Towards Accurate Prediction of Protonation Equilibrium of Nucleic Acids

ARTICLE *in* JOURNAL OF PHYSICAL CHEMISTRY LETTERS · MARCH 2013

Impact Factor: 7.46 · DOI: 10.1021/jz400078d · Source: PubMed

THE JOURNAL OF PHYSICAL CHEMISTRY Letters

Letter

pubs.acs.org/JPCL

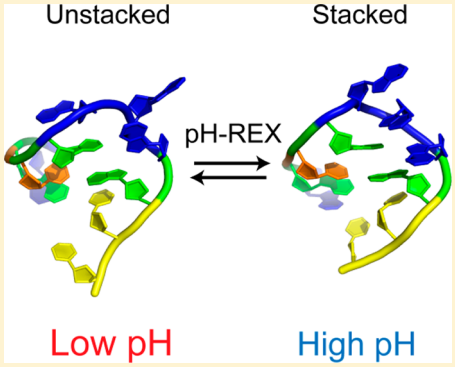
Toward Accurate Prediction of the Protonation Equilibrium of Nucleic Acids

Garrett B. Goh,[†] Jennifer L. Knight,[†] and Charles L. Brooks, III*[,†,‡]

[†]Department of Chemistry and [‡]Biophysics Program, University of Michigan, 930 North University, Ann Arbor, Michigan 48109, United States

Supporting Information

ABSTRACT: The role of protonated nucleotides in modulating the pH-dependent properties of nucleic acids is one of the emerging frontiers in the field of nucleic acid biology. The recent development of a constant pH molecular dynamics simulation (CPHMD$^{MS\lambda D}$) framework for simulating nucleic acids has provided a tool for realistic simulations of pH-dependent dynamics. We enhanced the CPHMD$^{MS\lambda D}$ framework with pH-based replica exchange (pH-REX), which significantly improves the sampling of both titration and spatial coordinates. The results from our pK_a calculations for the GAAA tetraloop, which was predicted with lower accuracy previously due to sampling challenges, demonstrates that pH-REX reduces the average unsigned error (AUE) to 0.7 pK_a units, and the error of the most poorly predicted residue A17 was drastically reduced from 2.9 to 1.2 pK_a units. Lastly, we show that pH-REX CPHMD$^{MS\lambda D}$ simulations can be used to identify the dominant conformation of nucleic acid structures in alternate pH environments. This work suggests that pH-REX CPHMD$^{MS\lambda D}$ simulations provide a practical tool for predicting nucleic acid protonation equilibrium from first-principles and offering structural and mechanistic insight into the study of pH-dependent properties of nucleic acids.



SECTION: Biophysical Chemistry and Biomolecules

The role of protonated nucleotides in modulating the pH-dependent properties of nucleic acids is at the frontier of questions in the field of nucleic acid biology.[1,2] Protonated nucleotides serve as key catalytic residues in many ribozymes,[3−10] and protonated base pairs are known to stabilize RNA loop structures.[11,12] They also control the pH-dependent dynamics of numerous RNA systems, such as the retrovirus pseudoknot,[13] the peptidyl-transferance center of the ribosome,[14−19] helix 69 of the 50S ribosomal subunit,[20,21] and the U6 intramolecular stem-loop of the spliceosome complex.[22] In DNA systems, the presence of protonated A$^+$·C base pairs is known to cause mutagenic and carcinogenic effects.[23,24] Measuring the pH dependence of such properties allows one to infer a macroscopic pK_a of the overall nucleic acid system, but limited information can be obtained about the specific residues that control its activity. Given the ubiquity of protonated nucleotides in regulating pH-dependent activity, the ability to measure the microscopic pK_a value associated with the protonation event at a specific residue is invaluable in identifying key nucleotides of interest, which is often necessary to obtain a deeper mechanistic insight into these pH-dependent processes.

Recently, Al-Hashimi and co-workers discovered the existence of low-population transient state conformations that are functionally important in both RNA and DNA systems, some of which are known to exhibit pH-dependent behavior.[25,26] Conventional experimental techniques such as NMR spectroscopy,[27−29] pH-dependent fluorescent nucleobase analogues,[30−32] and Raman spectroscopy[33,34] have not been able to directly characterize such pH-dependent transient states, although progress has been made through the development of novel NMR techniques.[35] Computational methods such as molecular dynamics (MD) simulations that have been traditionally used to provide mechanistic insight into the role of protonated residues[36−38] do not model realistic pH-dependent dynamics because the protonation states are fixed throughout the simulation. Therefore, there is a need for pH-coupled MD simulations, where the protonation state of titrating residues changes dynamically in response to their microenvironment. In addition, because no a priori information on the pK_a value of key titrating residues is required, pH-coupled MD simulations are uniquely suited to investigate pH-dependent transient states and other systems where there is limited experimental data.

One form of pH-coupled MD simulations, known as constant pH MD simulations, has been developed for proteins[39−43] and successfully applied to investigate numerous pH-dependent properties.[44−50] However, it was only recently that a newer framework of constant pH MD simulations based on multisite λ-dynamics (CPHMD$^{MS\lambda D}$) was established to address questions related to the pH-dependent properties of

Received: January 13, 2013
Accepted: February 12, 2013
Published: February 12, 2013

Table 1. Calculated pK_a Values from Conventional and pH-REX CPHMD$^{MS\lambda D}$ Simulations of the Lead-Dependent Ribozyme, Demonstrating a Similar Level of Accuracy

		conventional CPHMD$^{MS\lambda D}$ (3 × 5 ns)			pH-REX CPHMD$^{MS\lambda D}$ (3 ns)		
residue	exp pK_a	n	pK_a	error	n	pK_a	error
A4	<3.1	0.4 ± 0.1	0.6 ± 0.1		1.3 ± 0.5	0.9 ± 0.4	
A8	4.3 ± 0.3	0.7 ± 0.3	3.7 ± 0.3	−0.6	0.9 ± 0.4	3.8 ± 0.6	−0.5
A12	<3.1	1.1 ± 0.3	0.7 ± 0.3		1.0 ± 0.3	0.6 ± 0.2	
A16	3.8 ± 0.4	0.7 ± 0.1	2.6 ± 0.1	−1.2	0.7 ± 0.1	2.6 ± 0.0	−1.2
A17	3.8 ± 0.4	0.4 ± 0.0	0.9 ± 0.5	−2.9	1.0 ± 0.6	1.1 ± 0.5	−2.7
A18	3.5 ± 0.6	0.6 ± 0.0	3.8 ± 0.1	0.3	0.8 ± 0.1	3.6 ± 0.0	0.1
A25	6.5 ± 0.1	0.4 ± 0.1	4.8 ± 0.5	−1.7	0.5 ± 0.1	4.5 ± 0.2	−2.0
AUE				1.3			1.3

nucleic acids.[51] Blind pK_a prediction on the lead-dependent ribozyme has shown that the directions of pK_a shifts were accurately predicted, with an average unsigned error (AUE) of 1.3 pK_a units relative to experimental pK_a values.[52] However, for residues in a GAAA tetraloop, which presents significant sampling challenges because of conformation-dependent pK_a behavior and coupled titrating interactions, the calculated pK_a values were predicted with lower accuracy,[52] hampering the usefulness of CPHMD$^{MS\lambda D}$ simulations for predictive studies. In this Letter, we describe the application of pH-based replica exchange (pH-REX) to augment the sampling capabilities of CPHMD$^{MS\lambda D}$ simulations. Using pH-REX significantly improves sampling of titration and spatial coordinates of the residues in the GAAA tetraloop, reducing the error of A17, the most poorly predicted residue, from 2.9 to 1.2 pK_a units. Our work provides evidence that pH-REX CPHMD$^{MS\lambda D}$ simulations allow one to achieve accurate pK_a predictions, to around 1 pK_a unit, even for residues that are problematic in conventional CPHMD$^{MS\lambda D}$ simulations.

We first present our results on the performance of pH-REX CPHMD$^{MS\lambda D}$ simulations of the lead-dependent ribozyme. To judge the quality of a computational model, pH-dependent experimental observables, such as microscopic pK_a values, may be used as an indicator of how accurately the CPHMD$^{MS\lambda D}$ simulation reproduces pH-dependent properties. Unlike proteins, where the microscopic pK_a values of multiple ionizable residues for numerous proteins have been measured,[53] the literature of nucleic acid pK_a research is much sparser, with only a single pK_a value measured for a handful of RNA systems. To the best of our knowledge, the lead-dependent ribozyme is the most thoroughly studied RNA system as it has the largest number of experimentally measured pK_a values.[54] Therefore, we use it as a model system for benchmarking the performance of pH-REX CPHMD$^{MS\lambda D}$ simulations. The microscopic pK_a values calculated from pH-REX simulations, as summarized in Table 1, are consistent with previous work that utilized CPHMD$^{MS\lambda D}$ with conventional MD simulations.[52] As illustrated in Figure 1, up to an 8-fold improvement in the transition rates in λ-space is observed in our pH-REX simulations. The sampling improvement of titration coordinates results in faster convergence, which is demonstrated by fact that pH-REX sampling achieves the same level of accuracy using a total simulation time that is 5-fold shorter than that of conventional CPHMD$^{MS\lambda D}$ simulations. In addition, we also observe that the improvement in λ-space sampling for the residues of the lead-dependent ribozyme is higher than that of the 3-fold improvement in single nucleotide compounds (see Figure S1, Supporting Information).

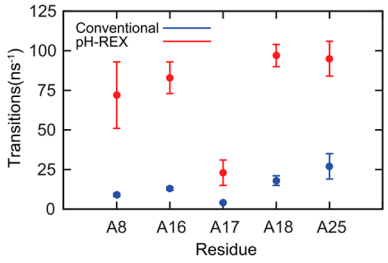


Figure 1. pH-REX CPHMD$^{MS\lambda D}$ simulations accelerate sampling of titration coordinates by up to 8-fold in the lead-dependent ribozyme.

In complex RNA structures, where multiple residues are titrated simultaneously, the coupled interactions between these titrating groups lead to slower convergence because the sampling of titration coordinates is hindered by the protonation states of adjacent interacting titrating groups.[52] The variable biases applied in conventional CPHMD$^{MS\lambda D}$ simulations only serve to flatten the potential energy surface of each λ variable, but the orthogonal barriers that arise from these coupled titrating interactions are not addressed. Unlike the recent methodological advances in enhanced sampling strategies reported by Yang and co-workers,[55,56] pH-REX sampling does not directly address these orthogonal barriers per se. However, it does periodically shuffle conformations to a higher or lower pH where all residues adopt a uniform protonation state. We suggest that this process effectively decouples the protonation states of interacting residues, allowing one to ameliorate the sampling issues related to these orthogonal barriers.

Having demonstrated that pH-REX accelerates sampling of titration coordinates, we now explore the apparent underprediction of the pK_a value of residue A17, which is situated in the GAAA tetraloop of the lead-dependent ribozyme. We performed an initial 15 ns simulation of the excised GAAA tetraloop for pH values between 1 and 4 and compared the results to conventional CPHMD$^{MS\lambda D}$ simulations. As summarized in Table 2 and Table S1 (Supporting Information), the calculated pK_a of residue A17 from the conventional simulations is 1.4, compared to the pK_a of 2.3 obtained using pH-REX sampling. Extending our simulations for an additional 15 ns confirmed that the pK_a value has converged (Figure S2, Supporting Information). On the whole, pH-REX sampling yields systematic improvement of the predicted pK_a values of the GAAA tetraloop, where the AUE was reduced to 0.7 pK_a units, which is 50% lower than that from our previous work.[52]

The apparent underprediction of the pK_a value of residue A17 originates from the anticooperative interactions between

dx.doi.org/10.1021/jz400078d | J. Phys. Chem. Lett. 2013, 4, 760−766

Table 2. pH-REX CPHMD$^{MS\lambda D}$ Simulations of the GAAA Tetraloop of Lead-Dependent Ribozyme that Improved the Accuracy of Calculated pK_a Values Compared to Those of Straightforward MD Simulations

	GAAA				AAA
residue	exp pK_a	conventional CPHMD$^{MS\lambda D}$ (0−15 ns)	pH-REX CPHMD$^{MS\lambda D}$ (0−15 ns)	pH-REX CPHMD$^{MS\lambda D}$ (16−30 ns)	pH-REX CPHMD$^{MS\lambda D}$ (0−15 ns)
A16	3.8 ± 0.4	3.2 ± 0.2	3.1 ± 0.1	3.3 ± 0.1	3.5 ± 0.1
A17	3.8 ± 0.4	1.4 ± 0.3	2.3 ± 0.6	2.6 ± 0.4	3.5 ± 0.1
A18	3.5 ± 0.6	3.9 ± 0.1	3.9 ± 0.1	4.0 ± 0.1	3.9 ± 0.1
AUE		1.1	0.9	0.7	

residues A17 and A18, which artificially suppresses the ability of A17 to adopt the protonated state at low pH conditions.[52] This arises from the triply stacked conformation of the GAAA tetraloop, which is characterized by short interatomic distances between the N1 atoms of the two residues. We analyzed this interatomic distance in our simulations of the GAAA tetraloop at pH 1, in the context of another reference simulation of the AAA trinucleotide sequence, which has no structural elements imposing conformational restrictions on it. As shown in Figure 2, the N1−N1 distances sampled in our pH-REX simulations of

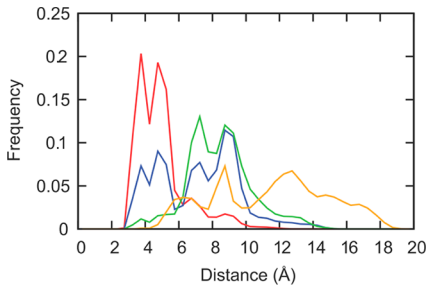


Figure 2. Distribution of the interatomic distance of the N1 atoms of residues A17 and A18 of the GAAA tetraloop at pH 1 for a conventional 15 ns MD simulation (red trace), the first 15 ns pH-REX simulation (blue trace), and the next 15 ns pH-REX simulation (green trace) compared to that of a 15 ns pH-REX simulation of the AAA trinucleotide (orange trace).

the GAAA tetraloop ranged from 2 to 10 Å, which are intermediate between the those of conventional GAAA tetraloop simulations (2−6 Å) and the AAA trinucleotide simulations (6−18 Å). The conformational space sampled using pH-REX is reasonable as it does not exhibit more dynamical behavior than the free AAA trinucleotide, but it also samples more conformations than conventional CPHMD$^{MS\lambda D}$ simulations of the GAAA tetraloop. This trend of exploring progressively larger N1−N1 distances results in more weakly coupled interactions, which is reflected in the pK_a value of the central adenine residue, which increases from 1.4 to 2.3 to 3.5 (Table 2). In addition, we also observed a slight difference in the distribution of the N1−N1 distances between the first 15 ns and the subsequent 15 ns trajectory of the excised GAAA tetraloop. Specifically, the "close contact" region of 3−6 Å that describes the initial stacked conformation was partially populated in the first 15 ns, which suggests that the system was still equilibrating. This suggests that sufficient equilibration on the order of a few nanoseconds may be required, and metrics such as RMSD relative to the initial structure may be used to determine when equilibration is complete, particularly if one is expecting a significant conformational change in an alternate pH environment.

The experimentally measured pK_a values are a superposition of the microscopic pK_a values of the various conformations visited by the GAAA tetraloop on the time scale accessible to NMR measurements,[52] and the various pH values at which such measurements were recorded. We clustered the conformations sampled by pH-REX simulations at pH 1 (low pH) and 4 (high pH), and the representative structures are illustrated in Figure 3. The initial triply stacked conformation

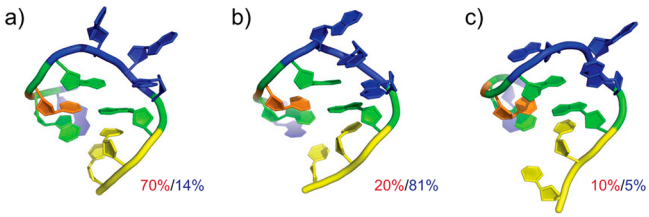


Figure 3. Representative conformations from a cluster analysis of the pH-REX trajectory of the GAAA tetraloop and their relative populations sampled at pH 1 (in red) and 4 (in blue).

(Figure 3b), which is representative of the NMR structure solved at physiological pH, is known to lower the pK_a value of residue A17. While it may be the dominant conformation sampled at high pH, this conformation is populated only 20% of the time at low pH. We observe that the dominant conformation sampled at low pH is partially unstacked, where the N1−N1 distance is increased to 9.3 Å (Figure 3a), and a fully unstacked conformation is observed 10% of the time (Figure 3c). Interestingly, these unstacked conformations are populated 21% of the time at higher pH. The significant improvement in our pK_a predictions for residues in the GAAA tetraloop corresponds to the sampling of these alternative conformations, suggesting that sampling using pH-REX provides a more accurate model of the tetraloop's pH-dependent dynamics. Lastly, our results also indicate that pH-REX CPHMD$^{MS\lambda D}$ simulations can be used to identify the dominant conformation of nucleic acid systems in different pH environments or low-population conformations at physiological pH. With the discovery of pH-dependent transient states in both RNA and DNA systems that have been suggested to be functionally important,[25,26] we anticipate that pH-REX CPHMD$^{MS\lambda D}$ simulations will provide further structural and mechanistic insight into the findings gleaned from experimental studies, especially in situations where direct experimental characterization of such transient states is challenging.

Thus far, we have shown that pH-REX CPHMD$^{MS\lambda D}$ simulations are effective in modeling accurate pH-dependent dynamics of small nucleic acid motifs like the GAAA tetraloop. In this final section, we demonstrate the scalability of pH-REX CPHMD$^{MS\lambda D}$ simulations to larger systems. Our initial pK_a calculations on the full-length ribozyme (Table 1) yielded

dx.doi.org/10.1021/jz400078d | J. Phys. Chem. Lett. 2013, 4, 760−766

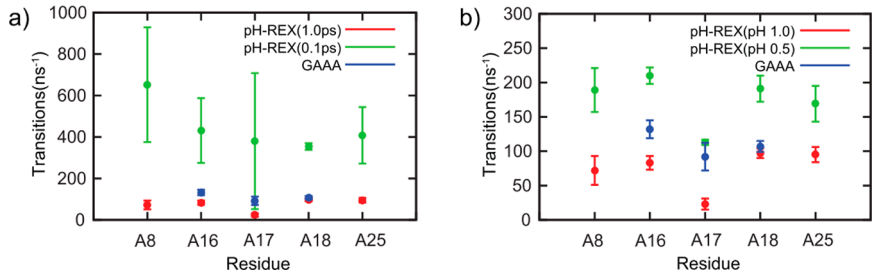


Figure 4. Effects on titration coordinate sampling by (a) increasing the MC attempt frequency and (b) reducing the pH window spacing from 1.0 to 0.5.

Table 3. pH-REX CPHMD$^{MS\lambda D}$ Simulations of the Full-Length Lead-Dependent Ribozyme at a 0.5 pH Window Spacing, Demonstrating Comparable Results to the GAAA Tetraloop within 13 ns

		pH-REX CPHMD$^{MS\lambda D}$				
residue	exp pK_a	(0–3 ns)	(3–8 ns)	(8–13 ns)	(13–18 ns)	(18–23 ns)
A16	3.8 ± 0.4	2.6 ± 0.1	2.7 ± 0.2	2.9 ± 0.2	2.9 ± 0.1	2.8 ± 0.1
A17	3.8 ± 0.4	1.4 ± 0.3	1.5 ± 0.6	1.8 ± 0.6	2.4 ± 0.1	2.4 ± 0.1
A18	3.5 ± 0.6	3.8 ± 0.1	3.8 ± 0.1	3.9 ± 0.1	3.8 ± 0.1	3.7 ± 0.0

similar results to conventional CPHMD$^{MS\lambda D}$ simulations, which suggest that the sampling efficiency is not as high as that in our simulations of the excised GAAA tetraloop. Because conformational diffusion across pH space is responsible for enhancing sampling, increasing the total number of accepted Monte Carlo (MC) moves should improve the accuracy of calculated pK_a values. In the full-length ribozyme, the majority (i.e., 10 out of 15) of the residues are base-paired and have pK_a values of less than 3.[52] Thus, unlike high pH conditions where most of the titrating residues adopt a uniform protonation state, at low pH conditions, the majority of residues would be titrating. This implies that there would be a more pronounced potential energy difference between adjacent replicas, and consequently, a lower MC exchange rate is expected at low pH. Under low pH conditions, the MC exchange rate of the excised tetraloop was at least 30%, which is 3 times higher than that of the full-length ribozyme (Supporting Information Figure S3). This lower exchange rate in the full-length ribozyme is correlated with a reduction in the sampling of titration coordinates, particularly for residue A17 (Figure 4). As shown in Figure 4a, increasing the frequency of MC attempts from every 1.0 to 0.1 ps significantly increased λ-transitions to ∼350–650 ns^{-1}. Prior work by Baptista and co-workers has shown that re-equilibration of the solvent induced by the introduction of a charged protonation state requires 1–3 ps,[39] and we have observed a similar solvent reorganization triggered by a protonation state change in explicit solvent CPHMD$^{MS\lambda D}$ simulation.[51] Using the mean solvent relaxation time of 2 ps, one arrives at a transition rate of ∼500 ns^{-1}, which balances the need for titration coordinate sampling and solvent relaxation. In some instances, such as residue A8, the transition rate was above this value. We acknowledge that the pH-REX metropolis criterion maintains detailed balance, and the results should, in principle, be unaffected by the MC exchange frequency. However, we observed significant variation in transition rates between the three independent runs, which could cause issues in reproducibility and convergence. Therefore, in practice, it may be useful to allow for sufficient solvent relaxation occur.

Our observations on increasing the MC exchange frequency differ with the findings reported by Roitberg and co-workers, where no performance degradation was observed at higher MC exchange frequencies.[57] This difference is likely due to the fact that our model uses an explicit solvent representation where solvent reorganization needs to be accounted for, whereas the work of Roitberg and co-workers was performed in implicit solvent, which adiabatically adjusts to the protein conformation. Instead of attempting to increase the MC exchange frequency, one may also increase the probability of exchange by reducing the potential energy difference between adjacent windows (i.e., reduce the pH spacing). In simulations using a smaller spacing of 0.5 pH units, the exchange rate for the full-length ribozyme was increased to 40%, as shown in Supporting Information Figure S4. As illustrated in Figure 4b, reducing the pH spacing more than doubled the transition rate in λ-space. We observed the most significant improvement in the transition rate of residue A17, which increased from 23 to 113 ns^{-1}. This is on par with the transition rate of 92 ns^{-1} observed in the GAAA tetraloop. The transition rate was also uniformly consistent across the three independent simulation runs, which ensures the robustness of the calculations. Qualitatively, the titration curves obtained across three independent runs also demonstrated better convergence for pH-REX simulations with smaller pH spacing (Figure S5, Supporting Information). Finally, using this smaller pH spacing of 0.5, we reran pH-REX CPHMD$^{MS\lambda D}$ simulations on the full-length ribozyme. After an initial ∼10 ns of equilibration, the calculated pK_a values started to converge, and results comparable to the GAAA tetraloop were achieved within 13 ns (Table 3, Table S2, and Figure S6 (Supporting Information)), demonstrating that pH-REX CPHMD$^{MS\lambda D}$ simulation scales well to simulate pH-dependent properties of full-sized nucleic acid systems.

In conclusion, we have enhanced the framework of constant pH molecular dynamics (CPHMD$^{MS\lambda D}$) for nucleic acids with pH-based replica exchange (pH-REX) sampling. Using pH-REX significantly improves sampling of titration and spatial coordinates, and the shuffling of conformations across pH space has the effect of decoupling interactions between titrating residues. This allows us to ameliorate some of the sampling issues related to orthogonal barriers that originate from the coupled protonation equilibrium and conformational-dependent pK_a behavior, as illustrated in our example of the GAAA tetraloop motif. Our pK_a calculations on the GAAA tetraloop

dx.doi.org/10.1021/jz400078d | J. Phys. Chem. Lett. 2013, 4, 760–766

indicate that pH-REX reduced the AUE to 0.7 pK_a units, which is 50% lower than that from previous work,[52] and the error of the most poorly predicted residue was drastically reduced from 2.9 to 1.2 pK_a units. The scalability of pH-REX sampling was also demonstrated by showing that similarly accurate pK_a values could be achieved when simulating full-sized nucleic acid systems, such as the lead-dependent ribozyme. Our results affirm that pH-REX CPHMD$^{MS\lambda D}$ simulation is maturing into a practical tool for accurate first-principles prediction of nucleic acid protonation equilibrium, where accurate modeling of pH-dependent properties is achieved, even for residues that have hampered convergence due to sampling issues (e.g., slow dynamics and/or coupled titration behavior). Finally, we highlighted that pH-REX CPHMD$^{MS\lambda D}$ simulations can be used to identify the dominant conformation of nucleic acid structures in alternate pH environments or to provide structural characterization of pH-dependent transient states, making it a useful tool to provide structural and mechanistic insight into the study of pH-dependent properties of nucleic acids.

METHODS

Structure Preparation. Input structures of the nucleotide compounds were generated from the CHARMM topology files using the *IC* facility in CHARMM. The input structure for the lead-dependent ribozyme was generated from the PDB file (Accession code: 1LDZ).[58] Hydrogen atoms were added using the *HBUILD* facility in CHARMM.[59] The systems were solvated in a cubic box of explicit TIP3P water[60] using the convpdb.pl tool from the MMTSB toolset,[61] and the ionic strength was simulated by adding the appropriate number of Na^+ and Cl^- counterions to match the experimental salt concentration (100−150 mM NaCl). The two isomers of the test nucleotides were constructed using the patch keywords *5PHO* and *3PHO*. All other nucleic acid structures were hydroxylated using the *5TER* and *3TER* patches in CHARMM. The excised GAAA tetraloop was constructed by extracting residues 12−21 from the lead-dependent ribozyme, and harmonic distance restraints were applied to enforce base pairing between residues A12 and U21. Patches for the protonated forms of adenine and cytosine were constructed as previously reported (see Supporting Information Methods).[51]

Simulation Details. MD simulations were performed within the CHARMM macromolecular modeling program (version c36a6) using the CHARMM36 all-atom force field for RNA[62] and TIP3P water.[60] The SHAKE algorithm[63] was used to constrain the hydrogen-heavy atom bond lengths. The Leapfrog Verlet integrator was used with an integration time step of 2 fs. A nonbonded cutoff of 12 Å was used with an electrostatic force shifting function and a van der Waals switching function between 10 and 12 Å. The distance cutoff in generating the list of pairwise interactions was 15 Å. CPHMD simulations utilize an extended Hamiltonian approach, where the protonation state of the residue is described by a continuous variable, λ, which is propagated simultaneously with the spatial coordinates at a specified external pH using multisite λ-dynamics (see Supporting Information Methods).[64,65] After an initial minimization, the system was heated for 100 ps and equilibrated for 100 (for nucleotide compounds) or 400 ps (for RNA structures), followed by a production run of 3 ns, unless otherwise stated. A pH-REX sampling protocol for CPHMD simulations in hybrid and implicit solvent has been previously reported by Wallace and Shen,[66] and we have implemented their formalism with our explicit solvent CPHMD$^{MS\lambda D}$ simulations. Replicas that were simulated at various pH conditions are exchanged based on the following Metropolis criterion.

$$P = \begin{cases} 1; & \text{if } \Delta \le 0 \\ \exp(-\Delta); & \text{otherwise} \end{cases} \quad \text{where}$$

$$\Delta = \beta \begin{bmatrix} U^{pH}(\{\lambda_i\};\ pH') + U^{pH}(\{\lambda_i'\};\ pH) \\ -U^{pH}(\{\lambda_i\};\ pH) - U^{pH}(\{\lambda_i'\};\ pH') \end{bmatrix} \tag{1}$$

Exchange attempts were made at every 1 ps unless otherwise stated. CPHMD$^{MS\lambda D}$ simulations were performed from pH 1 to 4 for the GAAA tetraloop and pH 1 to 8 for the full-length lead-dependent ribozyme, with integer value pH spacing (unless otherwise specified). All CPHMD$^{MS\lambda D}$ simulations were performed in triplicate.

Calculating pK_a Values. The populations of unprotonated (N^{unprot}) and protonated (N^{prot}) states are defined as the total number of times in the trajectory where conditions $\lambda_{\alpha,1} > 0.8$ and $\lambda_{\alpha,2} > 0.8$ are satisfied, respectively, and are used to derive the unprotonated fraction (S^{unprot}).

$$S^{unprot}(pH) = \frac{N^{unprot}(pH)}{N^{unprot}(pH) + N^{prot}(pH)} \tag{2}$$

These S^{unprot} ratios that were computed across the entire pH range were then fitted to a generalized version of the Henderson−Hasselbalch formula to obtain a single pK_a value.

$$S^{unprot}(pH) = \frac{1}{1 + 10^{-n(pH - pK_a)}} \tag{3}$$

The reported pK_a value and its error correspond to the mean and standard deviation calculated from three sets of independent runs. In the calculation of transition rates, a transition is defined as a move in λ-space between physical protonation states using the same definitions for calculating N^{unprot} and N^{prot} (i.e., moving between $\lambda_{\alpha,1} > 0.8$ and $\lambda_{\alpha,1} < 0.2$ constitutes a valid transition). The transition rate statistics reported for each residue were calculated from the simulation where the external pH was closest to the pK_a value of the residue.

ASSOCIATED CONTENT

Supporting Information

Additional titration curves, table of Hill coefficients, and analysis of the MC exchange rate. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health (GM037554 and GM057513). We thank Dr. Harish Vashisth for critically reviewing the manuscript and providing helpful comments.

REFERENCES

(1) Wilcox, J. L.; Ahluwalia, A. K.; Bevilacqua, P. C. Charged Nucleobases and Their Potential for RNA Catalysis. *Acc. Chem. Res.* **2011**, *44*, 1270−1279.

(2) Krishnamurthy, R. Role of pK(a) of Nucleobases in the Origins of Chemical Evolution. *Acc. Chem. Res.* **2012**, *45*, 2035−2044.
(3) Shih, I. H.; Been, M. D. Involvement of a Cytosine Side Chain in Proton Transfer in the Rate-Determining Step of Ribozyme Self-Cleavage. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 1489−1494.
(4) Wadkins, T. S.; Shih, I.; Perrotta, A. T.; Been, M. D. A pH-Sensitive RNA Tertiary Interaction Affects Self-Cleavage Activity of the HDV Ribozymes in the Absence of Added Divalent Metal Ion. *J. Mol. Biol.* **2001**, *305*, 1045−1055.
(5) Ke, A.; Zhou, K.; Ding, F.; Cate, J. H.; Doudna, J. A. A Conformational Switch Controls Hepatitis Delta Virus Ribozyme Catalysis. *Nature* **2004**, *429*, 201−205.
(6) Cerrone-Szakal, A. L.; Siegfried, N. A.; Bevilacqua, P. C. Mechanistic Characterization of the HDV Genomic Ribozyme: Solvent Isotope Effects and Proton Inventories in the Absence of Divalent Metal Ions Support C75 as the General Acid. *J. Am. Chem. Soc.* **2008**, *130*, 14504−14520.
(7) Ravindranathan, S.; Butcher, S. E.; Feigon, J. Adenine Protonation in Domain B of the Hairpin Ribozyme. *Biochemistry* **2000**, *39*, 16026−16032.
(8) Ryder, S. P.; Oyelere, A. K.; Padilla, J. L.; Klostermeier, D.; Millar, D. P.; Strobel, S. A. Investigation of Adenosine Base Ionization in the Hairpin Ribozyme by Nucleotide Analog Interference Mapping. *RNA* **2001**, *7*, 1454−1463.
(9) Kuzmin, Y. I.; Da Costa, C. P.; Cottrell, J. W.; Fedor, M. J. Role of an Active Site Adenine in Hairpin Ribozyme Catalysis. *J. Mol. Biol.* **2005**, *349*, 989−1010.
(10) Nam, K.; Gao, J.; York, D. M. Quantum Mechanical/Molecular Mechanical Simulation Study of the Mechanism of Hairpin Ribozyme Catalysis. *J. Am. Chem. Soc.* **2008**, *130*, 4680−4691.
(11) Durant, P. C.; Davis, D. R. Stabilization of the Anticodon Stem-Loop of tRNALys,3 by an A^+−C Base-Pair and by Pseudouridine. *J. Mol. Biol.* **1999**, *285*, 115−131.
(12) Chen, G.; Kennedy, S. D.; Turner, D. H. A CA^+ Pair Adjacent to a Sheared GA or AA Pair Stabilizes Size-Symmetric RNA Internal Loops. *Biochemistry* **2009**, *48*, 5738−5752.
(13) Nixon, P. L.; Giedroc, D. P. Energetics of a Strongly pH Dependent RNA Tertiary Structure in a Frameshifting Pseudoknot. *J. Mol. Biol.* **2000**, *296*, 659−671.
(14) Bayfield, M. A.; Dahlberg, A. E.; Schulmeister, U.; Dorner, S.; Barta, A. A Conformational Change in the Ribosomal Peptidyl Transferase Center Upon Active/Inactive Transition. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10096−10101.
(15) Muth, G. W.; Chen, L.; Kosek, A. B.; Strobel, S. A. pH-Dependent Conformational Flexibility within the Ribosomal Peptidyl Transferase Center. *RNA* **2001**, *7*, 1403−1415.
(16) Xiong, L.; Polacek, N.; Sander, P.; Bottger, E. C.; Mankin, A. pKa of Adenine 2451 in the Ribosomal Peptidyl Transferase Center Remains Elusive. *RNA* **2001**, *7*, 1365−1369.
(17) Hesslein, A. E.; Katunin, V. I.; Beringer, M.; Kosek, A. B.; Rodnina, M. V.; Strobel, S. A. Exploration of the Conserved A+C Wobble Pair Within the Ribosomal Peptidyl Transferase Center using Affinity Purified Mutant Ribosomes. *Nucleic Acids Res.* **2004**, *32*, 3760−3770.
(18) Beringer, M.; Bruell, C.; Xiong, L.; Pfister, P.; Bieling, P.; Katunin, V. I.; Mankin, A. S.; Bottger, E. C.; Rodnina, M. V. Essential Mechanisms in the Catalysis of Peptide Bond Formation on the Ribosome. *J. Biol. Chem.* **2005**, *280*, 36065−36072.
(19) Beringer, M.; Rodnina, M. V. The Ribosomal Peptidyl Transferase. *Mol. Cell* **2007**, *26*, 311−321.
(20) Abeysirigunawardena, S. C.; Chow, C. S. pH-Dependent Structural Changes of Helix 69 from Escherichia Coli 23S Ribosomal RNA. *RNA* **2008**, *14*, 782−792.
(21) Sakakibara, Y.; Chow, C. S. Probing Conformational States of Modified Helix 69 in 50S Ribosomes. *J. Am. Chem. Soc.* **2011**, *133*, 8396−8399.
(22) Reiter, N. J.; Blad, H.; Abildgaard, F.; Butcher, S. E. Dynamics in the U6 RNA Intramolecular Stem-Loop: A Base Flipping Conformational Change. *Biochemistry* **2004**, *43*, 13739−13747.
(23) Kim, M.; Huang, T.; Miller, J. H. Competition Between MutY and Mismatch Repair at A × C Mispairs In Vivo. *J. Bacteriol.* **2003**, *185*, 4626−4629.
(24) Giri, I.; Stone, M. P. Wobble dC.dA Pairing 5′ to the Cationic Guanine N7 8,9-Dihydro-8-(N7-guanyl)-9-hydroxyaflatoxin B1 Adduct: Implications for Nontargeted AFB1Mutagenesis. *Biochemistry* **2003**, *42*, 7023−7034.
(25) Dethoff, E. A.; Chugh, J.; Mustoe, A. M.; Al-Hashimi, H. M. Functional Complexity and Regulation Through RNA Dynamics. *Nature* **2012**, *482*, 322−330.
(26) Nikolova, E. N.; Kim, E.; Wise, A. A.; O'Brien, P. J.; Andricioaei, I.; Al-Hashimi, H. M. Transient Hoogsteen Base Pairs in Canonical Duplex DNA. *Nature* **2011**, *470*, 498−502.
(27) Legault, P.; Pardi, A. In Situ Probing of Adenine Protonation in RNA by ^{13}C NMR. *J. Am. Chem. Soc.* **1994**, *116*, 8390−8391.
(28) Wang, C.; Gao, H.; Gaffney, B. L.; Jones, R. A. Nitrogen-15-Labeled Oligodeoxynucleotides. 3. Protonation of the Adenine N1 in the A·C and A·G Mispairs of the Duplexes $\{d[CG(^{15}N^1)$-$AGAATTCCCG]\}_2$ and $\{d[CGGGAATTC(^{15}N^1)ACG]\}_2$. *J. Am. Chem. Soc.* **1991**, *113*, 5486−5488.
(29) Moody, E. M.; Brown, T. S.; Bevilacqua, P. C. Simple Method for Determining Nucleobase pK_a Values by Indirect Labeling and Demonstration of a pK_a of Neutrality in dsDNA. *J. Am. Chem. Soc.* **2004**, *126*, 10200−10201.
(30) Liu, L.; Cottrell, J. W.; Scott, L. G.; Fedor, M. J. Direct Measurement of the Ionization State of an Essential Guanine in the Hairpin Ribozyme. *Nat. Chem. Biol.* **2009**, *5*, 351−357.
(31) Cottrell, J. W.; Scott, L. G.; Fedor, M. J. The pH Dependence of Hairpin Ribozyme Catalysis Reflects Ionization of an Active Site Adenine. *J. Biol. Chem.* **2011**, *286*, 17658−17664.
(32) Viladoms, J.; Scott, L. G.; Fedor, M. J. An Active-Site Guanine Participates in *glmS* Ribozyme Catalysis in its Protonated State. *J. Am. Chem. Soc.* **2011**, *133*, 18388−18396.
(33) Gong, B.; Chen, J. H.; Chase, E.; Chadalavada, D. M.; Yajima, R.; Golden, B. L.; Bevilacqua, P. C.; Carey, P. R. Direct Measurement of a pK_a Near Neutrality for the Catalytic Cytosine in the Genomic HDV Ribozyme using Raman Crystallography. *J. Am. Chem. Soc.* **2007**, *129*, 13335−13342.
(34) Guo, M.; Spitale, R. C.; Volpini, R.; Krucinska, J.; Cristalli, G.; Carey, P. R.; Wedekind, J. E. Direct Raman Measurement of an Elevated Base pK_a in the Active Site of a Small Ribozyme in a Precatalytic Conformation. *J. Am. Chem. Soc.* **2009**, *131*, 12908−12909.
(35) Nikolova, E. N.; Gottardo, F. L.; Al-Hashimi, H. M. Probing Transient Hoogsteen Hydrogen Bonds in Canonical Duplex DNA Using NMR Relaxation Dispersion and Single-Atom Substitution. *J. Am. Chem. Soc.* **2012**, *134*, 3667−3670.
(36) Ditzler, M. A.; Sponer, J.; Walter, N. G. Molecular Dynamics Suggest Multifunctionality of an Adenine Imino Group in Acid−Base Catalysis of the Hairpin Ribozyme. *RNA* **2009**, *15*, 560−575.
(37) Banas, P.; Walter, N. G.; Sponer, J.; Otyepka, M. Protonation States of the Key Active Site Residues and Structural Dynamics of the glmS Riboswitch As Revealed by Molecular Dynamics. *J. Phys. Chem. B* **2010**, *114*, 8701−8712.
(38) Mlýnský, V.; Banás, P.; Hollas, D.; Réblová, K.; Walter, N. G.; Sponer, J.; Otyepka, M. Extensive Molecular Dynamics Simulations Showing that Canonical G8 and Protonated $A38H^+$ Forms are Most Consistent with Crystal Structures of Hairpin Ribozyme. *J. Phys. Chem. B* **2010**, *114*, 6642−6652.
(39) Baptista, A. M.; Teixeira, V. H.; Soares, C. M. Constant pH Molecular Dynamics Using Stochastic Titration. *J. Chem. Phys.* **2002**, *117*, 4184−4200.
(40) Mongan, J.; Case, D. A.; McCammon, J. A. Constant pH Molecular Dynamics in Generalized Born Implicit Solvent. *J. Comput. Chem.* **2004**, *25*, 2038−2048.
(41) Messer, B. M.; Roca, M.; Chu, Z. T.; Vicatos, S.; Kilshtain, A. V.; Warshel, A. Multiscale Simulations of Protein Landscapes: Using Coarse-Grained Models as Reference Potentials to Full Explicit Models. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1212−1227.

(42) Lee, M. S.; Salsbury, F. R.; Brooks, C. L., III. Constant pH Molecular Dynamics Using Continuous Titration Coordinates. *Proteins: Struct., Funct., Bioinf.* **2004**, *56*, 738−752.
(43) Khandogin, J.; Brooks, C. L., III. Constant pH Molecular Dynamics with Proton Tautomerism. *Biophys. J.* **2005**, *89*, 141−157.
(44) Khandogin, J.; Chen, J.; Brooks, C. L., III. Exploring Atomistic Details of pH-Dependent Peptide Folding. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18546−18550.
(45) Khandogin, J.; Brooks, C. L., III. Linking Folding with Aggregation in Alzheimer's Beta-Amyloid Peptides. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 16880−16885.
(46) Khandogin, J.; Raleigh, D. P.; Brooks, C. L., III. Folding Intermediate in the Villin Headpiece Domain Arises From Disruption of a N-Terminal Hydrogen-Bonded Network. *J. Am. Chem. Soc.* **2007**, *129*, 3056−3057.
(47) Machuqueiro, M.; Baptista, A. M. The pH-Dependent Conformational States of Kyotorphin: A Constant pH Molecular Dynamics Study. *Biophys. J.* **2007**, *92*, 1836−1845.
(48) Zhang, B. W.; Brunetti, L.; Brooks, C. L., III. Probing pH-Dependent Dissociation of HdeA Dimers. *J. Am. Chem. Soc.* **2011**, *133*, 19393−19398.
(49) Wallace, J. A.; Shen, J. K. Unraveling a Trap-and-Trigger Mechanism in the pH-Sensitive Self-Assembly of Spider Silk Proteins. *J. Phys. Chem. Lett.* **2012**, *3*, 658−662.
(50) Russo, N. V. D.; Estrin, D. A.; Martí, M. A.; Roitberg, A. E. pH-Dependent Conformational Changes in Proteins and Their Effect on Experimental pK_as: The Case of Nitrophorin 4. *PLoS Comput. Biol.* **2012**, *8*, 1002761−1002761.
(51) Goh, G. B.; Knight, J. L.; Brooks, C. L., III. Constant pH Molecular Dynamics Simulations of Nucleic Acids in Explicit Solvent. *J. Chem. Theory Comput.* **2012**, *8*, 36−46.
(52) Goh, G. B.; Knight, J. L.; Brooks, C. L., III. pH-Dependent Dynamics of Complex RNA Macromolecules. *J. Chem. Theory Comput.* **2013**, *9*, 935−943.
(53) Nielsen, J. E.; Gunner, M. R.; Garcia-Moreno, E. B. The pK_a Cooperative: A Collaborative Effort to Advance Structure-Based Calculations of pK_a Values and Electrostatic Effects in Proteins. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 3249−3259.
(54) Legault, P.; Pardi, A. Unusual Dynamics and pK_a Shift at the Active Site of a Lead-Dependent Ribozyme. *J. Am. Chem. Soc.* **1997**, *119*, 6621−6628.
(55) Zheng, L.; Chen, M.; Yang, W. Random Walk in Orthogonal Space to Achieve Efficient Free-Energy Simulation of Complex Systems. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 20227−20232.
(56) Zheng, L. Q.; Yang, W. Practically Efficient and Robust Free Energy Calculations: Double-Integration Orthogonal Space Tempering. *J. Chem. Theory Comput.* **2012**, *8*, 810−823.
(57) Swails, J. M.; Roitberg, A. E. Enhancing Conformation and Protonation State Sampling of Hen Egg White Lysozyme Using pH Replica Exchange Molecular Dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 4393−4404.
(58) Hoogstraten, C. G.; Legault, P.; Pardi, A. NMR Solution Structure of the Lead-Dependent Ribozyme: Evidence for Dynamics in RNA Catalysis. *J. Mol. Biol.* **1998**, *284*, 337−350.
(59) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(60) 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*, 926−935.
(61) Feig, M.; Karanicolas, J.; Brooks, C. L., III. MMTSB Tool Set: Enhanced Sampling and Multiscale Modeling Methods for Applications in Structural Biology. *J. Mol. Graphics Modell.* **2004**, *22*, 377−395.
(62) Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; Mackerell, A. D., Jr. Impact of 2′-Hydroxyl Sampling on the Conformational Properties of RNA: Update of the CHARMM All-Atom Additive Force Field for RNA. *J. Comput. Chem.* **2011**, *32*, 1929−1943.
(63) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(64) Knight, J. L.; Brooks, C. L., III. Applying Efficient Implicit Non-geometric Constraints in Alchemical Free Energy Simulations. *J. Comput. Chem.* **2011**, *32*, 3423−3432.
(65) Knight, J. L.; Brooks, C. L., III. Multisite λ Dynamics for Simulated Structure−Activity Relationship Studies. *J. Chem. Theory Comput.* **2011**, *7*, 2728−2739.
(66) Wallace, J. A.; Shen, J. K. Continuous Constant pH Molecular Dynamics in Explicit Solvent with pH-Based Replica Exchange. *J. Chem. Theory Comput.* **2011**, *7*, 2617−2629.