Structures, Dynamics, and Ribozymes: An Investigation of RNA Structural Dynamics with the Hepatitis Delta Virus and Hairpin Ribozymes

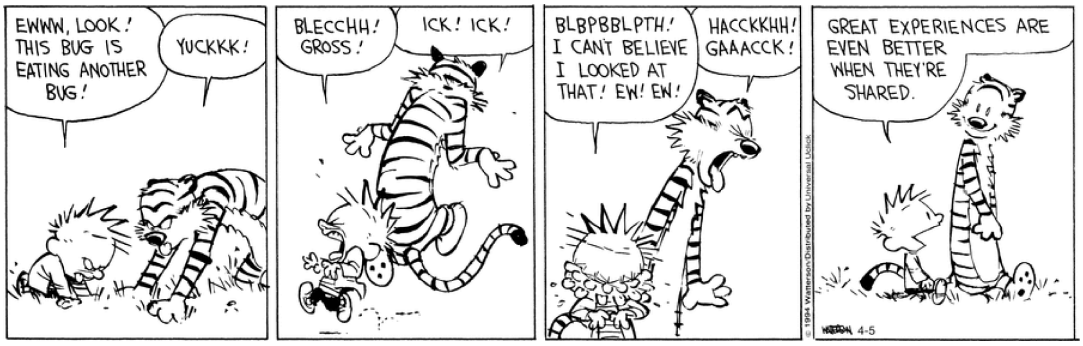
by

Wendy Wyn-Lee Tay

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in The University of Michigan
2014

Doctoral Committee:

Professor Nils G. Walter, Chair
Assistant Professor Julie S. Biteen
Professor Charles L. Brooks III
Professor Carol A. Fierke



"Calvin and Hobbes" by Bill Watterson

To my family and friends, near and far

ACKNOWLEDGEMENTS

Firstly, I would like to sincerely thank my thesis committee: Professors Nils Walter, Julie Biteen, Charles Brooks III and Carol Fierke. I thank them for all of their helpful, though sometimes harsh, feedback, advice and encouragement over the years. I could not be in this position without their assistance and patience. Things have not always gone smoothly, but I have learned much from my committee and I am truly grateful for the interactions I have had with each committee member. I particularly would like to thank Professor Nils Walter for taking me into his lab, for being very generous with his time and for the many productive suggestions over the years.

Next, a tremendous thank you to all Walter lab members, past and present. I have laughed with, cried with and been inspired by these great colleagues and friends. I thank my labmates for the comradeship and support that have made life in lab these past five years enriching. I would like to particularly thank Dr. Alex Johnson-Buck, Dr. Kamali Sripathi and Mr. Thomas Custer for their continued support and advice about science and the scientist.

I sincerely thank my collaborators Professor Jiří Šponer, Dr. Pavel Banáš, Professor Michal Otyepka, Dr. Nad'a Špačková and Dr. Kamila Réblová, without whom I would have suffered so much more when I was first learning molecular dynamics. I really am grateful for all the time they have generously spent explaining concepts and protocols to me. I also thank them for their kindness and hospitality in Brno; the research trip to the Czech Republic has definitely been one of the highlights of my PhD training.

I cannot proceed without next acknowledging my fellow chemical biologists, professors and students alike, in the Program for Chemical Biology. They have been such an awesome group to be a part of. I knew I was not making a mistake when I chose to pursue my PhD with such a unique, fun, creative and forward-thinking group of scientists. I am really thankful for this vibrant and amiable community that has made it a good place to be during this amazing, tough journey. I especially would like to acknowledge Dr. Amanda Dugan, Dr. Ningkun Wang, Dr. Elín Edwald and Professor Anna Mapp for their supportive roles during some of the most difficult times for me here as a PhD student. I also would like to acknowledge Ms. Laura Howe and Ms. Traci Swan for being immensely helpful and friendly whenever I had any administrative questions or concerns.

Next, a huge thank you to each one of my Wolverine friends and (non-academic) mentors. I am so thankful to know all of them and to receive their wonderful advice, perspectives and mentorship about the PhD process, careers and life in general. It has been really great to have these wonderful people as part of my professional and personal growth during the past five years.

Also, a thank you to the Natural Sciences and Engineering Council of Canada and Rackham Graduate School for funding, as well as to Dr. Erika Cline who, by sharing this LaTeX template with me, has made my life during the past two months a whole lot easier.

Finally, last but definitely not least, a thank you to my family. They have been so supportive of me pursuing my dreams and I do not take their support with a sense of entitlement. I especially would like to thank Suat Goh Chan for all the love and care throughout. I also would like to thank Lay Yong Chan for instilling in me the wisdom to make it through the difficulties of the PhD training as well as what lies beyond.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

LIST OF ABBREVIATIONS

RNA ribonucleic acid

mRNA messenger RNA

rRNA ribosomal RNA

tRNA transfer RNA

HDV hepatitis delta virus

ncRNA non-protein coding or non-coding RNA

VS Varkud satellite

FRET fluorescence resonance energy transfer

smFRET single molecule FRET

MD molecular dynamics

PEG polyethylene glycol

PAGE polyacrylamide gel electrophoresis

TMR tetramethylrhodamine

EMSA electrophoretic mobility shift assay

ssFRET steady-state FRET

trFRET time-resolved FRET

N1dA N1-deazaadenosine

4SU 4-thiouridine

TIRF total internal reflection of fluorescence

BG benzylguanine

ABSTRACT

Structures, Dynamics, and Ribozymes: An Investigation of RNA Structural Dynamics with the Hepatitis Delta Virus and Hairpin Ribozymes

by

Wendy Wyn-Lee Tay

Chair: Nils G. Walter

The function of catalytic ribonucleic acid (RNA), or ribozymes, depends on the structural dynamics of the molecule. Although the main function of RNA was once thought to be mainly in the production of proteins, it is now known to be involved in many crucial processes in biology. Small ribozymes, less than 200 nucleotides in length, are useful model systems for studying RNA due to their easily detectable functional activity. Knowledge gained from these small ribozymes is applicable to the wider range of non-protein coding RNAs and in the design of therapeutics.

In this thesis, the impact of structural dynamics on RNA function was investigated by studying two small ribozymes. The first aim analyzed the global and local structural dynamics of the hepatitis delta virus (HDV) ribozyme. From fluorescence resonance energy transfer (FRET) experiments, it is seen that at physiological pH, conversion of the precursor HDV ribozyme to the product form is accompanied by an approximately 8 angstrom global end-to-end lengthening and that the precursor population displays heterogeneity. Local, active site dynamics were probed using molecular dynamics (MD) simulations, which can be linked to observations from

solution experiments by the in-line fitness parameter. Together, experimental and MD data suggest that the HDV ribozyme has a rugged folding landscape along its catalytic pathway. The second aim involved MD simulations to determine hairpin ribozyme variants with a disrupted active site water chain. The robustness of the water chain is such that it displays disruption only for variants with two single-atom modified nucleobases. For the third aim, the impact on the hairpin ribozyme behavior by an intracellular-mimic buffer, macromolecular crowding and interactions with cellular components was explored. While the intracellular-mimic buffer favors the undocked state, both crowding by PEG-8000 and the presence of yeast or HeLa cell extract favors the docked state. The interactions with cellular components are also more potent in promoting the docked state compared to crowding by PEG. This thesis exemplifies two trends in RNA research: the integration of computational and experimental work, and an interest in understanding the effects on RNA by various features of the intracellular environment.

CHAPTER I

The Quest to Gain Insight into the Function of Catalytic RNA from a Structural Dynamics Point of View

1.1 Defying the Central Dogma: The Importance of Catalytic Non-Coding RNA Molecules in Biology

In the 1980s, a revolutionary discovery was made that challenged the role of ribonucleic acids (RNA) as understood by the central dogma of biology: RNA can be catalytic.[1,2] The famous "central dogma", initially proposed by Francis Crick in 1958, showcased proteins as the enzymatic workhorses in the cell.[3,4] RNA, on the other hand, was deemed as simply a messenger, relaying genetic information encoded by genes to the protein synthesis machinery (for it was assumed that genes generally encoded proteins).[3,5] It was later understood that, in addition to messenger RNA (mRNA), there are other types of RNA such as stable ribosomal RNA (rRNA), which is a significant component of ribosomes,[6] as well as transfer RNA (tRNA), which translates the nucleic acid genetic code to amino acids.[3,7,8] However, it wasn't until the 1980s when the unglamourous view of RNA as mere information relay molecules and infrastructural components in protein synthesis was overcome. It is about time

that we shift some of the limelight from proteins and appreciate the significance of catalytic RNA, or ribozymes, and other forms of RNA in molecular biology.

The most influential paradigm shift that came with the discovery of ribozymes was the realization that RNA, and not protein, is responsible for some of the most essential chemical reactions in the cell. Take, for example, genetic replication. Ribozymes have crucial roles in the replication of the Varkud mitochondrial plasmid in certain isolates of the fungi *Neurospora*[9] and in the genome replication of some pathogens like the satellite RNA of the tobacco ringspot virus and the hepatitis delta virus.[10,11] Even in protein synthesis, it is ribosomal RNA that catalyzes the vital reaction of peptide bond formation[12] and it is the RNA component of some variants of RNase P that processes the 5′ end of tRNA molecules.[2] Certain mechanisms of intron splicing, such as by group I and group II introns and possibly by the spliceosome as well, are also mediated by RNA.[13–15] Ribozymes are not catalyzing non-essential, peripheral reactions. They are involved in a range of fundamental biological processes. It is with this understanding that we can get further insight into the inner workings of the cell.

One interesting theory that arose from the discovery of ribozymes is the RNA world.[16] This theory suggests that RNA was the first biological precursor from which evolved a world that includes protein and deoxyribonucleic acids (DNA). After all, only RNA can by itself store genetic information and carry out catalytic functions that would allow for self-replication.[16] Also, both DNA and protein can be made from RNA.[17] Furthermore, RNA is heavily involved with very basic, fundamental biological processes such as protein synthesis (in the form of mRNA, tRNA, rRNA, RNase P).[18] The idea of a prebiotic RNA world is certainly a captivating one, though not without conundrums like the origin of the RNA itself, that still make it worthy of a good debate about the origin of life.[17] Regardless, undeniable is the widespread importance of RNA, especially non-protein coding RNA (ncRNA), in our modern world of molecular biology.[18,19] The quest to better understand ncRNA is thus a

worthy one to be a part of.

1.2 Ribozymes Are Useful Model Systems

There are several naturally-occurring ribozymes.[20–23] Larger, complex ribozymes are several hundreds of nucleotides in length.[21] These ribozymes, which include the group I and II introns, RNase P and the peptidyl transferase rRNA, are involved in a range of reactions from phosphoryl transfer to peptide bond formation.[21,22] On the other hand, small self-cleaving ribozymes are ~50 to ~150 nucleotides in length and include the hammerhead, hairpin, hepatitis delta virus (HDV), Varkud satellite (VS) and *glmS* ribozymes.[20] These small ribozymes catalyze the phosphodiester cleavage of an internal sequence-specific site, resulting in cleavage products with a 5′-OH and a 2′,3′-cyclic phosphate (Fig. 1.1).[20,21] The requirement of metal ions for catalysis varies depending on the ribozyme.[20–22] While only a handful of ribozymes are discussed in this chapter, it should be noted that the occurrence of ribozymes is widespread.[20,23,24] Recent bioinformatic searches have shown, for instance, that many HDV-like ribozymes are found in the genomes of several organisms ranging from nematodes to fungi to fish.[24]

The small self-cleaving ribozymes are especially useful model systems for structural and mechanistic research on ncRNAs.[25–27] Complex ncRNA molecules, like the ribosome, have complicated secondary and tertiary structures.[28] By comparison, the relatively small sizes and simpler folds of small ribozymes make them easier to work with and to understand.[25–27] The self-cleaving activity also provides a very characteristic readout for functionality.[20] Furthermore, small ribozymes consist of very common structural motifs, such as hairpins, hairpin junctions, pseudoknots and loop-loop interactions.[20,21] Therefore, the structural, dynamic and mechanistic knowledge gained from studying small ribozymes can also be applied to more complicated ncRNAs.

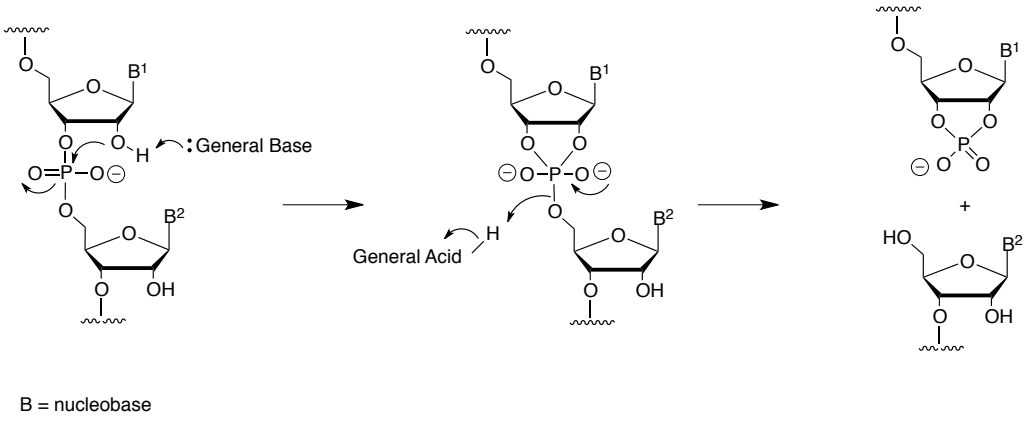


Figure 1.1: General cleavage mechanism of self-cleaving ribozymes. Cleavage is catalyzed by general base and general acid entities. Cleavage products contain a 5′-OH and a 2′,3′-cyclic phosphate in the RNA segments downstream and upstream respectively of the scissile bond.

The hepatitis delta virus ribozyme

Worldwide, it is estimated that 15 to 20 million people are infected with the hepatitis delta virus (HDV).[29] HDV is a satellite virus of the human hepatitis B virus (HBV), meaning that it requires the presence of HBV to form viral particles.[11] When co-infected with HBV, HDV exacerbates the symptoms of the HBV infection, leading to increased severity of the liver infection and even cirrhosis.[11] A particle of HDV contains a ~1.7 kb circular, single-stranded RNA genome complexed with ~70 HDAg protein components.[11] It is within the HDV genome that the HDV ribozyme is found.[11]

The HDV ribozyme is a small ribozyme ~85 nucleotides in length that is crucial for replication of the HDV genome (Fig. 1.2a).[21,32,33] HDV genome replication occurs by a double-rolling circle mechanism (Fig. 1.3).[11,20,34–36] During viral replication, a single concatemer containing the monomeric segments of the copied genome is synthesized.[20] To produce the individual genome monomers, the HDV ribozyme, at the interface of two monomeric segments, folds and self-cleaves, resulting in the formation of linear RNA products.[20] The cleaved genomic monomers then require ligation of the ends to be circularized.[20] It is suspected that the HDV ribozyme also catalyzes the ligation reaction based on the roles of other ribozymes, like the hairpin ribozyme, that are involved in similar genome replication mechanisms.[11] Yet, HDV ligation reactions are very slow and it is not clear whether the ligation activity is biologically relevant.[11] Because there is a genomic and an antigenomic version of the HDV genome, there is a genomic and an antigenomic version of the HDV ribozyme. The genomic and antigenomic HDV ribozymes share ~75% sequence identity as well as a double-pseudoknot secondary structure.[11,34,37] Both genomic and antigenomic HDV ribozymes have been the subject of many experimental studies but the focus of this thesis will be on the genomic ribozyme. Additionally, although the native HDV ribozyme is *cis*-acting (i.e. self-cleaving), there has been much work done to study

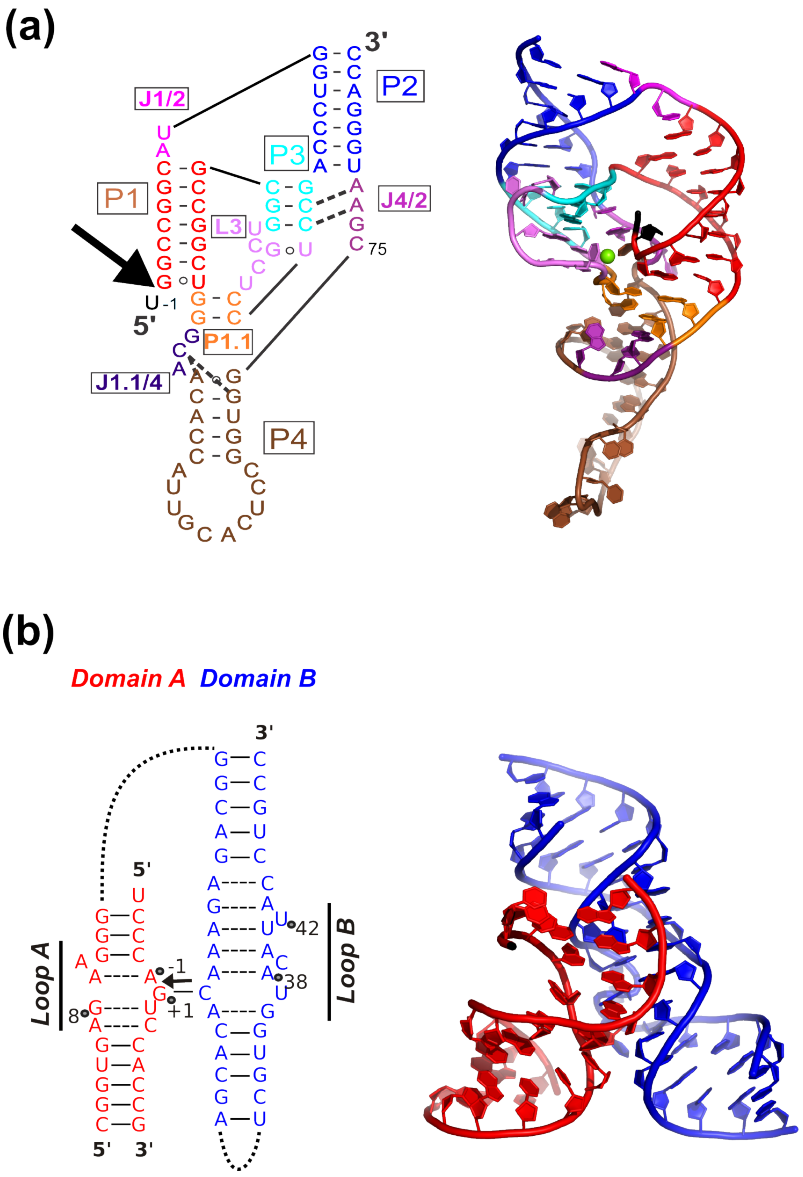


Figure 1.2: Secondary and tertiary structures of the HDV and hairpin ribozymes. **(a)** The HDV ribozyme. Sequence and crystal structure shown is from on a *cis*-acting ribozyme crystallized in 2004.[30] The crystal structure shows an active site magnesium ion (green) and has a C75U mutation to prevent cleavage during the crystallization process. The cleavage site is marked with a black arrow. Black dashed lines in the secondary structure schematic represent key tertiary interactions. Open circles represent wobble interactions. **(b)** The hairpin ribozyme. Sequence and crystal structure shown, representing the hairpin ribozyme in its docked state, is from Salter et al.[31] The black arrow indicates the cleavage site. Solid and longer dashed lines in the secondary structure schematic indicate canonical and non-canonical hydrogen bonding respectively. Shorter dashed lines indicate the regions typically present that were omitted to facilitate crystallization. The crystal structure contains a 2′-O-methyl at the A-1 position to prevent cleavage during crystallization.

trans-acting versions of this ribozyme, where the cleaved RNA strand is an external substrate, for kinetic and therapeutic purposes.[11,37–46] *Trans*-acting HDV ribozymes have been observed to cleave at rates 1 to 2 orders of magnitude more slowly than *cis*-acting versions, thought to be due to a less stable active site in the *trans*-acting version.[33,45]

The catalytic mechanism of the HDV ribozyme is still under debate. From pH-rate profiles, two functional groups were identified, one with a pK_a of ~6.4 and the second with a pK_a greater than 9.[47] It has been identified that the pK_a at ~6.4 belongs to the active site C75 (which corresponds to C76 in the antigenomic ribozyme).[47,48] The shifted pK_a of C75 allows C75 to act as a general acid or general base. Crystal structures of the HDV ribozyme show an active site cleft that is deeply electronegative, which is responsible for shifting the pK_a of C75 from ~4.2 to neutrality.[33,49] However, arguments can be made for C75 as a general acid or as a general base. If C75 were a general acid, it would be in its deprotonated form at neutral pH and the general base with the pK_a of >9 would be mostly protonated and weak.[47] By contrast, if C75 were a general base, the general acid with a pK_a of >9 would be weak.[47] However, hydrogen bonding of C75(N3) with G1(5′O) in the self-cleaved crystal structure[49] and experiments with a hyperactivated substrate containing a 5′-phosphorothiolate[44] both favor C75 as a general acid. The deeply electronegative active site also attracts metal ions that are necessary for cleavage.[50,51] Divalent metal ions are required for cleavage, but it is not fully understood what the role of the metal is.[50,52] In 2010, the location of the active site site magnesium was seen in a pre-cleaved HDV ribozyme crystal structure.[53] Additionally, upon conversion of the pre-cleaved ribozyme to the product in solution, the HDV ribozyme undergoes an end-to-end lengthening of ~2 to ~15 Å.[37,41,54,55] Pre-cleaved and cleaved crystal structures also show conformational changes where the metal ion has been ejected, the P1 and P3 helices collapse towards each other, G1 gets pushed closer to C75 and deeper into the active site and the

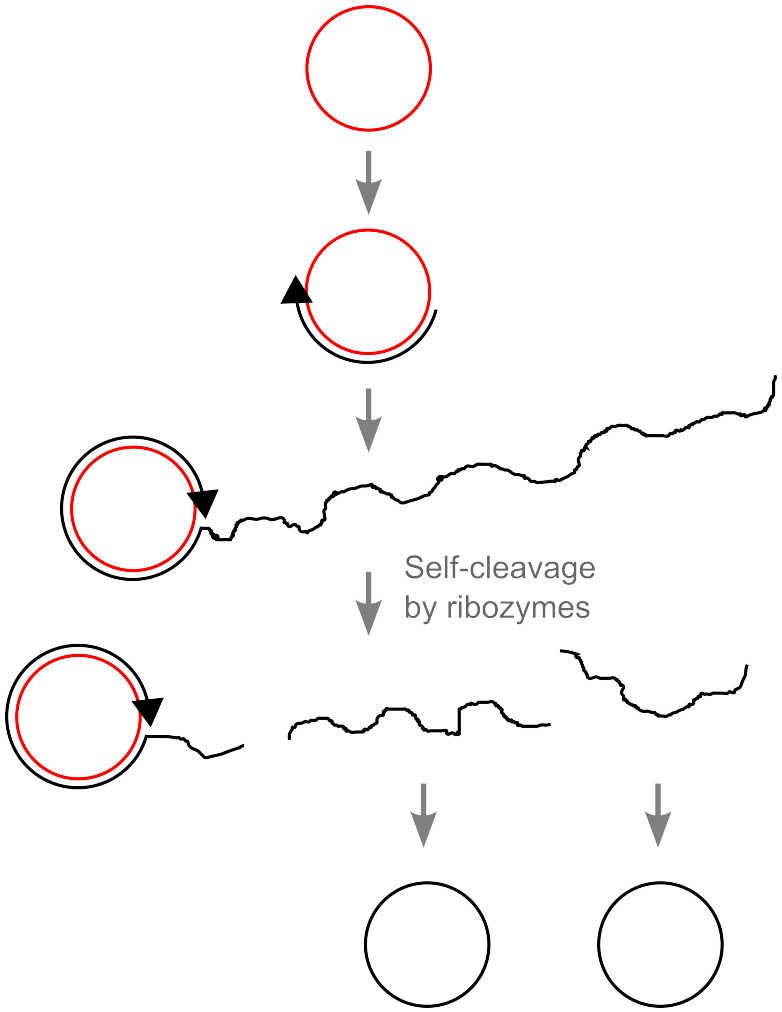


Figure 1.3: Double-rolling circle replication mechanism of subviral pathogens. Single-stranded circular RNA (red) acts as a template for transcription of the complementing sense of RNA (black) resulting in production of a concatemer of replicated monomeric segments. Ribozymes at the junction of the monomers fold and self-cleave to result in separated monomers. Ligation of the linear ends leads to circular single-stranded RNA monomers. Both genomic and antigenomic RNA can act as a circular template for replication of the complementing sense RNA; hence, there are ribozymes encoded in both the genome and antigenome.

G25U20 wobble base pair changes from an *anti* to a *syn* conformation and the active site magnesium ion is ejected.[30]

The hairpin ribozyme

Unlike the HDV ribozyme, which originates from a human pathogen, the hairpin ribozyme originates from a plant viroid pathogen.[56] The hairpin ribozyme is coded within the negative-sense strand of pathogenic satellite RNA that requires the tobacco ringspot virus as its helper virus for replication.[56] Like the HDV ribozyme, the hairpin ribozyme has a crucial role in the replication of the 359-nucleotide genome of the satellite RNA by cleaving the genome concatemer produced by double-rolling circle replication into its separate monomeric molecules.[10,57–59] The hairpin ribozyme also catalyzes the ligation reaction to circularize the genomic monomers (Fig. 1.3).[10,20,60–63] In nature, the structure of the hairpin ribozyme is ~113 nucleotides long and consists of four partially unpaired helices joined together at a four-way helical junction.[64,65] The minimal motif for activity, however, can be whittled down to ~50 nucleotides consisting of two helices, named domains A and B, joined together by a two-way helical junction (Fig. 1.2b).[31,66,67] Domains A and B each contain an internal loop.[31,66,67] Domains A and B interact with each other via interhelical loop-loop interactions (Fig. 1.2b).[31,66,67] Additionally, one-, two-, three- and four-strand versions of the two-way junction hairpin ribozymes have been made for various experiments.[10,31,63,66,68,69]

The catalytic mechanism of self-cleavage is different from that of the HDV ribozyme.[10,20,69–73] While the hairpin ribozyme also has an electronegative active site, it is secluded from bulk solvent and does not attract metal ions.[10,74–76] In fact, the hairpin ribozyme does not require metal ions for the actual catalytic reaction, as determined from cleavage experiments in the presence of cobalt(III) hexammine.[50,77,78,78–80] Thus, catalysis is mediated by the nucleobases themselves of the RNA.[50,77,78,78–80] From mutagenesis studies, G8 and A38 have been identified as the two residues within

the active that are most crucial for catalysis.[65,67,70,81,82] Mutation or deletion of G8 results in 100 to 1,000 times less activity without disruption of the overall structure.[70,82] In the case of A38, abasic mutations result in a >10,000 fold decrease in activity and substitution of N1-deazaadenosine at position 38 results in loss of activity.[83,84] The specific roles of G8 and A38 are still under debate with G8 proposed to be either a general base or transition state stabilizer and A38 proposed to have a number of roles such as general acid, general base and/or proton shuttle.[31,65,67,67,82,83,85–91] Within the hairpin ribozyme active site, there is also a chain of long-lived water molecules as observed from crystallography and molecular dynamics studies.[31,86,92] This water chain facilitates a network of hydrogen bonds throughout the active site that may be involved in communication between distal regions of the active site.[31,86,92] Further studies on the functional relevancy of the water chain and hydrogen bonding network are needed.

The hairpin ribozyme undergoes a distinct conformational change from the inactive state to the active state.[66,93] Although the hairpin ribozyme does not require metals to do the actual cleavage reaction, metal ions are necessary to form the active state.[76,77,93] In the active state, or "docked" conformation, domains A and B interact with each other such that an interhelical core forms, which leads to formation of the active site (Fig. 1.2b).[66,93] Without divalent metal ions present, stems A and B move away from each other and form what is known as the "undocked" conformation.[66,93] Because of the distinct conformational change of the hairpin ribozyme, it is an ideal system to study using single molecule fluorescence resonance energy transfer (smFRET) and thus many smFRET experiments have been performed on the hairpin ribozyme.[66,69,71] The docked state has a FRET value of around 0.8 and the undocked state has a FRET value of around 0.15.[66] Hairpin ribozymes missing the RNA strand that becomes cleaved have a FRET value of around 0.4.[66] From smFRET experiments, rate constants have been determined for all known kinetic steps of the hairpin

ribozyme cleavage mechanism (Fig. 1.4).[69,71]

Artificial ribozymes

In addition to naturally-occurring ribozymes, a number of artificial ribozymes have also been created.[94–97] Small ribozymes have been developed in the lab to perform a wide variety of reactions including aminoacylation of tRNAs and C-C bond formation via the Diels-Alder reaction.[95] Most notable of the artificial ribozymes are the ones with potential therapeutic applications, which most commonly are derived from hammerhead and hairpin ribozyme motifs.[96] These therapeutic ribozymes are engineered to specifically cleave at a particular site of interest. For instance, Angiozyme, currently licensed by Merck-Sirna and used for renal cancer, targets the mRNA of the vascular endothelial growth factor receptor-1 (VEGFR-1).[96] A number of other therapeutic ribozymes are being investigated as treatments for HIV.[96] The sequence specificity of ribozymes, combined with their ability to catalyze without proteins and the potential to improve function by *in vitro* selection,[94,95] makes ribozymes very attractive therapeutic candidates.

1.3 Understanding Biological Function: A Dynamic Consideration

While a picture says a thousand words, sometimes one picture is not enough to fully understand the story. Crystal structures do provide meaningful snapshots of biology. Yet, dynamic information is just as crucial as structural information if we are to truly understand how a molecule functions. Global dynamics, which occur on a millisecond or slower time scale, pertain to relative motions of sub-domains and tertiary structures.[98,99] These global dynamics influence the formation of functionally relevant substructures, such as active sites or subunit-subunit interactions.[98–101] In

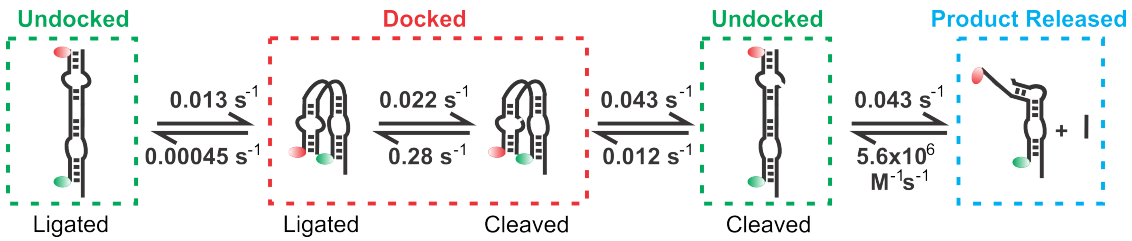


Figure 1.4: Known mechanistic steps of catalysis by the hairpin ribozyme. Docking and undocking rates are shown for a two-strand hairpin ribozyme as obtained by Liu et al.[69] Green and red ovals indicate the positions of the FRET pair of Cy3 and Cy5 fluorophores, respectively, used by Liu et al. to determine each of the kinetic constants shown.[69]

contrast, local dynamics take place on the picosecond (10^{-12} seconds) to microsecond (10^{-6} seconds) time scale and reflect motions such as bond vibrations, bond rotations and thermal fluctuations of residues.[98,99] There is interdependency between the range of dynamic motions as local dynamics collectively influence global dynamics, and the global dynamics affect the local environments where the local dynamics take place.[98] There are several instances where structural dynamics have been key to understanding biological processes, such as the maturation of viruses, the mechanism of the GroEL/GroES protein folding chaperone, substrate binding and release in RNase A, catalytic propensity in the hammerhead ribozyme and more.[102–109] The techniques I employed for the work in this thesis are focused on just that - observing structural dynamics. Single molecule and ensemble FRET experiments can give useful information about global dynamics whereas the interrogation of rapid local dynamics can be probed by computational simulations. If a picture tells us a thousand words, then a movie, which is made up of many pictures, can really tell us a more complete story.

1.4 Computational Work to Supplement and Support Experimental Work

All-atom molecular dynamics (MD) is a powerful tool for getting a glimpse at local structural dynamics, often reaching where physical experiments cannot easily go. In all-atom MD, molecules can be simulated with single-atom resolution based on empirical force fields that mathematically describe the theoretical behavior of the molecule.[110] Just less than a decade ago, MD simulations for nucleic acids were typically <100 nanoseconds long since longer simulations may have resulted in spurious force field artifacts.[111,112] Nowadays, improvements in the quality of force fields enable us to perform stable MD simulations on the microsecond scale.[113] Despite these

improvements, the type of dynamics best interrogated by MD are fast, on the order of picoseconds to nanoseconds.[113] Even if the actual catalytic reaction occurs on a slower time scale (microsecond or slower), it is the local, sub-microsecond pre-catalytic dynamics that collectively contribute to catalysis and thus are significant for understanding the slower dynamics.[98,99,113] Therefore, the combination of MD with experimental techniques bridges the range of local and global dynamics that can be studied and thereby enables scientists to gain a more enriched and comprehensive understanding of the biological system of interest.

As useful as MD simulations are, like with any technique, there are some limitations that should be kept in mind. Primarily, the quality of the simulation is greatly influenced by the quality of the force field.[110–114] There is no truly "correct" force field because force fields are empirical.[110] For example, some of the problems in previous force fields for RNA were the lack of accuracy in simulating the α and γ torsion angles of RNA and the formation of "ladder-like" structures in helical regions.[111,115] Currently, the most stable force field for simulating RNA, and hence the choice used for the work of this thesis, is the parmbsc0χ_{OL3} variant of the Cornell et al. AMBER force field.[111,115,116] The second relevant limitation to be aware of is that polarizability is still not well described by force fields.[114,117,118] This limitation means that divalent metal ions like Mg^{2+} are difficult to simulate accurately and can cause detrimental artifacts.[114,118] Thus, we use Na^{+} ions in our simulations, which are more accurately described by the force field.[114]

1.5 The Intracellular Environment: It's A Crowded World After All

Though ribozymes have been traditionally studied in buffered, minimal-species solutions, there is great interest in understanding the behavior of ribozymes in a

cellular environment. The natural environment of ribozymes are inside cells after all.[1,2,11,20,26,27,56] The environment inside a cell is very different from that of an *in vitro* buffer.[119–128] The inside of a cell has a reducing environment, a high concentration of monovalent ions, and a high concentration of macromolecules that can interact nonspecifically with RNA.[119–133] In fact, the combined concentration of protein and nucleic acids can get as crowded as 400 mg/ml inside cells and take up as much as 40% of the intracellular volume.[119–124,134–136] The effects of macromolecular crowding on the structure and activity of RNA have been the subject of several interesting studies.[137–145] Crowders have been shown to preferentially stabilize active ribozyme structures.[137–142] Studies on the hammerhead ribozyme have suggested that crowders like polyethylene glycol (PEG) destabilize RNA secondary structures while preferentially stabilizing the functionally relevant tertiary conformation via weak, nonspecific surface interactions.[139,140] More studies with the hammerhead ribozyme have demonstrated that crowders can promote efficient catalysis even at the low 0.1 to 3.5 mM magnesium concentrations found in cells.[141,142] Macromolecular crowders are also known to favor small, compact RNA structures due to the excluded volume effect.[137,143,144] However, crowding effects can depend on the size and electrostatic properties of the crowder.[140,145] For instance, unlike larger, inert crowders like PEG, small osmolyte crowders, such as sugars and methylamines, actually destabilize RNA structures.[145,146] There is no doubt that studying ribozymes in a cellular environment is much more complicated than studying them in simple buffers. However, because the world in which ribozymes natively function is crowded with many more species than minimal buffer components, the effort to pursue such studies is worthwhile.

1.6 Thesis Objectives

The HDV Ribozyme - Chapter II

A recent crystal structure of a precursor HDV ribozyme, obtained at low pH and by molecular replacement from a previous product conformation, shows a product-like active site.[53] This crystal structure raises the possibility that the conformation captured is an activated precursor. In **Chapter II**, I used time-resolved FRET, in conjunction with other solution FRET experiments, to investigate the global dynamics of this recently crystallized HDV ribozyme. I also used MD simulations to gain insight into the local dynamics of the recent crystal structure's active site and made particular use of the in-line fitness parameter to link the active site to catalytic propensity over the course of the simulations.

The Hairpin Ribozyme - Chapters III and IV

With regards to the hairpin ribozyme, there were two directions of study that I explored. The first direction pertained to learning more about the functional role of the active site water chain. Although MD simulations[92] and crystal structures[31] have shown the presence of these long-lived water molecules in an otherwise solvent-excluded active site, it is unclear what the role of this water chain is. In **Chapter III**, I used MD simulations to scan for potential hairpin ribozyme variants with a disrupted water chain, which can then be used for further experimental characterization to determine the role of the water chain. The second direction I explored was use of smFRET to determine the effect of various features of the intracellular environment on the structural distribution and dynamics of the hairpin ribozyme. Although there have been several smFRET studies performed on the hairpin ribozyme in buffered solutions, there is still little known about the structural and dynamic behavior of ribozymes in a cellular context. Using smFRET, I investigated in **Chapter IV** the

impact of three defining characteristics of the intracellular environment on the hairpin ribozyme: an intracellular-mimic buffer environment, macromolecular crowding, and interactions of the ribozyme with cellular components.

CHAPTER II

Disparate HDV Ribozyme Crystal Structures Represent Intermediates on a Rugged Folding Free Energy Landscape[1]

2.1 Introduction

The hepatitis delta virus (HDV) ribozyme (Fig. 2.1) is a small catalytic RNA originally found in the genome and antigenome of the hepatitis delta virus.[11,34–36] It has a minimal length of ~85 nucleotides and catalyzes the cleavage of its own phosphodiester backbone in *cis*, resulting in the formation of 5′-OH and 2′,3′-cyclic phosphate products. Self-cleavage of the genomic and antigenomic ribozymes, which share ~75% sequence identity, is an essential step in the double-rolling circle replication of the HDV.[11,34] The overall fold of the HDV ribozyme consists of a double-nested pseudoknot encompassing helices P1, P1.1, P2, P3, and P4 (Figs. 2.1a,b),[33,37,49] which leads to a tight-knit structure with a deeply electronegative active site cleft. This cleft

[1]Adapted from Sripathi, K. N.*, Tay, W. W.*, Banáš, P., Otyepka, M., Šponer, J. & Walter, N.G. Disparate HDV Ribozyme Crystal Structures Represent Intermediates on a Rugged Folding Free Energy Landscape *RNA* **2014**, just accepted. Dr. Kamali Sripathi performed experiments and analyses for the EMSAs and the steady-state FRET assays (Figs. 2.3a,b) Dr. Kamali Sripathi also performed the simulations and analyses for the 1SJ3_U-1+, U-1, dU-1+, dUGG and U-1G+ simulations as listed in Table 2.1. Wendy Tay performed all other experiments, simulations and analyses. Figs. 2.1b, 2.3a, 2.3b, 2.4, 2.5, 2.6d, 2.8c and the chemical structures in 2.9 were contributed by Dr. Kamali Sripathi. All other figures and tables were contributed by Wendy Tay. *These authors contributed equally to this work.

attracts metal ions and promotes cleavage by shifting the pK_a of an active site cytosine at position 75 (in the genomic ribozyme; C76 in the antigenomic form) towards neutrality.[33,49] Based on these structural constraints, numerous HDV-like ribozymes were recently discovered in the human and other eukaryotic genomes, suggesting that this RNA motif may have widespread functions in RNA processing and gene regulation.[23,24,147]

In addition to numerous studies of self-cleavage by the *cis*-acting genomic and antigenomic HDV ribozymes,[32–34,47,48,52,54,148–160] significant work has also been performed to investigate *trans*-acting forms of this ribozyme for enzymology and therapeutic purposes.[37–46] Generally, *trans*-acting HDV ribozymes exhibit a 1 to 2 order-of-magnitude slower catalytic rate constant compared to their *cis*-acting counterparts, likely due to a less tightly knit structure.[45] Both *cis*- and *trans*-acting ribozymes have been shown to exhibit in solution end-to-end lengthening when the catalytic precursor cleaves to form the product, with the *cis*-acting versions exhibiting more modest changes than the ~15 Å observed for *trans*-acting ribozymes.[37,41,54,55]

Since 1998, a number of crystal structures of the *cis*-acting genomic ribozyme have been solved of both precursor and product forms.[30,49,53,161] The emergence of the *cis*-acting precursor structure at up to 2.2 Å resolution was accompanied by the observation of a structural change upon release of the 5′ product and catalytic metal ion, as C75 moves deeper into the active site cleft to form hydrogen bonds with C22's nonbridging oxygen and the 5′ OH leaving group of G1, while the adjacent G25 rotates in conformation from *anti* to *syn*.[30,114] These changes result in a collapse of the P1 and P3 helices towards the active site and a more compact catalytic core and concomitant overall lengthening of the product (Fig. 2.1a),[30] consistent with the results from solution probing.[37,41,54]

As is typical, the *cis*-acting precursor was modified to prevent self-cleavage during crystallization in these studies, which was accomplished by either a C75U mutation,

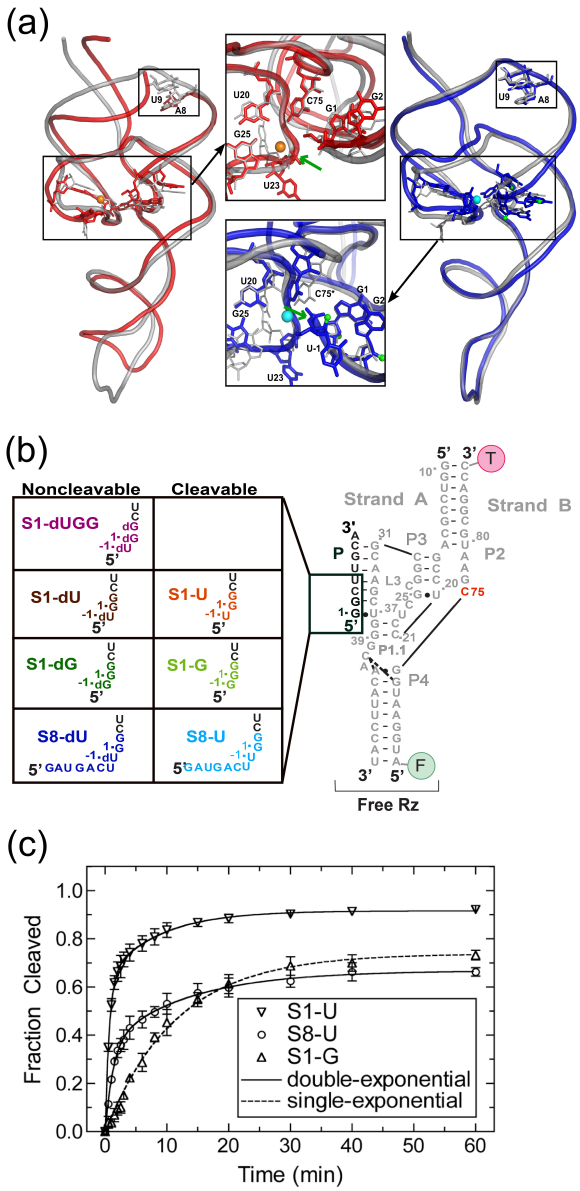


Figure 2.1: The *trans*-acting HDV ribozyme used in our solution studies. **(a)** Superpositions of *trans*-acting (red)[53] and *cis*-acting (blue)[30] precursor crystal structures on the *cis*-acting product structure (silver).[49] Active site magnesium ions are shown in orange and cyan for the *trans*- and *cis*-acting ribozymes, respectively. The smaller black boxes on both overlays highlight the region where there is a break in the RNA backbone for the *trans*-acting ribozyme. The green arrows indicate the cleavage site in both insets. The U-1 nucleotide and scissile phosphate were not resolved in the *trans*-acting precursor structure. **(b)** The 8 three-stranded precursor and product ribozymes used in our solution experiments. **(c)** Cleavage assay time courses for S1-U, S8-U and S1-G. Cleavage assays were conducted at 25°C in 40 mM Tris-HCl, pH 7.5, 25 mM DTT and 11 mM $MgCl_2$. Data for S1-G were fit with a single-exponential increase function, while the data for S1-U and S8-U were fit best with a double-exponential increase function. Standard deviations from at least three independent assays are shown as error bars.

crystallization in the absence of Mg^{2+}, or incorporation of a single 2′-deoxy modification at the cleavage site, although the latter structure was low in resolution.[30] In 2010, a different team solved a low-pH, 1.9 Å-resolution crystal structure of a *trans*-acting precursor HDV ribozyme in the presence of Mg^{2+}, inactivated with three 2′-deoxy modifications at the -1, 1 and 2 positions around the cleavage site (Fig. 2.1a, bottom inset).[53] Consistent with the fact that molecular replacement with the *cis*-acting product was used to elucidate its structure, the new *trans*-acting structure features both active-site and global conformations similar to those of the product (Fig. 2.1a).[49,53] For example, G25 is in the *syn* conformation and the relative positions of the P1 and P3 helices are more similar to those in the *cis*-acting product than the precursor ribozyme (Fig. 2.1a). Such close resemblance to the product suggests that this ribozyme may capture an activated conformer past the conformational change needed to activate all previous ribozyme versions. This hypothesis would predict a fast cleavage rate constant; however, the newly crystallized HDV ribozyme cleaves similarly slowly to other *trans*-acting versions.[53] Additionally, the substrate strand was sufficiently flexible (and/or nicked) so that no electron density was observed for the scissile phosphate and upstream U-1 nucleotide, further increasing resemblance with the cleavage product. The authors modeled these moieties using the cleavage site of the hammerhead ribozyme.[53] Consequently, questions remain concerning the relationship of the various proposed and experimentally observed conformations to each other and to the functional activity of the HDV ribozyme.

To probe the functional dynamics of the newly crystallized *trans*-acting HDV ribozyme in solution, we have employed a combination of experimental and computational approaches. Fluorescence resonance energy transfer (FRET) gel shift assays revealed a heterogeneous population of molecules in solution at physiological pH, suggesting that the crystal structure represents a conformation selected by crystallization. Steady-state and time-resolved FRET measurements consistently detect an

~8 Å lengthening of the end-to-end distance along the P1-P4 axis upon cleavage in solution, an elongation between those previously found for *trans*- and *cis*-acting HDV ribozymes.[37,41,54,55] We also performed a total of 1.8 μs of molecular dynamics (MD) simulations that showed that the hammerhead ribozyme model of the substrate conformation in the active site is compatible with a favorable catalytic in-line fitness (a measure of the poise to undergo catalytic transesterification), as long as C75 is protonated. By contrast, intermolecular crystal contacts and 2′-deoxyribose modifications found in the crystal structure near the active site result in unfavorable in-line fitness, offering an explanation for the lack of experimental electron density in this region. A *cis*-acting version of the crystallized ribozyme interconverts more frequently between unfavorable and favorable fitness, consistent with a lowered kinetic barrier towards cleavage. Finally, a U-1G mutation results in long-lived states of unfavorable fitness, consistent with its low cleavage activity and the conservation of U-1 in clinical isolates of the virus. Taken together, our data are consistent with a parsimonious model wherein the existing precursor crystal structures, determined for distinct constructs at varying pH, represent alternate RNA conformations on a rugged folding free energy landscape.

2.2 Materials and Methods

Preparation of RNA molecules

All sequences used here were based on the recent *trans*-acting HDV ribozyme crystal structure.[53] The GAAA loop of the P4 stem was replaced with an extension of the P4 helix to generate a three-stranded version of the ribozyme (Fig. 2.1b), as done before.[37] Additionally, U55 in the P4 helix was omitted to prevent potential misfolding. All RNA oligonucleotides were purchased from the Keck Biotechnology Resource Laboratory at the Yale University School of Medicine (http://keck.med.yale.edu/oligo).

Some strands were ordered already 2′-deprotected and cartridge-(Glen-Pak) purified, while others were 2′-deprotected in-house according to the manufacturer's instructions and then purified as described[162] by denaturing, 8 M urea, 20% polyacrylamide (19:1 acrylamide:bisacrylamide) gel electrophoresis. The bands of interest were cut out, diffusion-eluted into either 1 mM EDTA or "Crush 'n Soak" buffer (0.5 M NH_4OAc, 0.1 mM EDTA), ethanol-precipitated and resuspended in autoclaved double-distilled water, similar to in previous protocols.[162] Oligonucleotides that were cartridge purified by Keck were analyzed by denaturing, 8 M urea, 20% polyacrylamide gel electrophoresis to ensure high purity. RNA concentrations were calculated from absorption at 260 nm.

Strand B was purchased already labeled with a 5′-fluorescein and a 3′-C7-amino modifier in preparation for further in-house labeling with 5-carboxytetramethylrhodamine (TMR) succinimidyl ester (Invitrogen) according to the manufacturer's protocol, using overnight incubation at room temperature (22°C). The doubly-labeled RNA was then purified by denaturing, 8 M urea, 20% polyacrylamide gel electrophoresis. The bands of interest were cut, diffusion-eluted either into 1 mM EDTA or Crush 'n Soak buffer. The RNA was then ethanol-precipitated and resuspended using autoclaved, double-distilled water, similar to previous protocols.[162] The concentration of the doubly-labeled Strand B was determined from the absorption spectrum and estimated to contain ~35% of singly-labeled starting material.

Radioactive cleavage assays

Radioactive cleavage assays to determine ribozyme activity were performed similar to those previously described.[37] Cleavable substrate strands were 5′-end-labeled using T4 polynucleotide kinase (NEB) and [γ-^{32}P] ATP. All cleavage assays were carried out under single-turnover conditions, where free ribozyme concentration was ~125-fold above the substrate concentration, unless otherwise noted. Standard assay

conditions were a buffer of 40 mM Tris-HCl, pH 7.5, 25 mM DTT and 11 mM $MgCl_2$ at 25°C, unless otherwise stated. Ribozymes were preannealed by mixing Strand A and twice the concentration of Strand B (labeled with fluorescein and TMR) under standard assay conditions without magnesium, heating to 90°C for 2 min, and then cooling to 25°C for 10 min using a water bath. $MgCl_2$ was added to a final concentration of 11 mM. The free ribozyme was incubated at 50°C for 10 min, followed by cooling to 25°C for 10 min. The final ribozyme concentration, based on the concentration of Strand A, was 750 nM. Cleavage assays were initiated by adding a trace amount (<6 nM) of 5′-^{32}P-labeled substrate. Alternatively, the conditions of a previous study[157] were followed by mixing a 250-fold excess of ribozyme over the ^{32}P-labeled substrate in 25 mM Tris-HCl, pH 7.0, heating at 90°C for 2 min, and cooling to room temperature for 10 min. The mixture was then incubated at 37°C for 2 min. Magnesium chloride was added to initiate the assay to be carried out at 37°C. In either case, aliquots of 5 μl were taken at appropriate time intervals and quenched with 10 μl of stop solution (75% (v/v) formamide, 150 mM EDTA and 0.1% (w/v) each of bromophenol blue and xylene cyanol dyes). The 5′ cleavage product was separated from the uncleaved substrate by gel electrophoresis using denaturing, 8 M urea, 20% polyacrylamide gel electrophoresis at a limiting power of 25 W for 1.5-2 h at room temperature. The gel was then exposed to a phosphorimager screen for 18-22 h and scanned using a Typhoon 9410 Variable Mode Imager. Bands were quantified using ImageJ[163] and the calculated fraction cleaved at each time point was normalized to the sum of the substrate and product bands. Error bars were derived from 3-5 independent assays. Time traces of product formation were fitted with either a single-exponential ($y = A(1 - e^{(-k_{obs}t)})$) or double-exponential increase function ($y = A_1(1 - e^{(-k_{obs1}t)}) + A_2(1 - e^{(-k_{obs2}t)})$), utilizing the Marquardt-Levenberg nonlinear least-squares regression method (GraphPad Prism version 4.0b for Mac, GraphPad Software), where A is the pre-exponential factor and k_{obs} is the observed rate con-

stant for the reaction. Single- versus double-exponential behavior was determined using R-squared values to assess the quality of fit. R-squared values for shown fits were between 0.974 to 0.999.

Gel electrophoretic mobility shift assays

Similar to previous protocols,[37] 50 pmols of substrate or product strand was incubated with 30 pmols of Strand A and 8 pmols of Strand B in standard buffer without magnesium (final concentration of 40 mM Tris-HCl, pH 7.5, 25 mM DTT), supplemented with 10% (v/v) glycerol. The samples were then incubated using the protocol outlined for cleavage assays, with two minor changes: the initial 90°C incubation lasted 1 instead of 2 min and the two 10-min incubations were conducted at room temperature on the bench top instead of in a 25°C water bath. $MgCl_2$ was added to a final concentration of 11 mM and the solution incubated for at least 10 min at 50°C. The electrophoresis unit had been allowed to pre-equilibrate at 4°C for at least 2 h. After incubation, the samples were loaded onto a nondenaturing 10% polyacrylamide (19:1 acrylamide:bisacrylamide) gel containing 11 mM $Mg(OAc)_2$ and 40 mM Tris(OAc), pH 7.5. An electric field of ~10 V/cm was applied to the samples, which were electrophoresed for 12-14 h at 4°C. The gel was then imaged using a Typhoon 9410 Variable Mode Imager as described.[37] A dual-color image was generated by the scanner software and opened in ImageQuant. Relative band intensities were determined by analysis using ImageJ;[163] relative band percentages were obtained by summing the intensities from the fluorescein and TMR scans and normalizing them to their sum for each lane.

Steady-state FRET measurements

Solutions of 43-49 nM Strand B (final concentration after full assembly) in standard buffer were annealed as specified in the EMSA protocol. 140 μl of such a solution

were added to a Starna quartz fluorescence cuvette, its fluorescein label excited at 470 nm (bandwidth 8 nm) and emissions detected at 520 nm (bandwidth 16 nm) and 585 nm (bandwidth 16 nm) for fluorescein and TMR, respectively. After Strand B signal proved stable, 5 μl Strand A was added in 3.25-fold excess over Strand B for S8-U, S8-dU, S1-U, S1-dU, and S1-dUGG and P runs, and in a 3.8-fold excess over Strand B for S1-G and S1-dG runs. The resulting change in FRET was recorded for 30 min. Once this signal was stable, traces were further followed upon the addition of a substrate or product strand in 5.42-fold excess over Strand B for S8-U, S8-dU, S1-U, S1-dU, S1-dUGG and P runs, and in 5.8-fold excess for the S1-G and S1-dG runs. All substrate and product strands were added in 5-μl volumes, for a total final volume of 150 μl. The resultant signal change was measured in all cases for at least 5,000 seconds. In the case of the free Rz runs, 5 μl of standard buffer was added instead of the third strand, so that the final reaction volume was consistent. The ratio of acceptor to donor signal was calculated and normalized to the starting baseline. Each trace was then smoothed using a running average with a window of 10, and further normalized to control runs. Average curves of three runs for substrates and control, and four runs for the product, are reported here. Results were plotted as the relative FRET ratio decrease using Prism (GraphPad Prism version 4.0b for Mac, GraphPad Software).

Time-resolved FRET measurements

Preparations of 53 nM of Strand B (6 singly-labeled with fluorescein and 6 doubly-labeled with fluorescein and TMR) were incubated in standard buffer without magnesium for 1 minute at 90°C, followed by 10 min at 25°C. $MgCl_2$ was added to a final concentration of 11 mM and the solution was incubated for 10 min at 50°C, followed by 10 min at 25°C. Strand A was then added so that its final concentration would be 3-fold that of Strand B. Strands A and B were incubated at 25°C for 30 min. The

third strand, i.e., one of the noncleavable substrate strands or the product strand, or buffer in the case of the free ribozyme sample, was then added to a final concentration that was 50-fold that of Strand B. Each of the 6 pairs of samples was incubated at 25°C for 2 h and then kept at 4°C when not being used to measure decay lifetimes.

An ISS Alba confocal fluorescence microscope (based on an Olympus IX-81 frame) was used to collect fluorescein decay lifetimes. 40 μl of sample were loaded onto a coverslip, which was placed on a 60X 1.2 numerical aperture water immersion objective. A blue-enhanced supercontinuum laser (Fianium SC 400-6-PP, operated at 6 W) excited fluorescein at 488 nm with 100 ps pulse width and 20 MHz pulse picking. Decays were collected at 531 nm using a photomultiplier tube (Hamamatsu H7422P-40) feeding into a time-correlated single-photon counting acquisition card SPCM-830 (Becker & Hickl), with 4096 bins at 12.20 ps/bin. An instrument response function was obtained by measuring the back reflection off a coverslip without sample. For each sample, decays of donor (fluorescein) only and donor with acceptor (TMR) present were collected. Decays were collected so that the maximum raw peak count was between 40,000 to 60,000 counts. Five consecutive measurement reads were collected for a single sample and there were three inter-day replicates for each sample.

The donor-only traces with fluorescence intensity I_D were fit with a sum of exponentials $I_D(t) = \sum_i \alpha_{Di} \exp(-t/\tau_{Di})$ to extract three intrinsic donor lifetimes (τ_{Di} is the intrinsic donor lifetime and α_{Di} is the fractional contribution of each lifetime). Using these donor lifetimes, the doubly-labeled samples decays with fluorescence intensity I_{DA} were fit with the following model for a distance distribution:

$$I_{DA}(t) = (1 - f_A) I_D^0 \sum_i \alpha_{Di} \exp(-t/\tau_{Di}) + f_A I_D^0 \int_0^\infty P(r) \sum_i \alpha_{Di} \exp\left\{-\frac{t}{\tau_i}\left[1 + \left(\frac{R_0}{R}\right)^6\right]\right\} dR$$

where I_D^0 is the donor intensity at time 0, f_A is the fraction of molecules labeled with acceptor (allowing to correct for any singly labeled RNA), R is the distance between donor and acceptor, R_0 is the Förster distance (55 Å for the fluorescein and TMR pair) and $P(r)$ is the distance distribution. Distance distributions as-

sumed a three-dimensional Gaussian function model based on the equation: $P(r) = 4\pi R^2 c \exp[-a(R-b)^2]$ where a and b are parameters that describe the shape of the distribution and c is the normalization constant.

Molecular dynamics simulations

All MD trajectories were generated using the AMBER10/11 suite of programs[164,165] and the ff99 force field with recent modifications to χ (essential for RNA) and α/γ torsions (ff99bsc0χ_{OL3} version, current AMBER default).[111,115] Structures were net-neutralized with Na^+ ions placed at locations of optimal electrostatics using the LEaP module of AMBER10/11 and solvated using a rectangular periodic TIP3P water box model[166] extended to a distance of ≥10 Å from any ribozyme atom or Na^+ ion. The ABC protocol[167,168] was used to minimize and equilibrate structures prior to production runs. The minimization and equilibration protocol as well as the production runs were performed using the PMEMD module of the AMBER suite of programs. For the production runs, the particle mesh Ewald (PME) method was employed with a grid spacing of 1.0 Å, cubic-spline interpolation and a heuristic pair list update with a Lennard-Jones cutoff of 9.0 Å. Production runs were carried out at 300 K with constant-pressure boundary conditions and the Berendsen temperature coupling algorithm[169] with a time constant of 1.0 ps. The SHAKE procedure[170] was applied to constrain bonds involving hydrogen. Parameters for the protonated cytosine were derived using the Cornell et al. procedure.[113,116] Partial atomic charges were calculated for a protonated N1-methylcytosine using restrained electrostatic potential (RESP) fitting[171] and the HF/6-31G(d) method. Ab initio calculations were done using the program Gaussian03.[172] These charges were then used for the protonated cytosine; phosphate and sugar charges were kept the same as in unprotonated cytidine phosphate. Residual charge was added to C1′ to ensure a net charge of 0 for the protonated cytosine nucleotide. The U-1+ starting structure was prepared by

globally aligning the hammerhead ribozyme 2OEU and *trans*-acting HDV ribozyme 3NKB[53] structures to copy the active site residues from the former to the latter. The 1SJ3_U-1+ structure was prepared by globally aligning the *cis*-acting 1SJ3[30] and *trans*-acting 3NKB[53] structures. The coordinates for U-1, the scissile phosphate, and G1 were then copied from 1SJ3[30] into 3NKB.[53] The resulting structures were minimized and equilibrated using a protocol that involved first restraining the newly incorporated nucleotides, and subsequently removing those restraints. After these stages, the structure was equilibrated using the ABC protocol. Following minimization and equilibration, the structure was carefully checked for the presence of unfavorable geometry before carrying out production runs. For the U-1G mutation, the U-1 residue was mutated to G-1 by first deleting the atoms unique to uracil that do not overlay with guanine. The AMBER LEaP module was then used to fill in the missing atoms for guanine. The *cis*-acting model of the *trans*-acting ribozyme was created similarly to that of the 1SJ3_U-1+ structure, by modeling in the J1/2 joiner nucleotides from the *cis*-acting product crystal structure (PDB ID 1SJ3).[49] First, the product and *trans*-acting structures were globally aligned, then the J1/2 joiner region was cut and pasted into the *trans*-acting structure using PyMol. The newly created *cis*-acting structure was minimized and equilibrated using a protocol that involved first restraining the newly incorporated J1/2 joiner region, and subsequently removing those restraints. Following minimization and equilibration, the structure was carefully checked for the presence of unfavorable geometry before carrying out production runs. The Sander module was used to set distance restraints for appropriate simulations (Table 2.1).

MD trajectories were analyzed using the AMBER ptraj module and the program Simulaid.[173] The ptraj module was used to calculate heavy-atom distances, angles and radial distribution functions. All snapshots in the figures were generated from 100 ps-averaged structures calculated using ptraj. Simulaid was used to determine hydrogen

Table 2.1: List of HDV ribozyme MD simulations performed

Simulation Label	C75 Protonation State	Modifications of Starting Structure	Number of Replicates	Length (ns) of Each Replicate
U-1+	protonated	all ribose, with cleavage site modeled in	6	40, 20, 20, 20, 20, 100
1SJ3_U-1+[1]	protonated	ribose, with cleavage site modeled in using 1SJ3	5	50, 50, 50, 50, 20
U-1	unprotonated	all ribose, with U-1 modeled in	5	100, 50, 50, 50, 20
Product	unprotonated	all ribose, no U-1	5	20, 20, 20, 20, 20
Xtal_U-1+[2]	protonated	all ribose	5	20, 20, 20, 20, 20
dU-1+	protonated	d(U-1)	5	100, 50, 50, 50, 20
dUGG+	protonated	d(U-1)G1G2	5	20, 20, 20, 20, 20
Xtal_dUGG+[2]	protonated	d(U-1)G1G2	5	20, 20, 20, 50, 20
dUGG	unprotonated	d(U-1)G1G2	5	20, 20, 20, 20, 20
Cis+	protonated	all ribose, no U-1	5	20, 20, 20, 20, 20
U-1G+	protonated	all ribose, no U-1	5	50, 50, 50, 20, 20

[1]U-1(N3)-C3(O2P) was restrained during equilibration and for the first 10 ns in replicates 1 and 2, and during equilibration of replicates 5 and 6.
[2]Six distances are restrained during entire simulation as indicated in Fig. 2.6a.

bonding time traces using parameters of ≥120° donor-hydrogen-acceptor angle and ≤3.5 Å donor-acceptor distance. In-line fitness and other time traces were plotted using Matlab (2010b) (The MathWorks Inc.).[174] In-line fitness (F) was calculated using the following equation[175]

$$F = \frac{(\tau - 45)}{(180 - 45)} \times \frac{3^3}{d^3_{O2',P}}$$

where τ is the angle between the 2′ oxygen nucleophile (O2′), the phosphorus electrophile (P) and the 5′ oxygen of the leaving group (O5′) and d is the distance between the 2′ oxygen nucleophile and the phosphorus electrophile.

2.3 Results

The *trans*-acting HDV ribozyme preferentially cleaves a short U-1 substrate with biphasic kinetics

The crystallized *trans*-acting HDV ribozyme consists of two strands: the substrate and ribozyme segments. To incorporate site-specific fluorophore labels for solution probing by FRET, we synthesized three RNA strands by opening the non-conserved capping loop on helix P4 and labeling the 5′ and 3′ ends of the downstream segment (Strand B) with fluorescein and tetramethylrhodamine, respectively (Fig.2.1b), following an established strategy.[37] We first tested the catalytic activity of the resulting two-strand ribozyme using radioactive single-turnover cleavage assays at 25°C on three cleavable *trans*-substrates (Fig. 2.1b): (i) the short, U-1-containing substrate with the sequence used in the crystal structure, termed here S1-U; (ii) the U-1G mutant substrate, S1-G, to check whether this mutation is as deleterious as previously observed for other HDV ribozymes;[160] and (iii) the longer, 8-nucleotide substrate, S8-U, with U-1 and a 5′ sequence selected for its predicted low tendency to form alternate secondary structures. For S1-U and S8-U, the time traces of fraction

cleaved were fit best with double-exponential functions (Fig. 2.1c). The observed biphasic rate constants k_{obs1} and k_{obs2} for S1-U were 1.5 min^{-1} (amplitude: 0.64) and 0.12 min^{-1} (0.28), respectively. For S8-U, k_{obs1} and k_{obs2} were ~2-fold slower with 0.77 min^{-1} (0.36) and 0.075 min^{-1} (0.30), respectively, consistent with a potential of the ribozyme to form weak alternative secondary structures as predicted by the program mfold.[176] A 5-fold increase in ribozyme concentration resulted in a modest 2-fold increase in the faster k_{obs} values for both S1-U and S8-U, suggesting that the reported values approach the inherent cleavage rate constant (Fig. 2.2). In the case of S1-G, cleavage time traces were best fit with a single-exponential function, yielding a k_{obs} of 0.087 min^{-1} with an endpoint of ~70% cleaved (Fig. 2.1c). A 5-fold increase in ribozyme concentration did not further increase this low value (Fig. 2.2). The ~17-fold lower activity of the S1-G compared to the S1-U substrate is consistent with previous observations that G-1 is much disfavored relative to U-1,[160] rationalizing the evolutionary conservation of U-1 in clinical isolates of the virus.[177] The observation of double-exponential cleavage time courses and final fractions cleaved of less than unity for the S1-U and S8-U substrates under single-turnover conditions suggests heterogeneity of the ribozyme population. Conversely, the fact that the S1-G substrate is cleaved in a single-exponential time course does not rule out such heterogeneity, especially in light of the significantly lower-than-unity cleavage extent (Fig. 2.1c).

For the crystallized parent ribozyme was reported to catalyze cleavage of the S1-U substrate with a single k_{obs} of ~6.5 min^{-1} at 37°C, in 25 mM Tris-HCl, pH 7.0, 10 mM $MgCl_2$,[157] slightly higher than our k_{obs1}. We therefore explained the impact of the higher temperature and different annealing protocol by repeating our assay under these conditions. We again observed double-exponential cleavage, yielding a k_{obs1} of 2.4 min^{-1} (0.63) and a k_{obs2} of 0.99 min^{-1} (0.30) (Fig. 2.2). The small (2.7-fold) difference between our k_{obs1} at 37°C and the previously reported k_{obs} suggests that our fluorophore-labeled variant is a relevant model system for the crystallized HDV

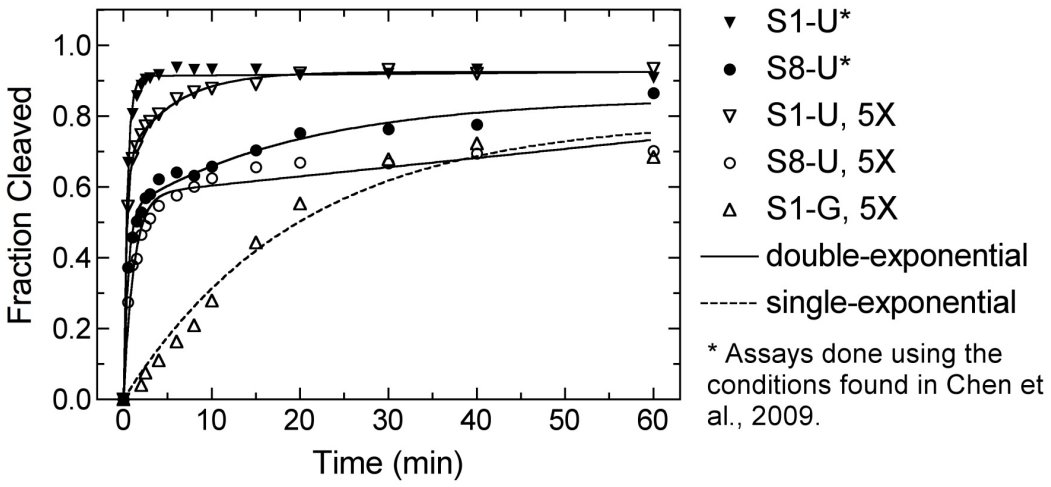


Figure 2.2: Additional HDV ribozyme cleavage assays. The asterisk (*) denotes cleavage assays following the protocol by Chen et al.,[53] described in the Materials and Methods section of the main text. The 5X denotes the use of ribozyme at a concentration 5-fold greater than that used for assays in the main text.

ribozyme.

The crystallized precursor ribozyme exhibits heterogeneous, catalytically active populations

Fluorescence resonance energy transfer (FRET) is a powerful tool to determine global conformations of RNA in solution.[178–180] To ask whether the observed heterogeneous cleavage kinetics of the crystallized *trans*-acting HDV ribozyme are reflected in folding heterogeneities, we used electrophoretic mobility shift assays (EMSAs). These assays are sensitive to the distinct charges and hydrodynamic radii of RNA conformers and can be monitored by FRET, as previously described.[37,41] In addition to our three cleavable substrates S1-U, S1-G and S8-U, which are expected to eventually convert mostly to product upon ribozyme cleavage during either sample preparation or gel electrophoresis, as well as the product strand P for reference, we generated four non-cleavable substrate versions that are expected to remain in the precursor state (Fig. 2.1b): one version of each substrate with a single 2′-deoxy modification at the cleavage site (i.e. modifying the U-1 or G-1 nucleotide), termed S1-dU, S1-dG and S8-dU, and one substrate identical to the crystallized one with three 2′-deoxy modifications at the -1, 1 and 2 positions, encompassing the cleavage site (S1-dUGG). We then formed substrate-free ribozyme (Strands A and B annealed, or "Free Rz", Fig. 2.1b) and added each one of the cleavable or non-cleavable substrates, or the product strand, to compare all complexes with Strand B alone on an EMSA gel (Fig. 2.3a).

The cleavable substrate complexes all co-migrate with the single band of the product form of the ribozyme, indicating that they cleave essentially to completeness before or during electrophoresis (Fig. 2.3a). Themuch longer time taken to do the EMSA experiment (~14 hours) compared to the time taken for the radioactive cleavage experiments (~4 hours) may be an explanation for a more complete cleavage reaction observed in the EMSA gel. This product band migrates only slightly more

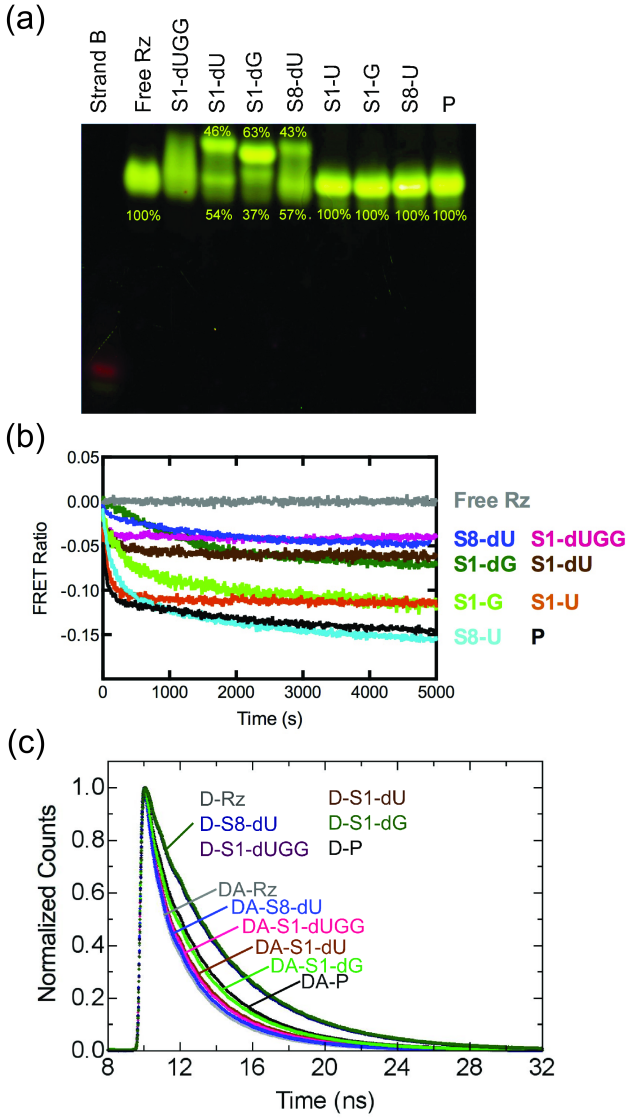


Figure 2.3: FRET probing of our 8 *trans*-acting precursor and product ribozymes. **(a)** Fluorescent EMSA of the individual complexes as indicated. Where appropriate, the relative fractions in the main bands of a lane are given. Strand B and free ribozyme (Rz) lanes are shown for reference. **(b)** Normalized ssFRET time courses upon substrate or product addition, as indicated. A decrease in FRET ratio indicates a lengthening of the fluorophore distance along the longitudinal axis of the ribozyme. Each trace is an average of at least three trials. **(c)** Representative fluorescein fluorescence decays for free ribozyme (Rz), ribozyme bound to noncleavable substrates, and ribozyme bound to product (P), as indicated. Traces for ribozymes labeled with donor-only (D) and with donor plus acceptor (DA) are plotted. Each trace is 1 of 3 replicates, but composed of an average of 5 reads.

slowly than the free ribozyme, suggesting that the addition of the product strand does not significantly increase the hydrodynamic radius of the free ribozyme. By contrast, each of the non-cleavable substrates shows more than one distinct band, where one band migrates similar to the product complex and another migrates significantly more slowly, with the relative contribution of each band indicated in Fig. 2.3a. The only exception is the S1-dUGG substrate used in the crystal structure, whose precursor ribozyme complex runs as a smear spanning the two bands observed for the other non-cleavable substrate complexes (Fig.2.3a). These observations suggest that each precursor ribozyme complex forms more than one conformation. These conformations either rapidly (compared to the electrophoretic migration time) interconvert or happen to migrate in a continuum of conformations of only slightly different hydrodynamic radius, leading to smearing. Such conformational heterogeneity is consistent with the catalytic heterogeneity observed in our cleavage assays.

The S1-dUGG bound ribozyme appears to exist in solution as a particularly broad set of conformations (Fig. 2.3a), suggesting that crystallization leads to the relative enrichment (purification) of just one of these conformers.[53] This observation, together with the fact that the bottom band in all precursor complexes almost co-migrates with the product complex, may provide an explanation for why the recent precursor crystal structure resembles the previous product crystal structures more than their precursor counterparts: a more product-like conformation was selected during crystal growth. Consistent with this hypothesis and the notion that the more product-like conformation is catalytically more activated, the non-cleavable S1-dG substrate leads to a smaller fractional population of the product-like band (Fig. 2.3a), as well as to lower catalytic activity (Fig. 2.1c).

All *trans*-acting ribozymes undergo a global conformational change upon conversion from precursor to product

To corroborate our EMSA results, we employed a previously developed steady-state FRET (ssFRET) assay,[37,41] in which we first form the substrate-free ribozyme, then add a cleavable or non-cleavable substrate or the product strand while monitoring in real-time the evolution of the relative FRET ratio. This ratio serves as a measure of changes in distance between the two fluorophores that mark the ends of helices P4 and P2 (Fig. 2.1b), thus allowing us to quantify the time dependence of global conformational changes. As observed for a majority of similar cases before,[41] binding of non-cleavable substrate to the free HDV ribozyme leads to a temporal decrease in FRET ratio, reflecting the substrate binding kinetics (Fig. 2.3b). When adding the product strand P instead, the decrease is more pronounced, suggesting that the P4-P2 end-to-end distance monitored by the FRET pair becomes longer in the product than in any of the precursor structures. The FRET ratio decreases over time in all cases but the S1-G ribozyme kinetics are characterized by a double-exponential decay, providing additional evidence for heterogeneous folding (Fig. 2.3b). Supporting the notion that progression from precursor to product leads to further lengthening of the P4-P2 axis, the combination of binding and subsequent (partial) cleavage upon addition of the cleavable substrates S1-U, S-1G and S8-U leads to FRET ratio decreases close to that associated with binding of product strand P (Fig. 2.3b). The fact that, for the S-1U and S-1G substrates, the overall decrease falls between those of binding the corresponding non-cleavable strands S1-dU/S1-dG and product strand P is further consistent with the observed incomplete conversion of these substrates into product in the radioactive cleavage assays (Fig. 2.1c). Taken together, these observations strongly support the result from our EMSAs that, in solution, the precursor and product forms of the most recent crystal structure display distinct global conformations. Thus, a global conformational change appears to accompany cleavage,

consistent with previous findings on other *trans*-acting HDV ribozymes.[37,40,41,43]

To quantify this conformational change and the resulting P4-P2 end-to-end lengthening, we performed time-resolved FRET (trFRET) measurements as described[37,41,181–183] on ribozyme free of substrate as well as bound to each one of our non-cleavable substrates or the product strand. As expected, the fluorescein-only singly labeled ribozyme complexes display nearly identical fluorescence decay curves, whereas all fluorescein-tetramethylrhodamine doubly-labeled ribozymes exhibit distinct, faster decays due to varying FRET efficiencies (Fig. 2.3c). Analysis of these decay curves yielded the single Gaussian inter-fluorophore distance distributions summarized in Table 2.2. The free ribozyme exhibits the most compact structure with a mean inter-fluorophore distance of ~51 Å. The precursor ribozymes adopt distances of between 53 Å (S8-dU) and 59 Å (S1-dUGG, Table 2.2), depending on the non-cleavable substrate. The product conformation, with a mean fluorescein-tetramethylrhodamine distance of ~67 Å, is indeed extended, by ~8 Å in the P4-P2 end-to-end distance, compared to the crystallized S1-dUGG complex (Table 2.2). Such an extension is consistent with our ssFRET results as well as previous studies of other *trans*-acting HDV ribozymes that showed extensions of between 2 and 15 Å upon conversion of the precursor to the product.[37,41]

Table 2.2: Distance distribution parameters obtained from time-resolved FRET data

	Mean Distance (Å)	FWHM (Å)	χ^2
DA-Rz	51 ± 2	30 ± 3	3.1 ± 0.8
DA-S1-dU	58 ± 1	48 ± 8	3.2 ± 1.2
DA-S1-dUGG	59 ± 3	19 ± 25	3.4 ± 1.3
DA-S8-dU	53 ± 3	30 ± 7	3.3 ± 0.9
DA-P	67 ± 4	38 ± 5	2.7 ± 0.3
DA-S1-G	59 ± 2	40 ± 3	3.0 ± 0.4

Taken together, our EMSA, ssFRET and trFRET data consistently show that, in solution, the *trans*-acting HDV ribozyme that was recently visualized through crys-

tallization undergoes a significant conformational change upon conversion from the precursor to the product. The ribozyme also displays folding heterogeneity in the precursor form that is removed upon cleavage. The finding that, in our EMSAs, all of our substrates cleave to completion (Fig. 2.3a) further indicates that all conformational isomers are catalytically active, although to varying degrees as evident from the biphasic kinetics of our cleavage assays (Fig. 2.1c). That the crystal structure shows a relatively well-defined single global conformation further suggests that crystallization may have purified a more product-like conformer. Consistent with this notion, the P4-P2 end-to-end distance in the crystal is ~70 Å,[53] close to that of the product conformer in solution as observed here (~67 Å).

In MD simulations, the modeled hammerhead ribozyme cleavage site results in favorable in-line fitness, as long as C75 is protonated

While a single product-like *trans*-acting HDV ribozyme conformer was crystallized,[53] our FRET studies show that in solution multiple catalytically active global conformations coexist. This discrepancy raises the question of whether the crystal structure itself may hint at such conformational divergence. Indeed, the U-1 and scissile phosphate were not resolved in the otherwise high-resolution electron density and had to be modeled in.[53] The two simplest explanations for this lack of electron density are that a significant fraction of molecules degraded during crystal growth or that the crystal contains a conformational ensemble with some flexibility and/or disparity around the active site, a possibility favored by the authors.[53] To test this notion and provide a window into the rapid (sub-microsecond), small-scale dynamics of the ribozyme, not easily accessible by experimental techniques,[184] we performed MD simulations on the crystallized *trans*-acting precursor using our established protocols.[92,106,114,159,160,184–186] We used the latest parmbsc0χ_{OL3}[111,115] variant of the Cornell et al. AMBER force field,[116] which is essential for stable RNA simulations.[187]

We analyzed 49 simulations with an aggregate 1.8 μs of simulation time (Table 2.1). With such an extent of simulations available, we focused our analysis on a parameter that monitors the most critical part of the underlying crystal structure, the modeled cleavage site, and is able to relate its conformation to a functional parameter, catalytic activity. To this end, we used the previously described in-line fitness that combines the distance between the attacking 2′-O nucleophile and the phosphorous atom with the in-line attack angle formed by the 2′-O nucleophile, phosphorous and 5′-O leaving group[188] into a single value reflecting the relative poise to undergo catalysis.[175] In essence, a "high" fitness value close to 1 indicates a phosphodiester bond geometry more amenable to cleavage than one described by a "low" fitness value closer to 0. Therefore, by tracking the fitness of the scissile bond over the course of the simulation, we can track how active the active site is.

The cleavage site of the crystallized *trans*-acting HDV ribozyme was proposed to resemble the cleavage site conformation of the crystal structure of the hammerhead ribozyme (PDB ID 2OEU).[53] Incorporating this model into an all-ribose form of the *trans*-acting crystal structure with an N3-protonated C75H+, we performed six simulations termed U-1+ (Table 2.1). We found that half of our simulations predominantly retained a high (>0.5) in-line fitness, whereas the other half showed low (<0.5) values with occasional excursions to higher values (Fig. 2.3a). Alternatively, we modeled the cleavage site using an alternative "U-turn" motif topology that we had identified in the previous *cis*-acting precursor crystal structure (PDB ID 1SJ3; Fig. 2.3b).[160] The resulting five 1SJ3_U-1+ simulations (Table 2.1) exhibited poor in-line fitness, with >99% of the total simulation time resulting in values of <0.5 (Fig. 2.3a). In addition, when restraining the U-1(N3)...C3(O2P) distance for the first 10 ns in two simulations (numbers 1 and 2) to that of a hydrogen bond characteristic of the U-turn motif,[160] the in-line fitness never increased beyond 0.5 (Fig. 2.3a). In almost all 1SJ3_U-1+ simulations, C75H+ resided far from G1(O5′), disrupting a

hydrogen bond that appears critical for retaining an optimal scissile phosphate geometry (Fig. 2.3c). Often, the two nonbridging oxygens of the scissile phosphodiester bond accepted a fluctuating, bifurcated hydrogen bond from the two exocyclic amino protons of A77(N6) (Fig. 2.3d). In turn, this interaction created a cation-binding pocket different from that observed in the hammerhead ribozyme model of the cleavage site, with on average one ion bound to G1(O1P), and a partially (~25%) occupied ion bound to A77(N7) and C75H+(O2) (Figs. 2.3e,f). Such Na^+ ion binding is often predictive of Mg^{2+} binding while avoiding the complications from insufficient sampling and lack of polarization associated with simulating a tight binding Mg^{2+}.[184,185] We conclude that, while the U-turn motif is consistent with the earlier *cis*-acting precursor crystal structure, it does not lend itself to high in-line fitness in the context of the product-like *trans*-acting ribozyme. In addition, combining the simulation data, while still far from full sampling, with our experimental findings suggests that the catalytic core of the HDV ribozyme occupies a rugged folding free energy landscape with multiple low-energy wells.[66,71,189]

In our simulations so far we used an N3-protonated C75H+, which we found leads to a C75H+(N3)...G1(5′O) hydrogen bond critical for maintaining a favorable in-line fitness. Protonation is justified since the crystal structure was obtained at a pH of 5.0,[53] however, since the pK_a of C75 of the HDV ribozyme was measured as ~6.4,[47,48] at the more physiological pH of 7.5 used in our solution probing, a protonated C75H+ will represent only a minor (<10%) species. To ask what impact an unprotonated C75 has on in-line fitness, we performed five U-1 simulations without protonating C75 in the starting structure (Table 2.1). Indeed, very few excursions to high (>0.5) in-line fitness occurred (Fig. 2.5a), and without the C75H+(N3)...G1(5′O) hydrogen bond this distance increased to typically 5-10 Å (Fig. 2.5b). Such distances are not typically seen in the U-1+ simulations (Fig. 2.5b). On average, a single Na^+ ion occupied a position close to the unprotonated C75(N3), as well as C75(O2), essentially adopt-

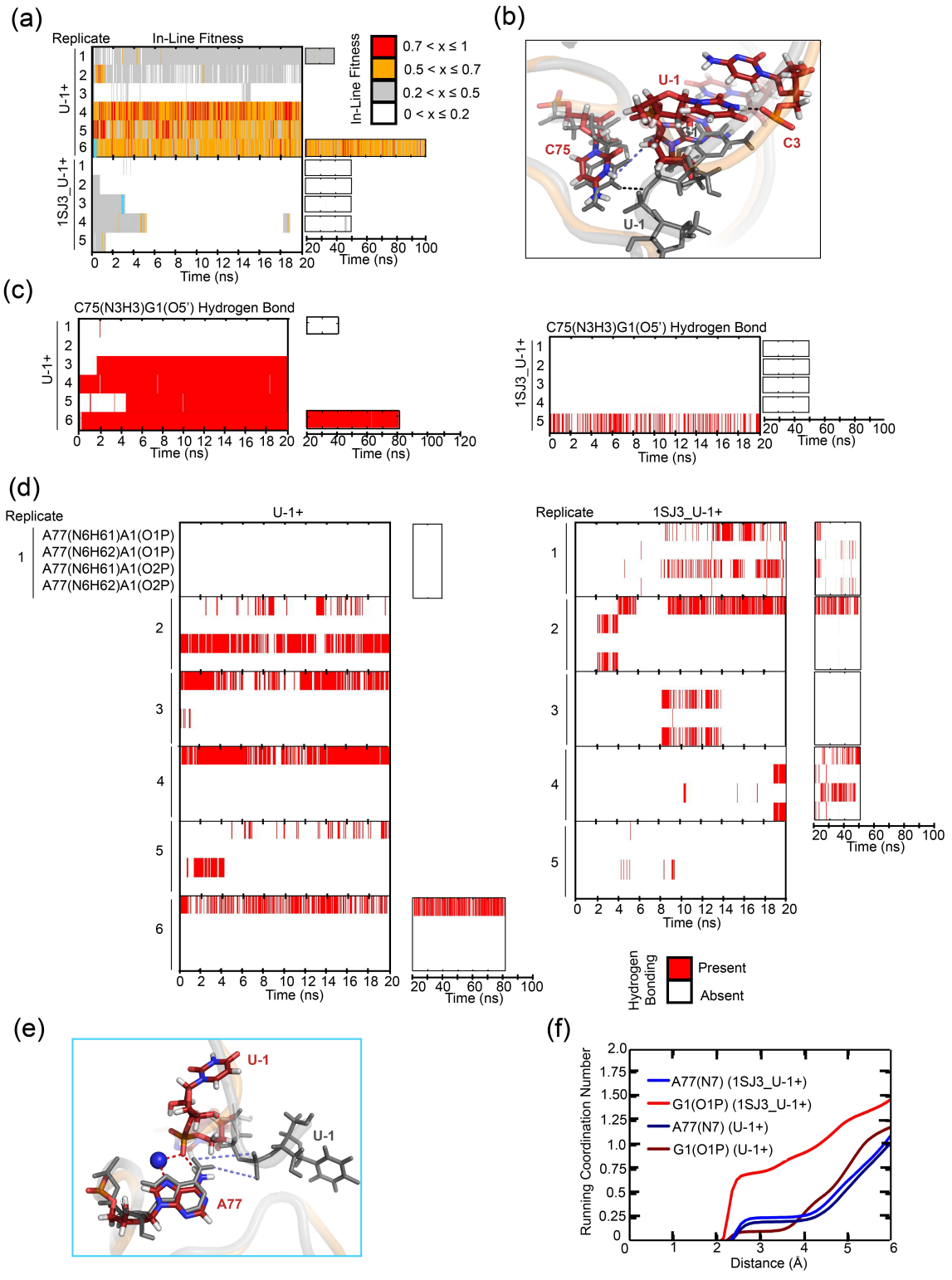


Figure 2.4: Testing the hammerhead ribozyme model for the cleavage site.
Caption continued on next page.

Figure 2.4: **(a)** In-line fitness of our U-1+ and 1SJ3_U-1+ simulations (Table 2.1). **(b)** Overlay of 1SJ3_U-1+ and U-1+ scissile phosphates. The 1SJ3_U-1+ structure is shown in color, while the U-1+ structure is shown in transparent cartoon with relevant residues shown as thickened lines. The U-1(N3)...C3(O2P) heavy atom distance restrained in some simulations (Table 2.1) is denoted with a black dashed line, while the destabilized C75(N3H3)...G1(O5′) distance is denoted with a light blue dashed line. For reference, the stable C75(N3H3)...G1(O5′) distance is denoted in the gray U-1+ structure with a black dashed line. In all stick representations of 1SJ3_U-1+ structures, carbons are colored dark red, nitrogens are blue, oxygens are red, hydrogens are white and sodium ions are blue. **(c)** Comparing C75H+(N3H3)...G1(O5′) hydrogen bonding occupancies of U-1+ and 1SJ3_U-1+ simulations. Traces were computed using the Simulaid program and in-house scripts, with cutoffs of ≥120° and ≤3.5 Å to indicate presence of a hydrogen bond. **(d)** Monitoring of alternate A77(N6)...G1(O1/O2P) hydrogen bonding interactions for U-1+ and 1SJ3_U-1+ simulations. **(e)** The alternative ion cage observed in 1SJ3_U-1+ simulations. Overlays of 1SJ3_U-1+ and U-1 are shown in the left-hand portion of the figure. As in panel (b), the 1SJ3_U-1+ structure is shown in color, while the U-1+ structure is in transparent gray cartoon for reference. Relevant residues are shown in sticks and thickened lines, respectively. Red dashed lines indicate hydrogen bonds/ionic interactions that form and are stable in 1SJ3_U-1+ trajectories, while blue dashed lines indicate hydrogen bonds in U-1+ simulations that are destabilized due to the new scissile phosphate conformation of the 1SJ3_U-1+ structure. The cyan lines in panel (a) denote from which replicate and at what times the snapshots were taken. **(f)** Coordination number of sodium ions at a given distance from each of the functional groups specified. The plotted curves are weighted averages from all available simulations of a given type.

ing the position of the C75N3 proton in the U-1+ simulations, occupying a pocket of the ribozyme rendered more negative in electrostatic surface potential due to deprotonation of C75H+ (Fig. 2.5c). The fractional residency of this ion was 101 ± 29% for C75(N3) (indicating that occasionally more than one Na^+ binds in the pocket) and 78 ± 21% for C75(O2) (Fig. 2.5d). These observations are consistent with previous experimental studies that suggested anti-cooperativity between protonation of C75 and relatively weak binding of an active-site magnesium ion.[185,190] Taken together, our data strongly suggest that protonation of C75 to form a C75H+(N3)...G1(5′O) hydrogen bond is critical for the *trans*-acting HDV ribozyme precursor to adopt a favorable in-line fitness value, positioning its scissile phosphate for catalysis.

Further supporting the notion that the C75H+(N3)...G1(5′O) hydrogen bond is a critical anchoring point for a product-like structure, we found that five simulations of a product form of the *trans*-acting crystal structure (with U-1 removed, G1(5′O) protonated and C75 left unprotonated, termed "Product" in Table 2.1) showed only a subtle difference within the active site relative to our U-1+ simulations. In this case of a less crowded active site without U-1, an even more stable G1(OH5′)...C75(N3) hydrogen bond forms. This hydrogen bond is present in all five Product simulations, with occupancies ranging from 30%-80% (Fig. 2.5e).

Intermolecular crystal contacts and 2′-deoxy modifications decrease in-line fitness and stacking of U-1 on U23

Upon inspection of the crystal structure, we noticed an intermolecular contact between the closing GAAA tetraloop of P4 of one molecule in the unit cell and the active site in another (Fig. 2.6a).[53] While the several hydrogen bonds from this contact may have fortuitously aided in stabilizing crystals, we sought to understand whether they may have an impact on the cleavage site. To this end, we prepared a starting structure of a ribozyme together with the adjacent P4 stem-loop from a second ribozyme

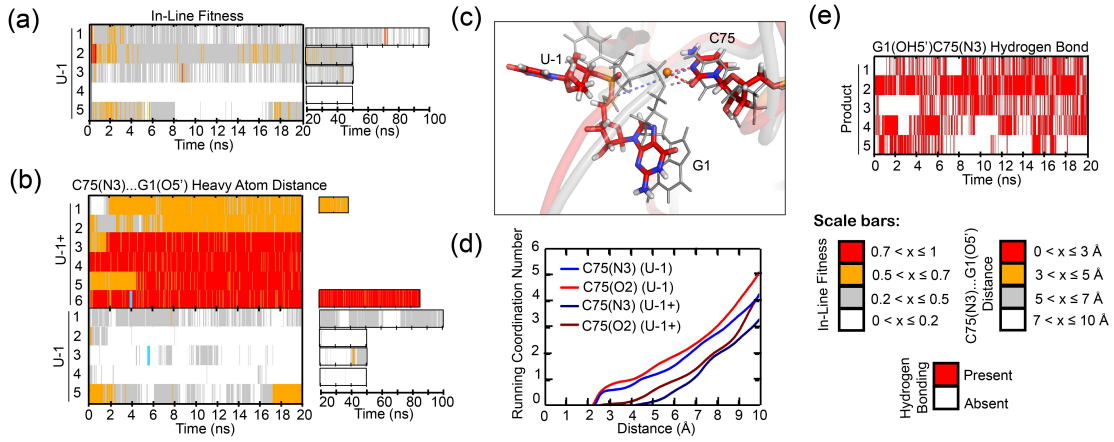


Figure 2.5: A C75(N3)...G1(O5′) hydrogen bond supports a product-like active site. **(a)** In-line fitness of U-1 trajectories. **(b)** Comparison of C75(N3)...G1(O5′) heavy atom distances in U-1+ and U-1 trajectories. **(c)** Structural overlay of U-1 and U-1+ active sites. The U-1 structure is in color, while the U-1+ structure is in transparent gray. Relevant residues are shown as sticks and thickened lines, respectively. Red dashed lines indicate stable coordination of the orange sodium ion. Light blue dashed lines indicate the C75H+(N3H3+)...G1(O5′) hydrogen bond lost in the U-1 structures. Colored atoms in stick representations are colored as indicated in Fig. 2.3. The cyan lines in panel (b) denote from which replicate and at what times the snapshots were taken. **(d)** Running coordination numbers of U-1+ and U-1 ion-binding residues. The plotted curves indicate the number of sodium ions coordinated by a given residue at a range of distances, and are weighted averages from all available simulations of a given set. **(e)** Time traces for the reverse G1(OH5′)...C75(N3) hydrogen bond in product trajectories.

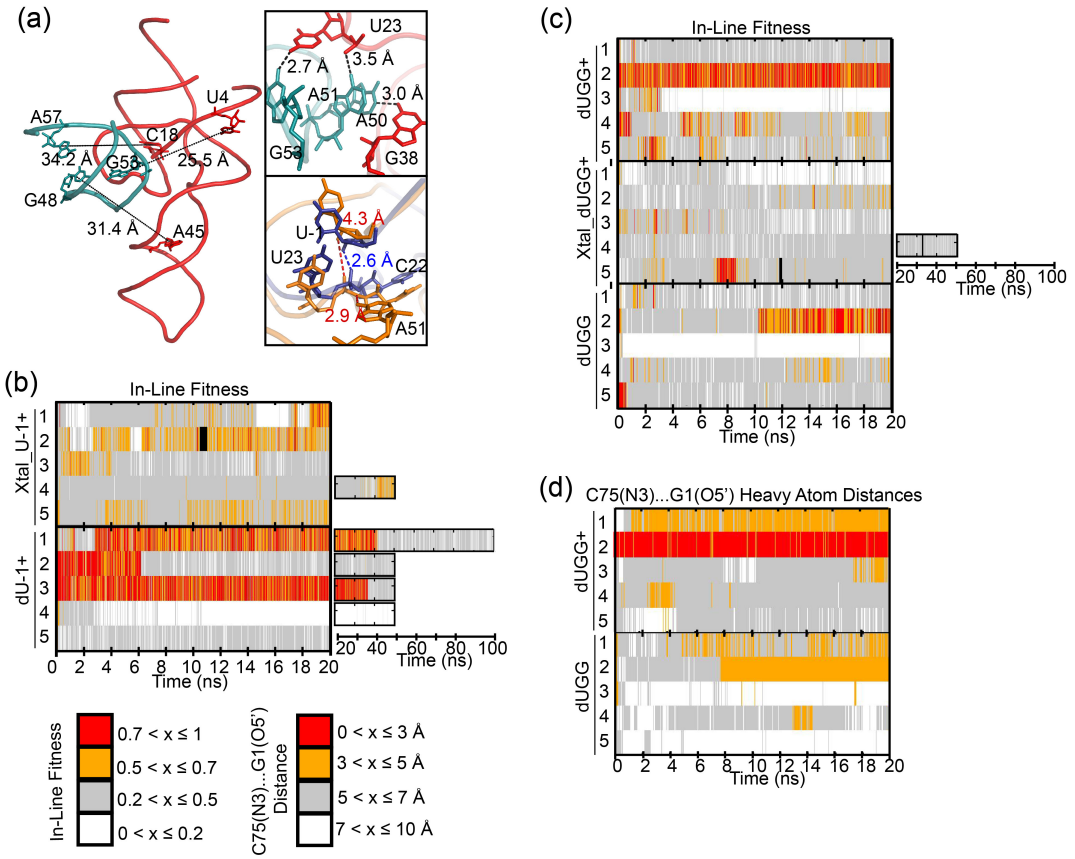


Figure 2.6: Effects of crystallization parameters on active site geometry. **(a)** Whole-molecule view of our crystal structure model, with length of long-range restraints and participating residues indicated. Top inset: Short-range restraints and participating residues in our Xtal simulations. Bottom inset: A 100-ps-averaged snapshot from a Xtal_U-1+ simulation (gold) is overlaid with a 100-ps-averaged snapshot from a U-1+ simulation (navy). The adjacent P4 loop interacts with U23 and C22 resulting in increased distances of these residues from U-1. As a result, interactions important for stabilizing favorable fitness cannot easily form. **(b)** Comparison of in-line fitness values in Xtal_dU-1+ and dU-1+ simulations. **(c)** Comparison of in-line fitness "equivalent" values (see Results section for definition) in dUGG+, Xtal_dUGG+ and dUGG simulations. (d) Comparison of C75(N3)…G1(O5′) heavy atom distances in dUGG+ and dUGG simulations.

in the unit cell, using six distance restraints to keep the latter P4 stem-loop from changing its relative positioning (Fig. 2.6a). The resulting five simulations termed Xtal_U-1+ (Table 2.1) exhibited considerably poorer fitness than our U-1+ simulations (Fig. 2.6b). A closer look revealed that the loop's formation of hydrogen bonds with U23 and C22 resulted in increased distances of U-1 from U23 and C22 (Fig. 2.6a). These increased distances hinder formation of a U-1(O2′)...U23(O1P) hydrogen bond, which in turn destabilizes a specific U-1 conformation wherein it stacks on U23, a stacking interaction that occurred frequently in our U-1+ simulations (Fig. 2.6a). The crystal packing contacts thus may contribute to the lack of U-1 electron density[53] since it confers a less well-defined positioning of U-1.

Another unusual feature of the crystal structure is the presence of three 2′-deoxy modifications in U-1, G1 and G2 flanking the cleavage site (Fig. 2.1a). While a single such modification of the N-1 nucleotide is often used to render a ribozyme substrate non-cleavable by removing the 2′-OH that serves as a nucleophile during cleavage, substrates with multiple 2′-deoxy modifications have been used less frequently. To ask what effect these extended modifications may have on the local structural dynamics of the cleavage site, we first performed five dU-1+ simulations with a single 2′-deoxy modification at U-1 (Table 2.1) and compared them to our all-ribose U-1+ simulations. Because there is no U-1(2′O) in the dU-1+ simulations, an in-line fitness "equivalent" was calculated by using U-1(2′H) in place of U-1(2′O). In two of our dU-1+ simulations, in-line fitness started off at favorable values, but later sampled unfavorable values as well (Fig. 2.6b). In two other dU-1+ simulations, fitness started with and remained at unfavorable values (Fig. 2.6b). This behavior resembles that of the all-ribose U-1+ simulations (Fig. 2.3a).

We then performed five dUGG+ simulations that includes all three 2′-deoxy modifications of the crystal structure (Table 2.1). In this case, less frequent adoption of favorable in-line fitness was observed and, if adopted, high in-line fitness values were

often rapidly lost (Fig. 2.6c). These observations suggest that, as with the intermolecular crystal contacts, 2′-deoxy modifications near the cleavage site may reduce catalytic propensity. Specifically the absence of a G1(2′OH)...U-1(5′O) interaction affects this difference, whereas hydrogen bonding by G2(2′OH) does not appear to have distinct effects. Interestingly, the dynamic exchange between high and low fitness values appears to be accelerated for the dUGG+ simulations (Fig. 2.6c), consistent with the large number of closely related (i.e. almost co-migrating) conformations we observed in our EMSAs for S1-dUGG ribozymes (Fig. 2.3a). Such an enhanced flexibility may further have contributed to the lack of experimental electron density for U-1 and the cleavage site in the 2′-deoxy-modified crystal structure.[53] When we performed simulations with the crystal contacts and all three 2′-deoxy modifications at the active site ("Xtal_dUGG+", Table 2.1), we observed a greater dominance of in-line fitness values <0.4 in each replicate compared to Xtal_U-1+ or dUGG+ (Fig. 2.6c). Finally, when performing five dUGG simulations with an unprotonated C75 we observed behavior similar to that of the C75 unprotonated dU simulations with particularly low in-line fitness. We also observed generally unfavorable C75(N3)...G1(5′O) heavy atom distances in both the dUGG+ and dUGG simulations (Fig. 2.6d), indicating that the incorporation of all three 2′-deoxy modifications contributes to further active site destabilization beyond simply unfavorable in-line fitness.

A *cis*-acting precursor model exhibits greater fluctuations in fitness whereas multimodal stacking of a U-1G mutation with U23 results in decreased fitness

A wealth of experimental evidence has indicated that *cis*-acting HDV ribozymes generally cleave ~10-100 times faster than their *trans*-acting counterparts.[33,37,45,157,159] It has been hypothesized that the loss in catalytic activity of *trans*-acting ribozymes is due to a less poised or stable cleavage site architecture as a result of the distal

discontinuity of the RNA backbone.[33,45] To test this hypothesis in the context of the *trans*-acting crystal structure, we modeled into it two J1/2 joiner nucleotides (A8 and U9) from the previous *cis*-acting product crystal structure[49] to arrive at a corresponding, C75(N3)-protonated (or C75H+) *cis*-acting precursor model (Cis+, see Table 2.1 and Fig. 2.8a). While generally similar in probability distribution of in-line fitness values (Fig. 2.7), we found this model to exhibit greater dynamics in fitness than our U-1+ simulations, as all five of our Cis+ simulations converted multiple times between high and low in-line fitness values (Fig. 2.8b). This finding contrasts with our U-1+ simulations that show either mostly high or mostly poor in-line fitness over entire 20 ns simulations (Fig. 2.3a). These more rapid fluctuations of Cis+ in-line fitness compared to that of the U-1+ simulations are consistent with a lower barrier to transitioning between inactive and active local conformations once C75 is protonated and may rationalize the generally higher catalytic activity of *cis*-acting HDV ribozymes.

Conversely, previous studies have shown that a G-1 upstream of the scissile phosphate results in a decrease in catalytic rate relative to the U-1 found conserved in clinical isolates of HDV.[160,177] To ask whether the *trans*-acting ribozyme structure may offer an explanation for this effect as well, we analyzed five simulations with a G-1 in place of the U-1, in the presence of an N3-protonated C75H+ (U-1G+, Table 2.1). We found 98.5% of the total frames of these trajectories to adopt in-line fitness <0.5 (Fig. 2.8c). This trend was mainly determined by unfavorable in-line attack angles, which in turn appeared to be due variable stacking of G-1 with U23; in addition to a sliding motion of U23 on the larger G-1, significant unstacking of the two bases was observed in 3 out of 5 trajectories (Fig. 2.8c). These observations rationalize the decrease in catalytic activity upon U-1G mutation of *trans*-acting HDV ribozymes[160] and further support the functional relevance of our MD simulations.

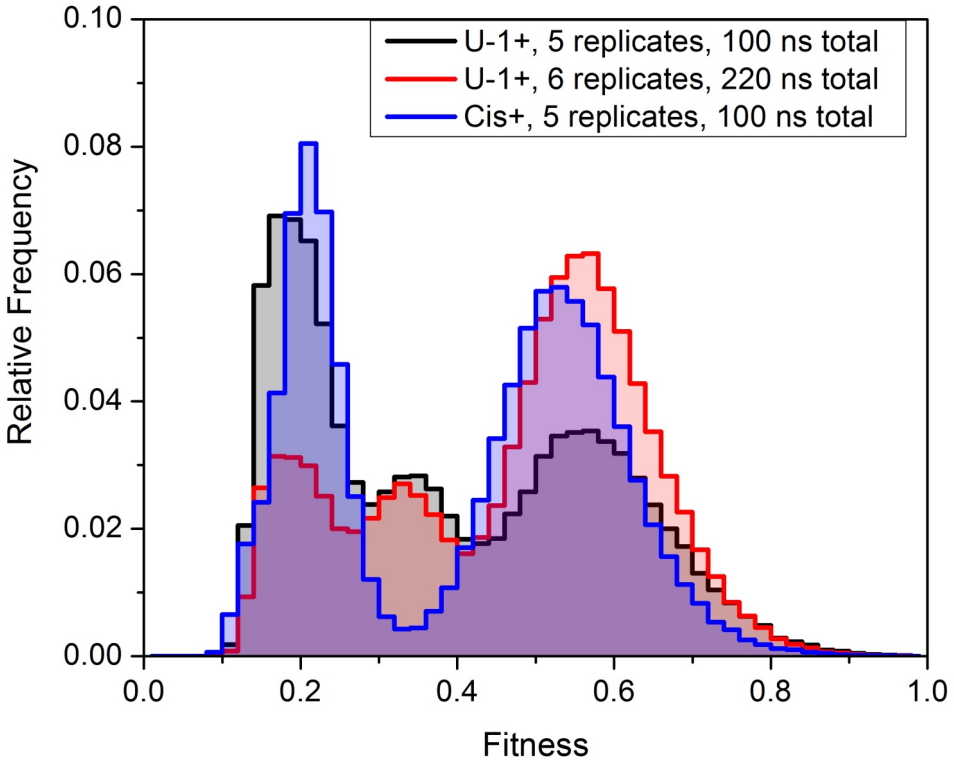


Figure 2.7: Probability density histograms of in-line fitness values for U-1+ and Cis+ simulations. For U-1+, histograms were made for the first 5 replicates (first 20 ns from each) as well as for all 6 trajectories together (including trajectory 6 of 120 ns length). For Cis+, there are 5 replicates of 20 ns each and all of these data were used to generate the histogram.

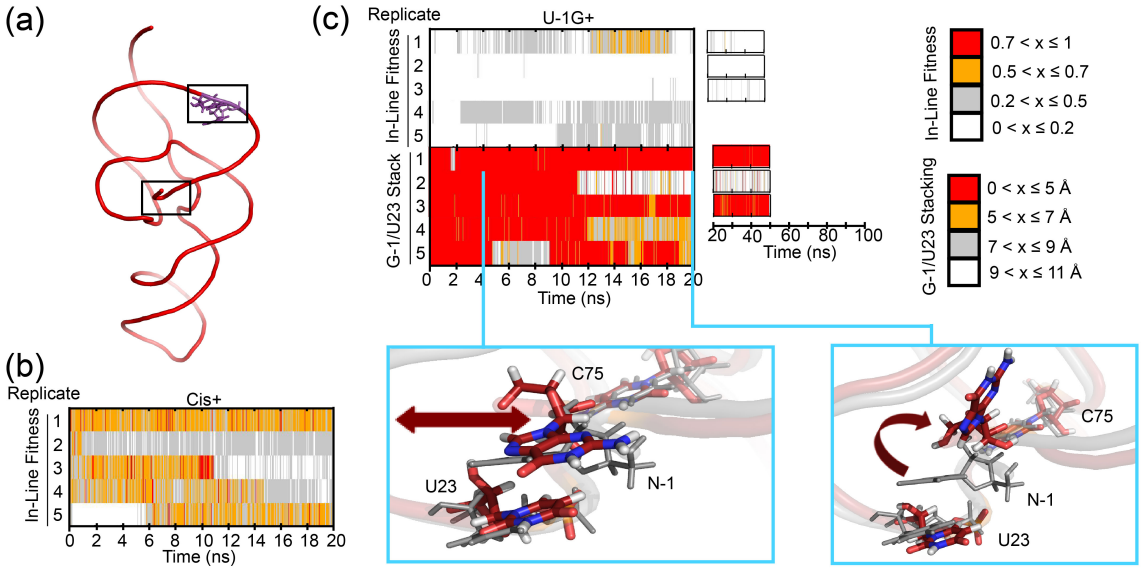


Figure 2.8: Simulating a *cis*-acting precursor model and a U-1G mutant. **(a)** Global structure of our Cis+ simulations. Modeled-in residues are denoted in purple. **(b)** In-line fitness of our Cis+ simulations. **(c)** In-line fitness (top), stacking behavior (middle) and representative stacking snapshots (bottom) of our U-1G+ simulations. Stacking distances were calculated as the distance between the center of mass of U23 heavy ring atoms, and the center of mass of G-1 atoms C4 and C5. Both snapshots overlay a snapshot from the U-1+ simulation to contrast the stacking behavior. The red double-headed arrow in the left snapshot indicates the ability of G-1 to slide. The curved arrow in the right snapshot depicts the "flipping up" movement of G-1 as it unstacks. The U-1G+ snapshots are in color, while the U-1+ snapshot is in transparent gray. Relevant residues are shown in sticks and thickened lines. Atom colors are as described in Fig. 2.3. The cyan lines in the top graph denote from which replicate and at what times the snapshots were taken.

2.4 Discussion

Disparate crystal structures of the cleavage precursor of the HDV ribozyme have raised questions about the validity of the particular modifications needed for each specific construct to crystallize. In the case of the *cis*-acting precursor, either the catalytic C75 was mutated to U or divalent metal ions were omitted,[30] both of which may have led to trapping of an off-pathway conformation. In the case of a more recent *trans*-acting precursor structure,[53] three 2′-deoxy modifications surrounding the cleavage site and the establishment of intermolecular contacts during crystal growth may have led to the poorly defined cleavage site, which lacks electron density for the U-1 nucleotide and scissile phosphate. In addition, this structure was grown at pH 5.0, protonating C75 into an ionized species that would be rare at physiological pH given C75's pK_a of ~6.4;[47,48] this protonation, in turn, likely trapped a rare, more product-like conformation that would be slow to react, as cleavage is disfavored at low pH due to the scarcity of the required general base.[47,149]

Our data support the unifying model summarized in Fig. 2.9 that resolves these questions. We find that the crystallized *trans*-acting ribozyme shows multiple distinct conformations in solution at physiological pH as evidenced by biphasic cleavage kinetics (Fig. 2.1c) and a smear upon native gel electrophoresis (EMSA, Fig. 2.3a). The fact that the EMSA smear is converted into a single product band upon cleavage in the gel (Fig. 2.3a) suggests that all of these conformations eventually cleave and hence are catalytically active even if with varying rate constants. Upon conversion of the precursor to the product in solution at physiological pH, the crystallized *trans*-acting HDV ribozyme undergoes a significant conformational change, as observed by ssFRET and trFRET in the form of a significant lengthening along the longitudinal axis (Figs. 2.3b,c). This finding is similar to observations for all previous *trans*- and *cis*-acting ribozymes,[37,41,54,55] suggesting that it represents an immutable, common feature of the HDV ribozyme in general. Taken together, these observations suggest

that occasional, stochastic protonation of C75 at physiologic pH leads to the transient adoption of an ionized state that moves C75H+ closer to the cleavage site so that its N3 donates a hydrogen bond to the G1(O5′) leaving group (while its exocyclic amine donates a hydrogen bond to one of C22's nonbridging oxygens; it remains unclear whether the adjacent G25 rotates from *anti* to *syn* before, during or after this event). This conformational adjustment sets the stage for the adoption of favorable in-line fitness. If concomitantly a Mg^{2+} ion bound to the adjacent electronegative cleft carries a deprotonated water molecule in its inner-coordination sphere, cleavage can proceed with C75H+ and $Mg(OH)^+$ serving as the general acid and base, respectively (Fig. 2.9). Since the opposing protonation states of C75H+ and $Mg(OH)^+$ are rare and transient, cleavage is slow compared to protein-based ribonucleases.[85,191,192] According to this model, the conformational change accompanying catalysis[30] is induced by the, at physiological pH, rare protonation of C75. The resulting, transient adoption of an active site conformation similar to that modeled into the *trans*-acting ribozyme structure was enriched by crystallization at low pH. Conversely, our solution probing at physiological pH detects a dominant conformation more similar to the *cis*-acting precursor crystal structure, prior to the conformational change, which represents a legitimate alternate conformation and on-pathway catalytic intermediate. (We note that there must also be other, more globally distinct conformations present in solution to explain the broad smear and multiple bands of the various *trans*-acting ribozyme observed during EMSA, as well as the biphasic cleavage kinetics, which points to deeply trapped conformations.[66,71,189])

Energetically, we have evidence that the *cis*-acting HDV ribozyme has a lower barrier towards catalysis than the *trans*-acting version (Fig. 2.9). First, all *cis*-acting ribozymes are 1-2 orders of magnitude faster in cleavage.[45] Second, in our MD simulations a *cis*-acting precursor model appears to exhibit more rapid fluctuations between conformations of high and low fitness (Fig. 2.8b), suggesting that the barrier between

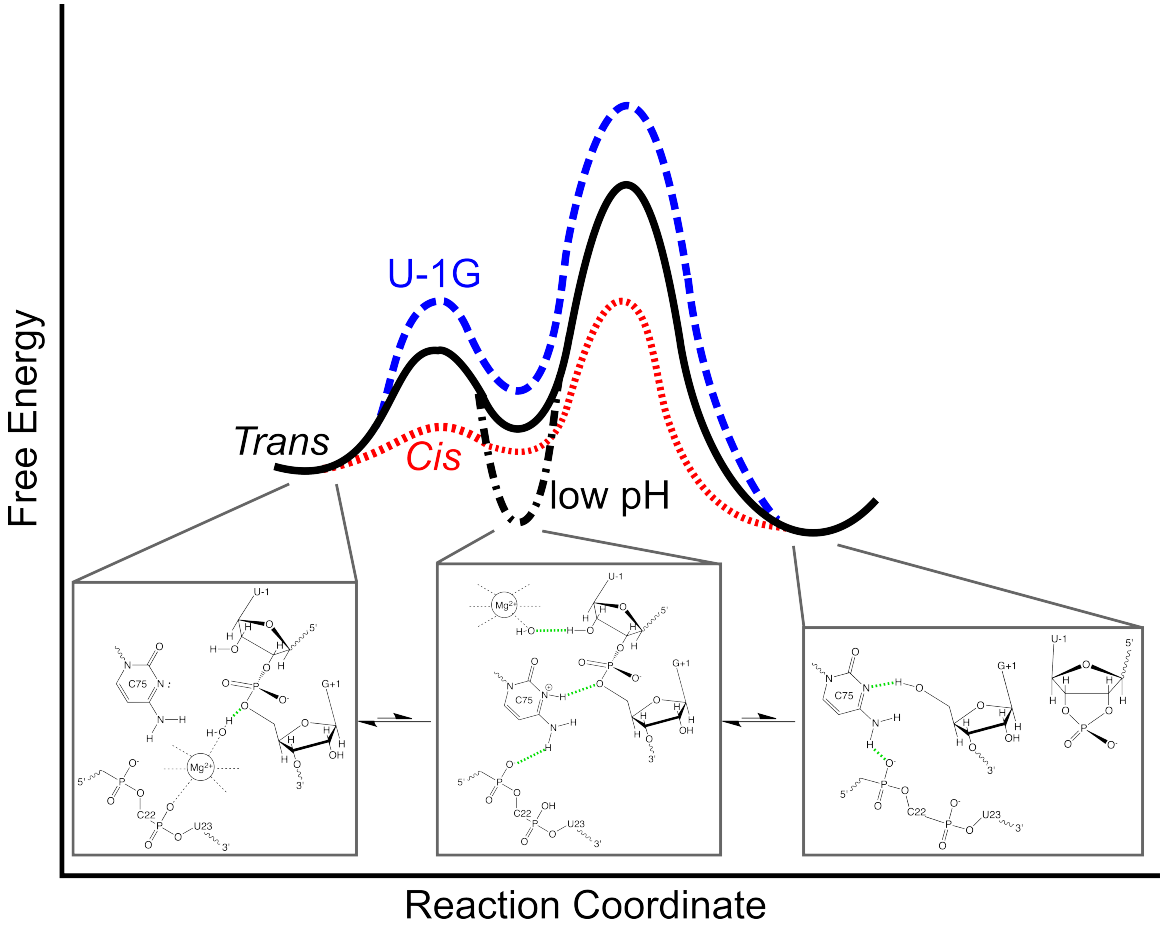


Figure 2.9: A parsimonious, unifying model of the HDV ribozyme folding free energy landscape along the catalytic reaction coordinate, indicating the divergence of the *trans*-acting, *cis*-acting and U-1G mutant ribozymes. Below each of the main on-pathway intermediates, structural models as derived from the available crystal structures are depicted. The *cis*-acting precursor crystal structure[30] resembles most closely an early intermediate dominant at physiological pH (left). This intermediate still needs to undergo a conformational change that positions the C75H+ for general acid catalysis and was trapped at low pH in the *trans*-acting precursor structure[53] (middle). If a magnesium hydroxide is bound to the active site simultaneously, cleavage into the product leads to loss of the 5′-precursor sequence and magnesium ion (right).

the pre- and post-conformational change states and, as a consequence, also the overall energetic barrier of catalysis is lowered (Fig. 2.9). However, at physiological pH the pre-conformational change state is still thermodynamically favored, leading to the observed pre-conformational *cis*-acting crystal structure. Conversely, a U-1G mutation leads to generally low in-line fitness and thus a higher catalytic barrier by disfavoring the conformational change needed to adopt a suitable active site conformation (Fig. 2.9), at least partially due to trapping of G-1 in a shifted stacking interaction with U23 (Fig. 2.8c). Our data and resulting parsimonious model thus offer rationalizations for a number of independent experimental observations. In addition, they further advance a widely held view of RNA folding free energy landscapes as particularly rugged (compared to, for example, those of proteins).[189,193]

Our MD simulations provide a glimpse of how subtle some of the conformational rearrangements may be that control a catalytically critical parameter such as the in-line fitness, which needs to reach a value of >0.5 for RNA cleavage to become significant.[175] Efficient ribozymes are likely to stabilize a favorable in-line fitness, as well as position general acid and base catalysts. The *cis*-acting HDV ribozyme is one of the most efficient self-cleaving ribozymes known, yet is still many orders of magnitude slower than RNase A.[192] One reason may be the electrostatic competition between C75 protonation and Mg^{2+} binding to the active site.[190] Furthermore, while C75H+ is a rare protonated species needed for catalysis, the hydrated Mg^{2+} ion with its high pKa of ~12 needs to adopt a rare deprotonated state to act as the general base, contrasting events that will have a low probability to coincide. Protonation of C75 in turn leads to a C75H+(N3)...G1(O5′) hydrogen bond that aids in adopting a favorable in-line fitness. It appears that it is the required temporal coincidence of these partially competing, dynamic low-probability protonation and conformational change events that slows RNA self-cleavage to at best some 10s to 100s min^{-1}.[160,186] Future studies will likely reveal more such examples for the complex relationship between folding

dynamics and function of RNA. Structures from MD simulations and crystallization are reasonable models that can be used to provide further insight into experimental results. Linking results from experimental structural and functional probing with those from computational MD simulations, as accomplished here through the use of in-line fitness as a parameter transcending the gap in accessible timescales, is likely to pave the way for more advances in this direction.

2.5 Acknowledgements

The authors thank Dr. Tristan Tabouillot, Dr. Kaushik Gunrunathan and Dr. May Daher for protocols, training and assistance with the time-resolved FRET experiments. We also thank Dr. Nad'a Špačková and Dr. Kamila Réblová for protocols and assistance with MD simulations. The authors acknowledge funding from NIH grant R01 GM062357 to N.G.W., grant P208/12/1878 from the Grant Agency of the Czech Republic to J.Š. and M.O., Operational Program Research and Development for Innovations European Regional Development Fund (project CZ.1.05/2.1.00/03.0058) and the Operational Program Education for Competitiveness European Social Fund (CZ.1.07/2.3.00/20.0017 and CZ.1.07/2.3.00/20.0058) of the Ministry of Education, Youth and Sports of the Czech Republic to M.O. and P.B., and a Canadian NSERC Postgraduate Scholarship to W.W.T.

CHAPTER III

Using Single-Atom Modifications to Disrupt the Water Chain in the Hairpin Ribozyme Active Site[1]

3.1 Introduction

Originally isolated from a plant viral pathogen, the hairpin ribozyme has become a well-studied non-coding RNA. It is a small self-cleaving RNA first discovered in the negative-sense strand of the 359-nucleotide satellite RNA of the tobacco ringspot virus, where it plays a crucial role in the replication of this viroid.[10,56–59] The minimally-active form of the hairpin ribozyme is ~50 nucleotides long and takes the form of two helices, called domains A and B, connected to each other via a two-way helical junction.[31,66,67] Domains A and B both contain internal loops, named loops A and B, respectively, that interact with each other to form the active site.[31,66,67] Many biochemical, biophysical and computational experiments have been performed with the hairpin ribozyme as a model[10,31,63,66,68–73,78–80,86,87,91,92,194] and from these experiments, four key characteristics of this ribozyme have been established. First, the active ribozyme must be in a "docked" conformation, where loops A and B physically interact with each other, which requires the presence of metal ions.[60,77,93] Second, there is a distinct conformational change between docked and undocked (inactive)

[1]The partial charge library for protonated adenosine was created by Dr. Pavel Banáš from Palacký University in the Czech Republic. All other work and analyses was done by Wendy Tay.

states that takes place on the seconds to minutes timescale.[66,69,71,93] Third, the active site has a deep negative electrostatic potential but does not attract metal ions as it is secluded from the bulk solvent.[10,74–76] The fourth key characteristic of the hairpin ribozyme is that its catalysis does not involve metal ions.[60,77,79,93] In other words, the catalytic activity of the hairpin ribozyme must consequently be mediated by the nucleobases themselves.[60,77,79,93] While the exact roles of the nucleobases are still under debate,[31,65,67,70,82,83,85–91,194] G8 and A38 have been identified as two very important residues for catalysis.[65,67,70,81,82,84] Abasic substitutions of G8 or A38 led to 10^2- to 10^3-fold and 10^4-fold losses of activity respectively.[70,83]

Part of the complexity in determining the roles of G8 and A38 can be seen in the varied observations of hydrogen bonding interactions with A-1 and G+1.[31,65,67,82,83,86,89] Self-cleavage of the hairpin ribozyme requires, first, activation of the A-1(O2′) nucleophile by a yet unidentified general base, the subsequent nucleophilic attack of this functional group on the G+1(O5′) and, finally, the formation of an upstream product containing a 2′,3′-cyclic phosphate and a downstream product containing a 5′-OH group.[10,194] In crystal structures of product and transition state mimics, G8(N1) was observed to hydrogen bond with A-1(O2′), suggesting that G8 may be the general base.[31,65,67] In molecular dynamics (MD) simulations, however, G8 preferentially hydrogen bonds with the nonbridging oxygens of G+1, suggesting an alternative role of G8 in stabilizing the negative charge of the oxyphosphorane transition state.[86,91] A38(N1) was observed in MD simulations to hydrogen bond with A-1(O2′) as well, sparking suggestions that A38 is actually the general base.[86] However, experimental and computational studies have also indicated that a protonated A38(N1) is likely present and can interact with G+1(O5′) as a general acid.[67,74,75,83,86–88,90,195–197] Consequently, the plausibility of a bifunctional A38 that shuttles a proton from the A-1(2′OH) to the G+1(O5′) has also been discussed.[86] Although hairpin ribozyme catalysis is thought to be facilitated by only the RNA bases themselves, it has still

been very difficult to assign exact roles to the catalytic residues involved.

In 2006, an intriguing observation was made about the hairpin ribozyme active site that may be, along with G8 and A38, another catalytically important component: a chain of trapped water molecules.[31,92] MD simulations and crystallographic studies indicated that 4 to 6 long-residing water molecules span the length of an otherwise solvent-excluded active site (Fig. 3.1).[31,92] The water chain makes hydrogen bonding interactions with a number of active site residues including A9, A10, A-1, G+1, A38, A40 and U42, resulting in the formation of an extended hydrogen bonding network (Fig. 3.1).[31,92] Crystallographic studies with N1-deazaadenosine at position 9 suggested that A9(N1) coordinates with these active site waters.[198] Since 2006, several studies have emerged in an attempt to eludciate the roles of the water chain and extended hydrogen bonding network.[84,86,87,198] One of the suggested roles of the active site waters, based on crystallized hairpin reaction intermediates and MD simulations, is that they stabilize the electrostatic charges of the transition state.[86,87] There have also been arguments made implicating the active site waters in a direct catalytic role.[31,92,199] Still others have argued that the active site waters were unlikely to play a role in catalysis but were rather better positioned to maintain structural integrity and to provide "backup" interactions in the event of RNA defects.[84,198] Finally, the extended nature of the hydrogen bonding network mediated by the water chain evokes the hypothesis that these waters act as a proton shuttle similar to a Grotthuss mechanism.[31,92,192,200,201] It is tempting to imagine the utility of such a proton shuttle in fast (femptosecond) long-range communication across the ~15 Å length of the hairpin ribozyme active site.[31,92,192,200,201] However, despite the several conjectures and stimulating discussions about the water chain in the active site of the hairpin ribozyme, its importance and exact function remain unclear.

The importance of internal waters is found in other biomolecules as well. Three other ribozymes, the hammerhead and *glmS* ribozymes as well as the ribosome, have

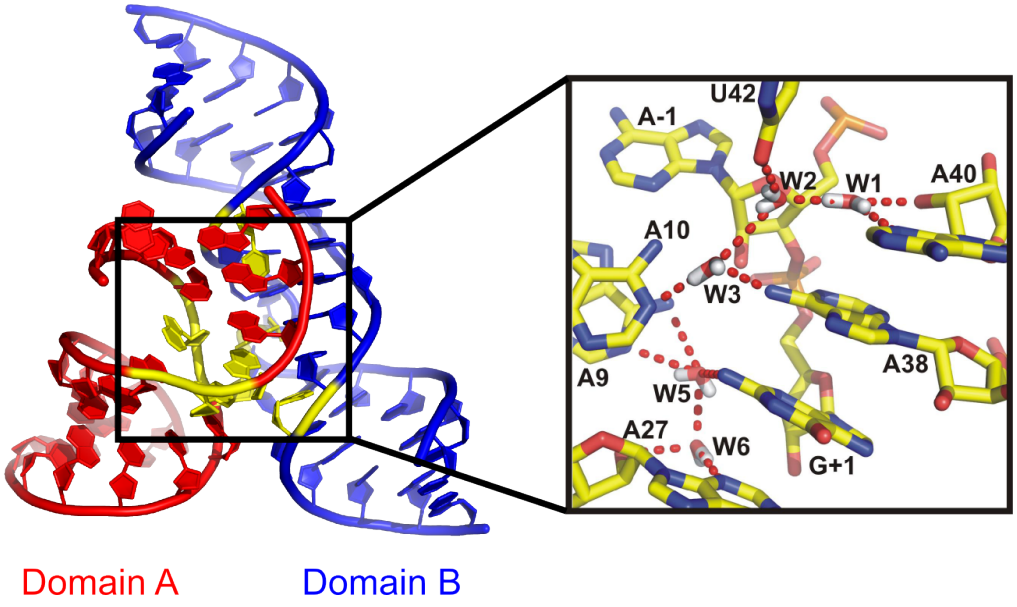


Figure 3.1: The water chain within the hairpin ribozyme active site. The docked crystal structure (PDB ID: 2OUE) of the minimally-active hairpin ribozyme (left).[31] The yellow residues indicate those residues involved in the active site water chain and are also shown in the active site close-up (right). The active site close-up is reprinted with permission from Ditzler et al.[86]

also been shown to have long-residing waters within their active sites.[192,202–204] In the case of the hammerhead ribozyme, it is thought that a hydronium ion within the active site provides a proton to the general acid residue to replenish the proton that was donated to the leaving group.[192,202] In contrast, active site waters in the *glmS* ribozyme are thought to act as a medium through which a remote general base abstracts a proton from the nucleophile.[192,203] In the ribosome, an active site water plays yet another role as a way to stabilize the oxyanion intermediate formed during peptide bond formation.[204] But active site waters are not just found in ribozymes. For instance, water is directly involved in the catalytic mechanism of the HIV-1 protease by acting as a nucleophile.[205] In cytochrome *c* oxidase, a proton pump, internal waters were proposed to have key roles in the translocation of the protons through the enzyme.[206] As demonstrated by these examples, active site waters have the potential to play a variety of functional roles. The question is now this: what is the role of the internal waters in the hairpin ribozyme?

To begin to address this question, we used all-atom MD simulations to gain insight into the dynamic behavior of the water chain and to find variants of the hairpin ribozyme that display a disrupted water chains. We also analyzed the behavior of various active-site hydrogen bonds and the in-line fitness[175] for wild type and variant ribozymes. To create variant ribozymes, we made use of N1-deazaadenosine (N1dA) and 4-thiouridine (4SU) because they have subtle single-atom substitutions for adenosine and uridine, respectively, that theoretically hinder hydrogen bonding with water. Additionally, these nucleotides can be physically incorporated into synthesized RNA, allowing for the potential for future experimental studies of the variants identified in this study.[84,198,207] N1dA has in fact already been used in kinetic and crystallographic studies by Wedekind and coworkers at positions 9, 10 and 38.[84,198] Incorporation of N1dA at these positions serves as a good starting point for further investigation of their effects on the water chain. Furthermore, because the protonation state of A38

prior to cleavage is not certain,[67,74,75,83,86–88,90,91,194–197] MD simulations were performed for both protonated and unprotonated versions of A38, which are denoted as *A38+* and *A38*, respectively. Our simulations suggest that the water chain is robust and that a variant with two single-atom modifications, N1dA9 and 4SU42, is needed to cause disruption in the water chain.

3.2 Materials and Methods

Non-canonical nucleotides and molecular dynamics simulations

The libraries of partial atomic charges for N1-deazaadenosine (N1dA), 4-thiouridine (4SU) and protonated adenosine (at the N1 atom) were derived following the procedure from Cornell et al.[113,116] Geometry of the nucleotide bases was first optimized in Gaussian03[172] using the Hartree-Fock theory 6-31G(d) basis set. Electrostatic potential (ESP) fits were then calculated in Gaussian in preparation for restrained electrostatic potential (RESP) fitting[171] (calculated in antechamber module of AMBER11[165]). From the antechamber RESP output, charges could be assigned to the base atoms. Phosphate and sugar charges were kept the same as in canonical adenosine or uridine. Residual charge was added to C1′ to ensure a net charge of -1 for the N1dA and 4SU and a net charge of 0 for protonated adenosine.

Hairpin ribozyme simulations were generated using the AMBER11 suite of programs[165] and the ff99bsc0χ_{OL3} force field.[111,115] Starting structures for the MD simulations were based on the crystal structure of the pre-cleaved, docked hairpin ribozyme with just domains A and B (PDB ID 2OUE).[31] This hairpin ribozyme has an A-1 2′-O-methyl to prevent cleavage and a U39C mutation that prevents conformational heterogeneity.[31] For the simulations, the A-1 2′-O-methyl was replaced with a 2′-OH, while the U39C mutation was left as it is in the crystal structure. The structures were charge-neutralized with Na^+ ions placed at locations of optimal electrostatics using

the LEaP module of AMBER11 and solvated using a rectangular periodic TIP3P water box model[166] extended to a distance of ≥10 Å from any ribozyme or Na^+ atom. The ABC protocol[167,168] was used to minimize and equilibrate structures prior to production runs. The minimization and equilibration protocol as well as the production runs were performed using the PMEMD implementation of the AMBER sander module. For the production runs, the particle mesh Ewald (PME) method was employed with a grid spacing of 1.0 Å, cubic-spline interpolation and a heuristic pair list update with a Lennard-Jones cutoff of 9.0 Å. Production runs were carried out at 300 K with constant-pressure boundary conditions and the Berendsen temperature coupling algorithm[169] with a time constant of 1.0 ps. The SHAKE procedure[170] was applied to constrain bonds involving hydrogen. The list of simulations performed is summarized in Table 3.1.

Table 3.1: List of hairpin ribozyme MD simulations

Simulation Label	A38 Protonation State	Number of Replicates	Length (ns) of Each Replicate
wild type[1]	A38	3	50, 50, 50
N1dA9	A38	3	50, 50, 50
N1dA38	n/a	3	50, 50, 50
N1dA9_N1dA10	A38	3	50, 50, 50
N1dA9_4SU42	A38	3	50, 50, 50
wild type	A38+	3	50, 50, 50
N1dA9	A38+	3	50, 50, 50
N1dA9_N1dA10	A38+	3	50, 50, 50
N1dA9_4SU42	A38+	3	50, 50, 50

[1]Starting structures of the hairpin ribozyme were based on PDB ID 2OUE[31] but A-1(2′O-methyl) was replaced with A-1(2′OH).

Three replicates of 50 ns each were performed for each type of simulation to obtain a reasonable sampling size. MD trajectories were analyzed using the AMBER ptraj module to calculate heavy-atom distances (used as an indication of the hydrogen bonding that occurred in the simulations), in-line angles and the water density grids. Images of the water density grids were averaged over 10 ns increments. Matlab (The

MathWorks)[174] and VMD[208] were used to visualize results from ptraj. In-line fitness (F) was calculated using the following definition[175]

$$F = \frac{(\tau - 45)}{(180 - 45)} \times \frac{3^3}{d^3_{O2',P}}$$

where τ is the angle between the 2′ oxygen nucleophile (O2′), the phosphorus electrophile (P) and the 5′ oxygen of the leaving group (O5′), and d is the distance between the 2′ oxygen nucleophile and the phosphorus electrophile.

3.3 Results

Unprotonated A38 results in better in-line fitness but less stable docking interactions

To assess whether protonated A38 (*A38+*) or unprotonated A38(*A38*) is more catalytically relevant for our MD simulations, we compared the catalytic in-line fitness from the simulations of wild type and variant ribozymes containing *A38+* against those containing *A38*. In-line fitness, a parameter originally discussed by Soukup and Breaker in 1999, is an indication of the potential for a phosphodiester linkage to undergo self-cleavage based on its geometry.[175] It is a way to link the active site geometry observed in crystal structures and MD simulations of ribozymes to the potential catalytic relevance of that structure. In-line fitness values range from 0 to 1, 0 being poor fitness with the least likelihood of cleavage, and 1 being excellent fitness with the most likelihood of cleavage. By tracking the in-line fitness of the scissile bond over the course of the simulation, we are able to essentially track how active the active site is. Among our hairpin ribozyme simulations, *A38* simulations sampled better fitness (Fig. 3.2a). This result indicates that the geometry at the scissile bond is primed better for catalysis in the *A38* simulations than in the *A38+* simulations, which also lends support to the notion that A38 is likely to act as a base

at some point during catalysis.[86] Furthermore, this trend is seen regardless of whether a wild type or modified ribozyme was simulated, indicating that the protonation state of A38 has a more substantial effect on catalytic propensity than subtle modifications at A9, A10 or U42.

However, in addition to suitable catalytic geometry, the hairpin ribozyme must also be in the docked conformation to be catalytically active.[60,77,93] Thus, hydrogen bonds involved in docking were monitored to assess the stability of the docked conformation. Two hydrogen bonds involved in docking that show different behavior between the *A38+* and *A38* simulations are shown in Fig. 3.2b. Interestingly, although *A38* simulations show better in-line fitness, many of the *A38+* simulations show more stable docking hydrogen bonding for these two docking interactions compared to their *A38* counterparts (Fig. 3.2b). Specifically, all *A38+* simulations show more consistent hydrogen bonding distances (≤3.5 Å) and less excursions to ≥4.0 Å throughout the representative trajectories compared with *A38* simulations (Fig. 3.2b). This observation, in conjunction with the better fitness seen for *A38* simulations, further supports the hypothesis that both protonation states of A38 are catalytically relevant.[86] It should be noted that the N1dA9_4SU42 simulations actually contain a U42(S4), which is a poorer hydrogen bond acceptor than U42(O4). The reduced hydrogen bonding capability from this atomic substitution, and thus a reduced docking capability, may be a contributing factor to why protonation at A38 does not improve the hydrogen bonding distance (Fig. 3.2b). Because *A38* promotes better fitness sampling but *A38+* generally promotes more stable docking interactions, we considered both sets of simulations in subsequent analyses.

N1dA9 and N1dA38 do not disrupt the water chain

We first checked the single-atom modifications N1dA9 and N1dA38 for water chain disruption. These variants are of particular interest because of previous experiments

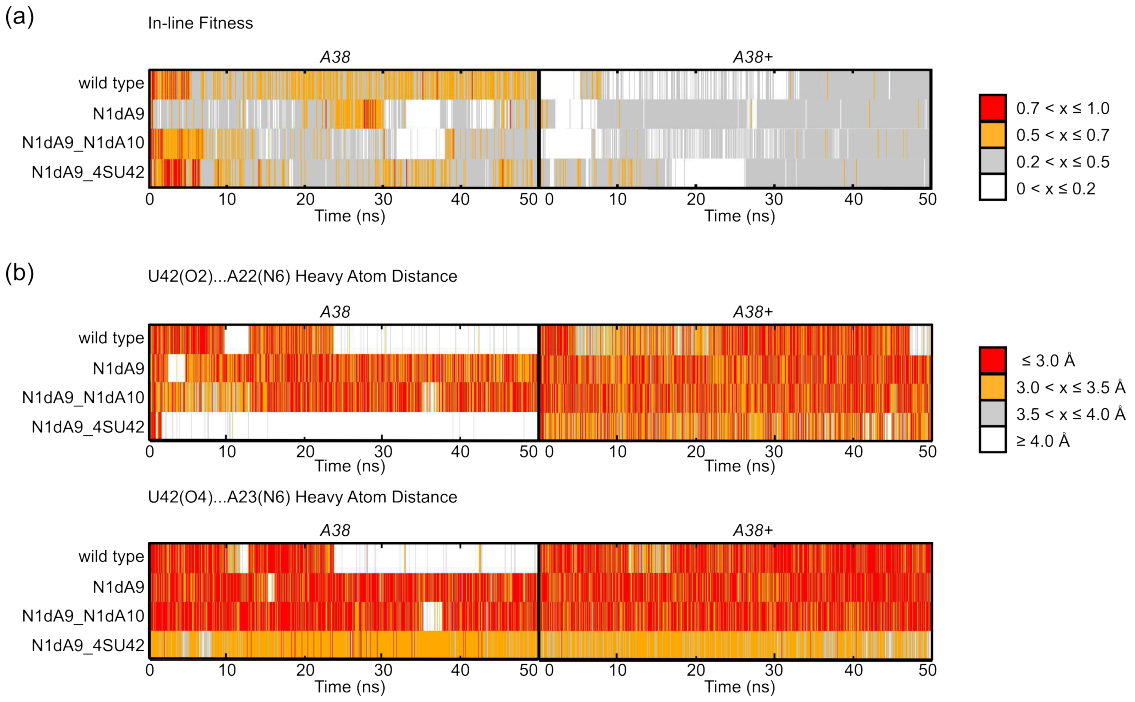


Figure 3.2: In-line fitness and docking hydrogen bonding distances over time. (a) Representative fitness time traces for simulations with unprotonated A38 (*A38*) or protonated A38 (*A38+*). **(b)** Hydrogen bonding heavy-atom distances for a two interactions involved in docking. Again, representative traces for each type of simulation are shown. The U42...A23 distance tracked for N1dA9_4SU42 was U42(S4)...A23(N6).

by Wedekind and coworkers.[84,198] The Wedekind group showed that A9(N1) may be important in coordinating with active site waters and that the N1dA9 variant resulted in an 8-fold loss of activity, making the N1dA9 variant an attractive candidate for further exploring the functional role of the water chain.[198] A38(N1), in contrast, was not seen to coordinate with a water molecule in a previous crystal structure but the N1dA38 variant resulted in complete loss of activity.[84] To ensure that the dramatic loss of activity is not somehow related to the water chain and because it is possible that a water interacting with A38(N1) is simply not crystallographically resolved, we also tested the N1dA38 variant. However, both N1dA9 and N1dA38 variants, for *A38* and *A38+* simulations, do not show a clear disruption of the water chain (Figs. 3.3b,g,j). The N1dA9 variant, for both *A38* and *A38+*, show a slight gap of water density around the A9(C1) (Figs. 3.3b,g), as expected due to the loss of the hydrogen bond acceptor A9(N1). Yet, the water chain and extended hydrogen bonding network are still mostly intact for N1dA9. N1dA38 simulations are consistent with previous observations from crystal structures in that the water chain is not affected (Fig. 3.3j).[84] In our MD simulations, A38(N1) and A38(C1) consistently face away from water chain, decreasing the potential for a N1dA38 modification to directly affect the waters (Fig. 3.3j). Both N1dA9 and N1dA38 simulations are similar to what is seen for wild type simulations (Figs. 3.3a,f). Therefore, we do not classify N1dA9 or N1dA38 to be effective water chain disruptors. Despite the detrimental effects of N1dA9 and N1dA38 on catalytic activity, we suggest that disruption of the water chain is not a contributing cause.

Heavy atom distances for active site hydrogen bonds involving A-1 and G+1 were tracked over the course of the simulations as well. Differences in hydrogen bonding behavior are observed when comparing wild type and modified ribozymes (Fig. 3.4). For example, in both the N1dA9 and N1dA38 variants for the *A38* simulations, G8(N1) and G+1(pro-R) hydrogen bond less with A-1(O2′) and more with each

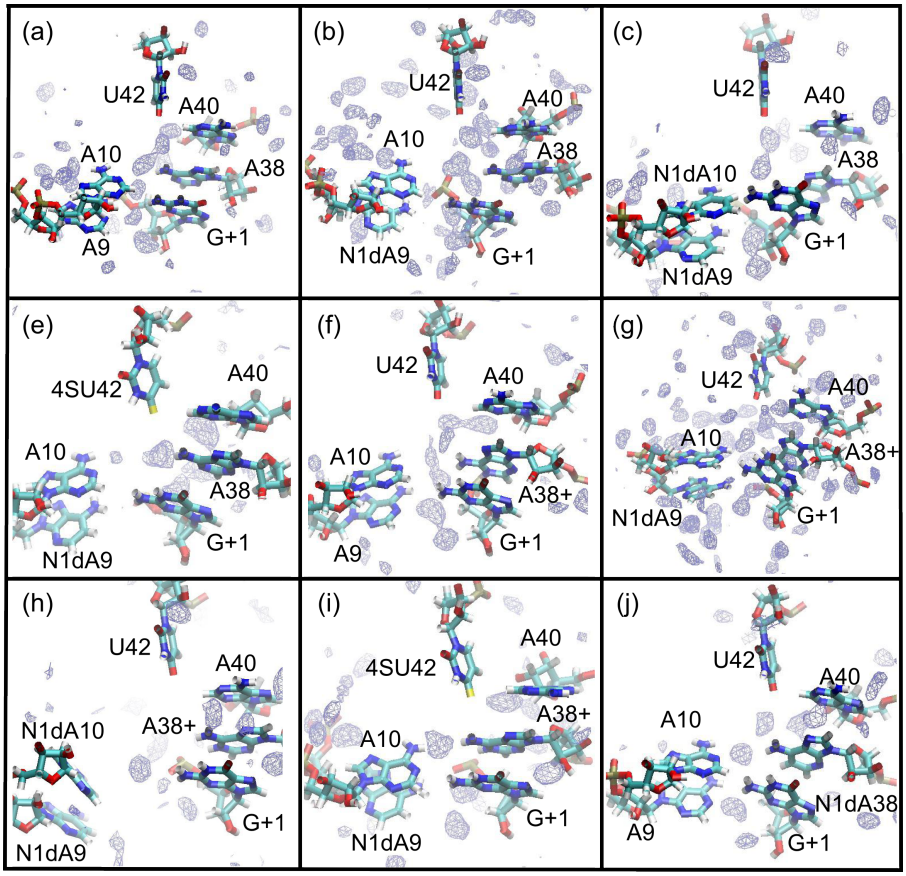


Figure 3.3: Representative atomic density maps of water in and around the active site of the wild type and variant ribozymes. Atomic density maps of water were made to indicate the location of water (indicated by the blue mesh) in and around the active site for the following versions of the hairpin ribozyme: **(a)** wild type, *A38* **(b)** N1dA9, *A38*, **(c)** N1dA9_N1dA10, *A38*, **(d)** N1dA9_4SU42, *A38*, **(f)** wild type, *A38+*, **(g)** N1dA9, *A38+*, **(h)** N1dA9_N1dA10, *A38+*, **(i)** N1dA9_4SU42, *A38+*, **(j)** N1dA38. Atomic density maps were made by averaging over 10 ns.

other compared to in the wild type ribozyme (Fig. 3.4). These observations suggest that the single-atom modifications do have a noticeable effect on local interactions, despite not causing a significant impact on the water chain. These data also indicate a robustness of the water chain that prevent it from being easily perturbed.

A38 protonation state influences the effect of water chain disruption by the N1dA9_N1dA10 variant

The results from the N1dA9 and N1dA38 modifications pose an important consideration: the subtlety of single-atom modifications. The water chain is more robust than initially hypothesized to be. Subtle, single-atom modifications that inhibit hydrogen bonding ability are perhaps not enough to disrupt the water chain. Therefore, in an attempt to disrupt the water chain using a more drastically modified variant, we simulated a ribozyme that contains single-atom modifications at positions 9 and 10 simultaneously (N1dA9_N1dA10). Interestingly, the outcome differs depending on the protonation state of A38. For *A38* simulations, the N1dA9_N1dA10 variant does not have a significant impact on the water chain (Fig. 3.3c). However, for *A38+* simulations, the N1dA9_N1dA10 variant does show a clearly disrupted water network with most of the water missing from the active site (Fig. 3.3h). The active site in these N1dA9_N1dA10 simulations is also a lot wider than in the wild type (Figs. 3.3f,h). We conclude that disruption of the water chain can be achieved using two single-atom modifications although, such as with the N1dA9_N1dA10 variant, disruption may also depend on the protonation state of A38.

Hydrogen bonding heavy atom distances involving A-1 and G+1 were also tracked for the N1dA9_N1dA10 simulations. Most significantly, for the *A38+* N1dA9_N1dA10 simulations, hydrogen bonding interactions involving G8 are completely disrupted (Fig. 3.5, bottom left). Closer inspection of the hydrogen bonding heavy atom distances reveal that distances range from ~6 to 10 Å between G8 and A-1 or G+1.

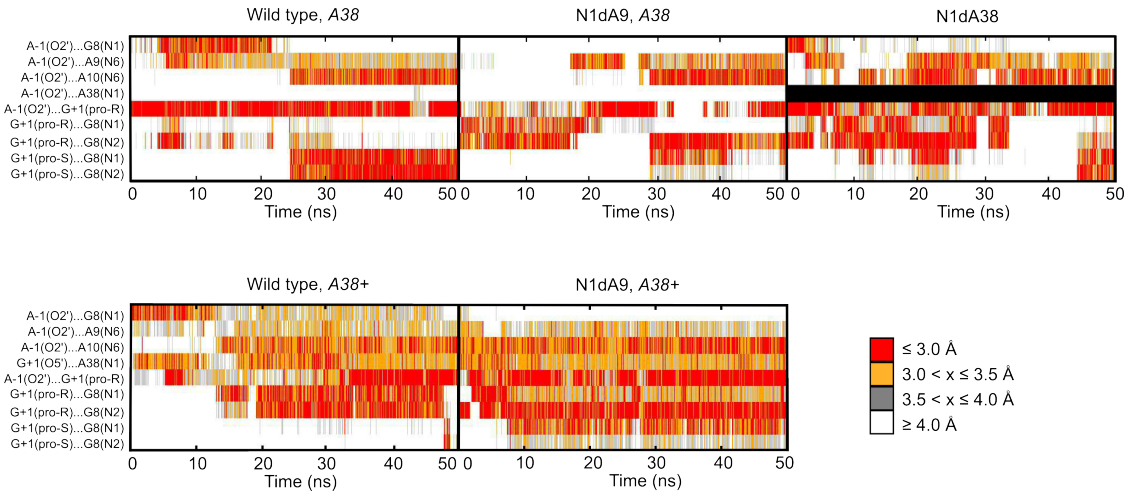


Figure 3.4: Evolution of hydrogen bonding heavy atom distances for wild type and single-atom variants. Representative traces are shown for a number of the active site hydrogen bonding interactions tracked over the course of the simulations. In the *A38* condition, A-1(O2′)...A38(N1) interaction was monitored that reflected the potential for A38 to act as a general base. In the *A38+* condition, the G+1(O5′)...A38(N1) was monitored instead, reflecting the potential for A38+ to act as a general acid. The black bar in the N1dA38 simulations indicates no value as that interaction could not be measured.

These data are consistent with the perturbed wider active site seen in Fig. 3.3h and the poor in-line fitness seen in Fig. 3.2a.

Single-atom modifications simultaneously at two distinct sites most consistently disrupt the water chain

The N1dA9_4SU42 variant was simulated to determine the effect of two single-atom modifications at different locations within the active site. For comparison, the N1dA9_N1dA10 variant contains two single-atom modifications localized around the same vicinity. Independent of A38's protonation state, the N1dA9_4SU42 variant shows disruption for the majority of the water chain for a period of at least 10 ns (Figs. 3.3e,i). The significant disruption seen in many of the replicates for a period of at least 10 ns strongly suggests that the N1dA9_4SU42 is consistently inducing disruption of the water chain. The strategic hindrance of hydrogen bonding interactions at two distinct locations in the active site is a key factor for disruption of such a robust water chain.

The active site hydrogen bonding interactions and in-line fitness were also analyzed for the N1dA9_4SU42 variant. The differences in hydrogen bonding frequency in the N1dA9_4SU42 variant active site compared to that of the wild type for both *A38* and *A38+* simulations are surprisingly subtle given the significant disruption of the water chain (Fig. 3.5, right side). Even the in-line fitness is not as poor for the N1dA9_4SU42 variant although it is poorer than that for the wild type (only for the *A38* simulations since the fitness for the *A38+* simulations is poor even for the wild type) (Fig. 3.2a). In fact, the water chain could be disrupted in the N1dA9_4SU42 variant, but there is not as substantial a difference in the hydrogen bonding patterns, fitness parameter or active site shown in Fig. 3.3 as that observed for the N1dA9_N1dA10 *A38+* variant. Together, these data suggest that perhaps the water chain may not be acting predominantly as a structural scaffold,[84,198] at least in the

sense of a "glue" that keeps the active site intact.

3.4 Discussion

The fascinating discovery of internal, long-lived waters within the hairpin ribozyme catalytic core sparked a new consideration regarding the ribozyme's mechanism.[31,84,86,87,92,198] Internal waters found inside other molecules, such as the hammerhead ribozyme, the ribosome, serine proteases and RNase A, are thought to have a functional role.[192,192,202–204,209] We used MD simulations to gain insight into the behavior of the hairpin ribozyme water chain and to discover active site modifications that are likely to disrupt the chain. The water chain is more robust than initially hypothesized. Single-atom modifications of N1dA9 and N1dA38 do not show a disrupted water chain despite having effects on active site hydrogen bonding behavior as well as causing significant losses of activity.[84,198] The N1dA9_N1dA10 variant, which has two single-atom modifications proximal to each other, only shows disruption of the water chain for the *A38+* simulations. The N1dA9_4SU42 variant has two single-atom modifications at two distinct locations in the active site and it shows disruption of the water chain regardless of A38 protonation state. Therefore, disruption of the water chain is most successful when simultaneously incorporating two active site modifications at distinct sites.

Both *A38* and *A38+* simulations were analyzed because we could not assess which protonation state was more relevant for our MD simulations. *A38* simulations give rise to better in-line fitness, but *A38+* simulations result in slightly more stable docking as shown by more consistent hydrogen bonding for two docking interactions. A38 has been proposed to have a bifunctional role by playing the general acid and the general base at different times throughout the catalytic mechanism.[86] Based on this idea, it is plausible that A38+ as a general acid is important for promoting more stable docking interactions, but upon deprotonation and conversion to A38, which

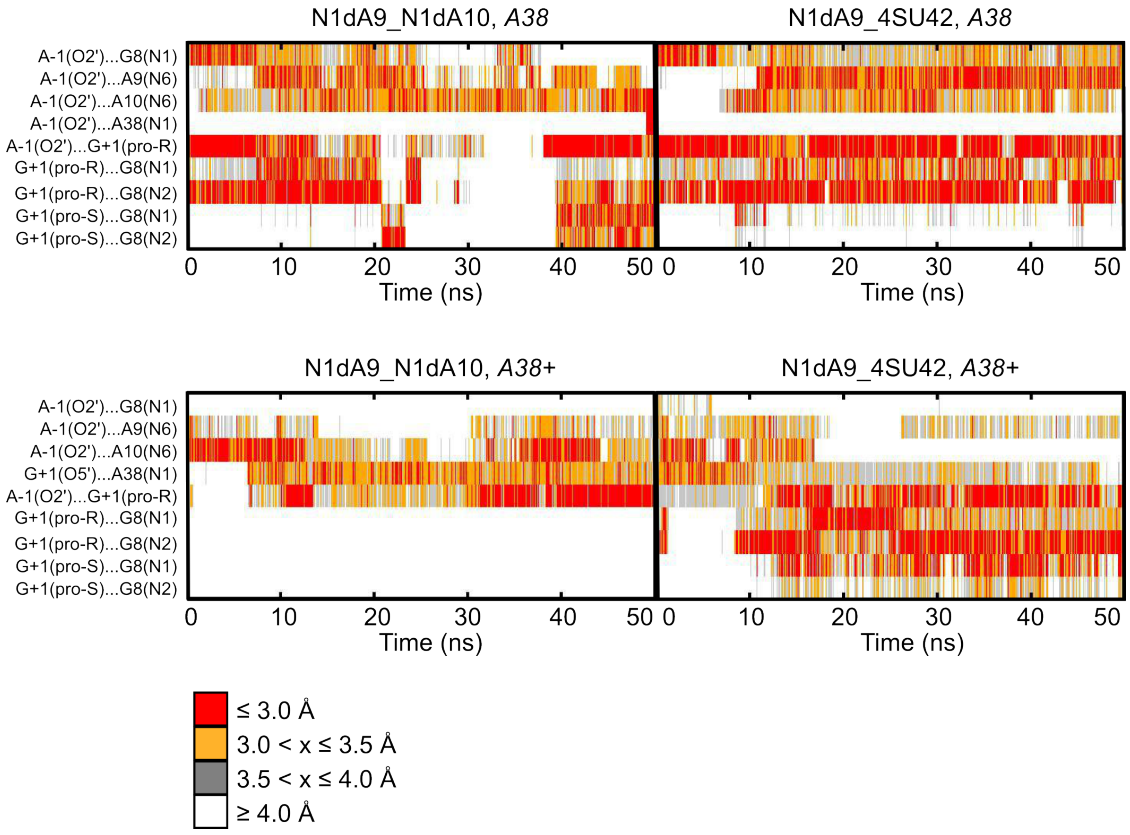


Figure 3.5: Evolution of hydrogen bonding heavy atom distances for variants with two single-atom modifications. Representative traces are shown for a number of the active site hydrogen bonding interactions tracked over the course of the simulations. In the *A38* condition, A-1(O2′)...A38(N1) interaction was monitored that reflected the potential for A38 to act as a general base. In the *A38+* condition, the G+1(O5′)...A38(N1) was monitored instead, reflecting the potential for A38+ to act as a general acid.

can act as a general base, it then is more important for promoting favorable in-line fitness conducive to cleavage. The protonation state of A38 can also be an important factor in disrupting the water chain, such as the case for the N1dA9_N1dA10 variant, where the N1dA9_N1dA10 *A38* simulations do not show a disrupted water chain but the N1dA9_N1dA10 *A38+* simulations do.

Coming up with a consistent functional role for the water chain has been made difficult by the observations from the two ribozyme variants that showed disrupted water chains in our MD simulations. N1dA9_N1dA10 *A38+* simulations initially hinted that the water chain may act as a structural scaffold role[84,198] because of the perturbation of the active site seen with the disrupted water chain (although we do acknowledge that it difficult from these observations to determine definitively whether the disrupted water chain is the cause or the effect of the perturbed active site). However, simulations of N1dA9_4SU42 (both *A38* and *A38+*) show, while with a disrupted water chain, a much less perturbed active site than in the N1dA9_N1dA10 variant. It is possible that the water chain may not be acting predominantly as a structural glue that keeps the active site from falling apart. The active site waters could be playing minor structural roles or a variety of non-structural roles as well.[192,202–206] Specifically for the hairpin ribozyme, other proposed functions of the active site waters include being directly involved in catalysis, helping with electrostatic stabilization of the transition state, and acting as a Grotthuss-like proton shuttle.[31,86,87,92,192,199–201] Variants, like N1dA9_4SU42, that have two single-atom modifications at distinct sites in the active site and consistently show disruption in the water chain in MD simulations, are predicted, based on the water chain not acting as a key structural player in keeping the active site from falling apart, to still form the active site but docking may occur to a lesser degree (suggested by the less favorable docking hydrogen bonding distances seen in Fig. 3.2) and there may be subtle differences in the various rate constants (such as