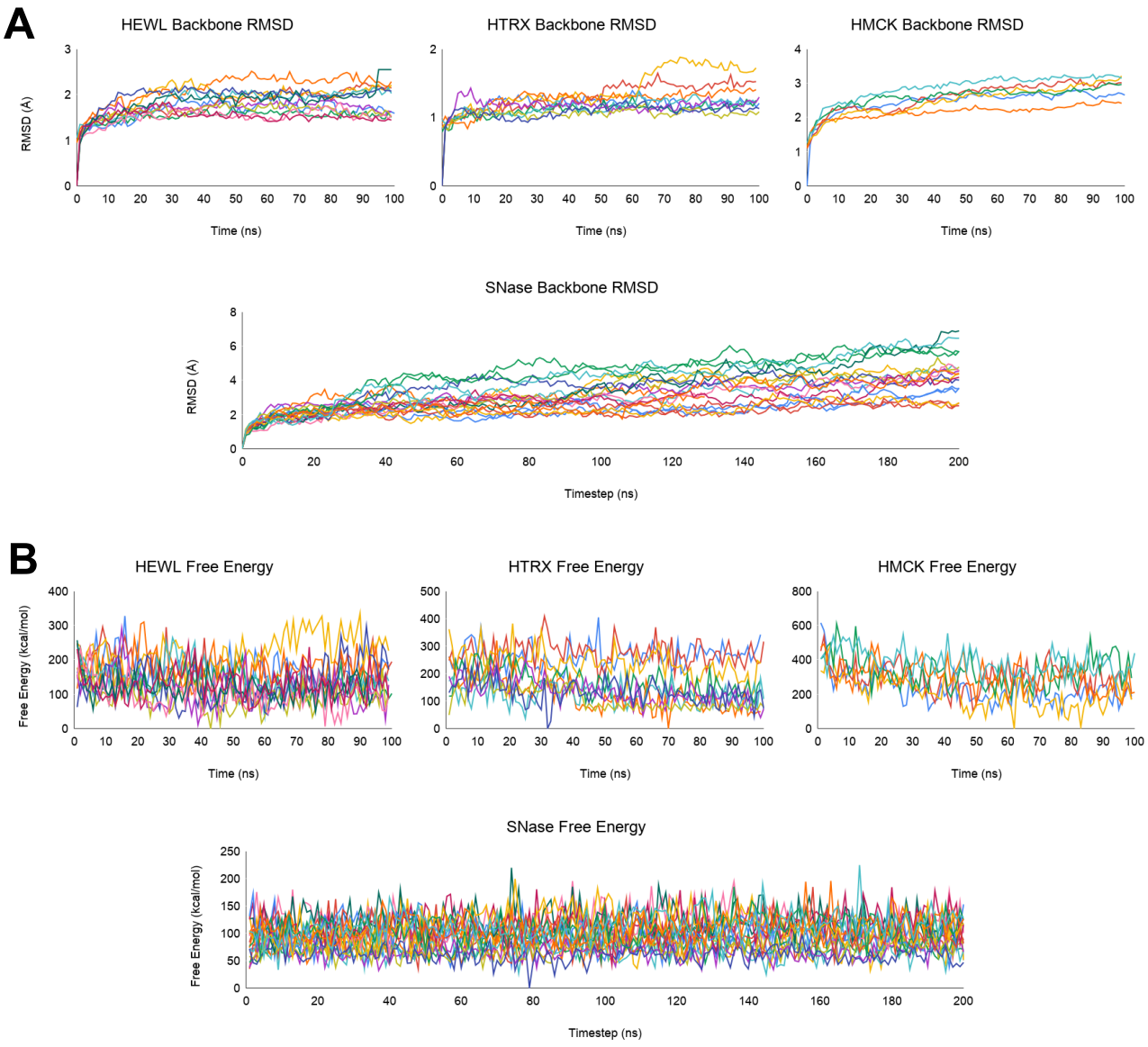


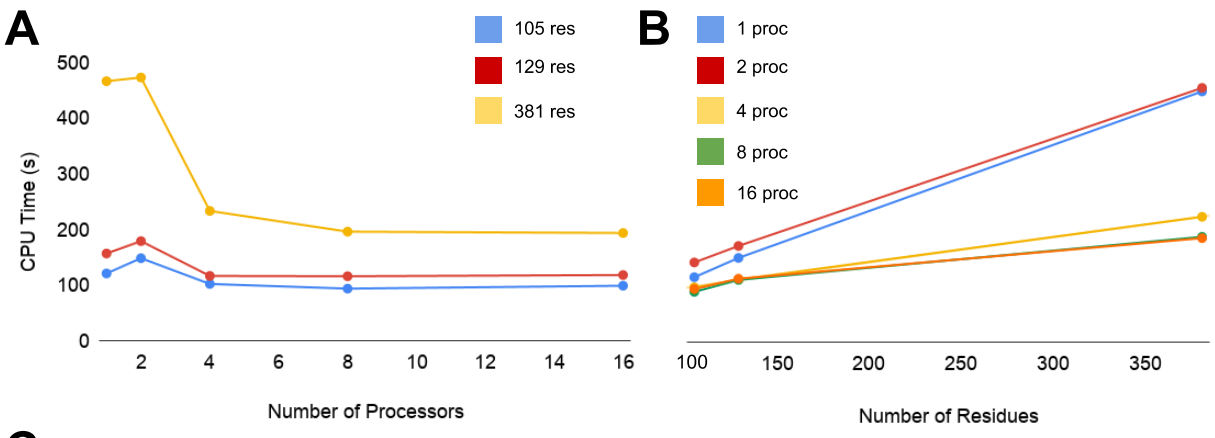
Figure 3: Convergence of Titr-DMD simulations tracked by (A) the backbone RMSD and (B) corrected DMD potential free energy. Note that by both metrics the results come to oscillate around fixed values by the end of the simulations, indicating convergence for the overall protein structures. The average RMSDs across all trajectories are 1.77 ± 0.29 Å for HEWL, 1.20 ± 0.18 Å for HTRX, 2.51 ± 0.48 Å for HMCK, and 3.67 ± 1.57 Å for SNase. The average energies across all trajectories are 140.18 ± 57.79 kcal/mol for HEWL, 174.93 ± 77.49 kcal/mol for HTRX, 393.52 ± 104.51 kcal/mol for HMCK, and 95.93 ± 29.94 kcal/mol for SNase.

much below, slightly below, and above the experimental pK_a of the LYS66 residue and its coupled dynamic behavior. All are above the denaturing point of the protein. Three replicates were performed for each system and *pH*. The four DMD simulations were run for SNase to provide a point of comparison. Two were run with LYS66 permanently deprotonated and two with it permanently protonated.

Convergence of the Titr-DMD simulations was attained according to a series of metrics. This is comprised chiefly of the backbone RMSD and the corrected Titr-DMD energy (Figure 3). The RMSD was calculated with the initial structure as the reference and with respect to the alpha carbon and amide nitrogen, carbon, and oxygen of each amino acid. All trajectories come to oscillate around fixed values, indicating convergence of the overall protein structures. Convergence for HMCK and SNase V66K, systems with just one amino acid of interest, was tracked by additional metrics covered in the Supporting Information: the average pK_a and the average protonation state of the titratable residue of interest.

Results and Discussion

Titr-DMD offers rapid sampling on limited resources. The combination of DMD and Propka in an implicit solvent makes it a fast and affordable method. We assessed the scaling of Titr-DMD through 1000 step (5 protonation assessments, about 50 ps) simulations of HEWL, HTRX, and HMCK executed with 1, 2, 4, 8, and 16 processors both with and without the Titr-feature. All simulations were run on the same node sequentially during a single submission to reduce the impact of the variability of other demands on the supercomputing cluster. Simulations were performed on AMD Opteron 2380 (2.5



C

	1 proc	2 proc	4 proc	8 proc	16 proc
105 res	4.50%	25.33%	39.16%	29.17%	40.61%
109 res	2.63%	16.64%	34.92%	33.71%	36.03%
381 res	3.38%	7.23%	18.97%	23.39%	22.13%

Figure 4: Computational resource scaling benchmark of Titr-DMD, plotted by (A) the number of processors and (B) the number of residues. Note the linear scaling with number of residues and that good performance is reached with four processors. (C) The percent increase of time for Titr-DMD over unmodified DMD. Note that the increase is relatively small and only becomes significant with many processors as the time DMD takes shortens.

GHz) cores on Hoffman2 at UCLA IDRE. This process was replicated five times, with the average of these results taken (Figure 4). Titr-DMD scales roughly linearly with the number of residues, and scales favorably out to four processors, with additional resources giving diminished returns. The Titr-feature does modestly increase the computational expense of DMD simulations, with the increase in relative runtime over base DMD growing some with the number of processors used. The increase largely derives from the need to initialize many short DMD simulations. However, Titr-DMD still runs quite well on limited resources; the CPU time for the four processor tests scales up to 500-1300 CPU hours (or 3-5.5 CPU hours per residue) to reach a 1 ms simulation.

With a couple exceptions, our Titr-DMD method successfully recapitulated the experimental pK_a of the test system residues with reasonable error. We calculated the average RMSE between the predicted and experimental values both by type of amino acid and by protein test system (Tables 1-4). The pK_a can be calculated two ways from Titr-DMD, therefore we calculated two average RMSE for each case. Propka-averaged pK_a is simply the average of the Propka predicted values from each timestep. The DMD-averaged pK_a for a residue is the natural logarithm of the fraction of timesteps in which the residue is protonated. That fraction is analogous to the K_a: the relative concentration of the protonated form of the residue. For solvent exposed residues (those that can freely change protonation state just based on their instantaneous pK_a) the two pK_a predictions should converge to the same values with appropriate sampling. The results show that indeed the pK_a are in good agreement between the two methods for each system, with the notable exception of SNase V66K LYS66 – a deeply buried residue. The overall maximum absolute error, mean average error, and root mean square error across the full dataset (excluding an outlier HMCK CYS283, discussed later) were 3.25, 0.77, and 1.03 for the DMD-averaged pK_a's and 2.43, 0.81, and 1.05 for the Propka-averaged pK_a's respectively, which is decent agreement with experiment for a CpHMD method.

Titr-DMD predictions of pK_a values are competitive with other CpHMD methods. Across nearly all of the benchmark systems, Titr-DMD outperforms the NULL model. The RMSEs by protein are lower at 0.82-0.83 versus 1.58 for HTRX and 1.19 versus 1.31 for HEWL. The absolute error for SNase V66K LYS66 is 0-1.5 versus 4.1 for the NULL model. HMCK is the one exception, as Propka predicts the pK_a of the single CYS283 residue poorly: 5.7 in absolute error from the experimental value versus 3.4. There was no DMD-averaged value for the residue as the Propka predicted pK_a's were too high and so the residue was rarely deprotonated in our simulations (all conducted at *pH* well below 11). Titr-DMD matches or outperforms more expensive CpHMD methods with HTRX and SNase V66K. For HTRX, Titr-DMD reports a smaller RMSE of 0.82-0.83 versus 0.95 for the Harris method, while the absolute error in the SNase V66K LYS66 pK_a is 0-1.5 compared to the 1.1 of the Wallace method. Titr-DMD performs worse than more expensive CpHMD methods with HEWL, but is comparable to another CG method. Its RMSE for a truncated dataset (comprised mostly of ASP and GLU residues) is 1.45-1.46, above the 0.82-0.89 of the more expensive Wallace, Goh, and Vila-Viçosa methods, but close to the 1.32 of the CG OPEP6.

The performance of Titr-DMD arises from its ability to accurately predict many large pK_a shifts. Our method generally does well with ASP and GLU residues that report pK_a's shifted to more basic values, but struggles to provide accurate pK_a's for those shifted to very acidic values (around 2.0 or below). Titr-DMD outperforms the more expensive method with

	Experiment				Titr-DMD		
Residue	Qin	Forman-Kay	Ave.	Harris[a]	DMD-Ave.	Propka-Ave.	NULL
GLU13	4.4	4.8	4.6	4.4	4.7	4.7	4.5
ASP16	4.0	3.7	3.9	4.0	3.9	3.8	3.8
ASP20	3.8	3.6	3.7	2.9	3.0	2.9	3.8
ASP26**	9.9	9.0	9.5	6.2	7.6	7.6	3.8
HIS43		5.5	5.5	6.1	6.2	6.2	6.5
GLU47	4.1		4.1	4.3	4.5	4.5	4.5
GLU56*	4.1		4.1	4.5	4.7	4.7	4.5
ASP58*	2.8	3.1	3.0	3.8	4.4	4.4	3.8
ASP61*	5.3		5.3	4.6	4.0	4.0	3.8
GLU68	4.9	4.2	4.6	4.3	3.9	3.9	4.5
GLU70	4.6		4.6	5.0	3.8	3.9	4.5
GLU88	3.7	3.9	3.8	3.8	4.0	4.0	4.5
GLU95	4.1		4.1	3.5	4.6	4.6	4.5
GLU98	3.9		3.9	3.9	4.0	4.0	4.5
GLU103	4.4	4.9	4.7	4.7	4.6	4.6	4.5
MAX				3.3	1.9	1.9	5.7
MAE				0.6	0.6	0.6	0.8
RMSE				0.95	0.83	0.82	1.58

*Table 1: Experimental and calculated pK_a values of HTRX. *Proximity of these residues meant that the exact experimental value in the Qin et al. study was unclear. **A series of possible pK_a were found for this residue in the Forman-Kay et al. study, the value of 9.0 was selected due to its consistency with the Qin et al. result. Experimental data from ref 63 (Qin) and ref 62 (Forman-Kay). Other calculated data from [a]ref 28.*

	Experiment								Titr-DMD		
Residue	Bartik	Webb	Ave.	Wallace[a]	Goh[b]	Vila-Viçosa[c]	OPEP6[d]	FPTS[e]	DMD-Ave.	Propka-Ave.	NULL
GLU7	2.85	2.60	2.73	2.58	2.70	3.34	3.36	3.30	4.26	4.21	4.5
LYS13		9.88	9.88						10.87	10.93	10.5
HIS15	5.36	5.50	5.43	5.34	6.00	5.43	5.99	5.60	5.74	5.76	6.5
ASP18	2.66	2.80	2.73	2.94	2.10	3.57	3.01	2.80	3.50	3.36	3.8
TYR20		10.30	10.30						9.45	9.55	10
TYR23		9.80	9.80						10.18	10.24	10
LYS33		9.92	9.92						9.91	10.01	10.5
GLU35	6.20	6.10	6.15	4.35	7.00	5.52	3.60	3.50	6.13	6.27	4.5
ASP48	1.60	1.40	1.50	2.84	1.30	1.95	3.31	3.40	3.52	3.61	3.8
ASP52	3.68	3.60	3.64	4.56	4.50	3.85	3.33	3.30	6.00	6.07	3.8
TYR53		12.10	12.10						11.39	11.82	10
ASP66	0.90	1.20	1.05	1.15	1.50	3.15	3.07	3.00	3.51	3.35	3.8
ASP87	2.07	2.20	2.14	2.03	1.30	2.31	3.22	3.20	3.57	3.50	3.8
LYS96		10.20	10.20						10.24	10.23	10.5
LYS97		9.64	9.64						10.88	11.10	10.5
ASP101	4.09	4.50	4.30	3.27	5.10	3.77	3.03	2.90	3.63	3.46	3.8
LYS116		9.76	9.76						10.40	10.38	10.5
ASP119	3.20	3.50	3.35	2.45	1.60	2.80	3.26	3.20	3.68	3.64	3.8
MAX									2.46	2.43	2.75
MAE									0.93	0.92	1.06
RMSE									1.19	1.19	1.31
MAX*				1.80	1.75	2.10	2.55	2.65	2.46	2.43	2.75
MAE*				0.66	0.70	0.61	1.06	1.03	1.19	1.19	1.34
RMSE*				0.89	0.83	0.82	1.32	1.34	1.46	1.45	1.56

*Table 2: Experimental and calculated pK_a values of HEWL. *Maximum absolute error, mean average error, and root mean square error were also calculated for a truncated set of ASP, GLU, and HIS residues so that Titr-DMD could be compared to referenced methods that only report those. Experimental data from ref 60 (Bartik) and ref 61 (Webb). Other calculated data from [a]ref 29, [b]ref 25, [c]ref 30, [d]ref 38, and [e]ref 34.*

		Titr-DMD		
Residue	Experiment	DMD-Ave.	Propka-Ave.	NULL
CYS283	5.6	-	11.3	9
Abs. Error		-	5.7	3.4

Table 3: Experimental and calculated pK_a for CYS283 in HMCK. Experimental data from ref 64.

				Titr-DMD		
Residue	Experiment	Wallace[a]	FPTS[b]	DMD-Ave.	Propka-Ave.	NULL
LYS66	6.4	7.5	11	6.4	7.9	10.5
Abs. Error		1.1	4.6	0	1.5	4.1

Table 4: Experimental and calculate pK_a for LYS66 in SNase V66K. Experimental data from ref 65. Other calculated data from [a]ref 29 and [b]ref 34.

HTRX largely through its more accurate prediction of the pK_a of ASP26, shifted according to experiment to the very basic 9.5. Conversely, the poorer performance of Titr-DMD with HEWL is due to the large number of ASP residues shifted to highly acidic values in that system. Titr-DMD struggles to predict the large pK_a shift of the cysteine residue in HMCK, the one case where it fails to beat the NULL model. However, as this was the one CYS residue considered in the test set, Titr-DMD may do better with other examples. Titr-DMD does quite well predicting the shifts of LYS and TYR residues, including the buried and highly shifted LYS66 of the SNase system. The Propka-averaged and DMD-averaged values give qualitative agreement with experiment, but the latter is quantitatively more accurate. Conformational sampling frequents solvent inaccessible states for this residue, with the result of LYS66 spending more time deprotonated than the Propka pK_a would suggest and correcting it toward the experimental value. Furthermore, for LYS66, both our CpHMD method and the Wallace et al. method outperform FPTS, which doesn't perform extensive backbone dynamics and does worse than the NULL model. This demonstrates how important *pH* dependent conformational dynamics are for particular residues and proteins.

Titr-DMD holds promise for the study of the effect of solution *pH* on protein structure. Simulations of SNase V66K are qualitatively consistent with rare, experimentally studied dynamics. With the Titr-feature, we observe partial unraveling of the first turn of the alpha helix on which K66 is localized on (residues 65-69), which is not apparent in DMD without titration (Table 5). Unraveling is only observed in 0.002-0.015% of structures in base DMD, while Titr-DMD simulations show it occurs in 3-8% of structures. We define an unraveled state as one where the ALA69-LYS66 and ASN68-MET65 hydrogen bonds are broken or breaking and the backbone RMSD of the loop is large relative to that of the full protein, indicative of significant, localized structural change (Figure 5). The criterion is

$$\frac{S_{RMSD}\,RMSD_T/RMSD_L}{((R_1-R_{HB})+(R_2-R_{HB}))/S_{HB}} \le 2 \quad (10)$$

where $RMSD_T$ is for the total protein and $RMSD_L$ is for the loop (residues 65-69), R_1 and R_2 are the backbone amide H to carbonyl O distances in Å of ALA69-LYS66 and ASN68-MET65 respectively, S_{RMSD} is 2, R_{HB} is 2.5 Å (for a long hydrogen bond length), and S_{HB} is 2 Å. We only consider structures where the ALA69-LYS66 and ASN68-MET65 backbone hydrogen bond distances are both at least 3 Å. While unraveling according to our criterion occurs in 3-8% of all states at the appropriate *pH*, it is not typically sustained for longer than

Titr-DMD	*pH 4.6*	*pH 5.7*	*pH 7*
45% cutoff	5.68%	7.06%	4.68%
65% cutoff	3.84%	5.64%	7.93%
	Prot. Start	*Deprot. Start*	
DMD	0.002%	0.015%	

Table 5: Frequency of SNase mutant V66K alpha helical loop 65-69 unraveling over the course of Titr-DMD (upper) and DMD (lower) simulations. Note that the frequency is much higher in the Titr-DMD simulations. The highest frequency that occurs a bit below the experimental pK_a of LYS66 (6.4) is when the solvent access cutoff is 45%.

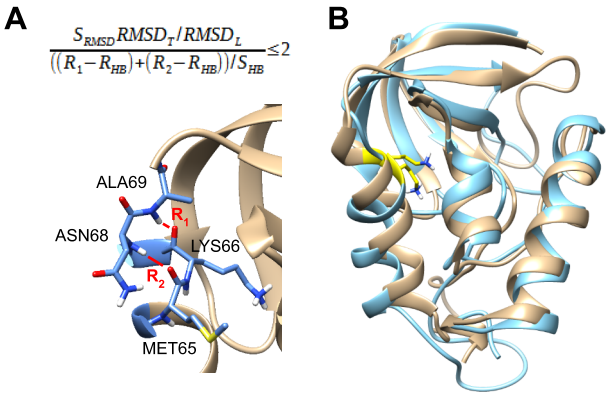


Figure 5: (A) Criterion for an unraveled 65-69 loop structure in our SNase simulations. This compares the RMSD of the loop ($RMSD_L$) to the RMSD of the full protein ($RMSD_T$) and compares the distances of important hydrogen bonding contacts (R_1, R_2) to standard values (R_{HB}) to determine structures where the conformation of the loop varies significantly from the original structure. We give the values of the other variables in the main text. (B) Example of a SNase conformation with an unraveled 65-69 loop by our criterion (light blue) overlaid on a structure where it is not unraveled (tan). LYS66 is colored yellow here.

about 1 ns at any one time. We surmise that our simulations do not have enough sampling to capture sustained loop unraveling, but do show the rare events that could lead to it.

Protonation and deprotonation of LYS66 is coupled with loop unraveling according to Titr-DMD. At *pH* 5.7, the percentage of unraveled states is significantly higher around LYS66 protonation state changes than the total simulation average (Table 6). Moreover, few events at *pH* 5.7 occur without contemporaneous unraveling. The coupling we observe in our simulations is thus consistent with the experimental hypothesis.[46]

Titr-DMD dynamics can predict the *pH* at which loop unraveling occurs. The Propka-averaged pK_a value of LYS66 is uniformly higher than the experimental 6.4, at an average of 7-8, but still shows a qualitatively correct large drop from the solution value of 10.5. However, as discussed before, the DMD-averaged pK_a is generally lower and close to the experimental value, representing the frequent solvent inaccessibility of the residue. The values in Table 4 are averages across the 45% cutoff and 65% cutoff simulations; their individual DMD-averaged pK_a are both close to the experimental result at 5.79 and 7.04 respectively. At these two cutoffs, unraveling is generally most common in the *pH* 5.7 simulations and nearly all protonation state changes occur alongside some unraveling. Titr-DMD can qualitatively model coupling between *pH* and protein structure, and when well calibrated can do so with more quantitative accuracy.

Conclusions

In this paper we demonstrate Titr-DMD as an effective new method to study *pH*-coupled protein dynamics. The challenges that face any CpHMD method are appropriate conformational and protonation state sampling, accuracy of protonation state changes, and whether the generated conformational ensemble is physically meaningful. Titr-DMD offers great sampling on just a few processors through atomic collision event calculations, implicit solvation, and semi-empirical pK_a prediction with Propka. Our method obtains reasonably accurate pK_a pre-

Near Event	*pH 4.6*	*pH 5.7*	*pH 7*
45% cutoff	1.11%	30.43%	0.54%
65% cutoff	3.20%	9.45%	10.97%
By Event			
45% cutoff	25%	85%	20%
65% cutoff	20%	80%	29%

Table 6: Frequency of unraveling of the SNase mutant 65-69 loop around LYS66 protonation and deprotonation events. 'Near event' refers to the percentage of structures within 25 timesteps (before and after) of an event that are unraveled. This value is roughly on par with the total simulation average except at pH 5.7, particularly during the simulation with a 45% solvent access cutoff. 'By event' refers to the percent of events that have at least one unraveled structure within 25 timesteps. Again, note that the pH 5.7 simulations show high coupling where protonation state changes nearly always occur alongside some contemporaneous unraveling.

dictions for its computational expense. Titr-DMD was successfully benchmarked on the partial unraveling of SNase mutant V66K: one of the few experimentally studied *pH* coupled conformational changes. Titr-DMD generates a conformational ensemble consistent with experiment, and this ensemble even reflects the experimental *pH* value of the conformational change. Our method is also modular to further improve sampling and accurate assignment of protonation states. Titr-DMD stands as a promising method to address questions of *pH* dynamics in industrial catalysis and medicine.

ASSOCIATED CONTENT

Supporting Information is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: ana@chem.ucla.edu. Phone: +1 310 8253769

Funding Sources

Financial support comes from the NSF CHE-1903808 grant to A.N.A., and from the Passan Foundation and US National Institutes of Health grant R35 GM134864 to N.V.D. We also acknowledge UCLA-IDRE and XSEDE for providing computational resources.

Notes

We declare no competing financial interest.

ACKNOWLEDGMENT

We thank the Institute for Digital Research and Education at UCLA and the Extreme Science and Engineering Discovery Environment for supercomputer time. We thank Matthew Hennefarth for advice on integration of the Titr-feature with DMD and a new implementation of DMD. We thank Prof. Feng Ding for advice on the heat exchange parameter of DMD.

ABBREVIATIONS

SNase, staphylococcal nuclease; HTRX, human thioredoxin; HEWL, hen egg-white lysozyme; HMCK, human muscle creatine kinase; SAS; solvent-accessible surface.

REFERENCES

(1) Dumorné, K.; Córdova, D. C.; Astorga-Eló, M.; Renganathan, P. Extremozymes: A Potential Source for Industrial Applications. *J Microbiol Biotechnol* **2017**, *27* (4), 649–659.

(2) Sarmiento, F.; Peralta, R.; Blamey, J. M. Cold and Hot Extremozymes: Industrial Relevance and Current Trends. *Front. Bioeng. Biotechnol.* **2015**, *3*, 148.

(3) Chen, G.-Q.; Jiang, X.-R. Next Generation Industrial Biotechnology Based on Extremophilic Bacteria. *Curr. Opin. Biotechnol.* **2018**, *50*, 94–100.

(4) Fernandes, P. Enzymes in Food and Feed Industries: Where Tradition Meets Innovation. In *Biocatalysis*; Springer, 2019; pp 233–253.

(5) Rivinoja, A.; Pujol, F. M.; Hassinen, A.; Kellokumpu, S. Golgi PH, Its Regulation and Roles in Human Disease. *Ann. Med.* **2012**, *44* (6), 542–554.

(6) Webb, B. A.; Chimenti, M.; Jacobson, M. P.; Barber, D. L. Dysregulated PH: A Perfect Storm for Cancer Progression. *Nat. Rev. Cancer* **2011**, *11* (9), 671–677.

(7) Fang, B.; Wang, D.; Huang, M.; Yu, G.; Li, H. Hypothesis on the Relationship between the Change in Intracellular PH and Incidence of Sporadic Alzheimer's Disease or Vascular Dementia. *Int. J. Neurosci.* **2010**, *120* (9), 591–595.

(8) Dokholyan, N. V. Experimentally-Driven Protein Structure Modeling. *J. Proteomics* **2020**, *220*, 103777.

(9) Fogolari, F.; Brigo, A.; Molinari, H. The Poisson–Boltzmann Equation for Biomolecular Electrostatics: A Tool for Structural Biology. *J. Mol. Recognit.* **2002**, *15* (6), 377–392.

(10) Bashford, D.; Gerwert, K. Electrostatic Calculations of the PKa Values of Ionizable Groups in Bacteriorhodopsin. *J. Mol. Biol.* **1992**, *224* (2), 473–486.

(11) Onufriev, A. V; Case, D. A. Generalized Born Implicit Solvent Models for Biomolecules. *Annu. Rev. Biophys.* **2019**, *48*, 275–296.

(12) Madura, J. D.; Briggs, J. M.; Wade, R. C.; Davis, M. E.; Luty, B. A.; Ilin, A.; Antosiewicz, J.; Gilson, M. K.; Bagheri, B.; Scott, L. R. Electrostatics and Diffusion of Molecules in Solution: Simulations with the University of Houston Brownian Dynamics Program. *Comput. Phys. Commun.* **1995**, *91* (1–3), 57–95.

(13) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: Automating p K Prediction and the Preparation of Biomolecular Structures for Atomistic Molecular Modeling and Simulations. *Nucleic Acids Res.* **2012**, *40* (W1), W537–W541.

(14) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K a Predictions. *J. Chem. Theory Comput.* **2011**, *7* (2), 525–537.

(15) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of p K a Values. *J. Chem. Theory Comput.* **2011**, *7* (7), 2284–2295.

(16) Beroza, P.; Fredkin, D. R.; Okamura, M. Y.; Feher, G. Protonation of Interacting Residues in a Protein by a Monte Carlo Method: Application to Lysozyme and the Photosynthetic Reaction Center of Rhodobacter Sphaeroides. *Proc. Natl. Acad. Sci.* **1991**, *88* (13), 5804–5808.

(17) Kesvatera, T.; Jönsson, B.; Thulin, E.; Linse, S. Focusing of the Electrostatic Potential at EF-hands of Calbindin D9k: Titration of Acidic Residues. *Proteins Struct. Funct. Bioinforma.* **2001**, *45* (2), 129–135.

(18) Teixeira, A. A. R.; Lund, M.; Barroso da Silva, F. L. Fast Proton Titration Scheme for Multiscale Modeling of Protein Solutions. *J. Chem. Theory Comput.* **2010**, *6* (10), 3259–3266.

(19) Barroso daSilva, F. L.; Dias, L. G. Development of Constant-PH Simulation Methods in Implicit Solvent and Applications in Biomolecular Systems. *Biophys. Rev.* **2017**, *9* (5), 699–728.

(20) Baptista, A. M.; Teixeira, V. H.; Soares, C. M. Constant-p H Molecular Dynamics Using Stochastic Titration. *J. Chem. Phys.* **2002**, *117* (9), 4184–4200.

(21) Fuzo, C. A.; Degrève, L. The PH Dependence of Flavivirus Envelope Protein Structure: Insights from Molecular Dynamics Simulations. *J. Biomol. Struct. Dyn.* **2014**, *32* (10), 1563–1574.

(22) Lee, M. S.; Salsbury Jr, F. R.; Brooks III, C. L. Constant-pH Molecular Dynamics Using Continuous Titration Coordinates. *Proteins Struct. Funct. Bioinforma.* **2004**, *56* (4), 738–752.

(23) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935.

(24) Tidor, B. Simulated Annealing on Free Energy Surfaces by a Combined Molecular Dynamics and Monte Carlo Approach. *J. Phys. Chem.* **1993**, *97* (5), 1069–1073.

(25) Goh, G. B.; Hulbert, B. S.; Zhou, H.; Brooks III, C. L. Constant PH Molecular Dynamics of Proteins in Explicit Solvent with Proton Tautomerism. *Proteins Struct. Funct. Bioinforma.* **2014**, *82* (7), 1319–1331.

(26) Chen, Y.; Roux, B. Constant-PH Hybrid Nonequilibrium Molecular Dynamics–Monte Carlo Simulation Method. *J. Chem. Theory Comput.* **2015**, *11* (8), 3919–3931.

(27) Huang, Y.; Harris, R. C.; Shen, J. Generalized Born Based Continuous Constant PH Molecular Dynamics in AMBER: Implementation, Benchmarking and Analysis. *J. Chem. Inf. Model.* **2018**, *58* (7), 1372–1383.

(28) Harris, R. C.; Shen, J. GPU-Accelerated Implementation of Continuous Constant PH Molecular Dynamics in Amber: P K a Predictions with Single-PH Simulations. *J. Chem. Inf. Model.* **2019**, *59* (11), 4821–4832.

(29) 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* (8), 2617–2629.

(30) Vila-Viçosa, D.; Reis, P. B. P. S.; 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* (5), 3108–3116.

(31) Kim, M. O.; Blachly, P. G.; McCammon, J. A. Conformational Dynamics and Binding Free Energies of Inhibitors of BACE-1: From the Perspective of Protonation Equilibria. *PLoS Comput Biol* **2015**, *11* (10), e1004341.

(32) 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. Bioinforma.* **2002**, *47* (4), 469–480.

(33) Mongan, J.; Case, D. A.; McCAMMON, J. A. Constant PH Molecular Dynamics in Generalized Born Implicit Solvent. *J. Comput. Chem.* **2004**, *25* (16), 2038–2048.

(34) Barroso da Silva, F. L.; MacKernan, D. Benchmarking a Fast Proton Titration Scheme in Implicit Solvent for Biomolecular Simulations. *J. Chem. Theory Comput.* **2017**, *13* (6), 2915–2929.

(35) Harris, R. C.; Liu, R.; Shen, J. Predicting Reactive Cysteines with Implicit-Solvent-Based Continuous Constant PH Molecular Dynamics in Amber. *J. Chem. Theory Comput.* **2020**, *16* (6), 3689–3698.

(36) Grünewald, F.; Souza, P. C. T.; Abdizadeh, H.; Barnoud, J.; de Vries, A. H.; Marrink, S. J. Titratable Martini Model for Constant PH Simulations. *J. Chem. Phys.* **2020**, *153* (2), 24118.

(37) Pasquali, S.; Frezza, E.; Barroso da Silva, F. L. Coarse-Grained Dynamic RNA Titration Simulations. *Interface Focus* **2019**, *9* (3), 20180066.

(38) Barroso da Silva, F. L.; Sterpone, F.; Derreumaux, P. OPEP6: A New Constant-PH Molecular Dynamics Simulation Scheme with OPEP Coarse-Grained Force Field. *J. Chem. Theory Comput.* **2019**, *15* (6), 3875–3888.

(39) Ding, F.; Tsao, D.; Nie, H.; Dokholyan, N. V. Ab Initio Folding of Proteins with All-Atom Discrete Molecular Dynamics. *Structure* **2008**, *16* (7), 1010–1018.

(40) Proctor, E. A.; Ding, F.; Dokholyan, N. V. Discrete Molecular Dynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1* (1), 80–92.

(41) Shirvanyants, D.; Ding, F.; Tsao, D.; Ramachandran, S.; Dokholyan, N. V. Discrete Molecular Dynamics: An Efficient and Versatile Simulation Method for Fine Protein Characterization. *J. Phys. Chem. B* **2012**, *116* (29), 8375–8382.

(42) Khandogin, J.; Chen, J.; Brooks, C. L. Exploring Atomistic Details of PH-Dependent Peptide Folding. *Proc. Natl. Acad. Sci.* **2006**, *103* (49), 18546–18550.

(43) Vila-Viçosa, D.; Campos, S. R. R.; Baptista, A. M.; Machuqueiro, M. Reversibility of Prion Misfolding: Insights from Constant-PH Molecular Dynamics Simulations. *J. Phys. Chem. B* **2012**, *116* (30), 8812–8821.

(44) Shi, C.; Wallace, J. A.; Shen, J. K. Thermodynamic Coupling of Protonation and Conformational Equilibria in Proteins: Theory and Simulation. *Biophys. J.* **2012**, *102* (7), 1590–1597.

(45) Chen, W.; Huang, Y.; Shen, J. Conformational Activation of a Transmembrane Proton Channel from Constant PH Molecular Dynamics. *J. Phys. Chem. Lett.* **2016**, *7* (19), 3961–3966.

(46) Isom, D. G.; Castañeda, C. A.; Cannon, B. R. Large Shifts in PKa Values of Lysine Residues Buried inside a Protein. *Proc. Natl. Acad. Sci.* **2011**, *108* (13), 5260–5265.

(47) Peck, M. T.; Ortega, G.; De Luca-Johnson, J. N.; Schlessman, J. L.; Robinson, A. C.; García-Moreno E, B. Local Backbone Flexibility as a Determinant of the Apparent p K a Values of Buried Ionizable Groups in Proteins. *Biochemistry* **2017**, *56* (40), 5338–5346.

(48) Schramm, V. L.; Schwartz, S. D. Promoting Vibrations and the Function of Enzymes. Emerging Theoretical and Experimental Convergence. *Biochemistry* **2018**, *57* (24), 3299–3308.

(49) Fang, C.; Frontiera, R. R.; Tran, R.; Mathies, R. A. Mapping GFP Structure Evolution during Proton Transfer with Femtosecond Raman Spectroscopy. *Nature* **2009**, *462* (7270), 200–204.

(50) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Aqueous Solvation Free Energies of Ions and Ion− Water Clusters Based on an Accurate Value for the Absolute Aqueous Solvation Free Energy of the Proton. *J. Phys. Chem. B* **2006**, *110* (32), 16066–16081.

(51) Zhan, C.-G.; Dixon, D. A. Absolute Hydration Free Energy of the Proton from First-Principles Electronic Structure Calculations. *J. Phys. Chem. A* **2001**, *105* (51), 11534–11540.

(52) Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V; Tuttle, T. R. The Proton's Absolute Aqueous Enthalpy and Gibbs Free Energy of Solvation from Cluster-Ion Solvation Data. *J. Phys. Chem. A* **1998**, *102* (40), 7787–7794.

(53) Huang, Y.; Yue, Z.; Tsai, C.-C.; Henderson, J. A.; Shen, J. Predicting Catalytic Proton Donors and Nucleophiles in Enzymes: How Adding Dynamics Helps Elucidate the Structure–Function Relationships. *J. Phys. Chem. Lett.* **2018**, *9* (6), 1179–1184.

(54) Sanner, M. F.; Olson, A. J.; Spehner, J. Reduced Surface: An Efficient Way to Compute Molecular Surfaces. *Biopolymers* **1996**, *38* (3), 305–320.

(55) Bon, C.; Lehmann, M. S.; Wilkinson, C. Quasi-Laue Neutron-Diffraction Study of the Water Arrangement in Crystals of Triclinic Hen Egg-White Lysozyme. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1999**, *55* (5), 978–987.

(56) Weichsel, A.; Gasdaska, J. R.; Powis, G.; Montfort, W. R. Crystal Structures of Reduced, Oxidized, and Mutated Human Thioredoxins: Evidence for a Regulatory Homodimer. *Structure* **1996**, *4* (6), 735–751.

(57) Ohren, J. F.; Kundracik, M. L.; Borders, C. L.; Edmiston, P.; Viola, R. E. Structural Asymmetry and Intersubunit Communication in Muscle Creatine Kinase. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2007**, *63* (3), 381–389.

(58) Bong, S. M.; Moon, J. H.; Nam, K. H.; Lee, K. S.; Chi, Y. M.; Hwang, K. Y. Structural Studies of Human Brain-Type Creatine Kinase Complexed with the ADP–Mg2+–NO3−–Creatine Transition-State Analogue Complex. *FEBS Lett.* **2008**, *582* (28), 3959–3965.

(59) Pahari, S.; Sun, L.; Alexov, E. PKAD: A Database of Experimentally Measured PKa Values of Ionizable Groups in Proteins. *Database* **2019**, *2019*.

(60) Bartik, K.; Redfield, C.; Dobson, C. M. Measurement of the Individual PKa Values of Acidic Residues of Hen and Turkey Lysozymes by Two-Dimensional 1H NMR. *Biophys. J.* **1994**, *66* (4), 1180–1184.

(61) Webb, H.; Tynan-Connolly, B. M.; Lee, G. M.; Farrell, D.; O'Meara, F.; Søndergaard, C. R.; Teilum, K.; Hewage, C.; McIntosh, L. P.; Nielsen, J. E. Remeasuring HEWL PKa Values by NMR Spectroscopy: Methods, Analysis, Accuracy, and Implications for Theoretical PKa Calculations. *Proteins Struct. Funct. Bioinforma.* **2011**, *79* (3), 685–702.

(62) Forman-Kay, J. D.; Clore, G. M.; Gronenborn, A. M. Relationship between Electrostatics and Redox Function in Human Thioredoxin: Characterization of PH Titration Shifts Using Two-Dimensional Homo-and Heteronuclear NMR. *Biochemistry* **1992**, *31* (13), 3442–3452.

(63) Qin, J.; Clore, G. M.; Gronenborn, A. M. Ionization Equilibria for Side-Chain Carboxyl Groups in Oxidized and Reduced Human Thioredoxin and in the Complex with Its Target Peptide from the Transcription Factor NFκB. *Biochemistry* **1996**, *35* (1), 7–13.

(64) Wang, P.-F.; McLeish, M. J.; Kneen, M. M.; Lee, G.; Kenyon, G. L. An Unusually Low p K a for Cys282 in the Active Site of Human Muscle Creatine Kinase. *Biochemistry* **2001**, *40* (39), 11698–11705.

(65) Fitch, C. A.; Karp, D. A.; Lee, K. K.; Stites, W. E.; Lattman, E. E.; García-Moreno, E. B. Experimental PKa Values of Buried Residues: Analysis with Continuum Methods and Role of Water Penetration. *Biophys. J.* **2002**, *82* (6), 3289–3304.

(66) Takayama, Y.; Castaneda, C. A.; Chimenti, M.; García-Moreno, B.; Iwahara, J. Direct Evidence for Deprotonation of a Lysine Side Chain Buried in the Hydrophobic Core of a Protein. *J. Am. Chem. Soc.* **2008**, *130* (21), 6714–6715.

(67) Karp, D. A.; Stahley, M. R.; García-Moreno E, B. Conformational Consequences of Ionization of Lys, Asp, and Glu Buried at Position 66 in Staphylococcal Nuclease. *Biochemistry* **2010**, *49* (19), 4138–4146.

(68) Chimenti, M. S.; Khangulov, V. S.; Robinson, A. C.; Heroux, A.; Majumdar, A.; Schlessman, J. L.; García-Moreno, B. Structural Reorganization Triggered by Charging of Lys Residues in the Hydrophobic Interior of a Protein. *Structure* **2012**, *20* (6), 1071–1085.

(69) Stites, W. E.; Gittis, A. G.; Lattman, E. E.; Shortle, D. In a Staphylococcal Nuclease Mutant the Side-Chain of a Lysine Replacing Valine 66 Is Fully Buried in the Hydrophobic Core. *J. Mol. Biol.* **1991**, *221* (1), 7–14.