## Supporting Information

# Quantum Mechanical/Molecular Mechanical Free Energy Simulations of the Self-Cleavage Reaction in the Hepatitis Delta Virus Ribozyme

Abir Ganguly,[†] Pallavi Thaplyal,[‡] Edina Rosta,[§] Philip C. Bevilacqua,[‡*] and Sharon Hammes-Schiffer[†*]

[†] Department of Chemistry, 600 South Mathews Avenue, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States

[‡] Department of Chemistry, 104 Chemistry Building, Center for RNA Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, United States

[§] Department of Chemistry, 7 Trinity Street, Britannia House, King's College London, London SE1 1DB, United Kingdom

E-mail: shs3@illinois.edu; pcb5@psu.edu

## Table of Contents

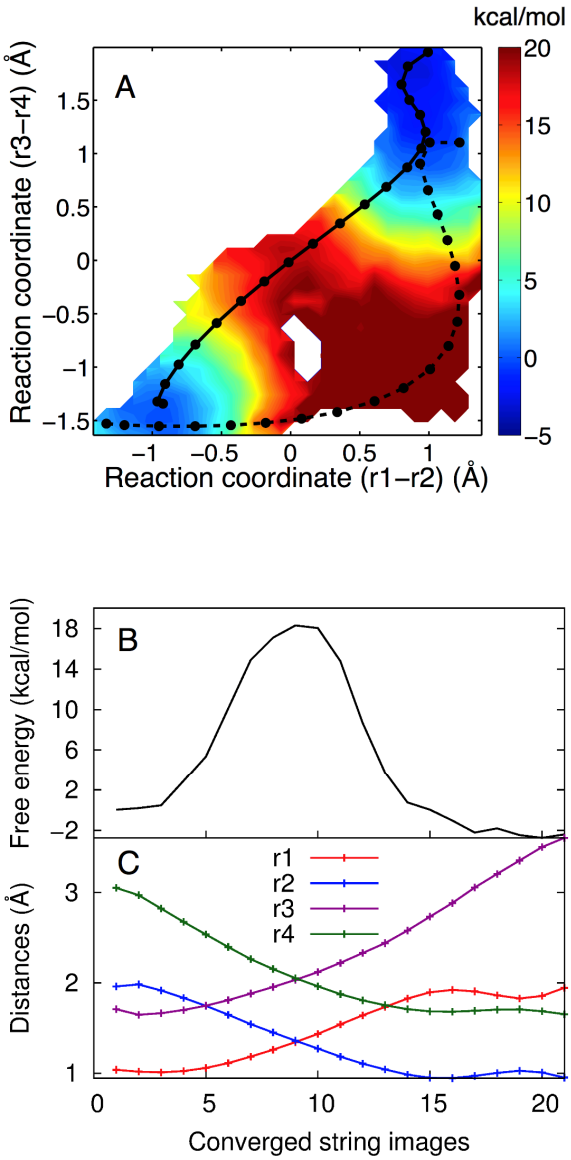


**Figure S1.** (A) The 2D free energy surface obtained from set B, where $Mg^{2+}$ is bound at the catalytic site, projected in the (r1−r2) and (r3−r4) space. The initial string (dashed black line) corresponds to an initial guess pathway in which the proton is transferred before the oxygen-phosphorus bond breaking/forming. To conserve computational resources, the string calculations in this set were terminated after 30 string iterations, when the reaction pathway associated with the final string (solid black line) was determined to be qualitatively similar to that obtained from sets A and C. The MFEP corresponds to a concerted mechanism with a phosphorane-like TS. Each circle is associated with an image along the string. The color scale denotes free energy in units of kcal/mol. (B) The 1D free energy profile along the MFEP obtained from set B, where $Mg^{2+}$ is bound at the catalytic site. The modestly higher free energy barrier (~5 kcal/mol) along this MFEP, as compared to the barrier along the MFEP obtained from set A, suggests that the string is still evolving but is nearing convergence. (C) Values of the most important reaction coordinates, r1, r2, r3, and r4, along the MFEP. Each circle corresponds to an image along the string.

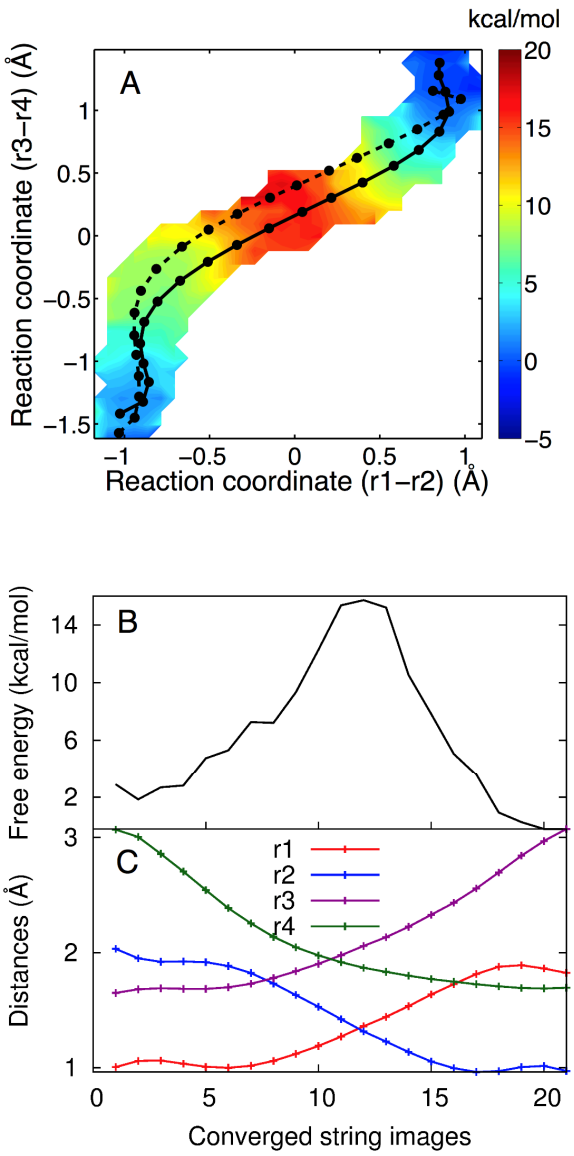


**Figure S2.** (A) The 2D free energy surface obtained from set C, where $Mg^{2+}$ is bound at the catalytic site, projected in the (r1−r2) and (r3−r4) space. The initial string (dashed black line) corresponds to an initial guess pathway passing through a phosphorane intermediate. The converged string (solid black line) represents the MFEP obtained from set C. The MFEP corresponds to a concerted mechanism with a phosphorane-like TS. Each circle is associated with an image along the string. The color scale denotes free energy in units of kcal/mol. (B) The 1D free energy profile along the MFEP obtained from set B, where $Mg^{2+}$ is bound at the catalytic site. The slightly higher free energy barrier (~2 kcal/mol) along this MFEP, as compared to the barrier along the MFEP obtained from set A, suggests that the string is nearly converged within the accuracy of the method. (C) Values of the most important reaction coordinates, r1, r2, r3, and r4, along the MFEP. Each circle corresponds to an image along the string.

## STRING CONVERGENCE

Description of string convergence criterion

The string in a given simulation set was considered to be converged when the root-mean-squared deviation (RMSD) of all coordinates of the latest string in that particular set from the mean value of the previous ten iterations, summed over all images, fell below a given threshold, which was chosen to be 0.1 Å. Mathematically, the RMSD of the reaction coordinate $k$ for the $j$th string update is calculated according to the equation

$$\mathrm{RMSD}_{j,k} = \sqrt{\frac{1}{N}\sum_{i}^{N}\left(f_j^k(\xi_i) - \overline{f_j^k(\xi_i)}\right)^2} \qquad (1)$$

where

$$\overline{f_j^k(\xi_i)} = \frac{1}{10}\sum_{j'=j-10}^{j-1} f_{j'}^k(\xi_i). \qquad (2)$$

Here $N$ is the number of images considered along the string, and a specific image is represented by the index $i$. String convergence is plotted for several sets in the figures S3-S6.

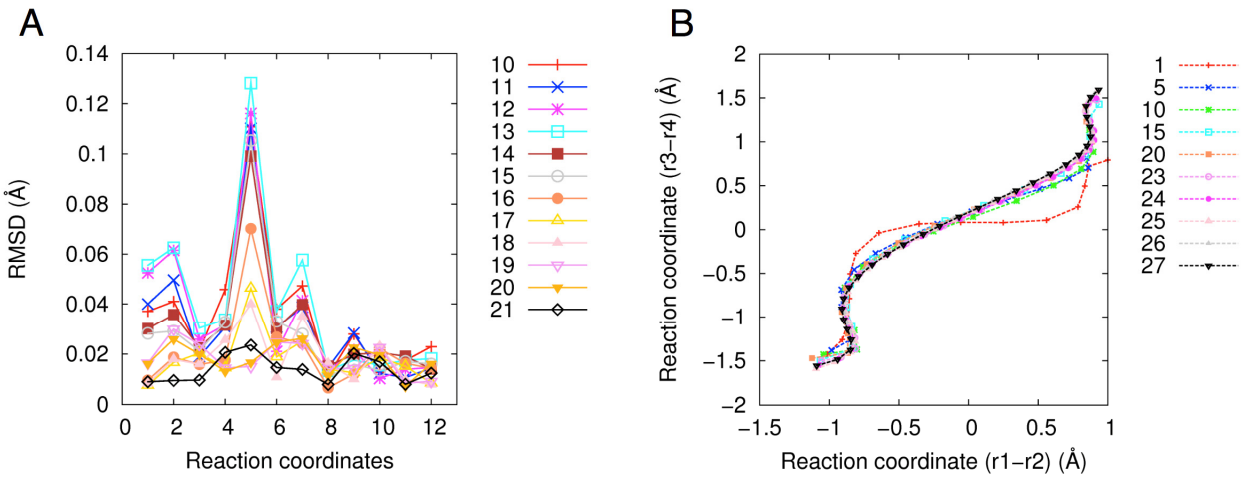


**Figure S3.** String convergence for set A, where $Mg^{2+}$ is bound at the active site. In total, 27 string iterations were performed. For iterations 1-21, 15 images were considered along the string. For iterations 22-27, the number of images along the string was increased to 30. (A) RMSD of all of the reaction coordinates is shown for iterations 10-21. Iterations 22-27 are excluded from the figure because the number of images in these iterations is not consistent with the rest of the iterations. Moreover, these iterations are essentially superimposable on one another. This figure indicates that the RMSD of the string for all reaction coordinates is well below 0.1 Å by iteration number 21. The additional iterations were performed to get a higher resolution of the MFEP and a smoother free energy profile. (B) The strings are plotted in the 2D space of the (r1−r2) and (r3−r4) coordinates for the selected iterations, as indicated in the figure key. This figure suggests that the string has stopped evolving in the space of the (r1−r2) and (r3−r4) coordinates.

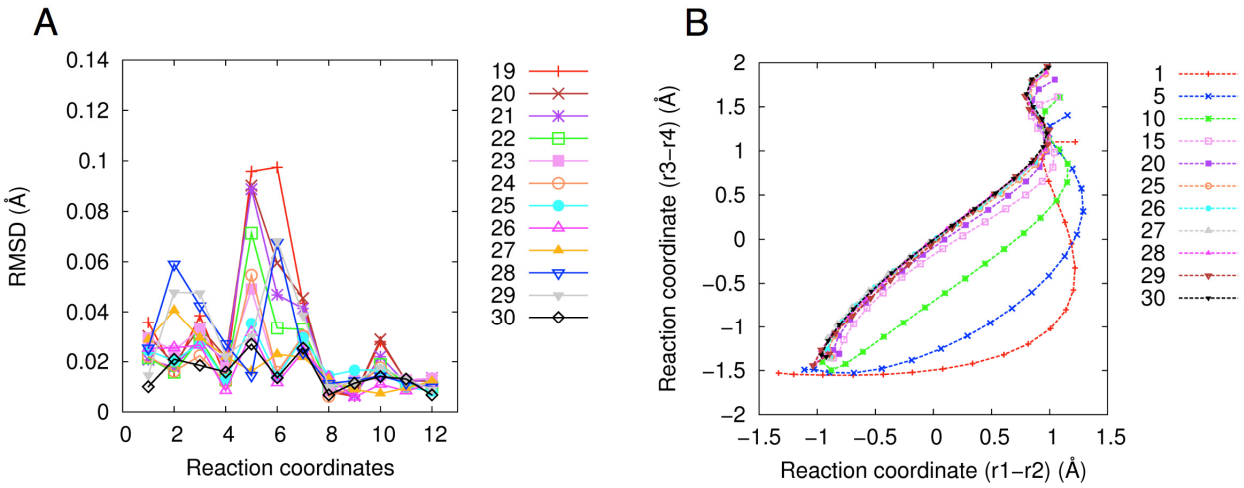


**Figure S4.** String convergence for set B, where $Mg^{2+}$ is bound at the active site. In total, 30 string iterations were performed. For all iterations, 21 images were considered along the string. (A) RMSD of all of the reaction coordinates is shown for the last 12 iterations. (B) The strings are plotted in the 2D space of the (r1−r2) and (r3−r4) coordinates for the selected iterations, as indicated in the figure key. As shown in the figure, the strings evolve significantly for the early iterations. The evolution of the strings for the later iterations is slow but is in the direction of the final strings obtained from simulation sets A and C. The string calculations for this set were terminated after iteration 30, when the reaction pathway, as indicated by the latest string (black curve), was determined to be qualitatively similar to that obtained from sets A and C.

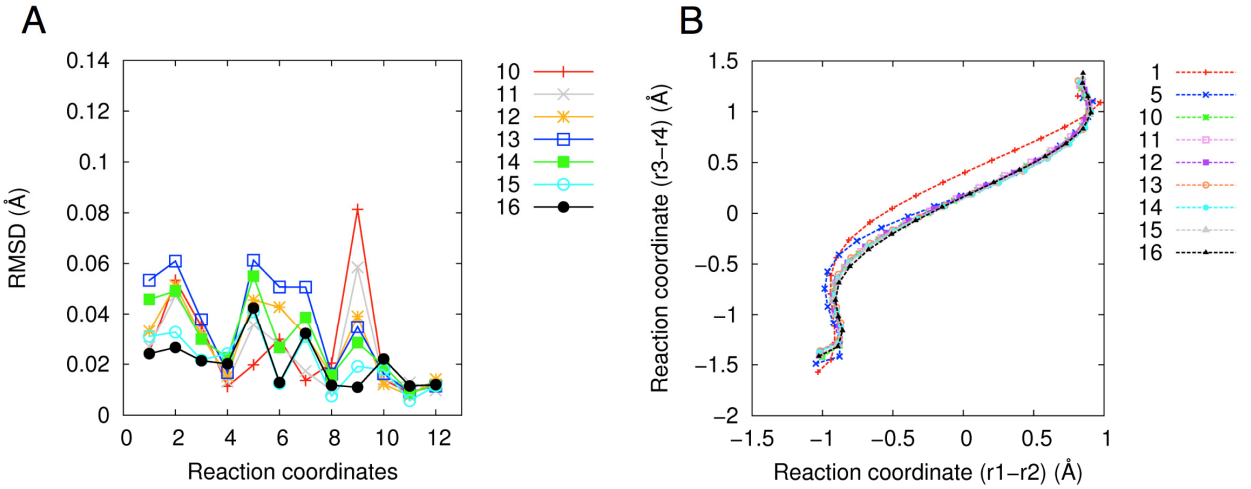


**Figure S5.** String convergence for set C, where $Mg^{2+}$ is bound at the active site. In total, 16 string iterations were performed. For all iterations, 21 images were considered along the string. (A) RMSD of all of the reaction coordinates is shown for the last 7 iterations. (B) The strings are plotted in the 2D space of the (r1−r2) and (r3−r4) coordinates for the selected iterations, as indicated in the figure key. The string calculations for this set were terminated after iteration 16, when the reaction pathway, as indicated by the latest string (black curve), was determined to be very similar to the converged string from simulation set A.

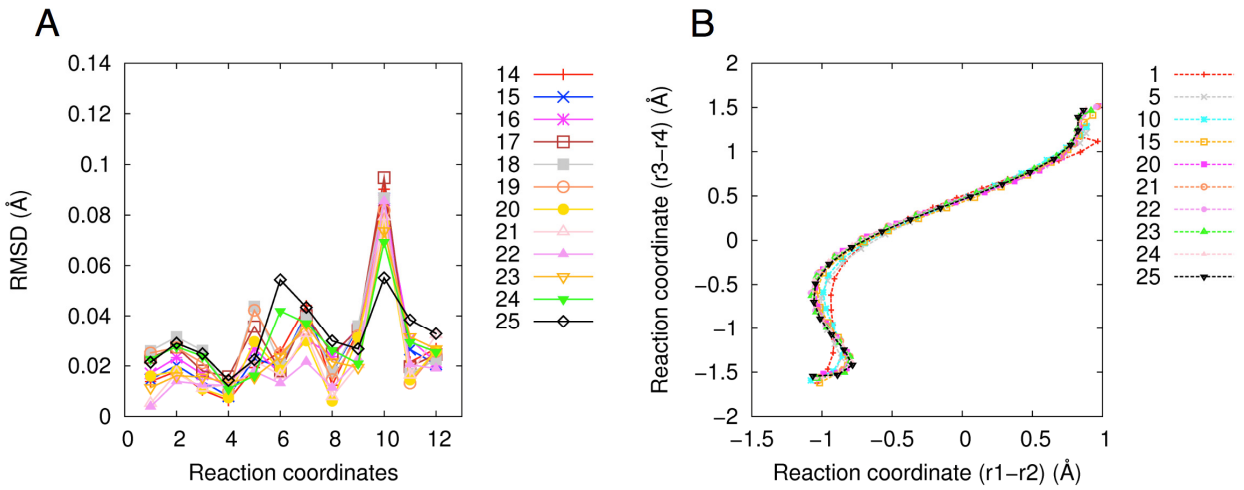


**Figure S6.** String convergence for set D, where $Na^+$ is bound at the active site. In total, 25 string iterations were performed. For all iterations, 21 images were considered along the string. (A) RMSD of all of the reaction coordinates is shown for the last 12 iterations. (B) The strings are plotted in the 2D space of the (r1−r2) and (r3−r4) coordinates for the selected iterations, as indicated in the figure key.

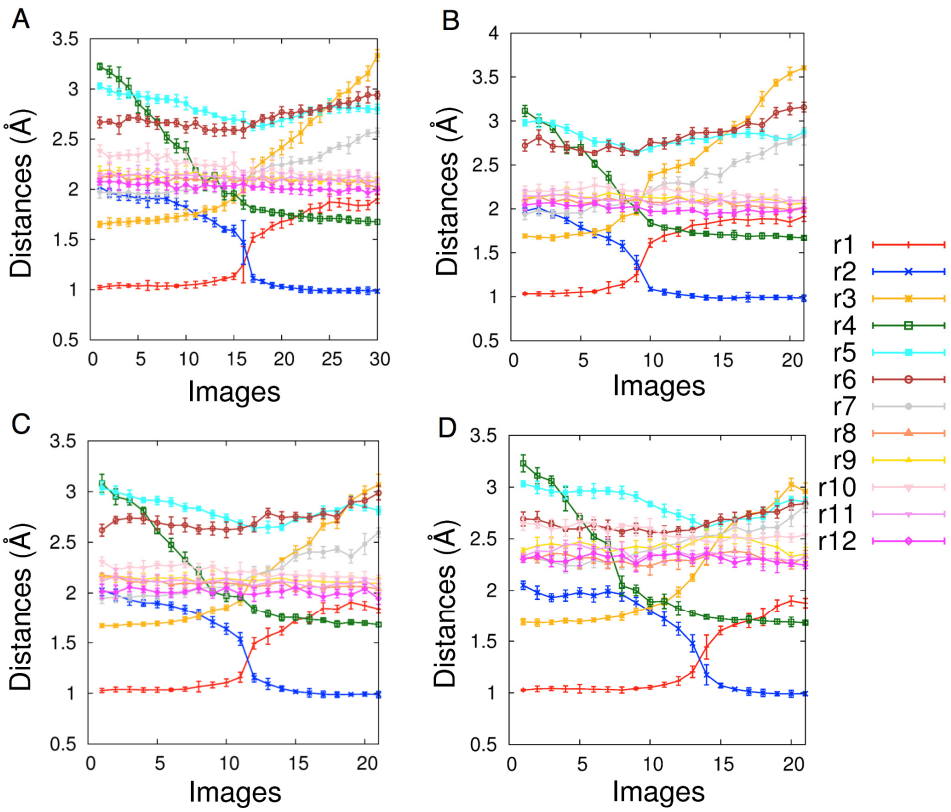


**Figure S7.** Average values of all reaction coordinates considered in the string simulations for the final iteration of simulation set A (Panel A), set B (Panel B), set C (Panel C), and set D (Panel D). The figure illustrates that in each of the simulation sets, the most important reaction coordinates are r1, r2, r3, and r4 because these coordinates change significantly along the MFEP.

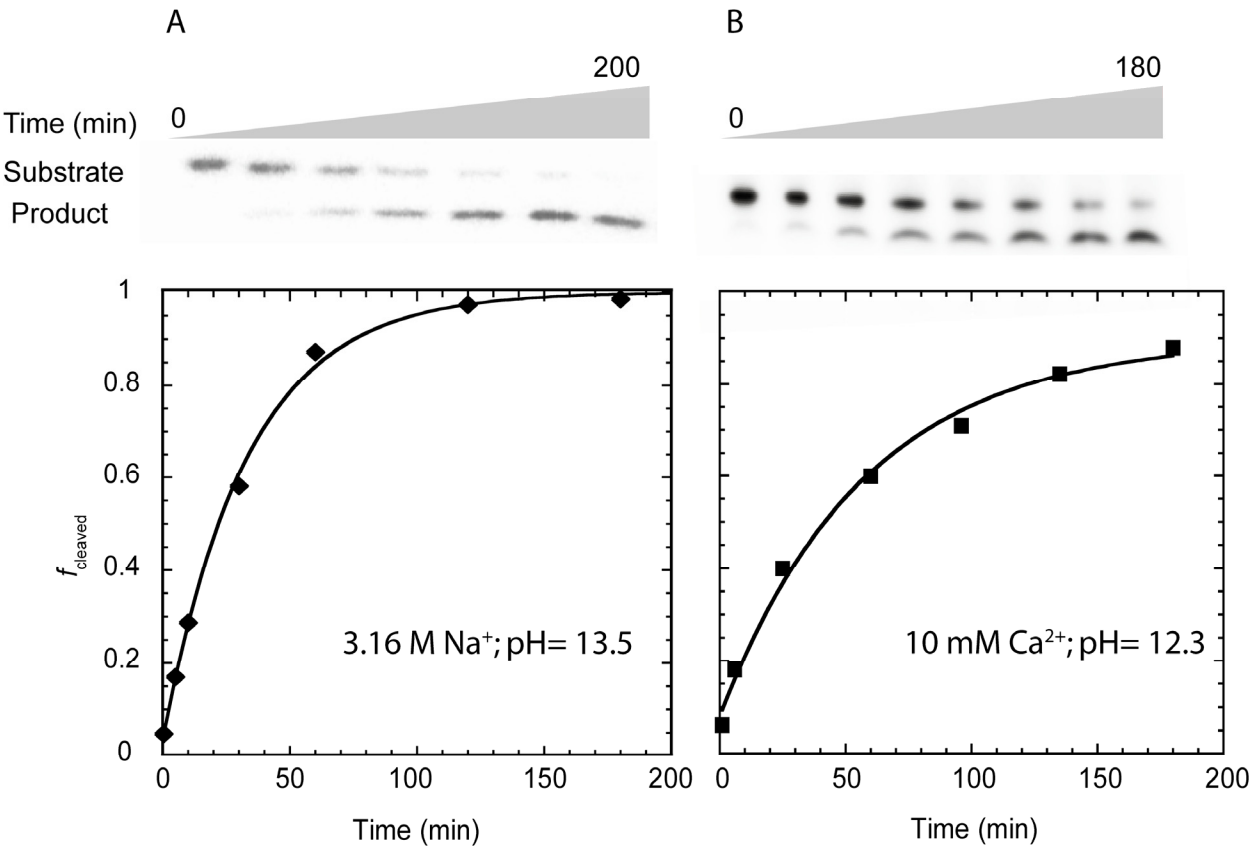


**Figure S8.** Autoradiograms of the chimeric substrate (upper row) and plots of fraction cleaved versus time (lower row) in the presence of A) 3.16 M $Na^+$ at pH 13.5, and B) 10 mM $Ca^{2+}$ at pH 12.3. The chimeric substrate undergoes complete cleavage yielding a single product under both of these conditions.

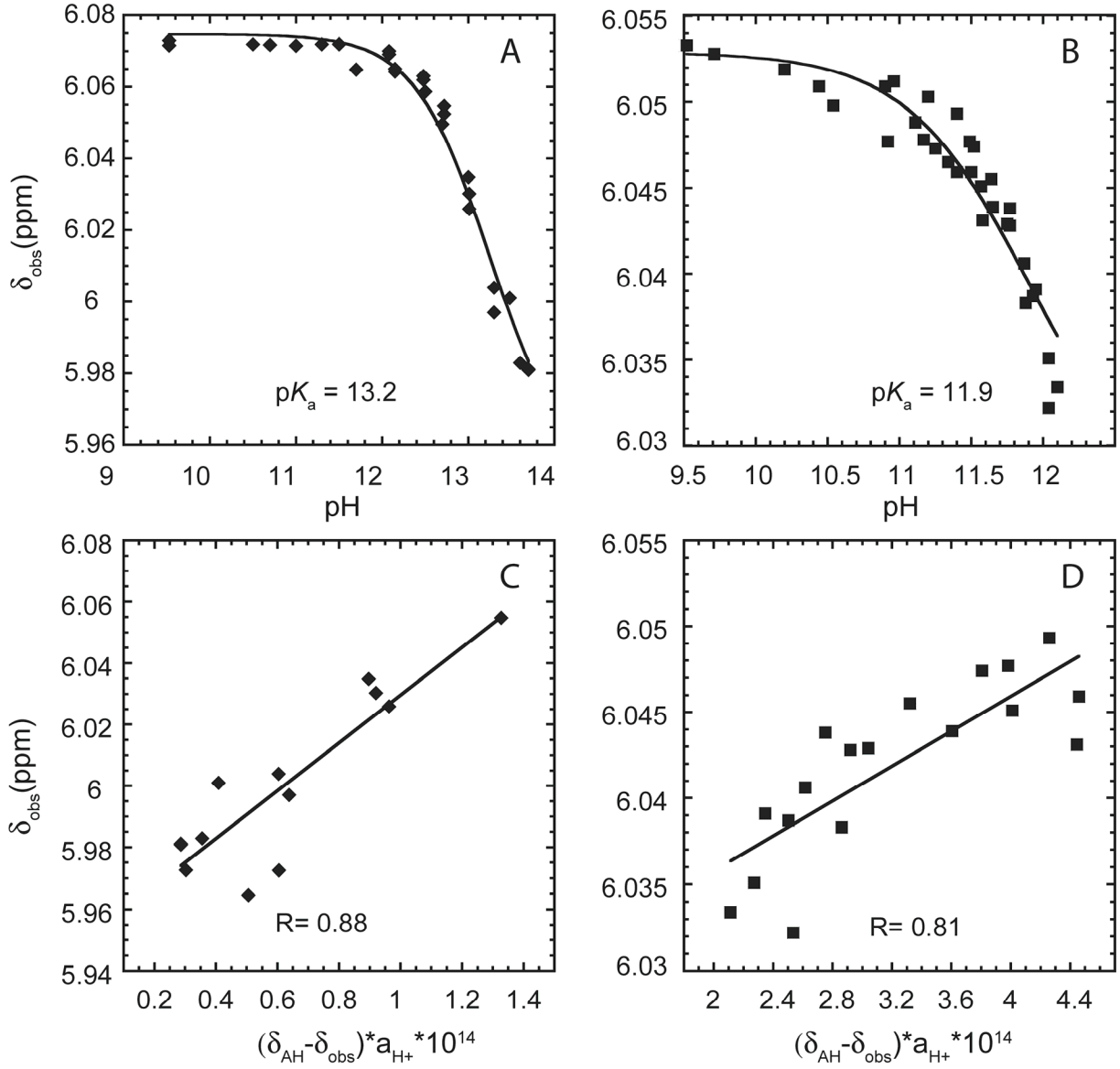


**Figure S9:** Determination of the p$K_a$ of the 2′OH of 3′AMP by $^1$H NMR. A) B) NMR titration monitoring H1′ chemical shift as a function of pH in the presence of A) 0.5 M $Na^+$ and B) 10 mM $Ca^{2+}$. The data were fit to Eq. 4. C) D) The value of $\delta_A$ used in Eq 4 was first determined by linearizing the sloping portion of the panel A and B data using Eq. 5 as shown in C) 0.5 M $Na^+$ and D) 10 mM $Ca^{2+}$. The p$K_a$ values were determined from fitting to panels A and B and found to be 13.2 and 11.9 in the presence of 0.5 M $Na^+$ and 10 mM $Ca^{2+}$, respectively.

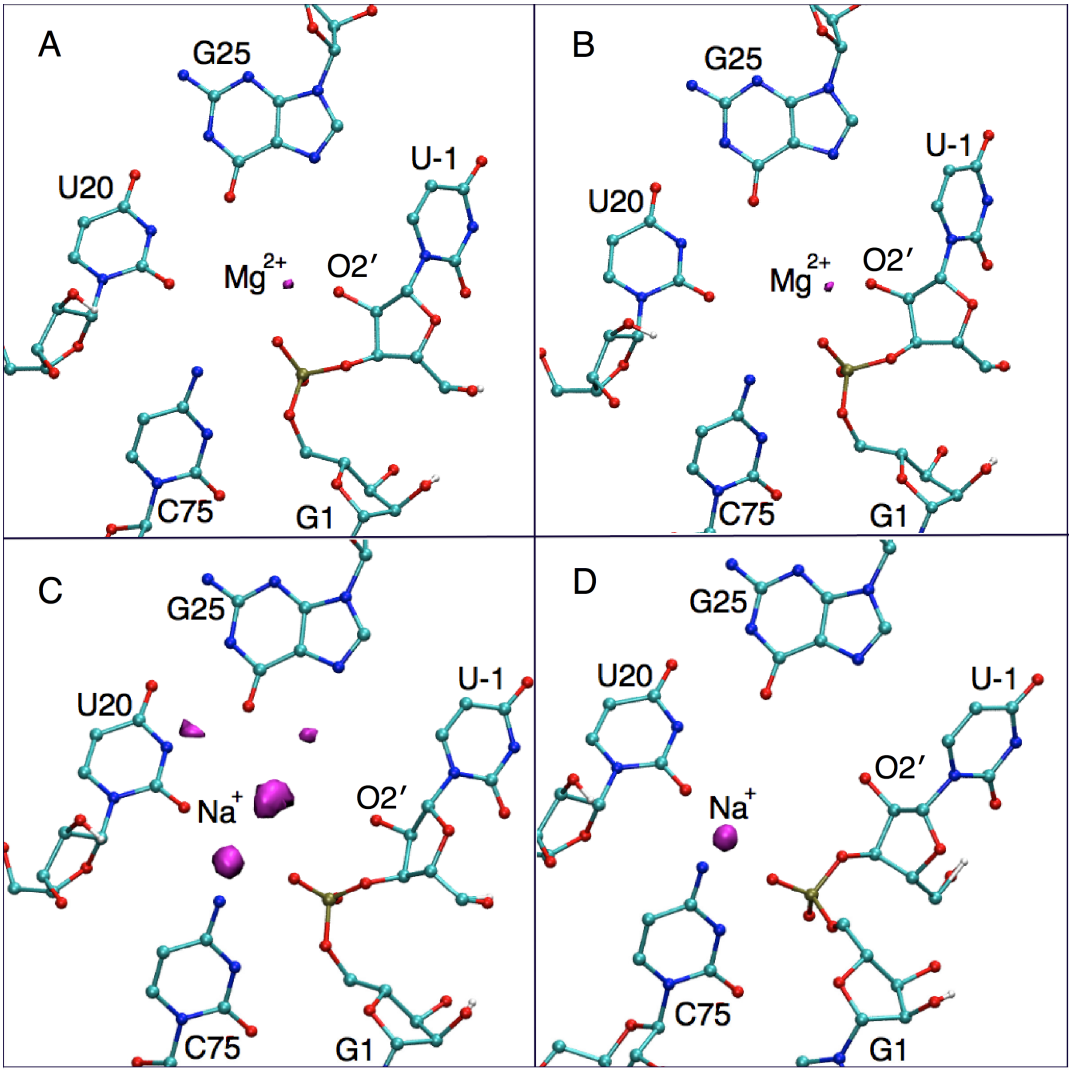


**Figure S10.** Charge isodensity plots illustrating positive charge in the active site region of the HDV ribozyme calculated from a series of 25 nanosecond classical MD trajectories. Panels A and B correspond to two independent trajectories of the WT ribozyme with a $Mg^{2+}$ ion at the catlytic site. Panels C and D correspond to two independent trajectories of the ribozyme in which the catalytic $Mg^{2+}$ ion was replaced by two $Na^{+}$ ions in the bulk, and $Na^{+}$ moves into the region of the catalytic site. The details pertaining to the MD simulations are provided in refs. 1 and 2. In each of the panels, the purple surface represents the greatest 30% of the positive charge density. In panels A and B, the positive charge is due to the $Mg^{2+}$ ion, while in panels C and D, the positive charge is due to a $Na^{+}$ ion. Panels A and B indicate that the $Mg^{2+}$ ion is highly localized and is positioned at a site where it interacts strongly with the O2′ of U-1. Panel C suggests a more delocalized nature of the $Na^{+}$ ion at the active site with multiple binding sites, only one of which allows the ion to interact with the O2′. Panel D suggests that the favored binding site of the $Na^{+}$ ion does not allow the ion to interact with O2′. All of these simulations were performed with the O2′ protonated, but the hydrogen atoms are not shown for clarity. The differences between Panels C and D arise from differences in conformational sampling for two independent MD trajectories starting at different initial configurations.

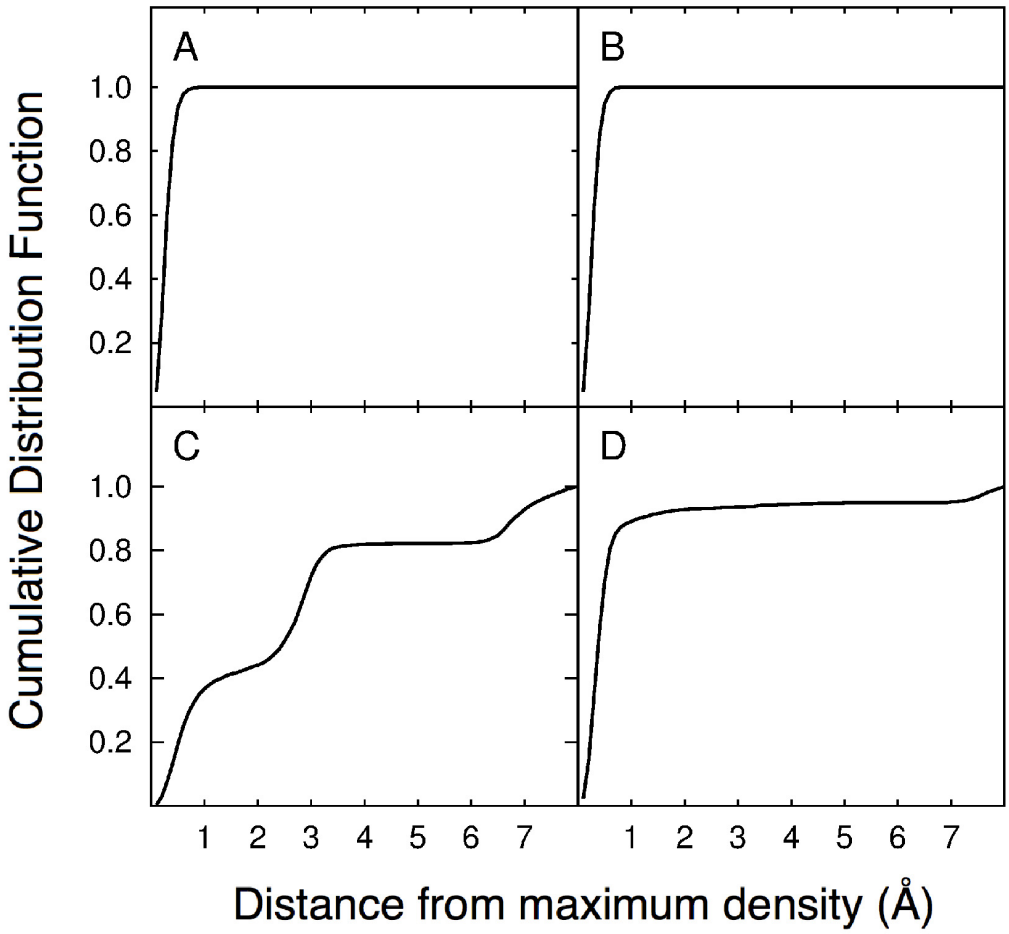


**Figure S11.** Cumulative distribution functions (CDFs) for metal ion occupancy at the active site region of the HDV ribozyme calculated from MD simulations. The CDF for a metal ion at a given distance *r* is the probability of finding that ion within a sphere of radius *r*. Panels A and B correspond to two independent trajectories of the WT ribozyme with a $Mg^{2+}$ ion at the catlytic site. Panels C and D correspond to two independent trajectories of the ribozyme in which the catalytic $Mg^{2+}$ ion was replaced by two $Na^{+}$ ions in the bulk, and $Na^{+}$ moves into the region of the catalytic site. These panels are associated with the respective panels depicted in Figure S10. In each case, the site of maximum positive charge density from the isodensity surfaces is chosen as the origin of the CDF, and only ions lying within a distance of 8 Å in all directions from the origin are considered. Thus, the CDFs measure the occupancy of the metal ions at the various binding sites that are revealed in Figure S10. In panels A and B, the CDF reaches unity within 1 Å, confirming that the binding sites of $Mg^{2+}$ revealed in the charge isodensity plots depicted in Figures S10A and S10B are highly localized and have 100% occupancy. Panel C indicates that the two prominent binding sites of $Na^{+}$ observed in Figure S9C each have ~40% occupancy. Panel D indicates that the occupancy of the binding site observed in Figure S9D has ~90% occupancy. The differences between Panels C and D arise from differences in conformational sampling for two independent MD trajectories starting at different initial configurations.

## REFERENCES

(1) Veeraraghavan, N.; Ganguly, A.; Chen, J.-H.; Bevilacqua, P. C.; Hammes-Schiffer, S.; Golden, B. L. *Biochemistry* **2011**, *50*, 2672-82.

(2) Veeraraghavan, N.; Ganguly, A.; Golden, B. L.; Bevilacqua, P. C.; Hammes-Schiffer, S. *J. Phys. Chem. B* **2011**, *115*, 8346-57.