Supporting Information

Towards Accurate Prediction of Protonation Equilibrium of Nucleic Acids

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

† Department of Chemistry, University of Michigan, 930 N. University, Ann Arbor, Michigan 48109, United States

‡ Biophysics Program, University of Michigan, 930 N. University, Ann Arbor, Michigan 48109, United States

Methods for Multi-Site λ-dynamics:

Additional patches were constructed to represent the protonated forms of adenine and cytosine. All of the associated bonds, angles and dihedrals were explicitly defined in the patch. Each titratable residue was simulated as a hybrid model that explicitly included atomic components of both the protonated and unprotonated forms. The titratable fragment included the nitrogen atom that is protonated, the protonated hydrogen and adjacent atoms whose partial charge differed according to the protonation state as reported previously by Goh *et. al.*[1] The environment atoms were defined as all atoms that were not included in the titratable fragments. λ dynamics was performed within the BLOCK facility using the MSλD framework (*MSLD*) and selecting the λ^{Nexp} functional form for λ (*FNEX*) using a coefficient of 5.5. Linear scaling by λ was applied to all energy terms except bond, angle and dihedral terms, which were treated at full strength regardless of λ value to retain physically reasonable geometries. In the MSλD λ^{Nexp} functional form, the dynamics of the fictitious λ particles that correspond to each residue's

protonation state are propagated by θ particles using the following equation, where α denotes the specific residue and *i/j* denotes the specific protonation state:

$$\lambda_{\alpha,i}^{N\exp} = \frac{e^{c\sin\theta_{\alpha,i}}}{\sum_{j=1}^{N} e^{c\sin\theta_{\alpha,j}}} \quad (1)$$

Each θ_α was assigned a fictitious mass of 12 amu•Å^2 and λ values were saved every 10 steps. Variable biases (F_{var}) were added to enhance transition rates between the two protonation states, and the associated force constant (k_{bias}) used were identical to the optimized values reported by by Goh *et. al.*[1] The temperature was maintained at 298K by coupling to a Langevin heatbath using a frictional coefficient of 10ps^{-1}.

Figure S1: (a) Transition rate and (b) fraction of physical ligand as a function of pH for model nucleosides demonstrate a 3-fold improvement in sampling of titration coordinates with pH-REX sampling, with a small 10% decrease in the fraction of physical states.

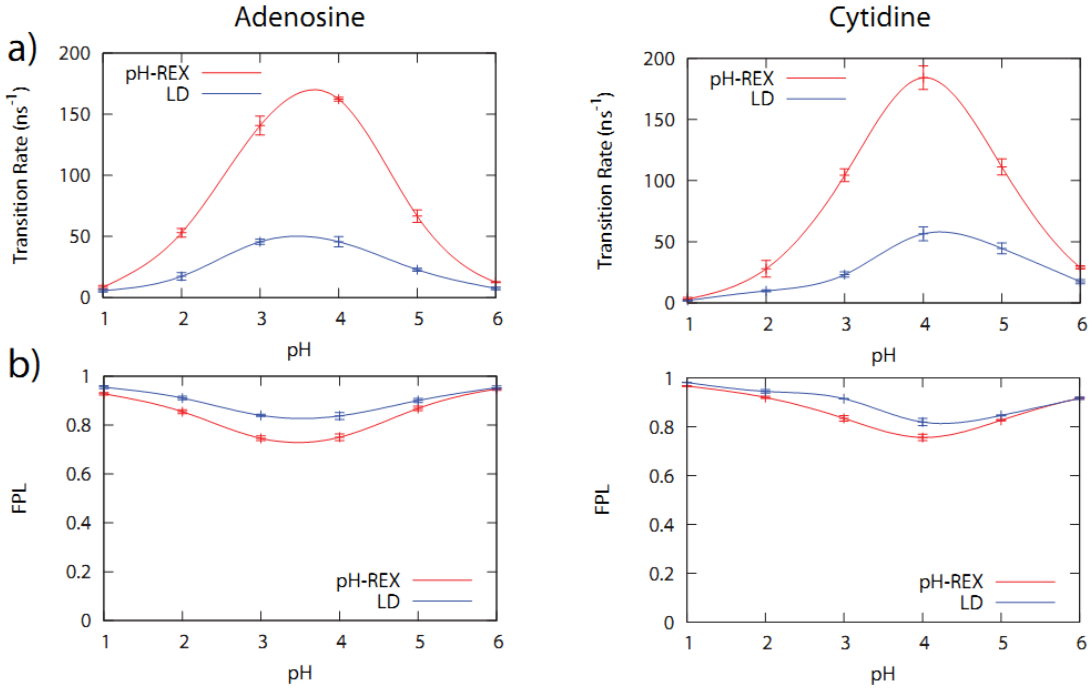


Table S1: Hill coefficients of the excised GAAA tetraloop and the AAA trinucleotide sequence.

Residue	GAAA			AAA
	Conventional CPHMD$^{MS\lambda D}$ (0-15ns)	pH-REX CPHMD$^{MS\lambda D}$ (0-15ns)	pH-REX CPHMD$^{MS\lambda D}$ (16-30ns)	pH-REX CPHMD$^{MS\lambda D}$ (0-15ns)
A16	1.1 ± 0.2	1.0 ± 0.1	1.1 ± 0.1	0.8 ± 0.1
A17	0.4 ± 0.1	1.2 ± 0.3	1.6 ± 0.2	0.8 ± 0.0
A18	0.8 ± 0.3	0.7 ± 0.1	0.8 ± 0.0	0.7 ± 0.1

Figure S2: Titration curves from the last 15 ns of pH-REX CPHMD$^{MS\lambda D}$ simulations of excised GAAA tetraloop for residues (a) A16, (b) A17 and (c) A18. Each color represents an independent run that have different seed numbers from one another.

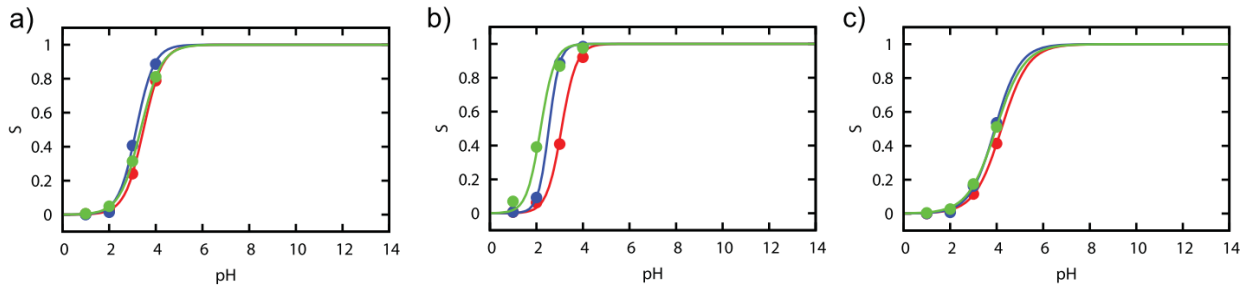


Table S2: Hill coefficients of the full-length lead-dependent ribozyme.

Residue	pH-REX CPHMD$^{MS\lambda D}$				
	(0-3ns)	(3-8ns)	(8-13ns)	(13-18ns)	(18-23ns)
A16	0.7 ± 0.1	0.7 ± 0.1	0.6 ± 0.2	0.8 ± 0.2	0.7 ± 0.0
A17	0.5 ± 0.1	0.5 ± 0.2	0.5 ± 0.3	0.5 ± 0.1	0.6 ± 0.1
A18	0.8 ± 0.1	0.8 ± 0.1	0.9 ± 0.2	0.8 ± 0.1	0.8 ± 0.1

Figure S3: MC exchange rate as a function of pH for (a) excised GAAA tetraloop and (b) full-length lead-dependent ribozyme. Error bars denote the standard deviation of exchange rate across 3 independent simulation runs.

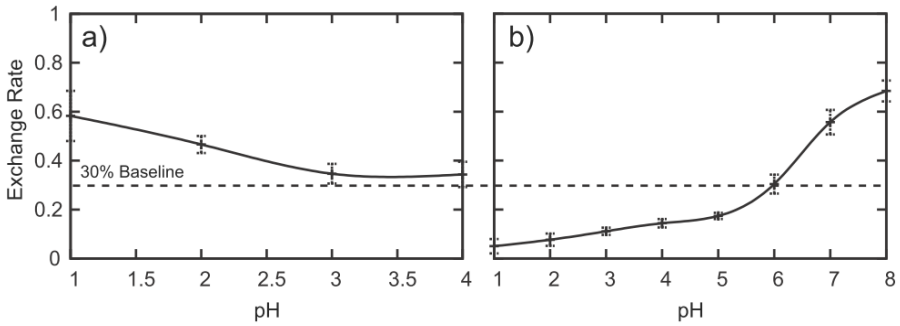


Figure S4: MC exchange rate as a function of pH window spacing, (a) 2.0 pH units, (b) 1.0 pH units, (c) 0.5 pH units for the full-length lead-dependent ribozyme. Error bars denote the standard deviation of exchange rate across 3 independent simulation runs.

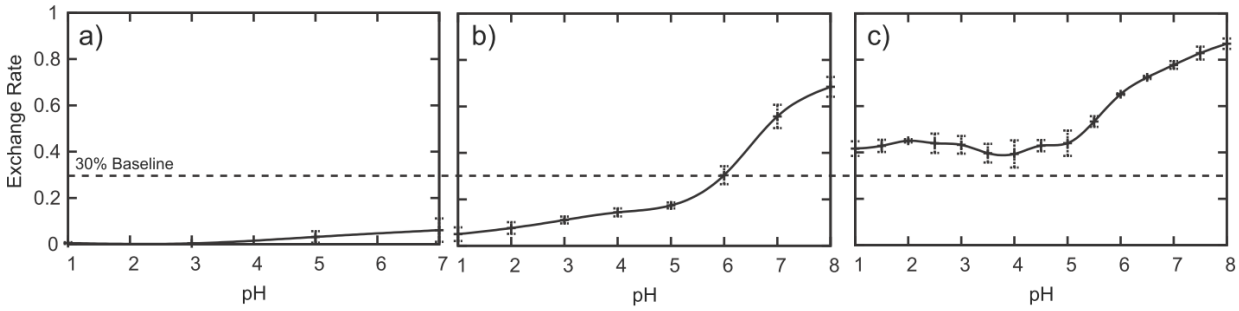


Figure S5: Titration curves for residues A16 to A18 calculated from the *first* 3ns of pH-REX CPHMD$^{MS\lambda D}$ simulations for the full-length lead-dependent ribozyme as a function of pH window spacing: (a) 0.5 pH units, (b) 1.0 pH units, (c) 2.0 pH units. Each color represents an independent run that have different seed numbers from one another.

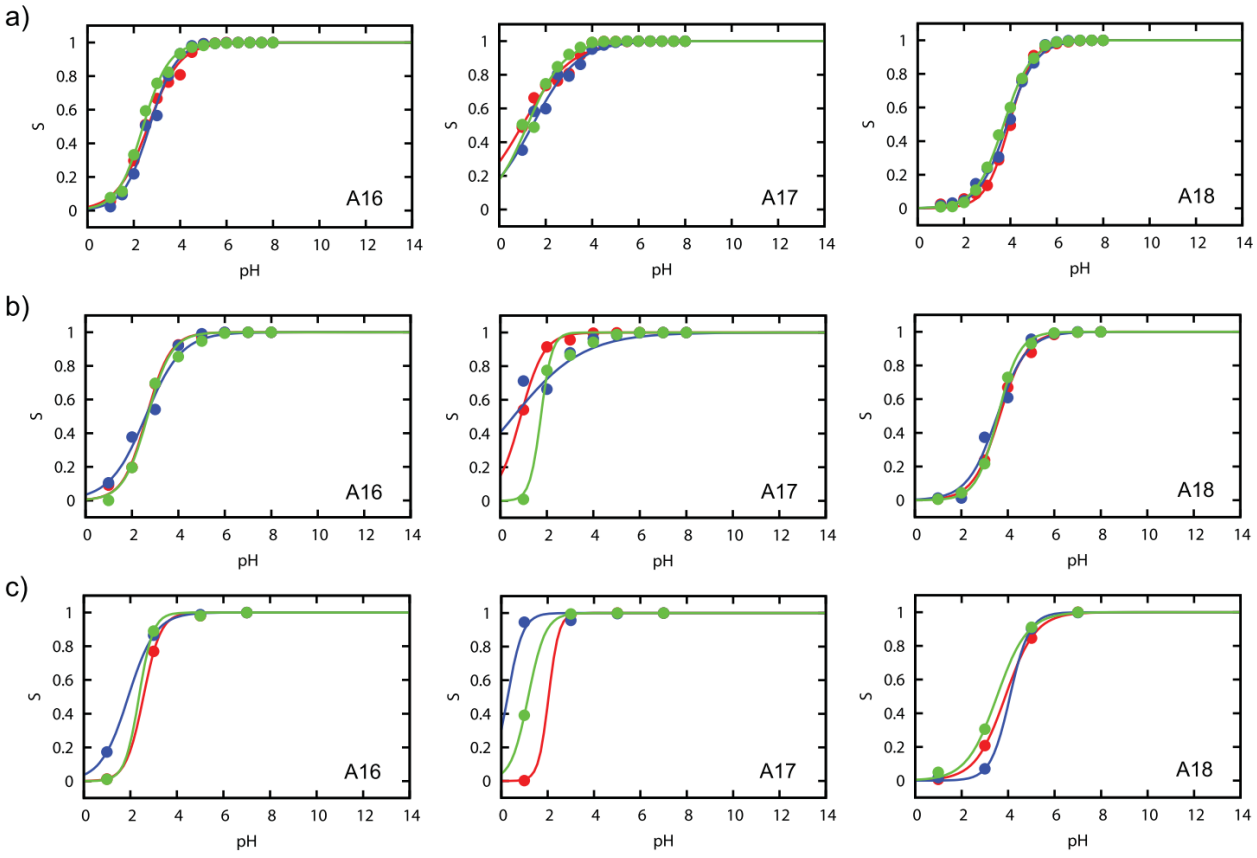
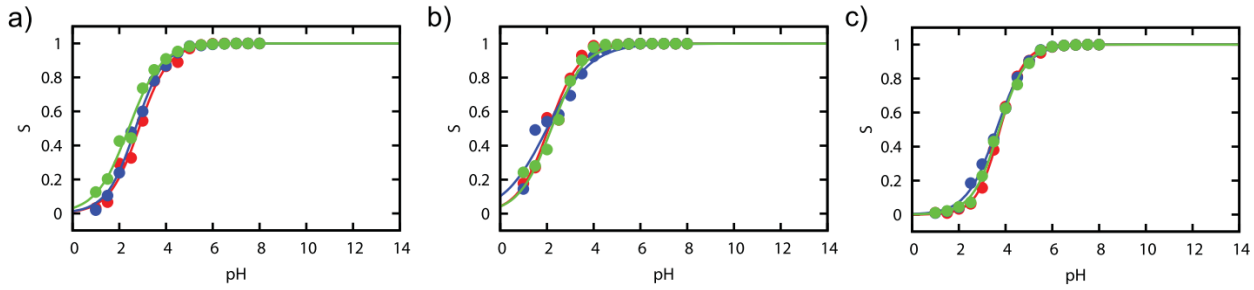


Figure S6: Titration curves from the last 5 ns of pH-REX CPHMD$^{MS\lambda D}$ simulations for the full-length lead-dependent ribozyme for residues (a) A16, (b) A17 and (c) A18. Each color represents an independent run that have different seed numbers from one another.



REFERENCES

(1) 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.